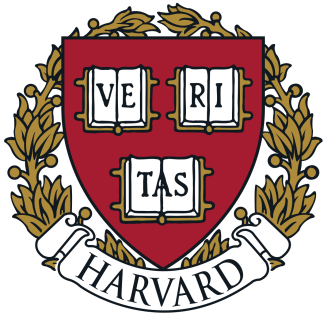
Machine Learning for Drug Development

Marinka Zitnik

Department of Biomedical Informatics
Broad Institute of Harvard and MIT
Harvard Data Science Initiative

marinka@hms.harvard.edu
https://zitniklab.hms.harvard.edu



HARVARD MEDICAL SCHOOL

HDSI | Harvard Data Science Initiative

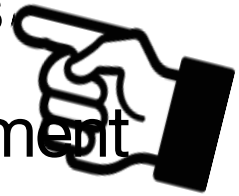
BROAD INSTITUTE

Outline

Overview and introduction

Part 1: Virtual drug screening and drug repurposing

Part 2: Adverse drug effects, drug-drug interactions



Part 3: Clinical trial site identification, patient recruitment

Part 4: Molecule optimization, molecular graph generation, multimodal graph-to-graph translation

Part 5: Molecular property prediction and transformers

Demos, resources, wrap-up & future directions

Drug-drug interactions and polypharmacy

Paper:

Zitnik, Marinka and Agrawal, Monica and Leskovec, Jure. Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics* 2018

Poly-Therapy

Patients take multiple drugs to treat complex or co-existing diseases

46% of people over 65 years take more than 5 drugs

Many take more than 20 drugs to treat heart diseases, depression or cancer

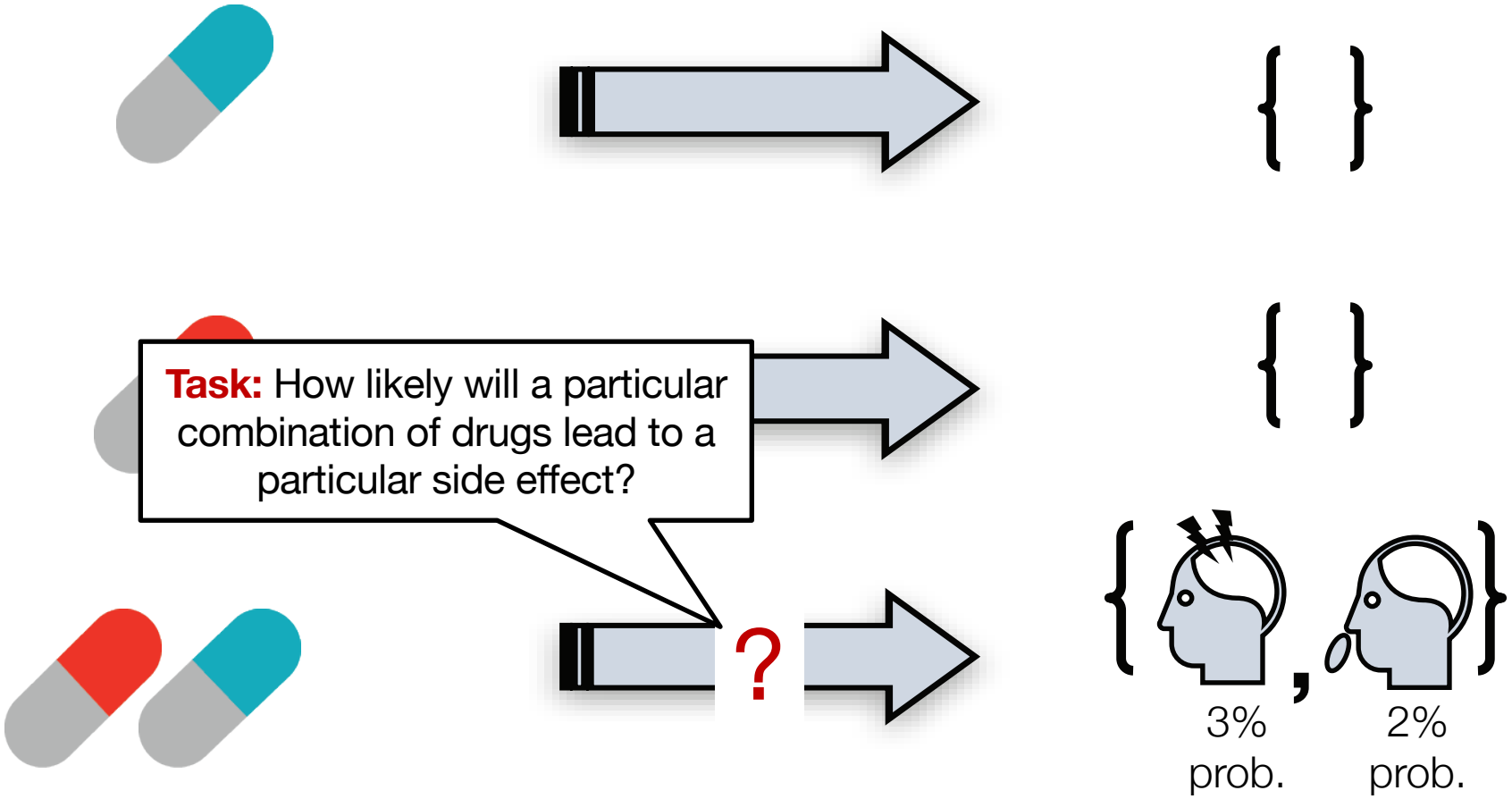
15% of the U.S. population affected by unwanted side effects

