Learning from Graphs

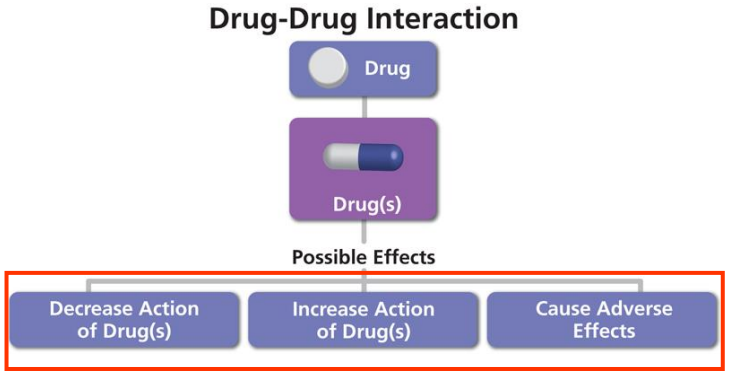
- Part 2: Applications in AI4Science

Quanming Yao

Assistant Prof. EE. Tsinghua

qyaoaa@tsinghua.edu.cn

Drug Interaction Prediction Tasks



Clinical drug combinations are common

Adverse reaction

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

Five major drug interaction combinations and their outcomes

Serious adverse drug reactions occurred in 6.7% of hospitalized patients, and the fatality rate was 0.32%.

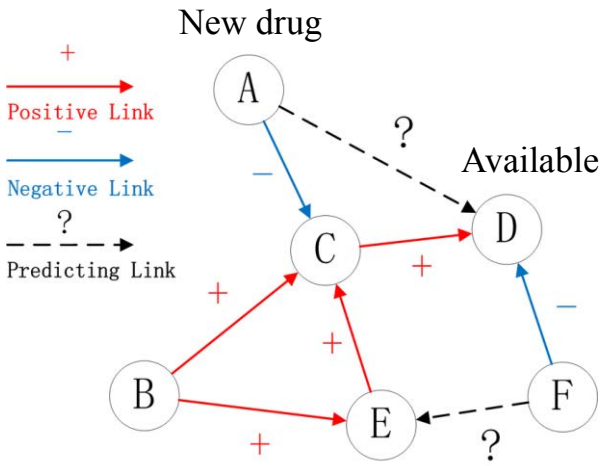
——FDA

Data sparsity V.S. Data hunger

↑ DDI Prediction ↑ Deep Learning

These problems are more serious for emerging drugs

Graph Learning Perspective



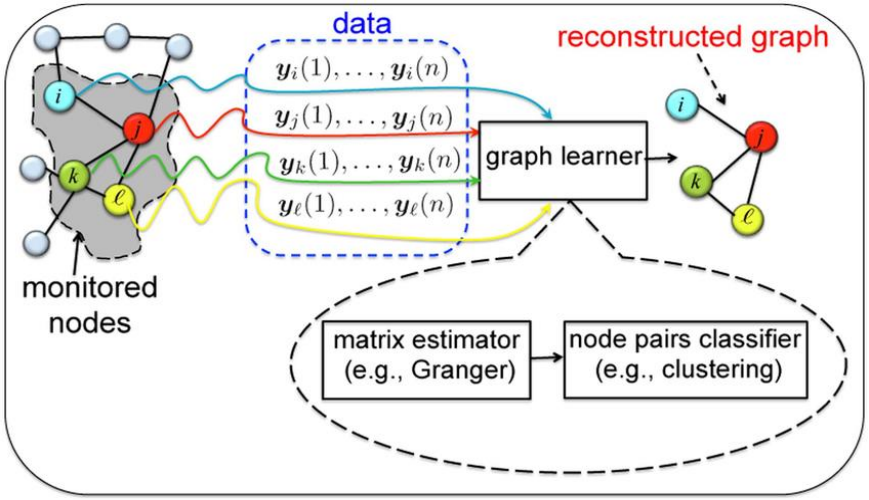
Problem definition

- Input:
  - Fingerprint characteristics of a given emerging drug u (f_u)
  - drug interactions (DDI) network
  - Biomedical network
- Output:
  - The type of interaction between a given emerging drug and a given existing drug
  - (New drugs,? , drugs)

Predictions on the links of the biomedical network (BN)

