DrugDAGT: Drug-Drug Interaction Prediction using Dual-Attention Graph Transformer

Paper by Chen et al.

BMC Biology (2024)

2024

Outline

The Drug-Drug Interaction Challenge

Critical Clinical Problem:

- Multi-drug treatments increasing
- Drug-Drug Interactions (DDIs) can result in:
  - Unexpected pharmacological outcomes
  - Adverse drug events
  - Treatment complications

Current Limitations:

- Clinical trials: expensive, time-consuming
- Limited coverage of drug combinations
- Need for predictive approaches

The Molecular Understanding Challenge

Why Molecules Are Complex:

- Multiple interaction points
- Various binding mechanisms
- Complex 3D structures
- Dynamic behaviors

Computational Challenges:

- Capturing local and global structure
- Maintaining chemical validity
- Balancing complexity vs. interpretability

Evolution of DDI Prediction

Historical Approaches:

1. **Rule-Based Systems**
   - Expert-crafted rules
   - Limited coverage
2. **Statistical Methods**
   - Pattern matching
   - Correlation analysis
3. **Machine Learning Era**
   - Feature engineering
   - Basic neural networks

Current State-of-Art:

- Graph Neural Networks
- Attention Mechanisms
- Deep Learning Models

DrugDAGT: A New Approach

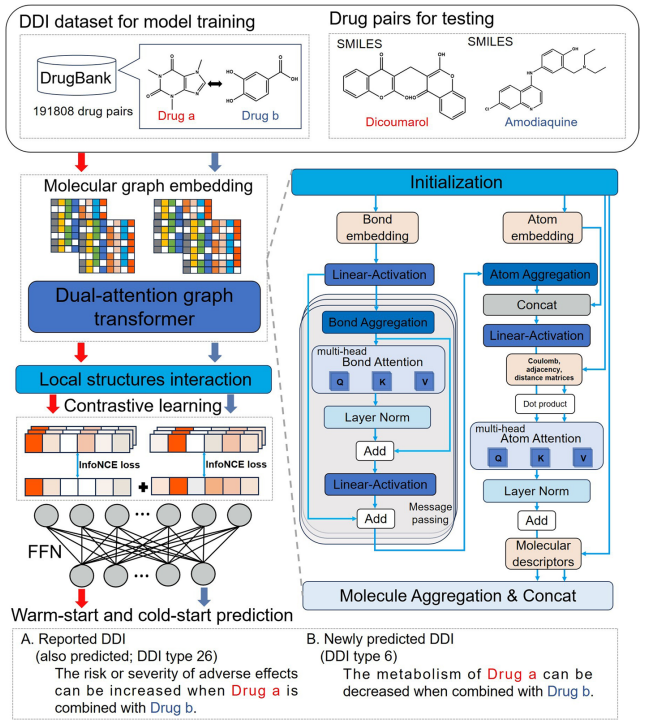


Figure: Complete pipeline of DrugDAGT

Key Innovations:

- Dual-attention mechanism for both local and global views
- Chemical knowledge integration via matrices
- Enhanced learning through contrastive approach

Outline

Why Molecular Structure Matters for DDIs

Drugs interact through their structural components:

- Atoms: Basic units with specific properties
- Bonds: Define molecular shape and flexibility
- Electron distributions: Determine interaction potential

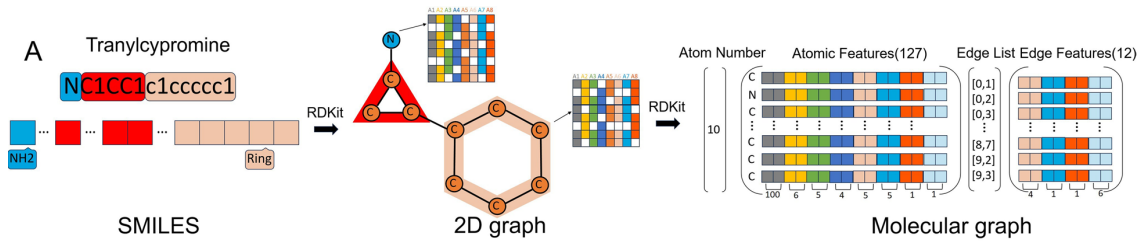
These properties influence:

