

**Electronic supplementary information for learning size-adaptive
molecular substructures for explainable drug-drug interaction
prediction by substructure-aware graph neural network**

Ziduo Yang^{a#}, Weihe Zhong^{a#}, Qiujie Lv^a, Calvin Yu-Chian Chen^{a,b,c}*

^aArtificial Intelligence Medical Center, School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, 510275, China

^bDepartment of Medical Research, China Medical University Hospital, Taichung 40447, Taiwan

^cDepartment of Bioinformatics and Medical Engineering, Asia University, Taichung, 41354, Taiwan

[#] Equal contribution

** Corresponding Authors*

Calvin Yu-Chian Chen, Ph.D.

Dean of Artificial Intelligence Medical Center, Professor of School of Intelligent Systems Engineering, Sun Yat-sen University.

TEL: 02039332153

E-mail: chenychian@mail.sysu.edu.cn

S1. Supplemental Methods

S1.1 Problem formulation

The DDI prediction task is to develop a computational model that receives two drugs with an interaction type as inputs and generates an output prediction indicating whether there exists an interaction between them. Formally, given a set of drugs \mathcal{D} , a set of interaction types $\mathcal{I} = \{I_i\}_{i=1}^K$, and a dataset $\mathcal{M} = \{(d_x, d_y, r)\}_{i=1}^N$, where d_x and d_y are sampled from \mathcal{D} with interaction type $r \in \mathcal{I}$, and K is the total interaction types and N is the sample numbers, our goal is to find a model $f: \mathcal{D} \times \mathcal{D} \times \mathcal{I} \rightarrow \{0,1\}$ that maps the input into the binary decision.

S1.2 Input representation

The model takes DDI tuples (d_x, d_y, r) as input. A drug d is represented as a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of nodes and \mathcal{E} is the set of edges/bonds. In a molecule, $v_i \in \mathcal{V}$ is the i -th atom and $e_{ij} \in \mathcal{E}$ is the chemical bond between i -th and j -th atoms. Each node v_i has a corresponding feature vector $x_i \in \mathbb{R}^d$ and each edge/bond e_{ij} has a feature vector $x_{ij} \in \mathbb{R}^{d'}$. The features used for atoms and bonds can be found in Tables S1 and S2 of ESI.

S1.3 Directed message passing neural network for substructures extraction

The D-MPNN¹ is a variant of the generic message passing neural network (MPNN)² architecture. Therefore, we first introduced the MPNN and then generalized it to D-MPNN.

S1.3.1 Message passing neural network

The MPNN is some type of GNN that maps an undirected graph \mathcal{G} to a graph-

level vector $h_{\mathcal{G}}$ usually with a message passing phase and a readout phase. The message passing phase updates each node by considering its neighboring nodes, and the readout phase computes a graph-level feature vector $h_{\mathcal{G}}$ for the whole graph. Concretely, the message passing phase consists of T steps. On each step t , node-level hidden features $h_i^{(t)}$ and messages $m_i^{(t)}$ associated with each node v_i are updated using the message function M_t and node update function U_t according to

$$m_i^{t+1} = \sum_{v_j \in N(v_i)} M_t(h_i^{(t)}, h_j^{(t)}, e_{ij}) \quad (1)$$

$$h_i^{(t+1)} = U_t(h_i^{(t)}, m_i^{(t+1)}) \quad (2)$$

where $N(v_i)$ represents the set of neighbors of v_i in the graph \mathcal{G} , and $h_i^{(0)}$ is set to the initial atom features x_i . The readout phase then uses a readout function R to obtain a graph-level feature vector based on the node-level features at the final step as follows

$$h_{\mathcal{G}} = R(\{h_i^{(T)} \mid v_i \in \mathcal{G}\}) \quad (3)$$

The message function M_t , node update function U_t , and readout function R are all learned differentiable functions.