Convert

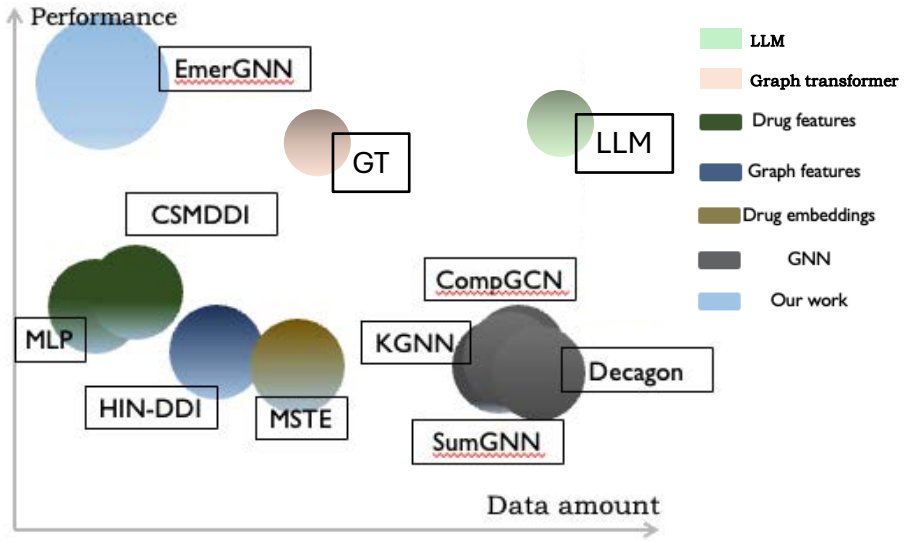
Graph learning problem

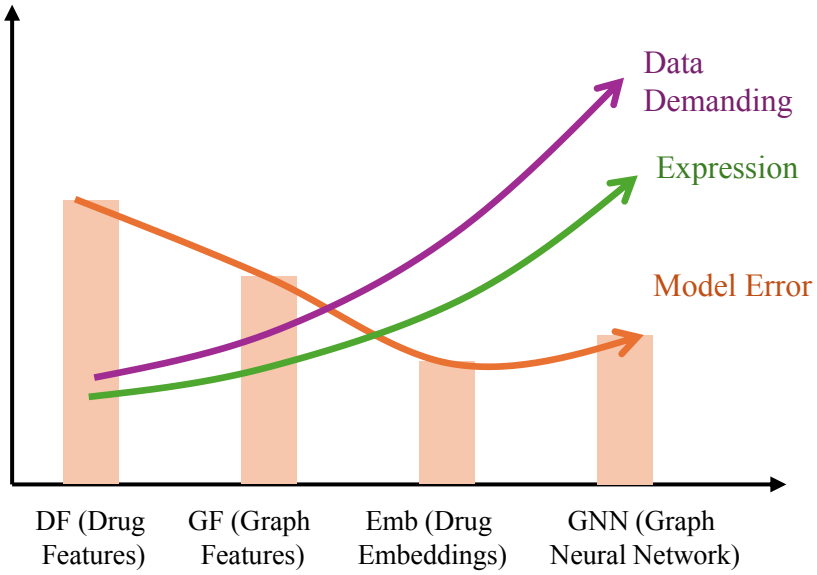


Two main types of DDI:

- Pharmacokinetic DDI: How one drug affects the absorption, distribution, metabolism, or excretion of another drug.
- Efficacy DDI: How two or more drugs affect the same receptor to produce synergistic or harmful effects.

