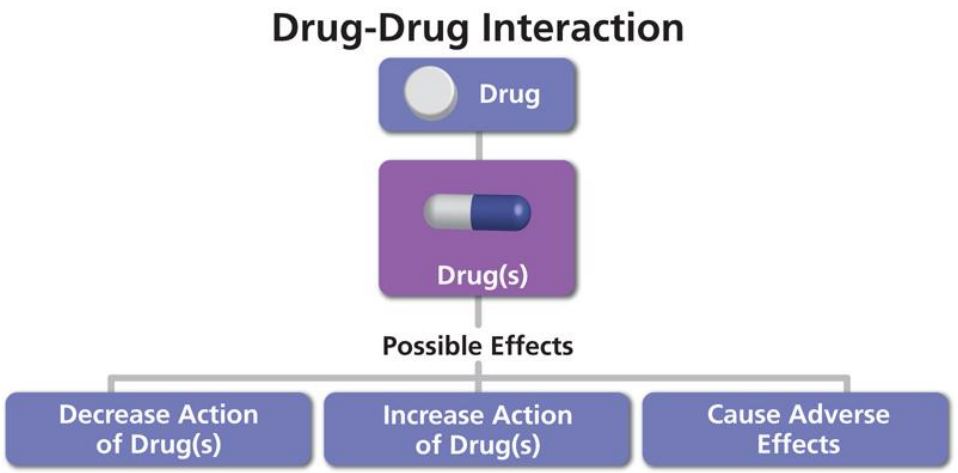
NEURAL INFORMATION PROCESSING SYSTEMS

Customized Subgraph Selection and Encoding for Drug-drug Interaction Prediction

Haotong Du, Quanming Yao, Juzheng Zhang, Yang Liu, Zhen Wang

NeurIPS 2024

Background – Drug-drug Interaction



Drug combinations are very common in clinical therapy

side effect

S. No.	Drugs Interaction Combination	Frequency	Outcome
1	Ceftriaxone + Calcium Gluconate	6	Precipitation of ceftriaxone-calcium salt
2	Furosemide + Amikacin	5	Potentiate the risk of oto- and nephrotoxicity
3	Atracurium + Amikacin	3	Severe and/or prolonged respiratory depression
4	Omeprazole + Clopidogril	2	Decreased effectiveness of Clopidogril
5	Aspirin + Clopidogril	2	Increased platelet inhibition effect

5 Common Drug Combinations and Consequences of Combinations

- 15% of the U.S. population is affected by unwanted side effects.
- About 0.32% (about 100,000 people) of hospitalised patients were exposed to life-threatening drug side effects.

- Annual costs in treating side effects exceed $177 billion in the U.S. alone.

——Source: U.S. Food and Drug Administration[1]

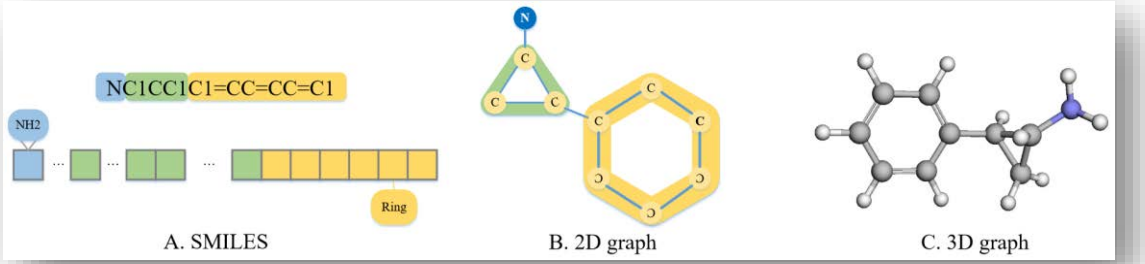
[1] https://www.fda.gov/drugs/drug-interactions-labeling/preventable-adverse-drug-reactions-focus-drug-interactions