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

Unexpected Drug Interactions

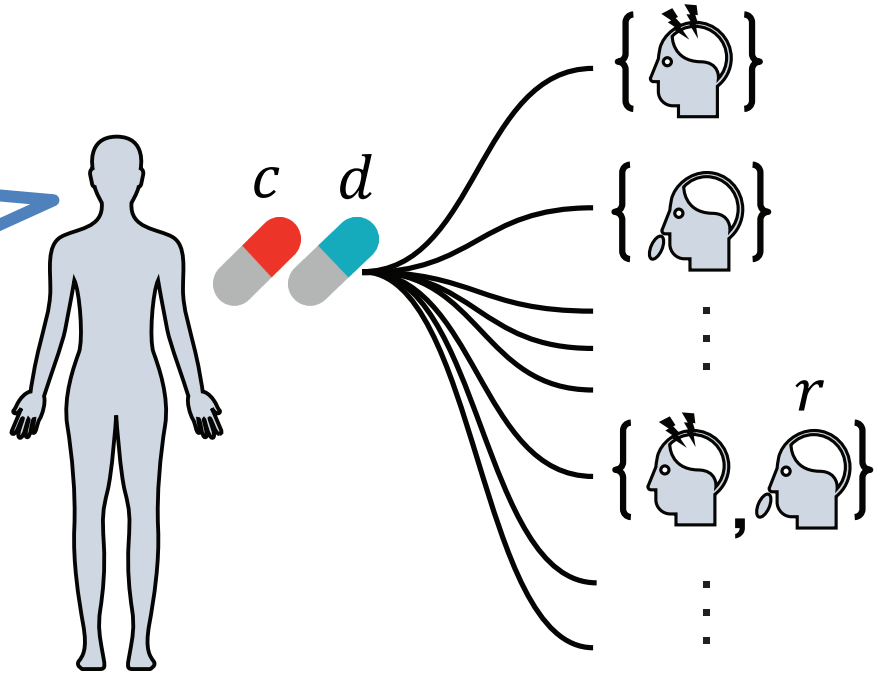
Co-prescribed drugs

Side Effects



Decagon

How likely with a pair of drugs c, d lead to side effect r?



Model and predict side effects of drug pairs

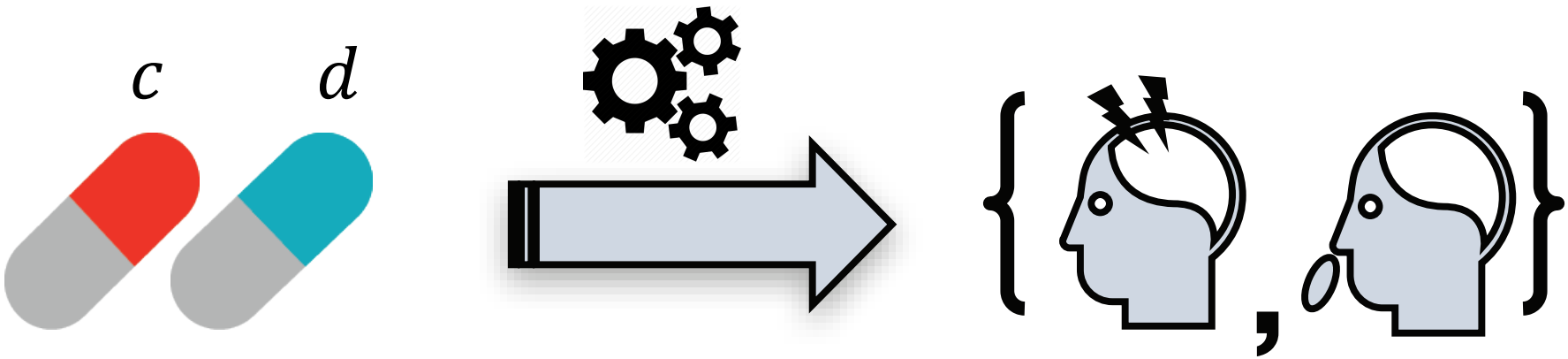
Challenges

- Large number of types of side effects:
  - Each occurs in a small subset of patients
  - Side effects are interdependent
- No information about drug pairs that are not yet used in patients
- Molecular, drug, and patient data:
  - Heterogeneous and multi-relational