- Binding site compatibility
- Interaction strength
- Metabolic processing

Molecular Graph Representation

From SMILES to Graphs:

- Nodes: Atoms
- Edges: Bonds
- Features: Chemical properties



Feature Encoding:

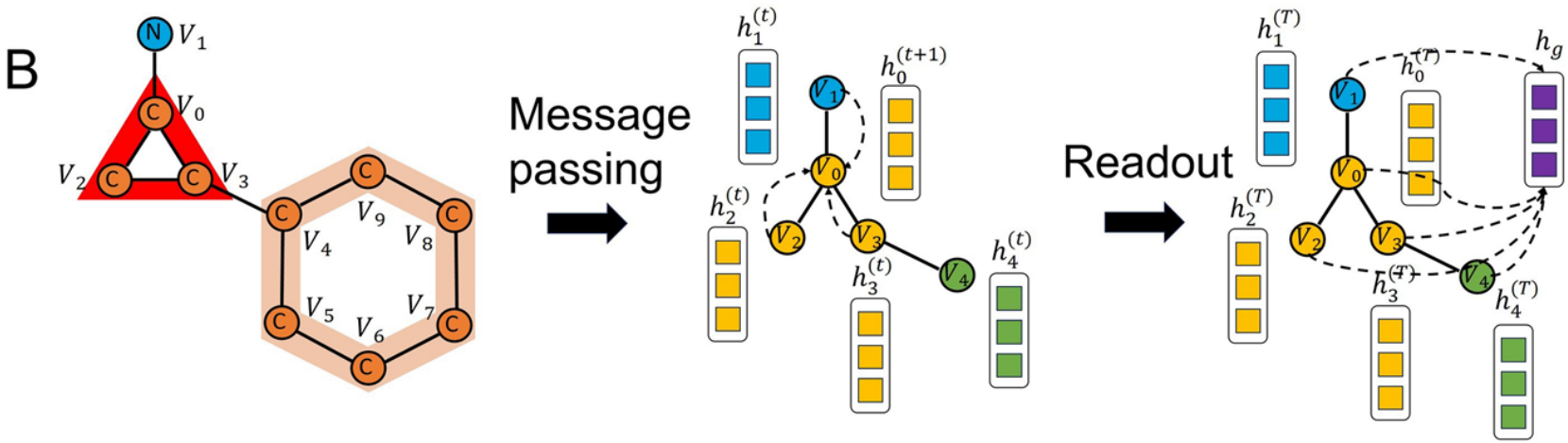
- Atom Features (127D (Charge, Hybridization, ...))
- Bond Features (14D (Bond Type, Ring membership, ...))
- Spatial Information

Message Passing Implementation

Core Algorithm:

---

for each iteration t **do**
  Update bond messages
  Apply bond attention
  Update atom features
  Apply atom attention
end for



Message Passing Details

1. Initialization Phase:

$$h_{ij}^0 = \sigma(W_0(M_i \| E_{ij}))$$

where M_i : atom features, E_{ij} : bond features

2. Message Passing Phase:

Bond Message Update:

$$r_{ij}^t = \sum_{x \in N(i)} h_{xi}^{t-1} - h_{ji}^{t-1}$$

Attention Application:

$$t_{ij}^t = \alpha_{bond}(r_{ij}^t) + r_{ij}^t$$

Hidden State Update:

$$h_{ij}^t = \sigma(h_{ij}^0 + W_t t_{ij}^t)$$

3. Atom-Level Updates:

$$x_i = \sigma(W_T \| (M_i, \sum_{j \in N(i)} h_{ij}^T))$$

$$h_i = \alpha_{atom}(x_i, A_c, A_a, A_d) + x_i$$

Chemical Matrix Integration

Three Key Matrices:

- Adjacency (A): Topology

$$A_{ij} = \begin{cases} 1 & \text{if bonded} \\ 0 & \text{otherwise} \end{cases}$$

- Distance (D): Spatial

$$D_{ij} = \text{spatial\_distance}(i, j)$$

- Coulomb (C): Electronic

$$C_{ij} = \begin{cases} 0.5 Z_i^{2.4} & i = j \\ \frac{Z_i Z_j}{|R_i - R_j|} & i \neq j \end{cases}$$

where:
Z_i is the atomic number corresponding to Cartesian coordinates R_i

Outline

DrugDAGT: Architecture Overview

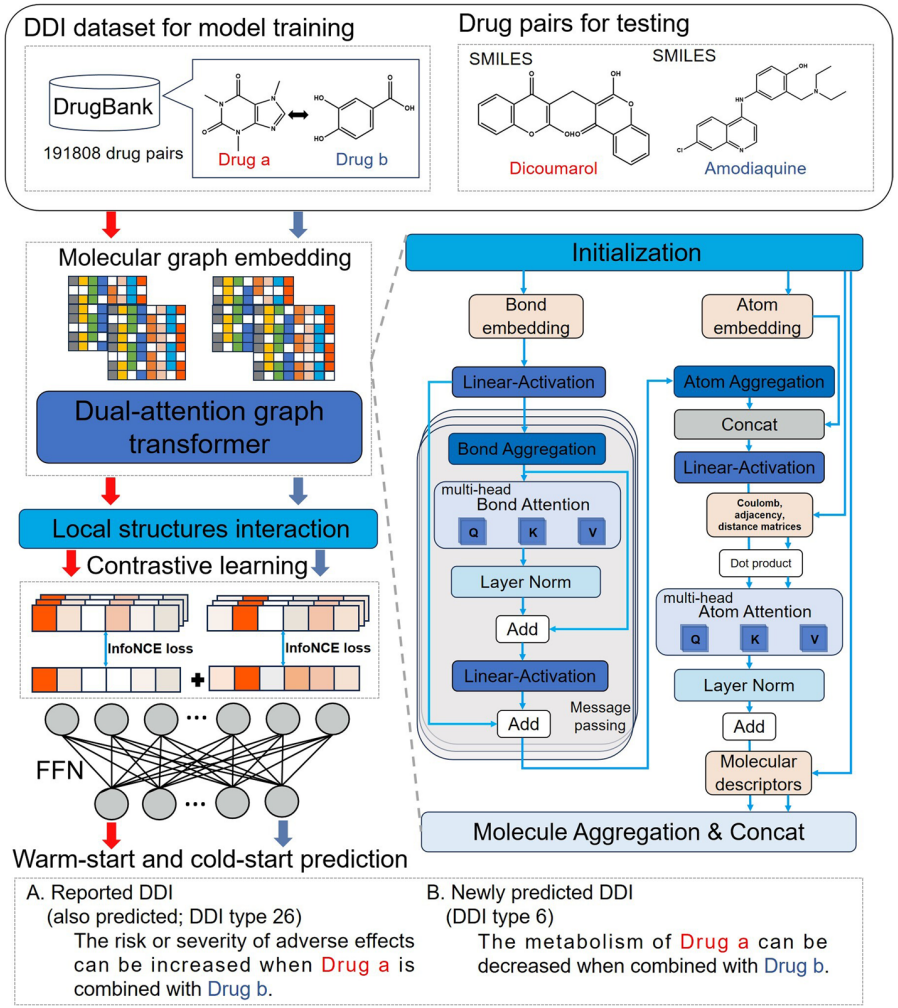


Figure: Complete DrugDAGT Framework

Three Core Components:

- Dual-attention graph transformer
- Chemical matrix integration
- Contrastive learning module

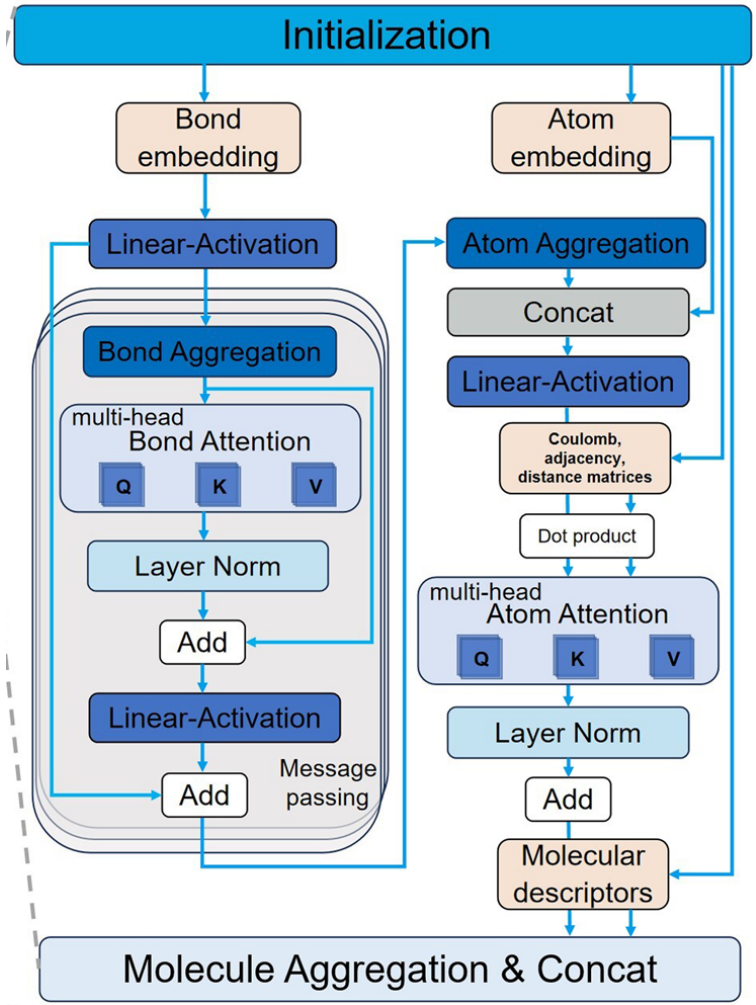
Dual-Attention Innovation

Bond-Level (Local View):

- Fast attention variant
- O(n) complexity
- Chemical environment focus

Atom-Level (Global View):

- Matrix-enhanced attention
- Long-range interactions
- Chemical property awareness



Dual-attention mechanism

Fast Bond Attention

Fast Attention Steps:

1. Global Query Formation:

$$\alpha_{b_k} = \text{softmax}(q_{b_k} w_q / \sqrt{d})$$

$$q_{global} = \sum_k \alpha_{b_k} q_{b_k}$$

2. Key-Value Interaction:

$$p_{b_k} = q_{global} \cdot k_{b_k}$$

$$\gamma_{b_k} = \text{softmax}(p_{b_k} w_{b_k} / \sqrt{d})$$

$$k_{global} = \sum_k \gamma_{b_k} q_{global}$$

3. Final Update:

$$c_{b_k} = g_{b_k} w_v + q_{b_k}$$

$$O_b = \text{LayerNorm}([c_1, ..., c_{2N}])$$

Key Benefits:

- Avoids quadratic complexity via global aggregation
- Maintains chemical meaning through residual connections
- Captures both local and global bond interactions
- Enables stable training through normalization

Matrix-Enhanced Atom Attention

Integration of Chemical Knowledge:

- Each attention head specializes
- Combines different chemical aspects
- Maintains interpretability

Head Specialization:

Heads 0,1: $A_{total} = \frac{QK^T}{\sqrt{d_k}} + f_s A_{adj}$

Heads 2,3: $A_{total} = \frac{QK^T}{\sqrt{d_k}} + f_s D_{dist}$

Heads 4,5: $A_{total} = \frac{QK^T}{\sqrt{d_k}} + f_s C_{coulomb}$

where $f_s = 0.45$ is the scaling factor

Benefits:

- Captures multiple interaction types
- Balances different chemical aspects
- Provides interpretable attention patterns

Multi-Class DDI Prediction Architecture

Problem Setup:

- 86 DDI types classification
- Multi-label prediction
- Balanced training strategy

Model Output Layer:

$$p(y|d_1, d_2) = \text{softmax}(\text{FFN}([h_{d1} \| h_{d2}]))$$

where:

- h_{di}: Drug representations
- FFN: 2-layer network ($800 \rightarrow 86$)
- Output: 86-dim probability vector

Loss Function:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + 0.05 \cdot \mathcal{L}_{contrast}$$

$$\mathcal{L}_{CE} = -\sum_{c=1}^{86} y_c \log(p_c)$$

Contrastive Learning Enhancement

Concept:

- Generate different views
- Maximize similarity of same drug
- Minimize similarity between different drugs

Implementation:

$$h' = h + 0.1 \cdot \text{sign}(h) \odot \frac{\epsilon}{|\epsilon|}$$

where $\epsilon \sim \mathcal{N}(0, I)$

Loss Function:

$$\mathcal{L}_{contrast} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_k \mathbf{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)}$$

Critical Details:

- Temperature $\tau = 0.1$
- Normalized features
- Batch-wise negative sampling
- Memory-efficient implementation

Outline

Implementation Overview

Architecture Details:

- Hidden size: 800
- Message passing steps: 5
- Attention heads: 6
- Head dimension: 133
- FFN layers: [800*2 $\rightarrow$ 800 $\rightarrow$ 86]

Training Configuration:

- Batch size: 200
- Learning rate schedule: NoamLR
- Gradient clipping: Yes
- Early stopping: AUPR

Data Statistics:

- 1706 unique drugs
- 191,808 DDI pairs
- 86 interaction types
- Train/Val/Test: 8:1:1

Critical Implementation Decisions

- **Memory Management:**
  - Fast attention for O(n) complexity
  - Batch size optimization
  - Efficient matrix operations
- **Chemical Validity:**
  - Sign-preserving noise
  - Matrix normalization
  - Structure-aware pooling
- **Numerical Stability:**
  - Layer normalization
  - Gradient clipping
  - Careful scaling factors

Outline

Performance Overview

Method	AUPR	F1	PRE	REC	AUC
GMPNN-CS	0.8347	0.7854	0.799	0.7915	0.9928
Molormer	0.8634	0.8113	0.8157	0.833	0.9941
SA-DDI	0.8693	0.8257	0.8408	0.8297	0.9959
DrugDAGT	**0.8959**	**0.8807**	**0.8941**	**0.8857**	**0.9975**