Background – Drug-drug Interaction

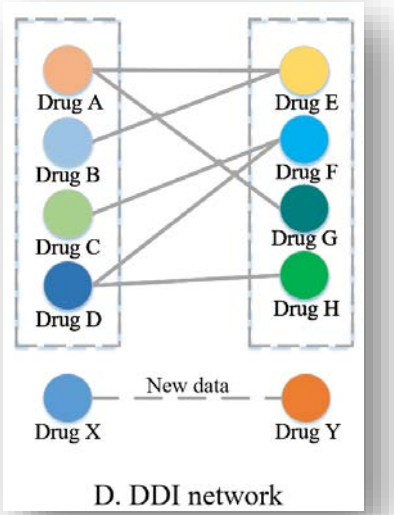
- Experimental methods
  - Drug pharmacokinetics parameters
  - Drug metabolism information (e.g., CYP enzymes)
- Disadvantages
  - Labor-intensive and time-consuming
  - Not scalable and often low throughput

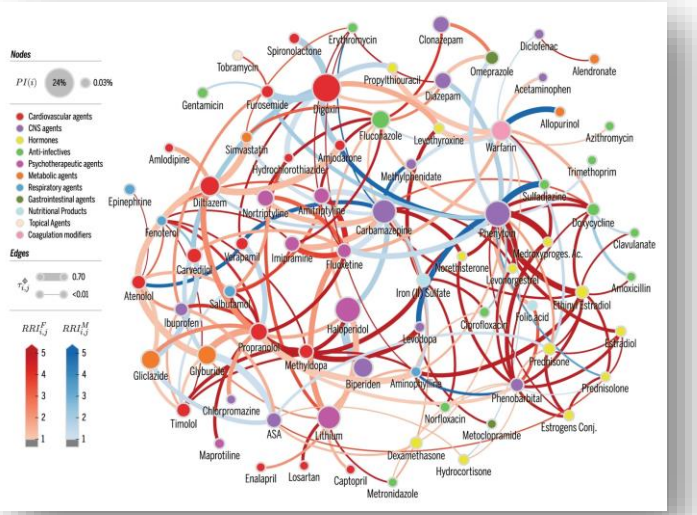
Background – Drug-drug Interaction

- Computational methods
  - Save the cost of biological experiments
  - Provide relevant guidance for combination therapy to some extent



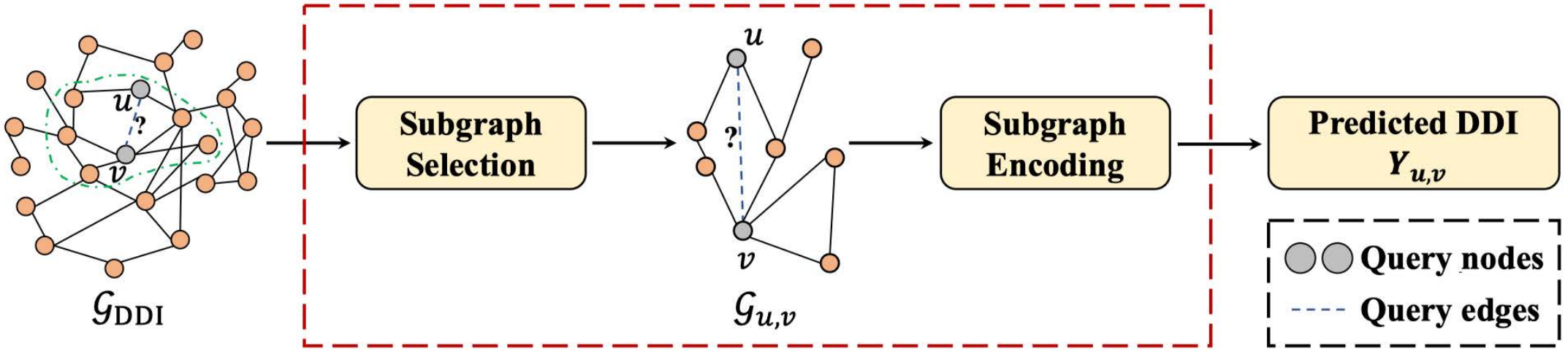
- Network-based methods(our focus)
  - Cast the prediction as a link prediction problem on DDI graph
  - Subgraph-based method is SOTA.