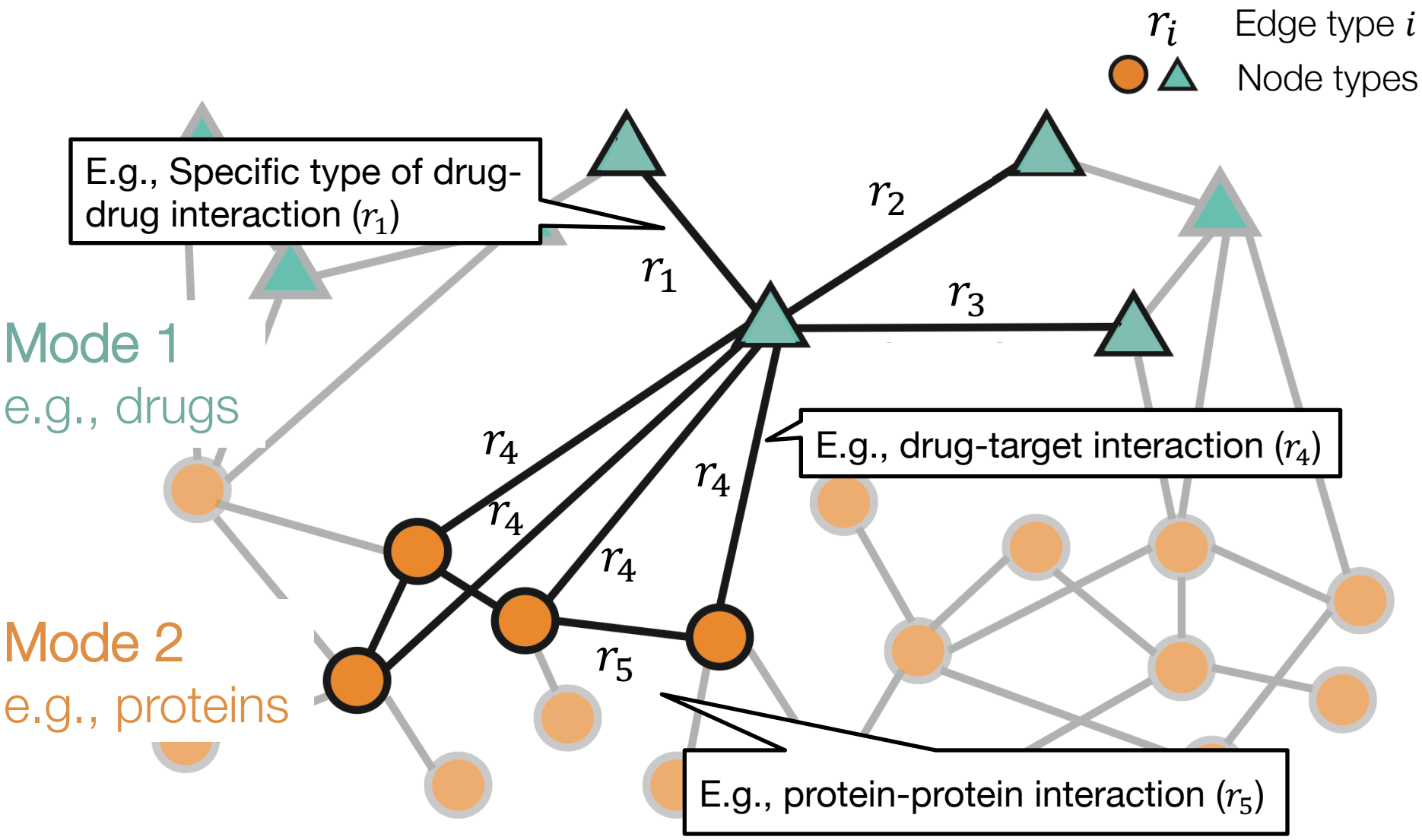
Decagon

In silico screening of drug combinations

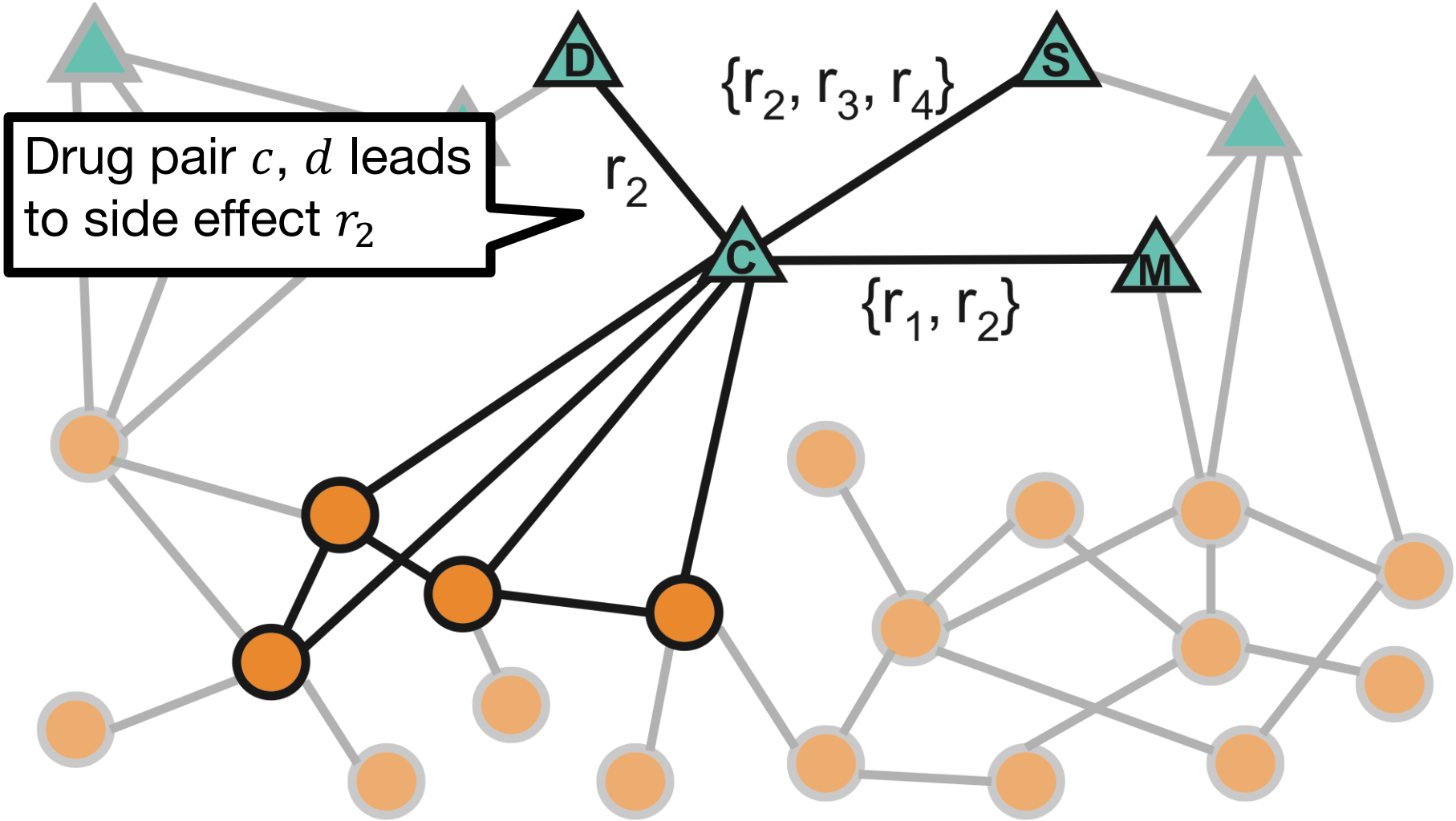
- Use molecular, drug, and patient data
- **Task:** Given a drug pair c, d, predict side effects of that drug pair