S1.3.2 Directed message passing neural network

In the MPNN, every node will have received a message from all of its neighbors. However, such a mechanism is likely to introduce noise into the graph representation¹. Using Figure S1 (a) as an illustration, in node-level message passing, the node v_1 aggregates the message from its neighbors (i.e., v_2 , v_3 , and v_4); however, in the next iteration, the message of the node v_1 will be propagated to the node v_4 that already contains the information of node v_1 in the previous iteration, which creates an

unnecessary loop in the message passing trajectory.

The D-MPNN propagates messages along with directed bonds instead of nodes, as shown in Figure S1 (b). To highlight the difference between e_{ij} and e_{ji} (i.e., two different directions of a bond), we renamed them as $e_{i \rightarrow j}$ and $e_{j \rightarrow i}$. Formally, the D-MPNN operates on bond-level hidden features $h_{ij}^{(t)}$ and messages $m_{ij}^{(t)}$ instead of $h_i^{(t)}$ and $m_i^{(t)}$ used in MPNN. Note that $h_{ij}^{(t)}$ and $m_{ij}^{(t)}$ are distinct from $h_{ji}^{(t)}$ and $m_{ji}^{(t)}$, where the former is bond-level features along with the bond $e_{i \rightarrow j}$ while the latter are bond-level features along with the bond $e_{j \rightarrow i}$. The corresponding message passing update equations are thus

$$m_{ij}^{(t+1)} = \sum_{v_k \in N(v_i) \setminus v_j} M_t(v_k, v_i, h_{ki}^{(t)}) \quad (4)$$

$$h_{ij}^{(t+1)} = U_t(h_{ij}^{(t)}, m_{ij}^{(t+1)}) \quad (5)$$

Note that $m_{ij}^{(t+1)}$ does not depend on its reverse message $m_{ji}^{(t)}$ from the previous iteration, which leads to a more efficient message passing compared with MPNN¹. We implemented functions M_t and U_t as follows

$$M_t(v_i, v_j, h_{ij}^{(t)}) = h_{ij}^{(t)} \quad (6)$$

$$U_t(h_{ij}^{(t)}, m_{ij}^{(t+1)}) = h_{ij}^{(0)} + m_{ij}^{t+1} \quad (7)$$

Note that Equations (6) and (7) do not apply any transformations to feature vectors at each iteration t , which can be viewed as an instance of the concept of simplifying graph convolutions³. Besides, we proposed to initialize the bond-level hidden features as

$$h_{ij}^{(0)} = W_i x_i + W_j x_j + W_{ij} x_{ij} \quad (8)$$

where $W_i \in \mathbb{R}^{h \times d}$, $W_j \in \mathbb{R}^{h \times d}$, and $W_{ij} \in \mathbb{R}^{h \times d'}$ are learnable weight matrices. After T

times updates of bond-level hidden features, we returned to the node-level hidden features of the molecule by summing the incoming bond-level features according to the following:

$$h_i^{(T)} = \sum_{v_j \in N(v_i)} h_{ji}^{(T)} \quad (9)$$

The readout phase of the D-MPNN is the same as the readout phase of a generic MPNN as described in Equation (3). The original D-MPNN used a global sum pooling function to obtain the graph-level representation $h_{\mathcal{G}}$ for a given molecule/graph \mathcal{G} .

Reference

- 1 K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea and others, *J. Chem. Inf. Model.*, 2019, **59**, 3370–3388.
- 2 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, in *International Conference on Machine Learning*, 2017, pp. 1263–1272.
- 3 F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu and K. Weinberger, in *International conference on machine learning*, 2019, pp. 6861–6871.

S2. Supplemental Tables

Table S1. Atom features.

Name	Description	Dim
Atom type	Heavy atom type (e.g., C, O, N, S, I)	Total number of heavy atoms in the dataset*
Degree	Number of covalent bonds [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	11*
Implicit valence	Implicit valence of the atom [0, 1, 2, 3, 4, 5, 6,]	7*
Hybridization	[sp, sp2, sp3, sp3d, sp3d2]	5*
Aromatic	Whether the atom is part of an aromatic system	1
Formal charge	Formal charge of the atom	1
Radical electrons	Number of radical electrons for the atom	1

*One-hot representation

Table S2. Bond features.