Background – Drug-drug Interaction

- Subgraph-based method

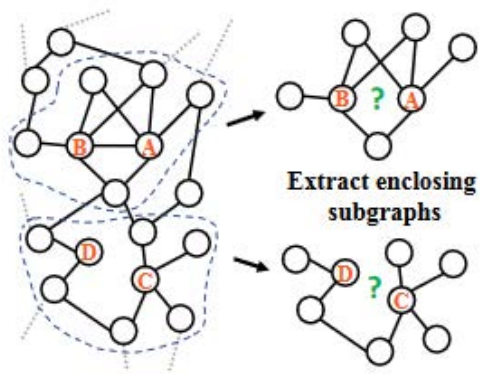


1. Select local subgraph as support context
2. Encode the local subgraph to reasoning

Motivation

- Subgraph selection
  - Existing methods sample subgraphs using a fixed subgraph range, which may lead to an imprecise collection of evidence for interaction reasoning

- Subgraph encoding
  - Manually designed encoding functions are limited in their ability to accommodate both types of distinct semantic patterns simultaneously

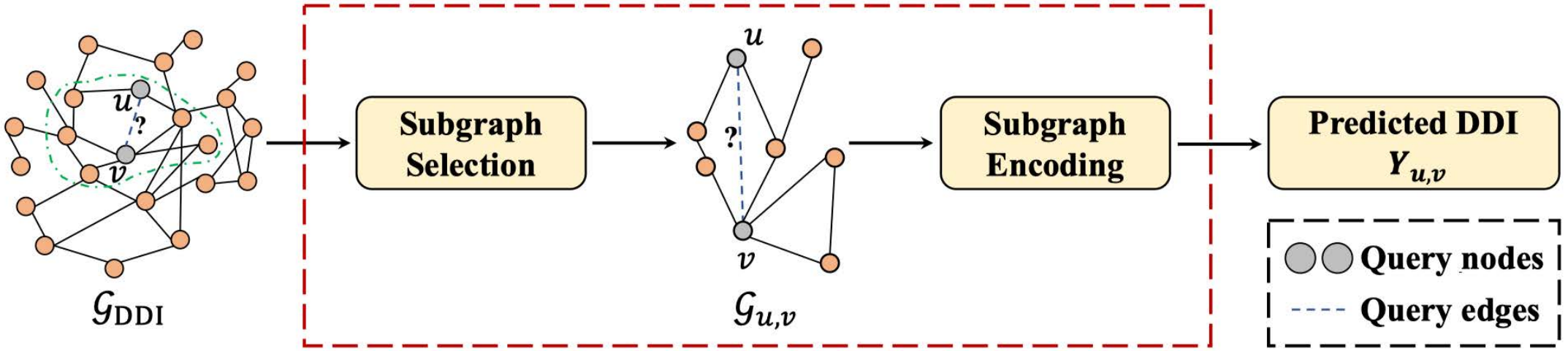


Diverse semantic properties in drug-drug interactions.

Dataset	Interaction Type	Examples	Semantic Property
DrugBank	Metabolic levels-based	#Drug1 may decrease the excretion rate of #Drug2	asymmetry $(r(x,y) \nRightarrow r(y,x))$
TWOSIDES	Phenotype-based	Combination of #Drug 1 and #Drug 2 may cause headaches	symmetry $(r(x,y) \Rightarrow r(y,x))$

Motivation

- Subgraph-based method



Customization by NAS

1. Select local subgraph as support context
2. Encode the local subgraph to reasoning

→ Select more informative subgraph

→ Design more expressive encoding function

Subgraph Selection Space

- We define a drug-pair subgraph selection space containing a range of subgraphs of different sizes for a given query (u, v):

$$\mathcal{S}_{u,v} = \{\mathcal{G}_{u,v}^{i,j} \mid 1 \leq i, j \leq \eta\}$$

  - where $\mathcal{G}_{u,v}^{i,j}$ is generated by taking the union of the i-hop ego-network of node u and the j-hop ego-network of node v, i.e., $\mathcal{G}_{u,v,}^{i,j} = \{z \in \mathcal{V} | z \in (u \cup \mathcal{N}_i(u) \cup v \cup \mathcal{N}_j(v))\}$
  - the threshold η constrains the maximum subgraph range
- Since each drug-pair has a specific subgraph selection space, the overall size of space in a whole graph is $\eta^{2|\mathcal{E}|}$, where $|\mathcal{E}|$ represents the number of edges in a drug interaction network

Subgraph Encoding Space

- We adopt a unified message passing framework comprising several key modules: the message-computing function MES, the aggregation function AGG, the combination function COM, and the activation function ACT, as follows:

$$\texttt{step 1: } \mathbf{m}_u \leftarrow \texttt{AGG}(\texttt{MES}(\mathbf{h}_v, \mathbf{h}_{r(u,v)})_{v \in \mathcal{N}_1(u)})$$

$$\texttt{step 2: } \mathbf{h}_u \leftarrow \texttt{ACT}(\texttt{COM}(\mathbf{h}_u, \mathbf{m}_u))$$