Setup: Multimodal Networks



Setup: Multimodal Networks

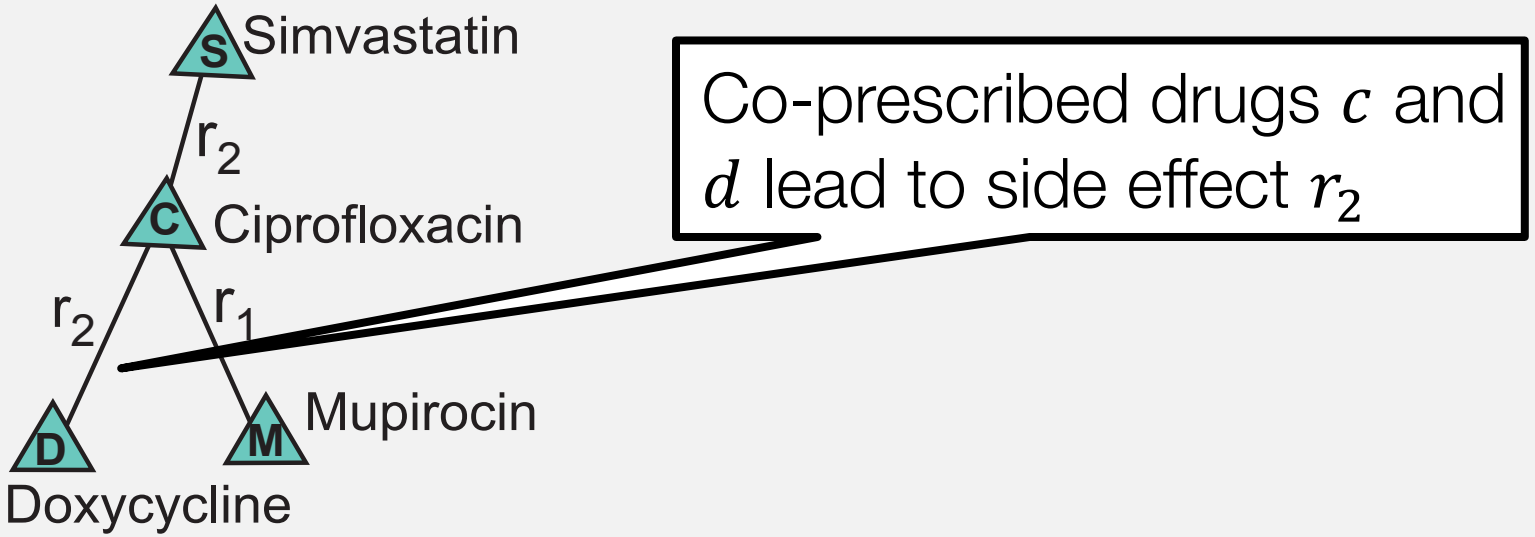


r_1 Gastrointestinal bleed side effect
r_2 Bradycardia side effect
r_3 Nausea side effect
r_4 Mumps side effect
Drug-protein interaction
Protein-protein interaction