Name	Description	Dim
Bone type	[single, double, triple, aromatic]	4*
Conjugated	Whether the bond is part of a conjugated system	1
Ring	Whether the bond is part of a ring	1

*One-hot representation

Table S3. Search range and selected values of hyperparameters for SA-DDI.

Hyperparameter	Search range	Selected value
Number of iterations T	[1, 2, 3, 4, 5, 6, 10, 15, 20, 25]	10
Number of hidden units for h_i/h_{ij}	[32, 64, 128]	64
Learning rate	[1e-4, 5e-4, 1e-3, 1e-2]	1e-3

S3. Supplemental Figures

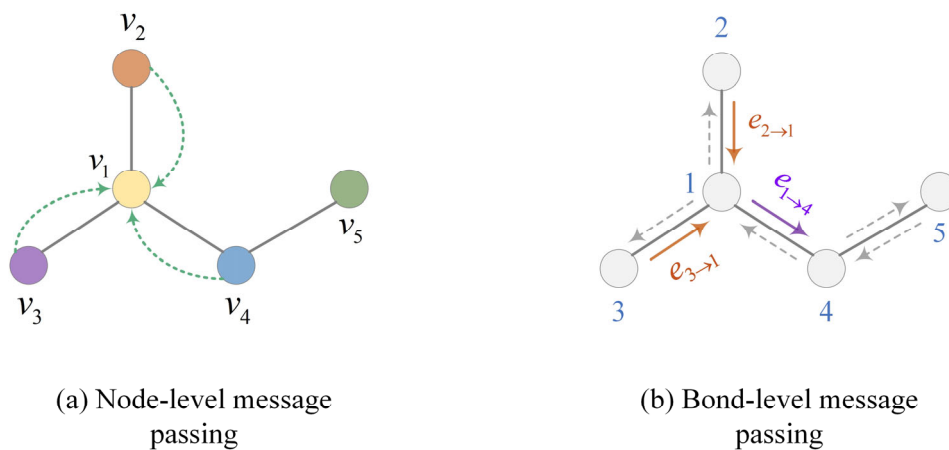


Figure S1. The difference between MPNN and D-MPNN. The messages are propagated through (a) nodes for MPNN and (b) bonds for D-MPNN.

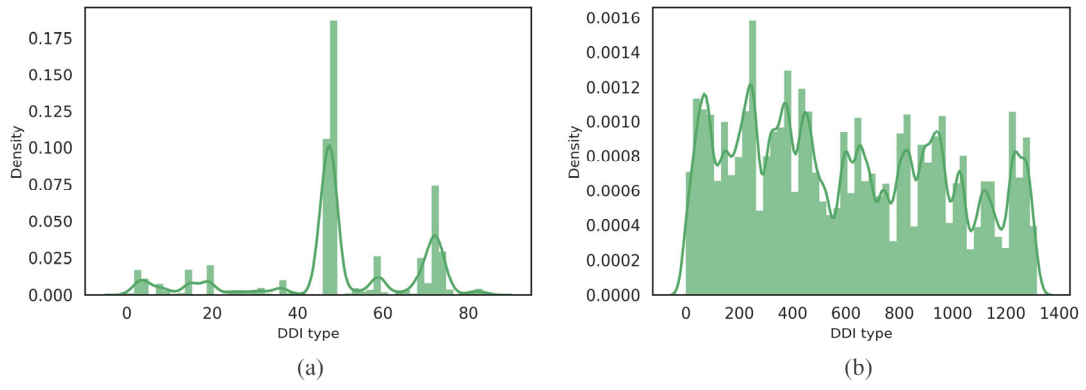


Figure S2. Distribution of DDI types of the (a) DrugBank dataset and (b) TWOSIDES dataset.