Related Works





Difficulties: Interpretability Training Model complexity

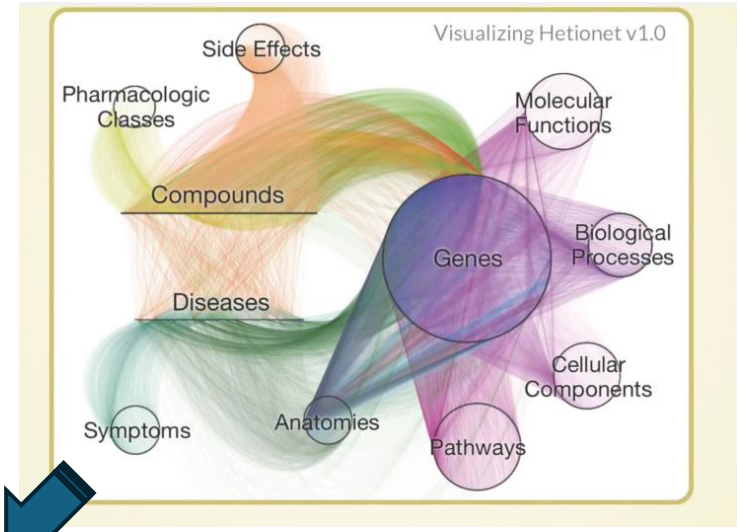
Appearance: **Data sparsity** **Data hunger**

The Breakthrough of Parsimonious

【药品名称】
通用名称：布洛芬颗粒
英文名称：Ibuprofen Granules
汉语拼音：Buluofen Keli
【成份】
本品每包含布洛芬 0.2 克。辅料为：糊精、蔗糖、甘露醇。
【性状】
本品为白色或黄色颗粒，气芳香；味甜。
【作用类别】
本品为解热镇痛类非处方药药品。
【适应症】
用于缓解轻至中度疼痛如头痛、关节痛、偏头痛、牙痛、肌肉痛、感冒或流行性感冒引起的发热。
【药物相互作用】
1. 本品与其他解热、镇痛、抗炎药物同用时可增加胃肠道的副作用，并可能导致溃疡。
2. 本品与肝素、双香豆素等抗凝药同用时，可导致凝血酶原时间延长，增加出血倾向。
3. 本品与地高辛、甲氨蝶呤、口服降血糖药物同用时，能使这些药物的血药浓度增高，不宜同用。

Ibuprofen granules

Relieve fever and headache



- Medication instructions

Emerging drugs often share the same entities as existing drugs, such as the same target gene or disease.

?

- How to extract relevant information?

An effective and efficient approach is needed.

Challenge:

- Proper utilization of biomedical networks can be challenging because these networks are not specifically developed for emerging drugs.
- Mismatches of goals can lead to machine learning models learning scattered knowledge.

Our Work: EmerGNN

EmerGNN

It captures relevant information in a BN by **learning the subgraph** between a new drug and an existing one, and **uses attention mechanisms** to measure the importance of edges.

Our Contributions:

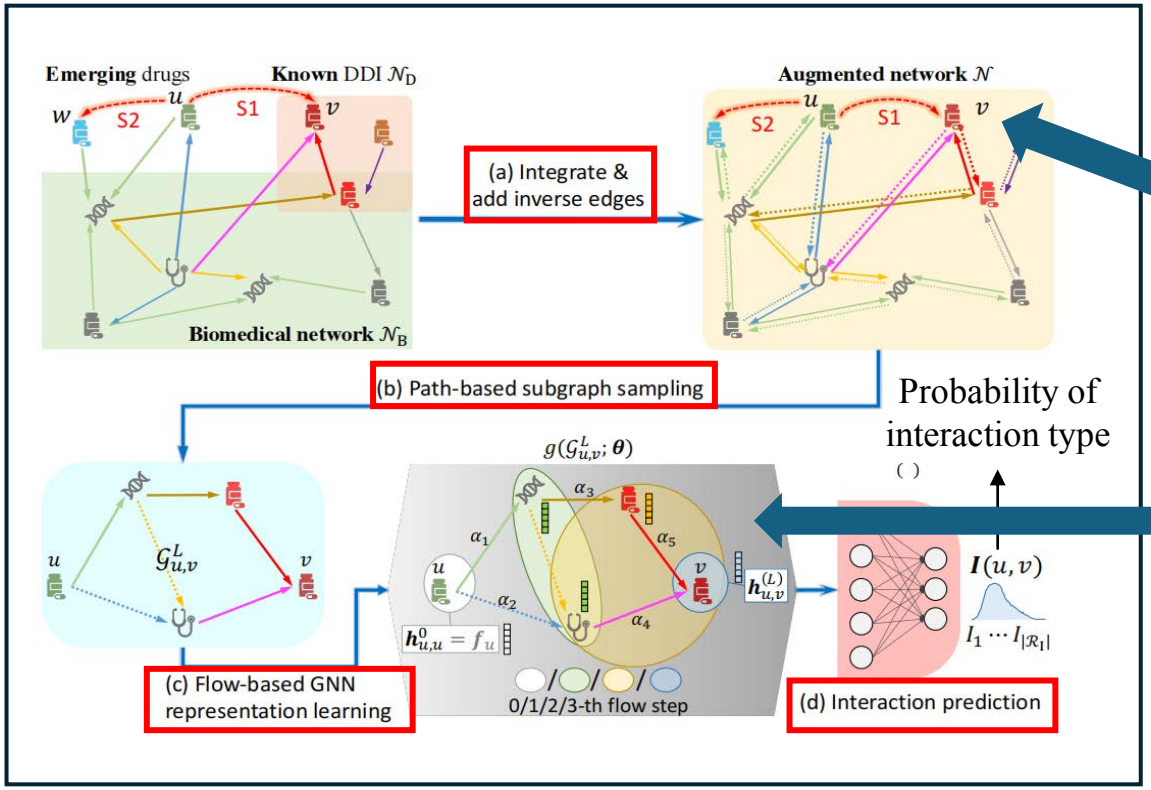
- **Novel GNN Architecture**
- **Path-based Subgraph Extraction**
  - Enhances prediction by capturing local structural information.
- **Effective Message Passing and Attention Mechanisms**
  - Improves aggregation and weighting of information.
- **Comprehensive Evaluation**
  - Demonstrates effectiveness across multiple datasets.