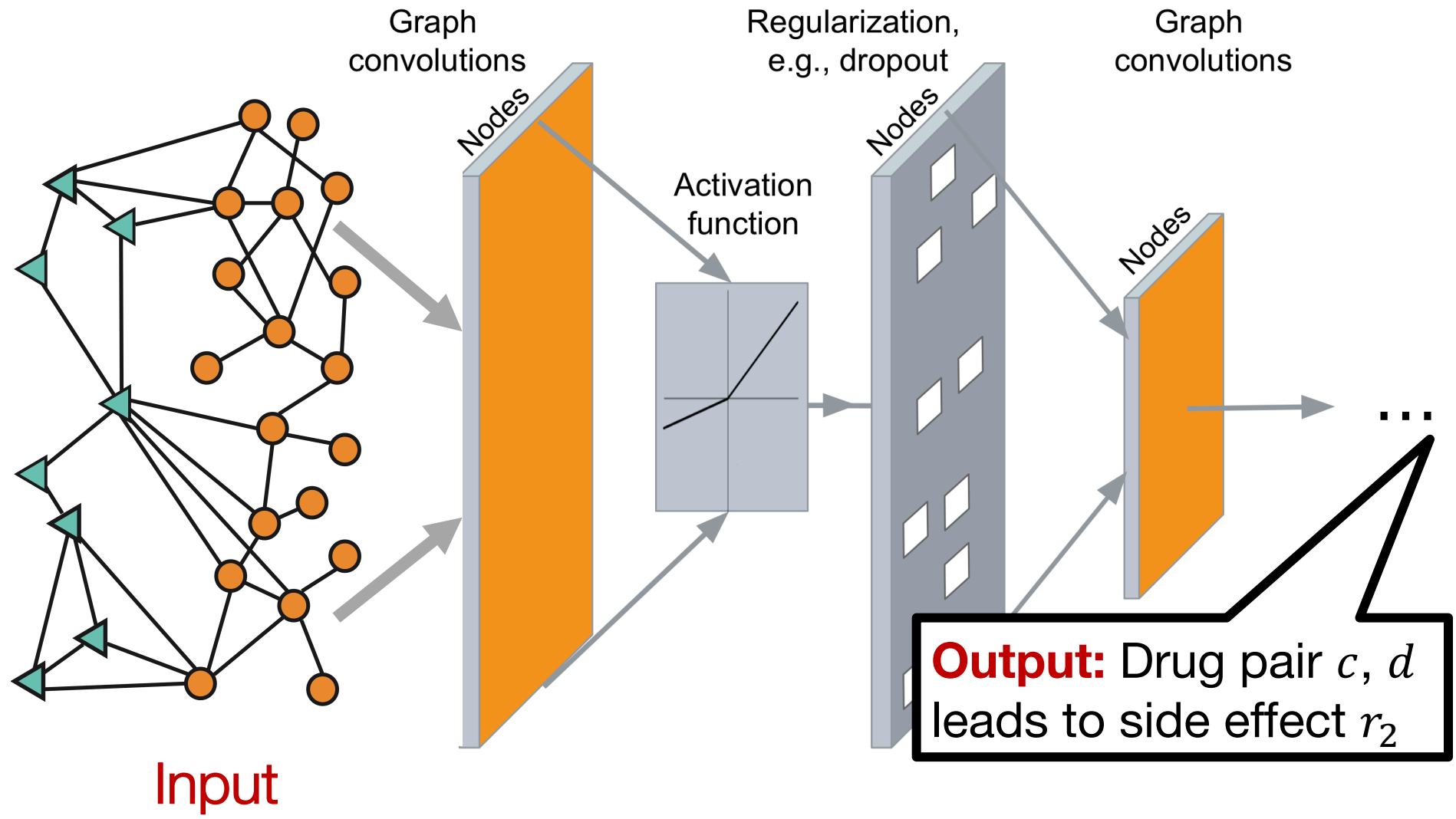
Problem Formulation: Predict

Goal: Given a partially observed graph, predict labeled edges between drug nodes

Query: Given a drug pair c, d, how likely does an edge (c, r_2, d) exist?