Function name	Operations
Message Computing Function	SUB, MULT, CORR, ROTATE
Aggregation Function	SUM, MAX,MEAN
Combination Function	MLP, CONCAT
Activation Function	RELU, TANH, IDENTITY

Customized Search Problem

- Based on the well-designed search space described above, we formulate a bi-level optimization problem to adaptively search for the optimal configuration of subgraph-based pipelines:

Definition 1 (Customized Subgraph-based Pipeline Search Problem). *Let $\mathcal{A}$ denote the subgraph encoding space, $\mathcal{S}_{u,v}$ represent the subgraph selection space for the query (u, v), $\boldsymbol{\alpha}$ be a candidate encoding function in $\mathcal{A}$, $\mathbf{W}$ represent the parameters of a model from the search space, and $\mathbf{W}^*(\mathcal{G}_{u,v}; \boldsymbol{\alpha})$ denote the trained operation parameters. Let $\mathcal{D}_{\text{tra}}$ and $\mathcal{D}_{\text{val}}$ denote the training and validation sets, respectively. The search problem is formulated as follows:*

$$\arg\max_{\boldsymbol{\alpha}\in\mathcal{A},\mathcal{G}_{u,v}\in\mathcal{S}_{u,v}} \sum_{(u,r,v)\in\mathcal{D}_{\text{val}}} \mathcal{M}(\mathbf{W}^*(\mathcal{G}_{u,v};\boldsymbol{\alpha});\mathcal{G}_{u,v};\boldsymbol{\alpha}),$$

$$s.t.\ \mathbf{W}^*(\mathcal{G}_{u,v};\boldsymbol{\alpha}) = \arg\min_{\mathbf{W}} \sum_{(u,r,v)\in\mathcal{D}_{\text{tra}}} \mathcal{L}(\mathbf{W};\mathcal{G}_{u,v};\boldsymbol{\alpha}),$$

where the classification loss $\mathcal{L}$ is minimized for all interactions, while the performance measurement $\mathcal{M}$ is expected to be maximized.

Customized Search Problem

- Solving the proposed bi-level optimization problem is non-trivial:
  - For the subgraph selection space, the traditional continuous relaxation strategy is not directly applicable due to the structural mismatch between graphs and vectors.
  - To enable searching within the subgraph selection space, we would need to first generate all subgraphs in the space. However, sampling such a large number of subgraphs is computationally intractable.

Subgraph Space Relaxation

- We first utilize encoding function $f(\cdot)$ to encode subgraphs with different scopes, making it feasible to implement a relaxation strategy
- Additionally, inspired by the reparameterization trick, we adopt the Gumbel-Softmax function to facilitate differentiable learning over a discrete space:

$$\hat{\mathbf{z}}_{u,v}^{i,j} = \sum_{1 \le i,j \le \eta} \frac{\exp(\log(g(f(\mathcal{G}_{u,v}^{i,j})) + \mathbf{G}_{i,j})/\tau)}{\sum_{i',j'=1}^{\eta} \exp(\log(g(f(\mathcal{G}_{u,v}^{i',j'})) + \mathbf{G}_{i',j'})/\tau)} f(\mathcal{G}_{u,v}^{i,j})$$

  - where $g(\cdot)$ scores the subgraph representations using multiple linear layers, $\mathbf{G}_{i,j} = -\log(-\log \mathbf{U}_{i,j})$ is the Gumbel random variable, $\mathbf{U}_{i,j}$ is a uniform random variable, and τ is the temperature parameter controlling sharpness. $\hat{z}_{u,v}^{i.j}$ is the mixed selection operation of subgraph $\mathcal{G}_{u,v}^{i,j}$ used to optimize searching process.

Subgraph Approximation Strategy

- Inspired by the k-subtree extractor, we apply an encoding function to the entire graph and use the resulting node representations of u and v as the ego-network representations for these nodes:

$$f(\mathcal{G}_{u,v}^{i,j}) \approx \texttt{CONCAT}(f(\mathcal{G}_{\text{DDI}}, u, i), f(\mathcal{G}_{\text{DDI}}, v, j))$$

  - No need to explicitly sampling subgraphs, improving the efficiency in solving the bi-level optimization problem

Robust Search Algorithm

- Sampling-based NAS paradigm and message-aware partitioned supernet training strategy improve the efficiency, consistency and accuracy of supernet training.