Emerging Drug Interaction Prediction Enabled by Flow-based Graph Neural Network with Biomedical Network. Nature Computational Science.

Architecture of EmerGNN

Key Concept:

- Construct a subgraph to extract knowledge related to emerging drugs
- Set edge attention weights to highlight important paths and design GNN

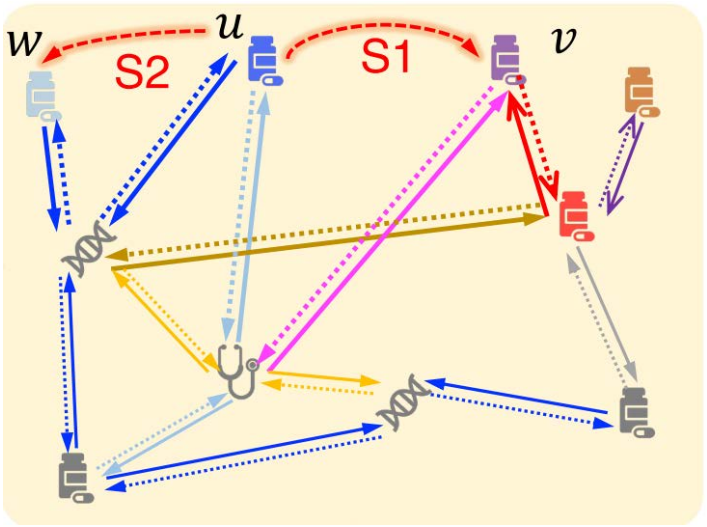


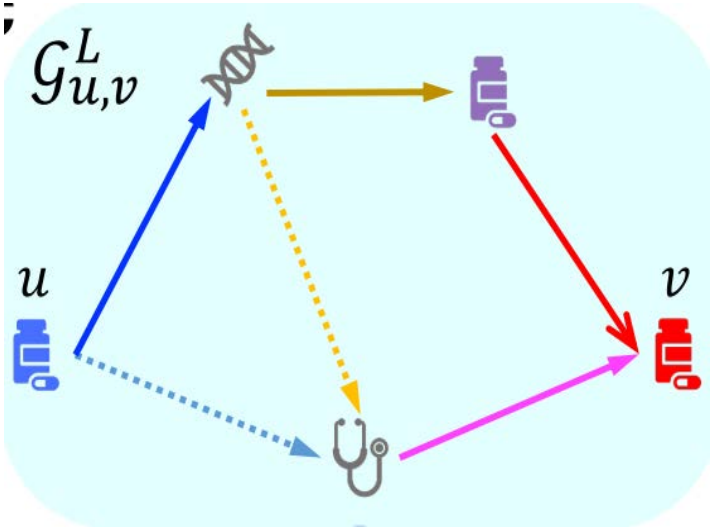
u (emerging drugs) and v (existing drugs) share some of the same entities, such as genes, side effects, and compounds.