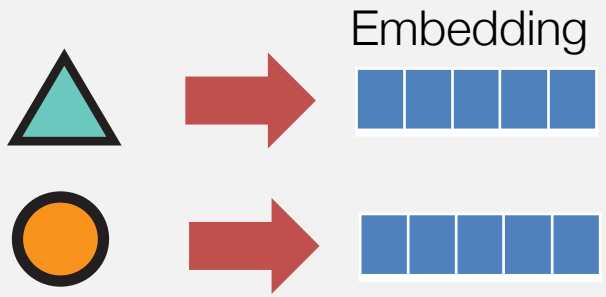


Graph Neural Network

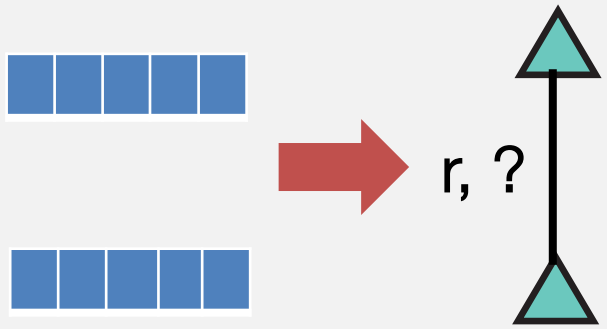


Decagon: Graph Neural Net

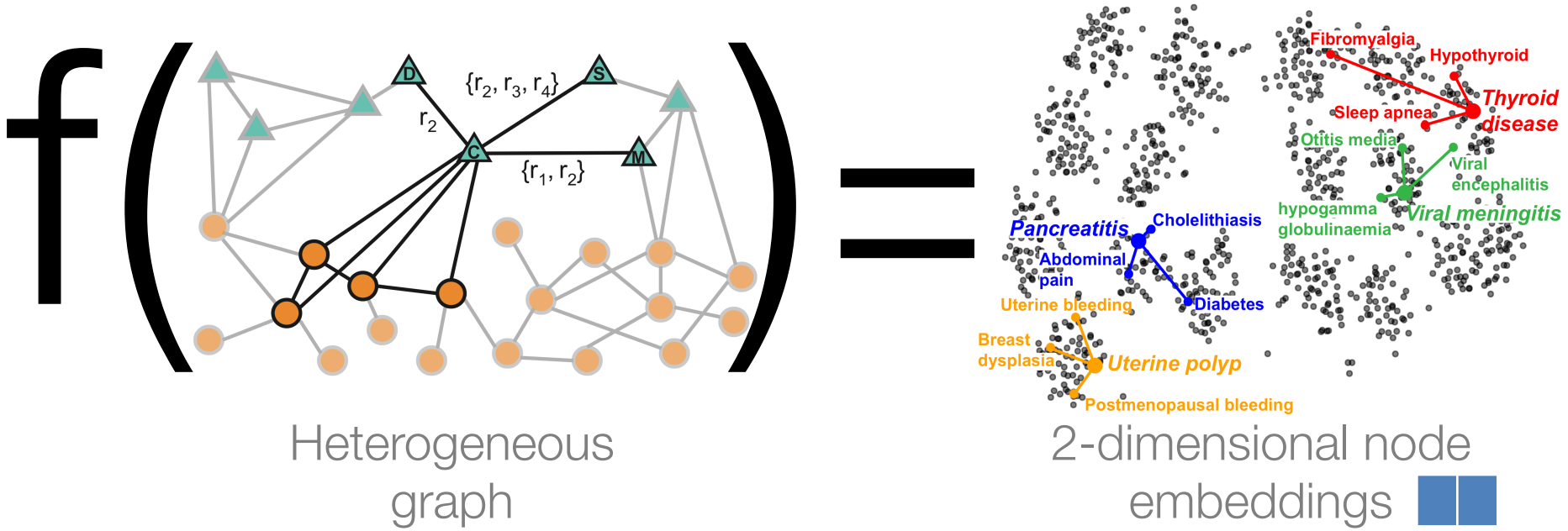
1. **Encoder:** Take the graph and learn an *embedding* for every node



2. **Decoder:** Use the learned embeddings to predict side effects



Embedding Nodes



Heterogeneous graph

2-dimensional node embeddings

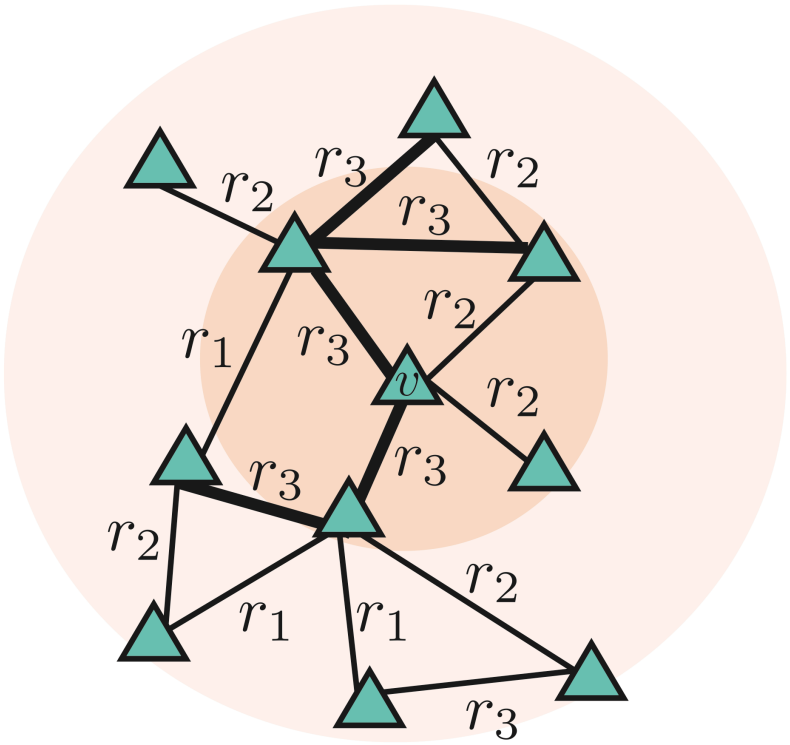
How to learn f?