Algorithm 1: The search algorithm of CSSE-DDI.

Input: Supernet $\mathcal{S}$, number of partitions based on message computing function categories M $(M = 4)$, subsupernet $\mathcal{S}_i, (i = 1, \cdots, M)$.

// supernet training phase
Train $\mathcal{S}$ by continuously sampling a single path until convergence;
// supernet partition phase
Partition $\mathcal{S}$ into M sub-supernets $\mathcal{S}_1, \cdots, \mathcal{S}_M$;
// sub-supernet training phase
forall $i = 1, \cdots, M$ **do**
    Initialize $\mathcal{S}_i$ with weights transferred from $\mathcal{S}$;
    Train $\mathcal{S}_i$ by continuously sampling a single path until convergence;
end
// searching phase
Search the optimal encoding function from sub-supernets $\mathcal{S}_1, \cdots, \mathcal{S}_M$ on validation data by natural gradient descent;
Select the optimal subgraphs from sub-supernets $\mathcal{S}_1, \cdots, \mathcal{S}_M$ on validation data by preserving the subgraphs with the largest probabilities;

Comparison with existing works

- Previous methods rely on fixed subgraph selection strategy to sample subgraphs and employ hand-designed functions for encoding, as summarized in Table 1.

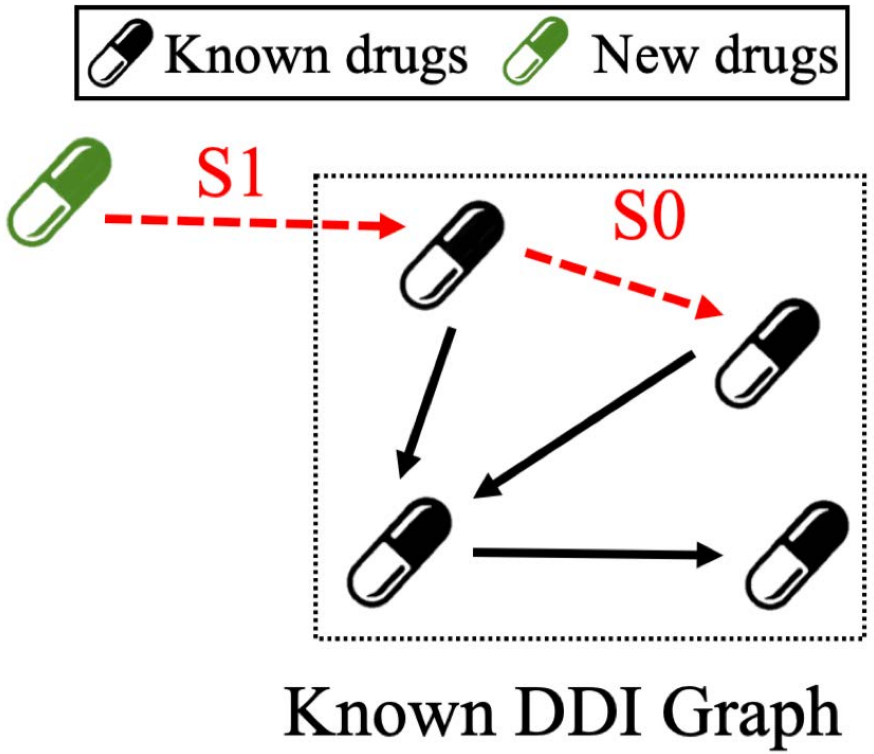
- CSSE-DDI is the first to customize the subgraph selection and encoding processes for subgraph-based DDI prediction

Table 1: Comparing with existing methods."-" represents not applicable.