- Weighing the different types of relationships in a biomedical network.
- More weighted edges on the path help interpretability.

Definition of Subgraphs

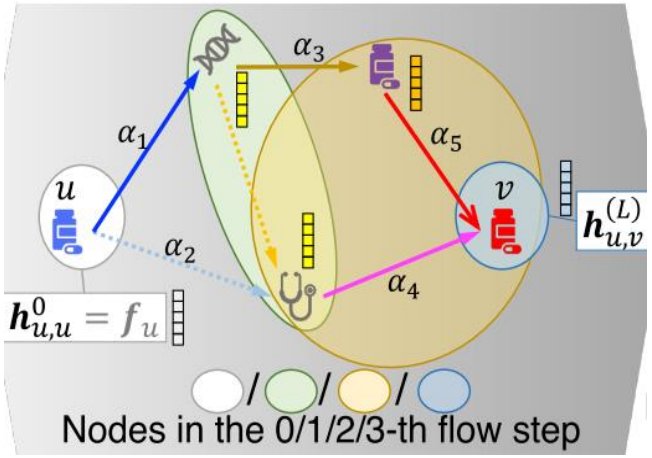
Subgraph: Given a graph G=(V,E), where V is the set of vertices and E is the set of edges, a subgraph H=(V_H,E_H) is a graph such that $V_H \subseteq V$ and $E_H \subseteq E$.





Path-based subgraph: Given a drug pair (u, v) to be predicted, extract all paths from u to v with a length no greater than L to construct a path-based subgraph.

Message Passing Functions



$$h_{u,e}^{(\ell)} = \delta\left(W^{(\ell)} \sum_{e' \in V_{u,v}^{(\ell-1)}} \left(h_{u,e}^{(\ell-1)} + \varphi(h_{u,e'}^{(\ell-1)}, h_r^{(\ell)})\right)\right)$$

- $W^{(\ell)} \in \mathbb{R}^{d \times d}$ is a learnable weighting matrix for step ℓ.
- $h_{u,e'}^{(\ell-1)}$ is the pair-wise representation of entity $e' \in V_{u,v}^{(\ell-1)}$.
- r is the relation type between e' and e.
- $h_r^{(\ell)} \in \mathbb{R}^d$ is the learnable representation of r in the ℓ-th step.
- $\varphi(\cdot,\cdot)$ is the function combining the two vectors.
- $\delta(\cdot)$ is the activation function, specifically ReLU.

Additionally, an attention mechanism is applied to control the importance of different edges:

$$\varphi(h_{u,e'}^{(\ell-1)}, h_r^{(\ell)}) = \alpha_r^{(\ell)} \cdot (h_{u,e'}^{(\ell-1)} \odot h_r^{(\ell)})$$

Attention weight: $\alpha_r^{(\ell)} = \sigma\left((w_r^{(\ell)})^\top [f_u; f_v]\right)$

Loss Function

The loss functions used in the EmerGNN model vary depending on the dataset:

DrugBank

Objective: Predict the interaction type between two drugs.

$$L_{\text{DB}} = -\sum_{(u,i,v)\in\mathcal{N}_{\text{D-train}}} y_i(u,v)\log I_i(u,v)$$

$$I_i(u,v) = \frac{\exp(l_i(u,v))}{\sum_{j\in\mathcal{R}_I}\exp(l_j(u,v))}$$

TWOSIDES

Objective: Predict whether there is an interaction ppp between two drugs.

$$L_{\text{TS}} = -\sum_{(u,i,v)\in\mathcal{N}_{\text{D-train}}}\left(\log I_i(u,v) + \sum_{(u',v')\in\mathcal{N}_i}\log(1-I_i(u',v'))\right)$$