Intuition: Map nodes to d-dimensional **embeddings** such that similar nodes in the graph are embedded close together

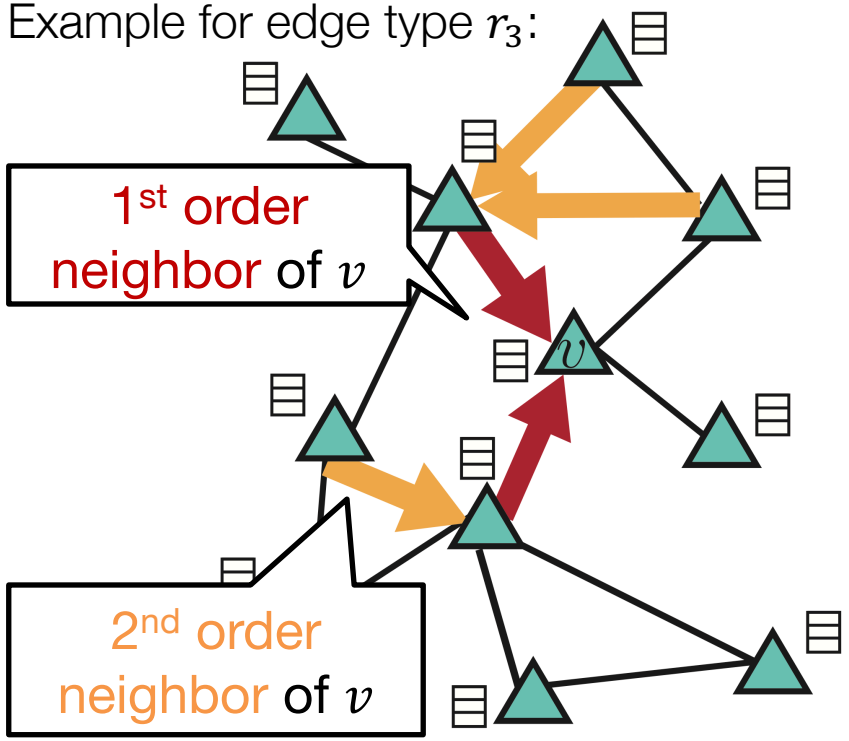
Key Idea: Aggregate Neighbors

Generate embeddings based on local network neighborhoods separated by edge type

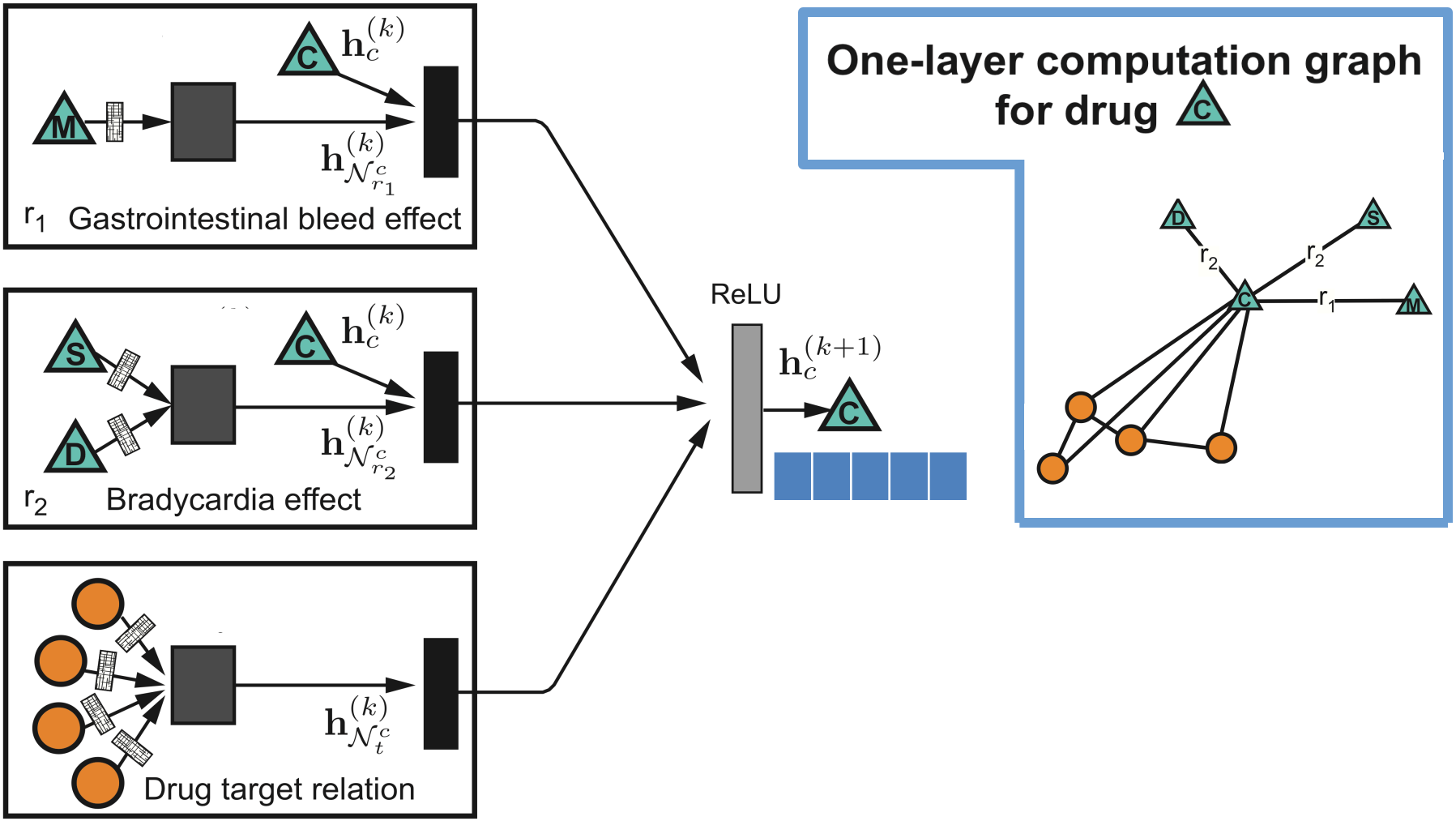
1) Determine a node's computation graph for each edge type