Method	Fine-grained Subgraph Selection	Data-specific Encoding Function
SEAL [18]	✗	✗
GraIL [19]	✗	✗
SumGNN [12]	✗	✗
SNRI [20]	✗	✗
KnowDDI [14]	✓	-
MR-GNAS [21]	-	✓
AutoGEL [22]	-	✓
CSSE-DDI	✓	✓

Experimental Setup

- Dataset: DrugBank and TWOSIDES
- We examine two DDI prediction task settings: S0 and S1.
  - S0 setting: both drug nodes are present in the known DDI graph
  - S1 setting: involves a pair where one drug is known and the other is a novel drug not represented in the known DDI graph



Main Results in S0 setting

- Compared with GNN-based methods, subgraph-based methods, and NAS-based methods

- CSSE-DDI consistently outperforms all baselines, demonstrating its effectiveness in searching for data-specific subgraph-based pipelines

Model Type	Dataset	Dataset 1: DrugBank			Dataset 2: TWOSIDES		
	Task Type	Multi-class			Multi-label		
	Methods	F1 Score	Accuracy	Cohen's κ	ROC-AUC	PR-AUC	AP@50
GNN-based	Decagon	57.35±0.26	87.19±0.28	86.07±0.08	91.72±0.04	90.60±0.12	82.06±0.45
	GAT	33.49±0.36	77.18±0.15	74.20±0.23	91.18±0.14	89.86±0.05	82.80±0.17
	SkipGNN	59.66±0.26	85.83±0.18	84.20±0.16	92.04±0.08	90.90±0.10	84.25±0.25
	CompGCN	71.20±0.70	88.30±0.29	86.15±0.35	93.00±0.07	91.26±0.07	86.18±0.10
	ACDGNN	86.24±0.93	90.53±0.38	87.81±0.33	93.69±0.47	92.12±0.21	87.45±0.24
	TransFOL	89.97±1.64	91.92±0.89	90.92±0.72	94.16±0.62	93.52±0.53	88.13±0.39
Subgraph-based	SEAL	48.82±0.98	76.61±0.26	71.91±0.59	90.74±0.22	90.11±0.17	84.13±0.13
	GraIL	73.20±0.69	85.40±0.39	82.70±0.47	92.93±0.10	91.69±0.14	87.43±0.09
	SumGNN	78.35±0.51	89.05±0.36	87.28±0.08	92.62±0.04	90.80±0.40	85.75±0.10
	SNRI	85.57±0.32	90.15±0.21	88.94±0.36	93.12±0.18	92.64±0.12	87.53±0.11
	KnowDDI	90.06±0.27	93.15±0.19	91.87±0.21	95.05±0.06	93.75±0.05	89.24±0.06
	LaGAT	81.63±0.56	86.21±0.18	85.38±0.23	89.78±0.21	86.33±0.15	83.75±0.36
NAS-based	MR-GNAS	74.24±0.45	88.17±0.24	87.31±0.11	93.85±0.07	91.80±0.03	87.16±0.05
	AutoGEL	76.87±0.63	89.35±0.59	86.14±0.41	94.11±0.32	92.35±0.29	88.13±0.41
	CSSE-DDI-FS	86.31±0.36	91.08±0.21	89.17±0.27	94.35±0.07	93.01±0.06	89.08±0.04
	CSSE-DDI-FF	80.96±0.65	90.27±0.23	88.69±0.31	94.26±0.08	92.74±0.06	88.91±0.09
	CSSE-DDI	**92.08**±0.22	**95.56**±0.15	**94.72**±0.26	**95.47**±0.02	**94.21**±0.05	**89.76**±0.05

Main Results in S1 setting

- A significant performance drop from the transductive setting (S0) to the inductive setting (S1) demonstrates that DDI prediction for new drugs is more challenging.