$$I_i(u,v) = \frac{1}{1+\exp(-l_i(u,v))}$$

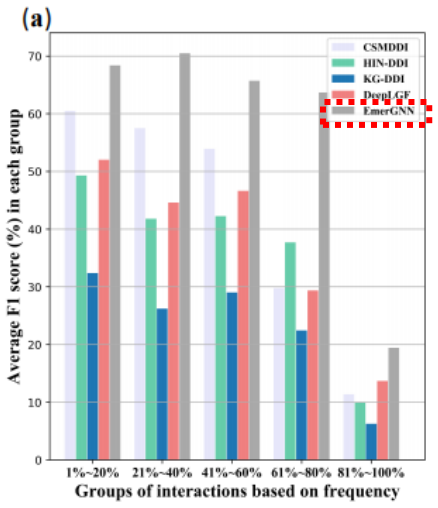
Experimental Result

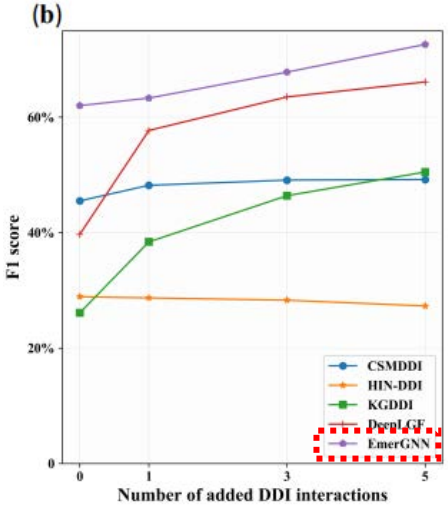
Datasets		Drugbank			Twosides		
Type	Methods	F1-score	Accuracy	Kappa	PR-AUC	ROC-AUC	Accuracy
DF	CSMDDI	45.5(1.8)	62.6(2.8)	55.0(3.2)	73.2(2.6)	74.2(2.9)	69.9(2.2)
GF	HIN-DDI	37.3(2.9)	58.9(1.4)	47.6(1.8)	81.9(0.6)	83.8(0.9)	79.3(1.1)
Emb	KG-DDI	26.1(0.9)	46.7(1.9)	35.2(2.5)	79.1(0.9)	77.7(1.0)	60.2(2.2)
GNN	DeepLGF	39.7(2.3)	60.7(2.4)	51.0(2.6)	81.4(2.1)	82.2(2.6)	72.8(2.8)
GF	TIGER	47.0(3.0)	60.5(2.8)	52.3(3.2)	86.0(0.5)	85.6(0.5)	77.9(1.0)
LLM	TextDDI	58.7(1.2)	66.3(0.3)	59.2(0.4)	86.5(0.6)	87.2(0.6)	79.0(0.2)
GNN	EmerGNN	62.0(2.0)	68.6(3.7)	62.4(4.3)	90.6(0.7)	91.5(1.0)	84.6(0.7)

Index:

- F1 Score (Macro) (Main)
- Accuracy
- Coenkappa coefficient(Cohen, 1960)

Overall, EmerGNN significantly outperforms all comparison methods with a small P-value.

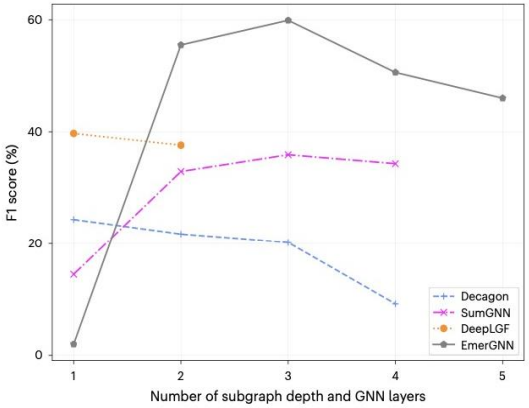




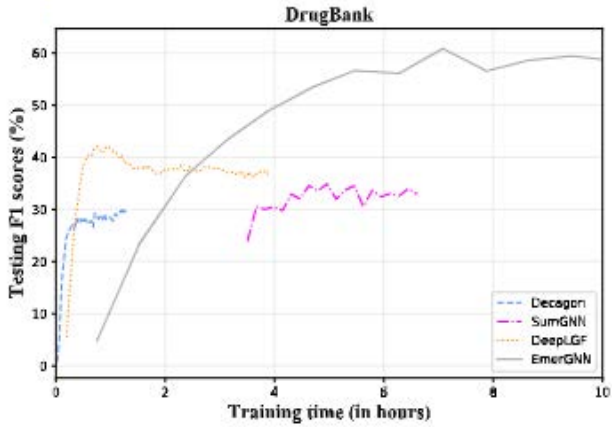
(a) EmerGNN outperforms baseline on all occurrence frequencies.