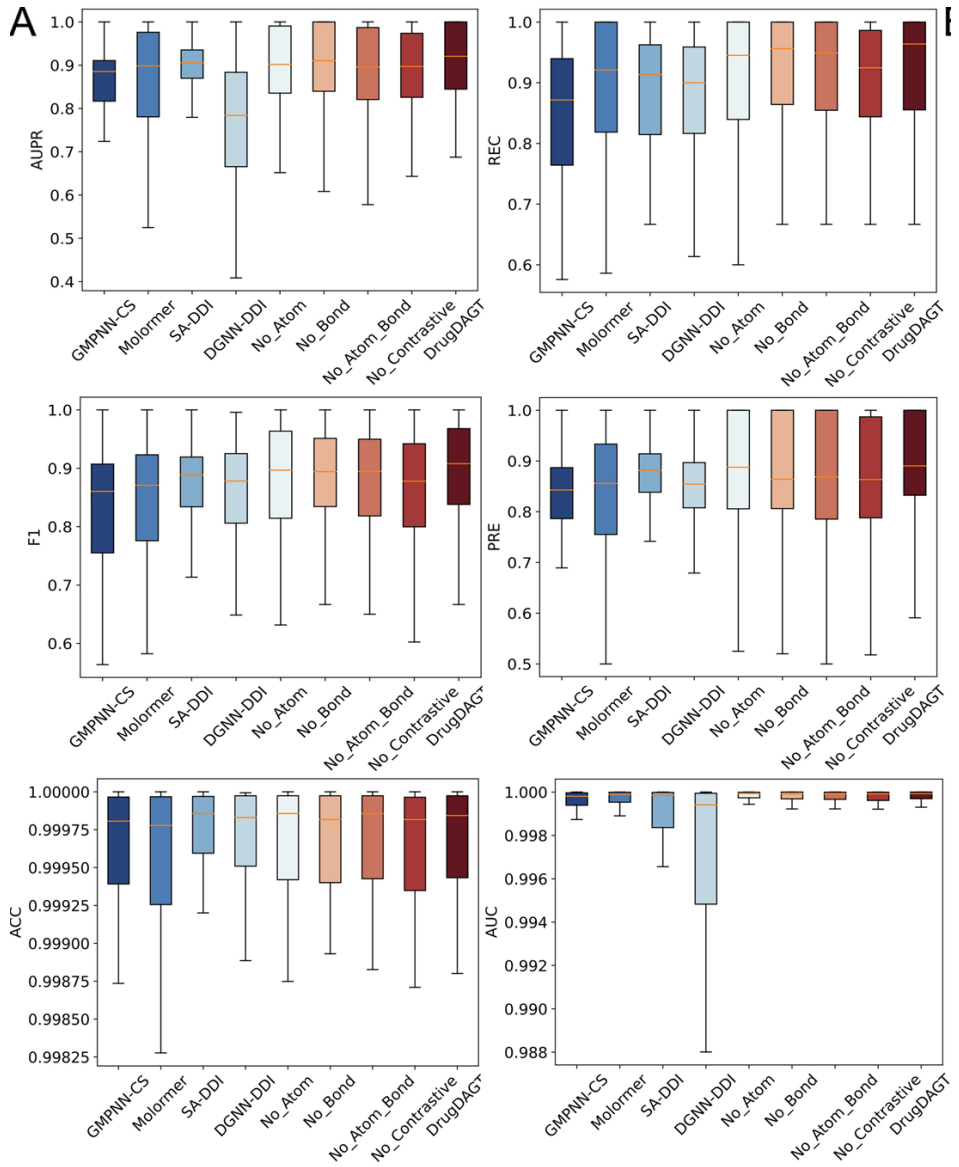
Table: Comparison with state-of-the-art methods

- Consistent improvement across all metrics
- Strong performance in both warm-start and cold-start scenarios
- Robust across different DDI types

Ablation Analysis

Component Contributions:

- Bond attention: +2.1
- Atom attention: +1.8
- Chemical matrices: +1.6
- Contrastive learning: +1.4



Impact of different components

Attention Analysis and Interpretation

Visualization Process:

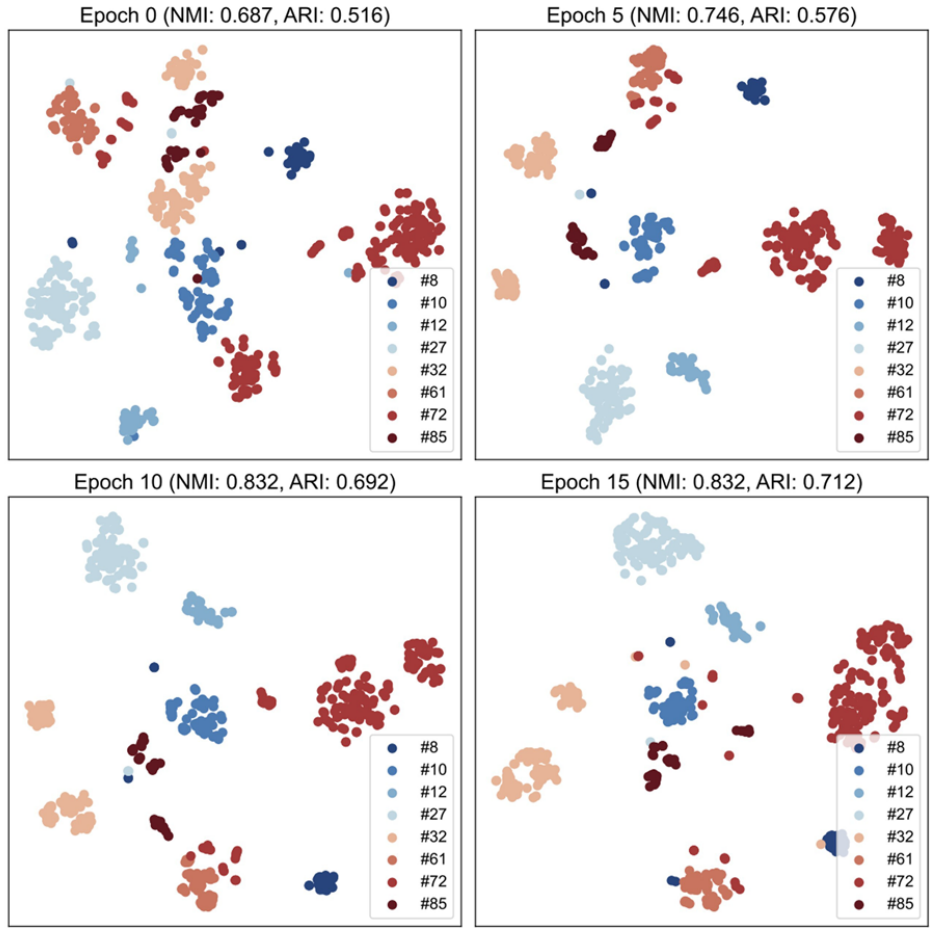
1. Extract attention weights
2. Normalize per atom/bond
3. Map to molecular structure

Interpretation Metrics:

- NMI: Cluster quality
- ARI: Clustering accuracy

Key Findings:

- Functional group focus
- Interaction site detection
- Chemical validity verification