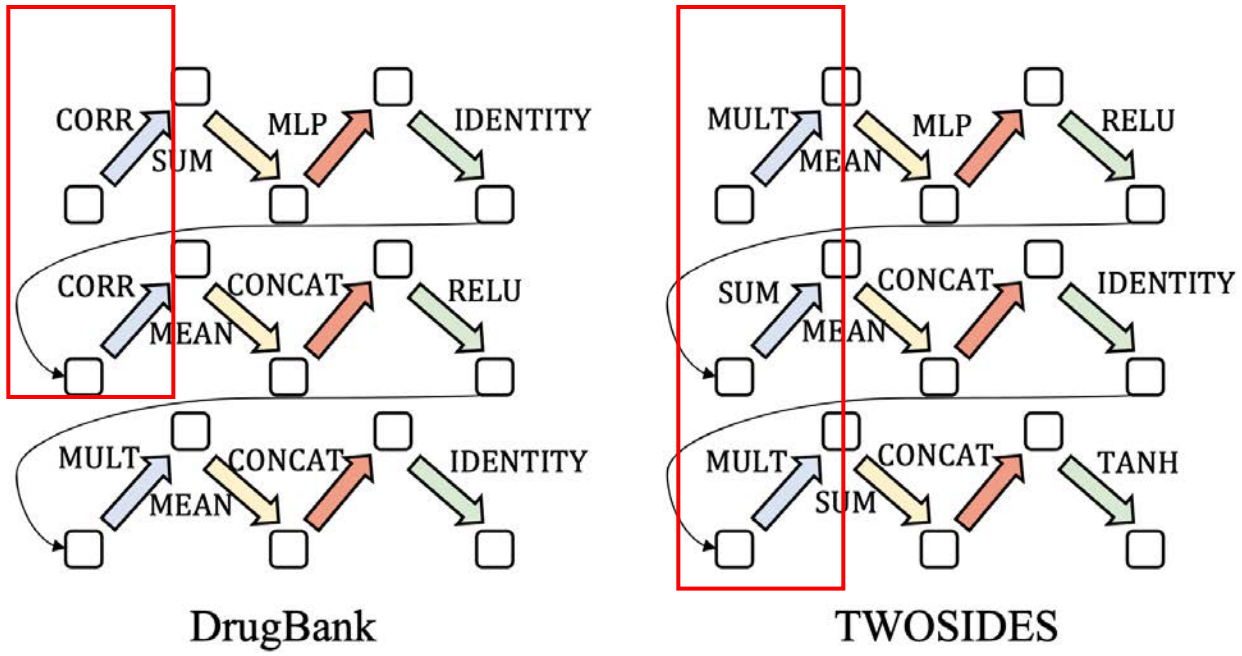
- CSSE-DDI still demonstrates impressive results, outperforming existing GNN-based and subgraph-based methods.

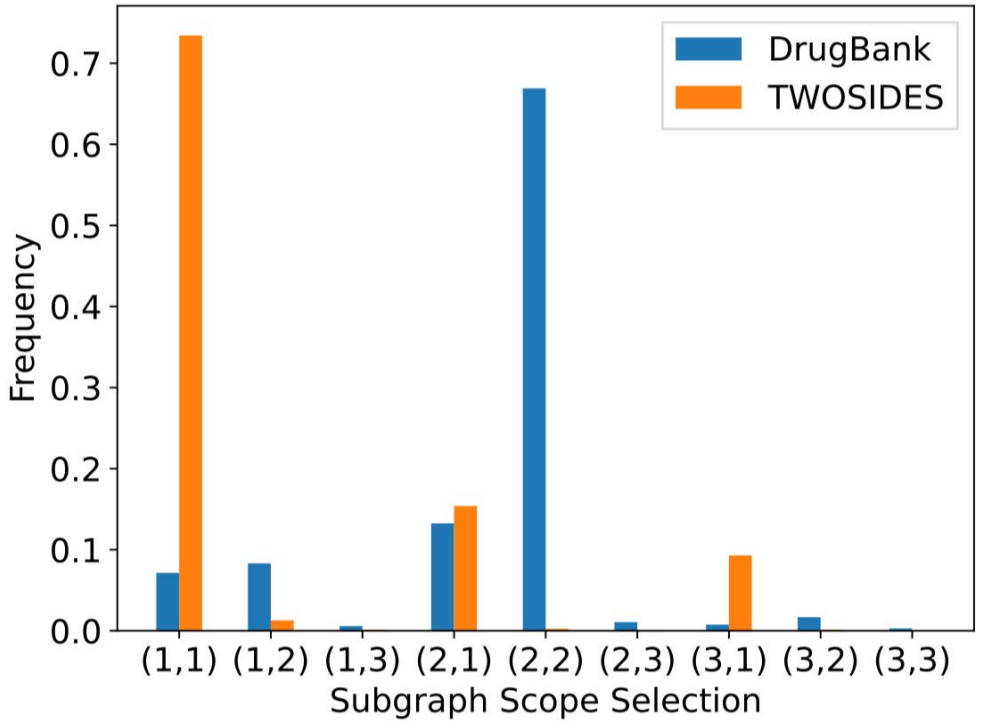
- This strong performance is largely due to the robust learning capability of NAS technology in handling unknown data.