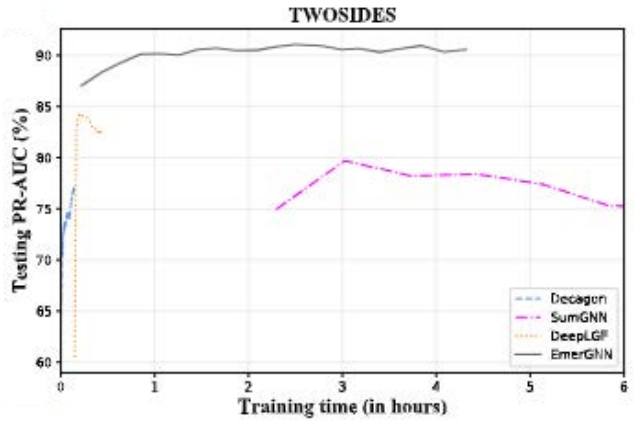
(b) Complementing emerging drug interactions: EmerGNN improves performance by adding more interactions and remains the best of all methods compared.

Model Performance

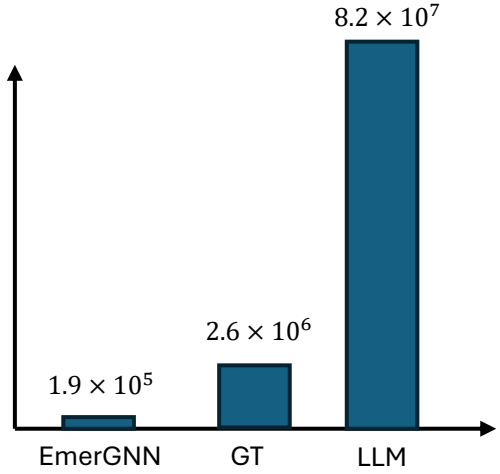


Performance under different layers

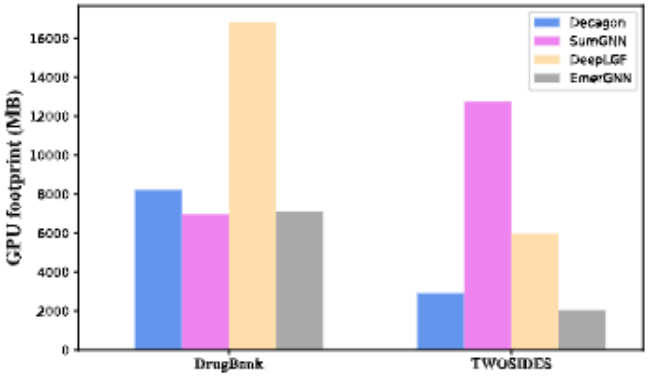




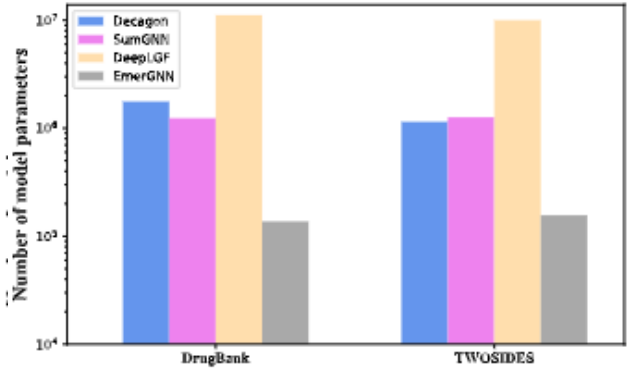
Training curves on different datasets



Parameters comparison

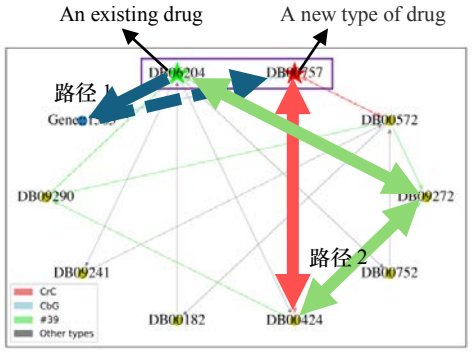


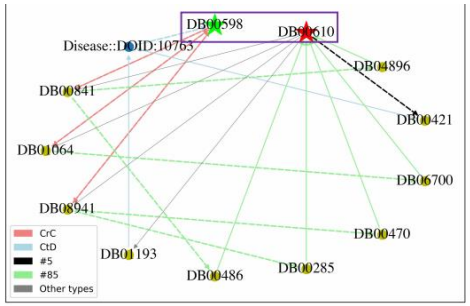
GPU memory usage



Number of mode parameters

Interpretability





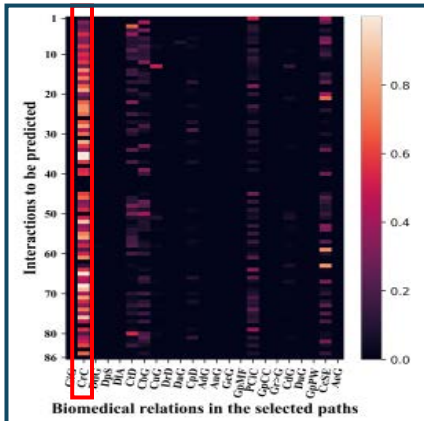
Target: Tapentadol (DB06204) may decrease the analgesic activity of Dolasetron (DB00757).
Path1 (0.6666): Tapentadol $\xrightarrow{\text{binds}}$ CYP2D6 (P450) $\xrightarrow{\text{binds\_inv}}$ Dolasetron
Explanation: Tapentadol can binds the P450 enzyme CYP2D6 (Gene::1565), which is vital for the metabolism of many drugs like Dolasetron (Estabrook, 2003). In addition, Binding of drug to plasma proteins is reversible, and changes in the ratio of bound to unbound drug may lead to drug-drug interactions (Kneip et. al. 2008).
Path2 (0.8977): Dolasetron $\xrightarrow{\text{resembles}}$ Hyoscyamine $\xrightarrow{\text{\#39:↑ constipating}}$ Eluxadoline $\xrightarrow{\text{\#39\_inv}}$ Tapentadol