t-SNE embedding plot

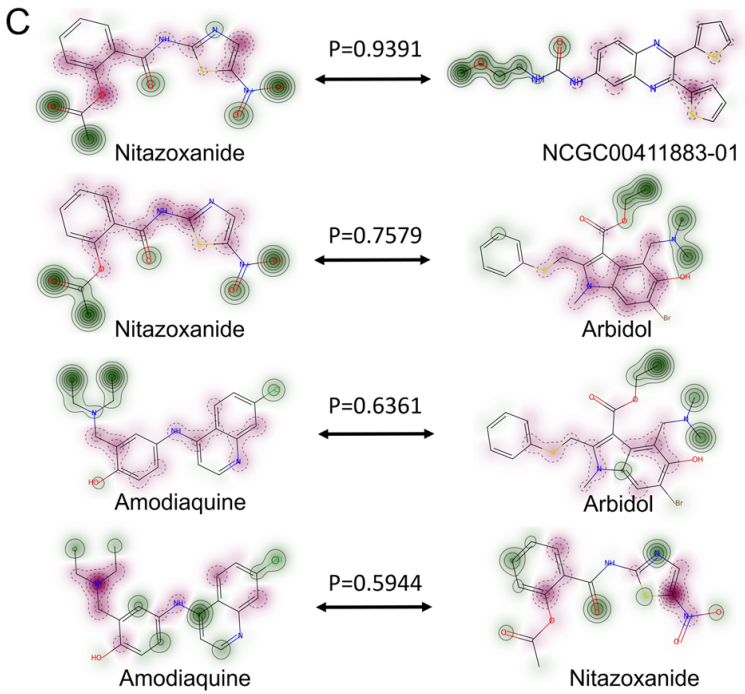
Case Study: COVID-19 Drug Combinations

Dataset:

- 73 drug interactions
- 12 validated synergistic pairs
- Focus on Nitazoxanide combinations

Key Findings:

- Identified known synergies
- Predicted new combinations
- Matched experimental validation



COVID-19 analysis

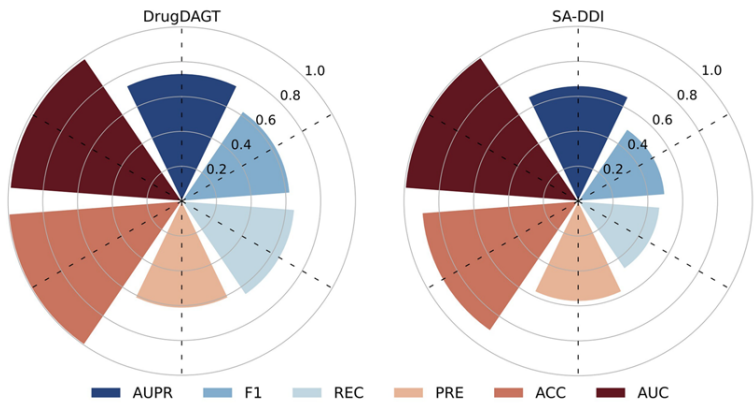
Impact & Future Directions

Current Impact:

- State-of-the-art DDI prediction
- Interpretable insights
- Practical clinical relevance

Future Work:

- 3D structural information
- Multi-drug combinations
- Temporal interaction effects



Performance summary

Questions?

Thank You for Your Attention

Code:
`github.com/codejiajia/DrugDAGT`

Outline

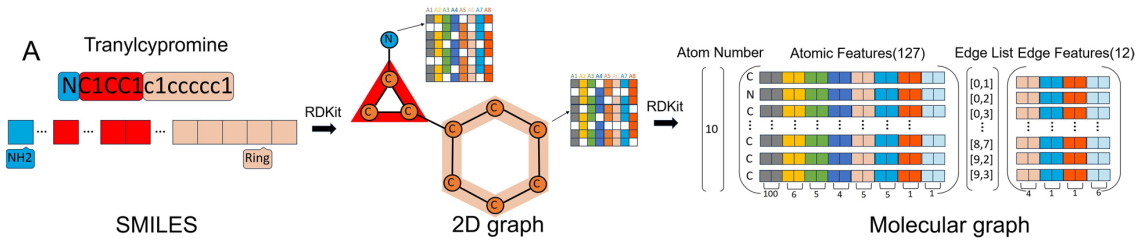
Feature Engineering Details

Atom Features (127D):

- Atomic number (1-100)
- Degree (0-5)
- Formal charge (-2 to +2)
- Hybridization state
- Aromaticity
- H-bond features

Bond Features (12D):

- Bond type
- Conjugation
- Ring membership
- Stereochemistry



Cold-Start Performance

Scenario Definition:

- New drugs not in training
- No historical interaction data
- Structure-based prediction only

Results:

- AUPR: 0.8247
- F1: 0.7954
- Comparison with baselines

Error Analysis and Limitations

Common Error Types:

- Rare interaction mechanisms
- Complex structural patterns
- Limited training examples

Current Limitations:

- 2D structure only
- Static representation
- Binary interaction prediction