Dataset	**Dataset 1: DrugBank**			**Dataset 2: TWOSIDES**		
Task Type	Multi-class			Multi-label		
Methods	F1 Score	Accuracy	Cohen's κ	ROC-AUC	PR-AUC	Accuracy
CompGCN	30.98±3.26	52.76±0.46	37.87±1.28	84.83±1.02	83.68±1.86	74.64±0.79
Decagon	11.39±0.79	32.56±0.92	20.29±1.33	57.49±1.75	59.38±1.09	52.27±1.48
SumGNN	26.57±1.59	44.30±1.04	40.24±1.26	80.02±2.17	78.42±1.62	69.81±1.77
KnowDDI	31.14±1.24	53.44±1.73	43.93±1.17	84.23±2.63	82.58±1.94	74.72±1.51
EmerGNN	**58.13±1.36**	**69.53±1.97**	**62.19±1.62**	87.42±0.39	86.20±0.71	79.23±0.54
CSSE-DDI	37.24±1.13	58.57±0.85	49.97±1.01	**88.33±0.52**	**86.47±0.27**	**80.01±0.39**

Visualization of Searched Results

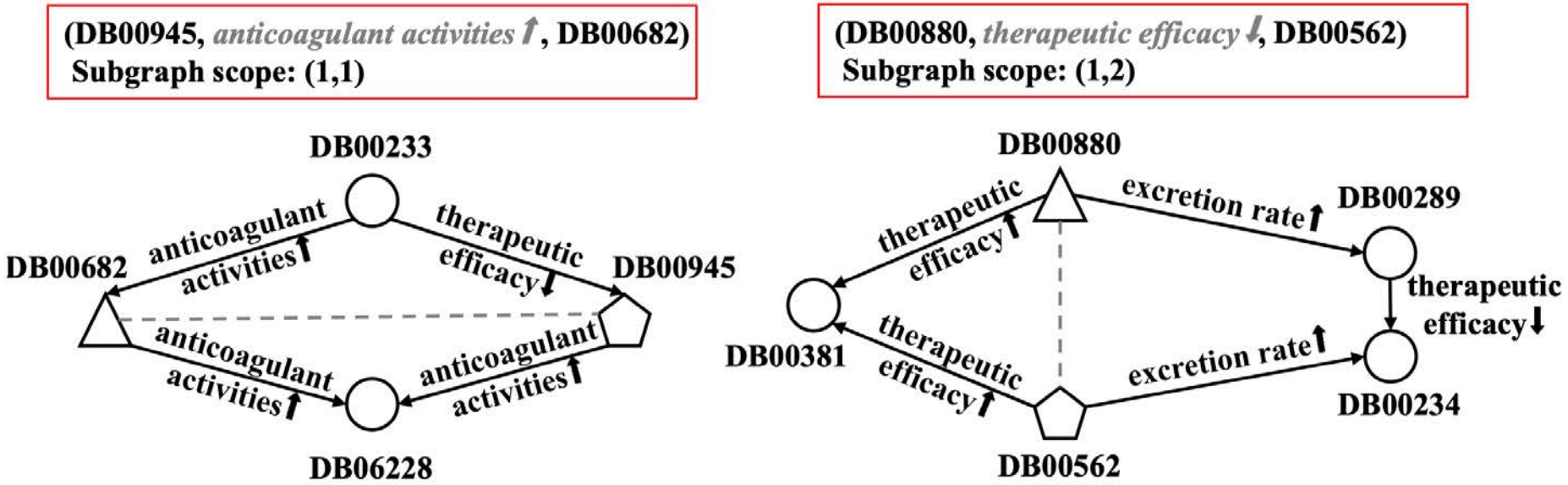
- Data-specific subgraph encoding functions are obtained.
- CSSE-DDI can effectively learn different subgraph scope distributions for various datasets.





Case Study

- CSSE-DDI can identify distinctive subgraphs containing semantic information to support inference for different queries, revealing pharmacokinetic and metabolic relationships



Summary

- We present CSSE-DDI, a searchable framework for DDI prediction that adaptively customizes the subgraph selection and encoding processes
  - Refined search spaces to enable fine-grained subgraph selection and data-specific encoding function optimization.
  - A relaxation mechanism that uses an approximation strategy to efficiently explore optimal subgraph configurations.

- The search results generated by CSSE-DDI offer interpretability in the context of drug interactions, revealing domain-specific concepts such as pharmacokinetics and metabolism.

Thanks for your listening!