Explanation: Dolasetron is similar to drug Hyoscyamine (DB00424). Hyoscyamine and Tapentadol can get some connection since they will both increase the constipating activity of Eluxadoline (DB09272). As suggested by Liu and Wittbrodt (2022), reversing opioid-induced constipation often causes the unwanted side effect of analgesia reversal.

Target: Tapentadol (DB06204) may reduce the analgesic activity of Dolacidol (DB00757)

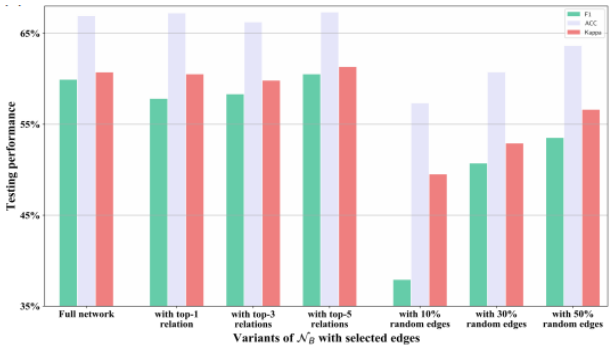
Path1: Tapentadol CYP2D6(P450) Dorasichon

Path declaration: Tapantadol binds to the P450 enzyme CYP2D6 (Gene: :1565), which is critical for the metabolism of many drugs, such as Dolacidone (Estabrook, 2003). In addition, the binding of drugs to plasma proteins is reversible, and changes in the ratio of bound to unbound drugs can lead to drug-drug interactions.

EmerGNN can find important pathways for emerging drug interactions



Correlation between i_{pred} and the biomedical relationship r



Use only Full network/First 1 Relationship (CrC)/First 3 relationships /... /10% random edge /30% random edge /...

EmerGNN can select important and relevant relationships in the biomedical network

Compare with Existing Works

	Parsimony GNN	Graph Transformer	LLM
Training	✅	➖	❌
Data requirement	✅	➖	❌
Interpretability	✅	➖	❌

Why can we solve the problem of data sparse v.s. data hunger?

- Introduce biomedical networks based on specially designed GNN to extract important information needed for attention mechanisms.