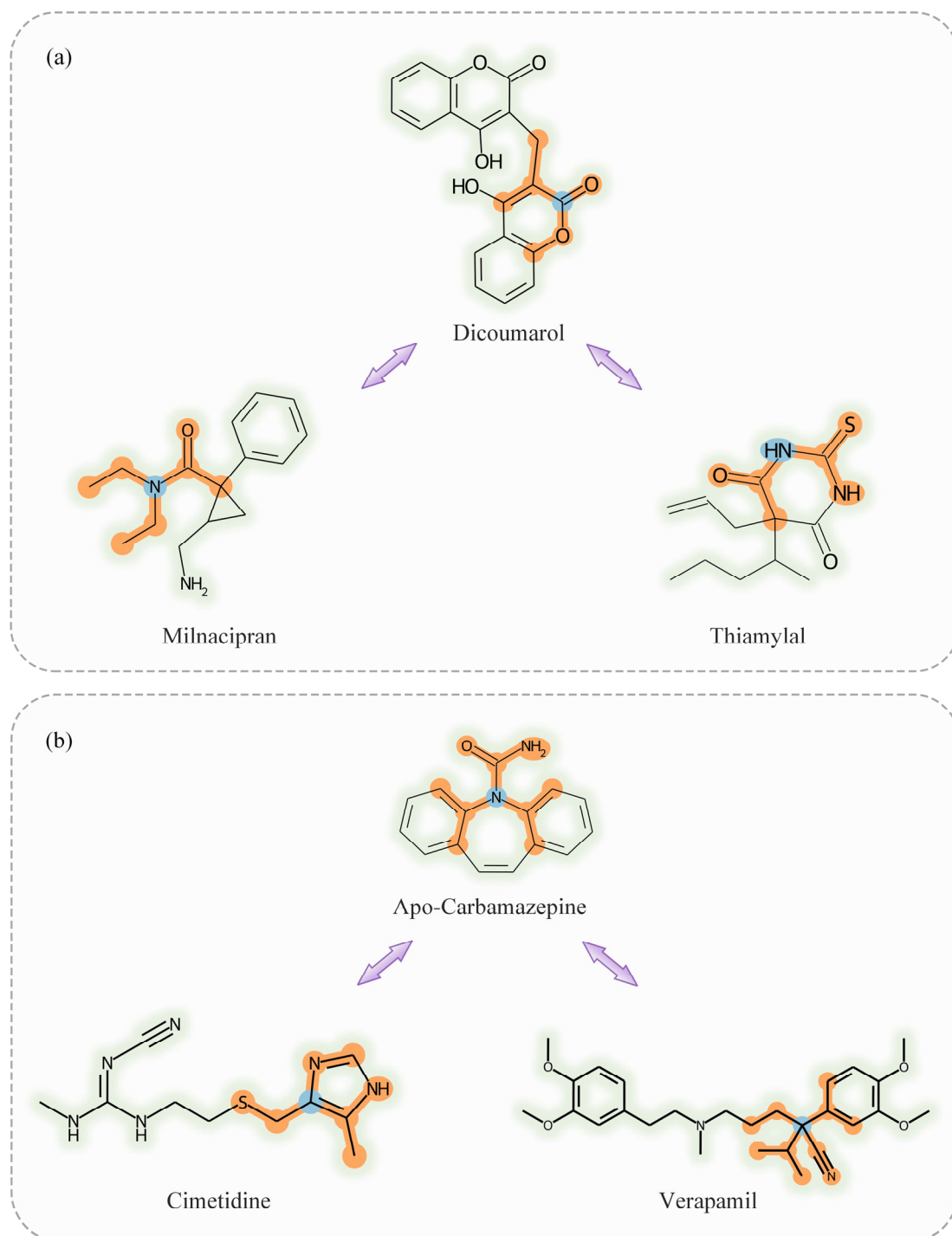


Figure S3. Visualization of the key substructures for DDIs (a) between dicoumarol and milnacipran, thiamylal, and (b) between apo-carbamazepine and cimetidine, verapamil.

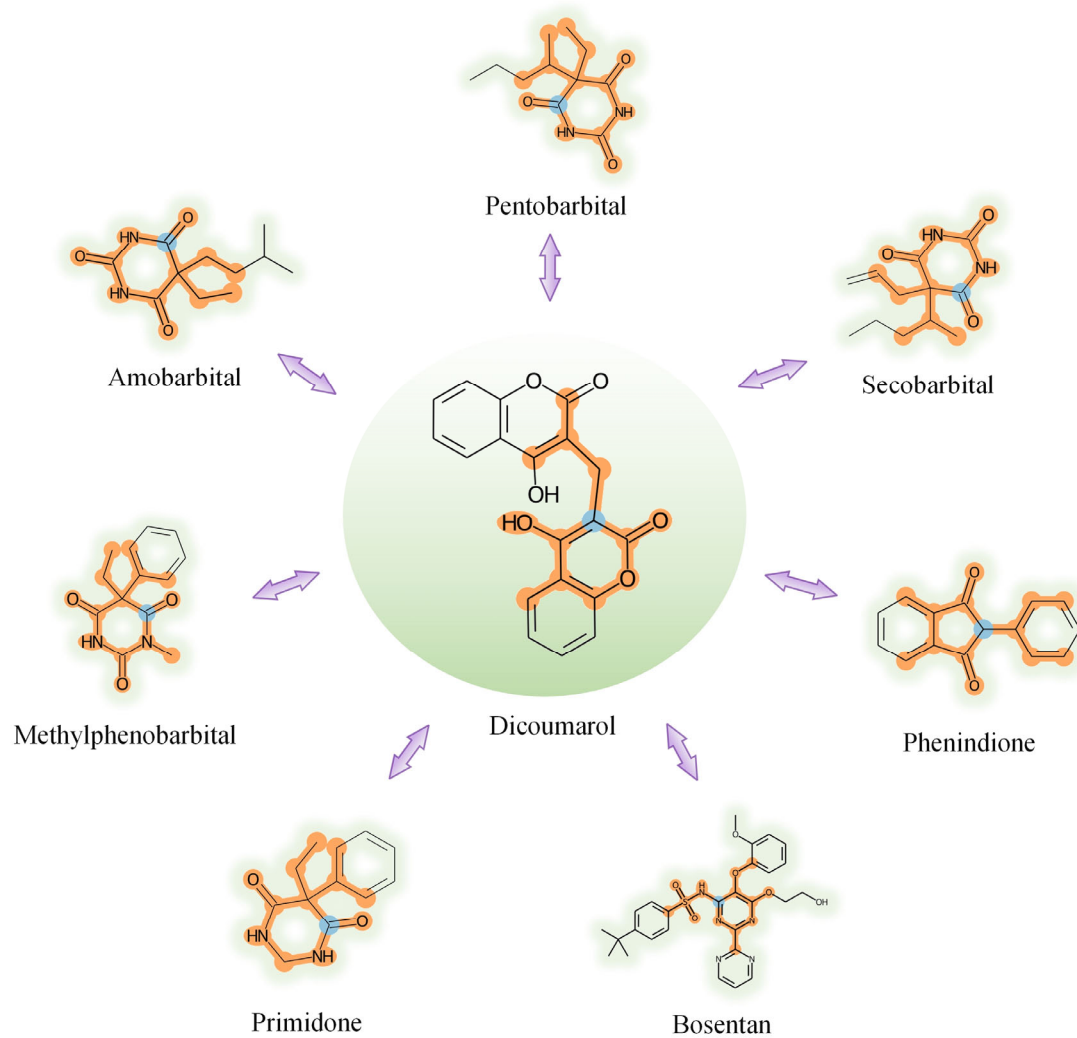


Figure S4. Visualization of the key substructures for DDIs between dicoumarol and the other seven drugs using the cold start setting which removes dicoumarol from the training set. The center of the most important substructure and its receptive field are shown as blue and orange colors respectively.