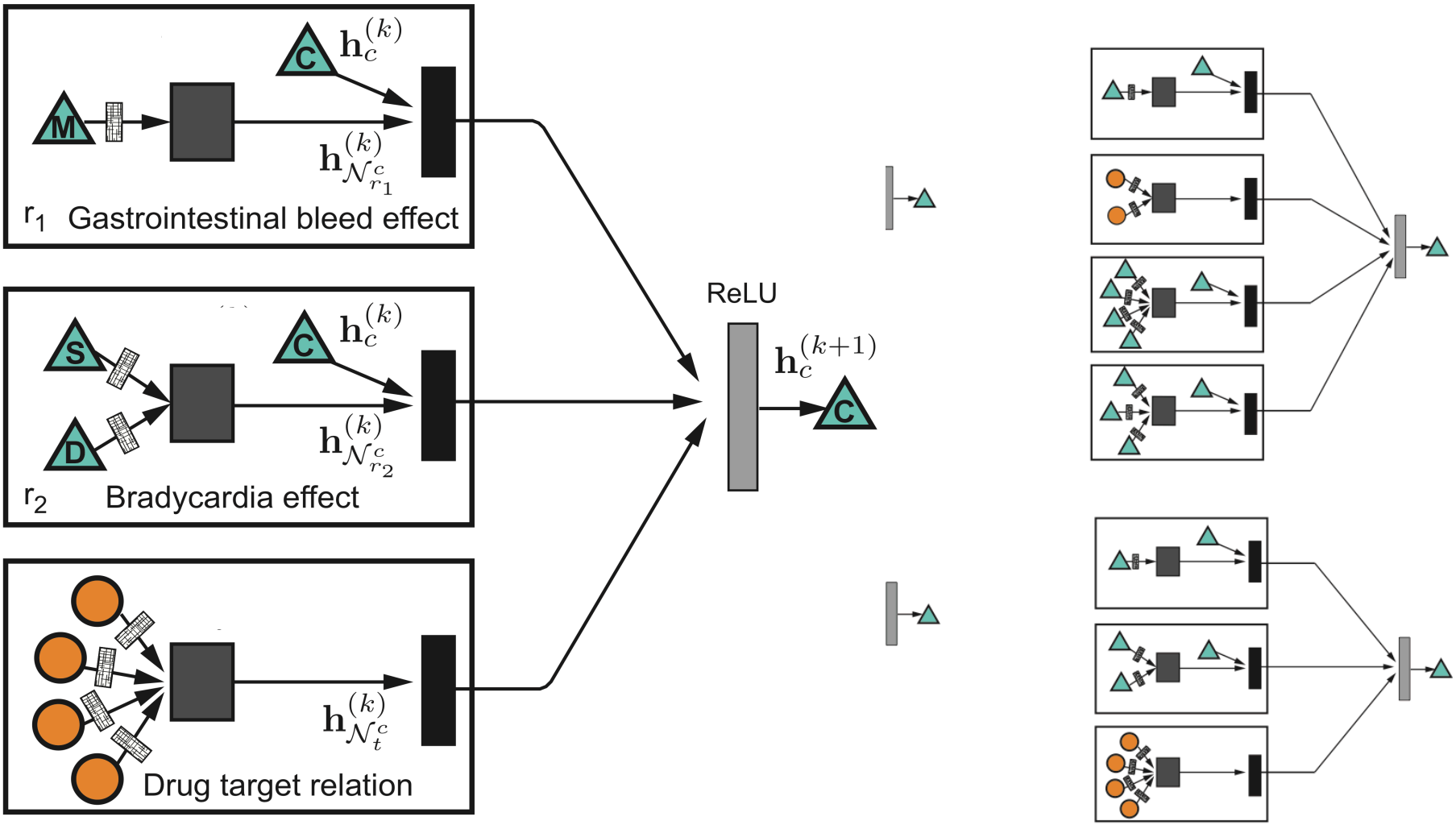
2) Learn how to transform and propagate information across computation graph



Encoder: Embeddings

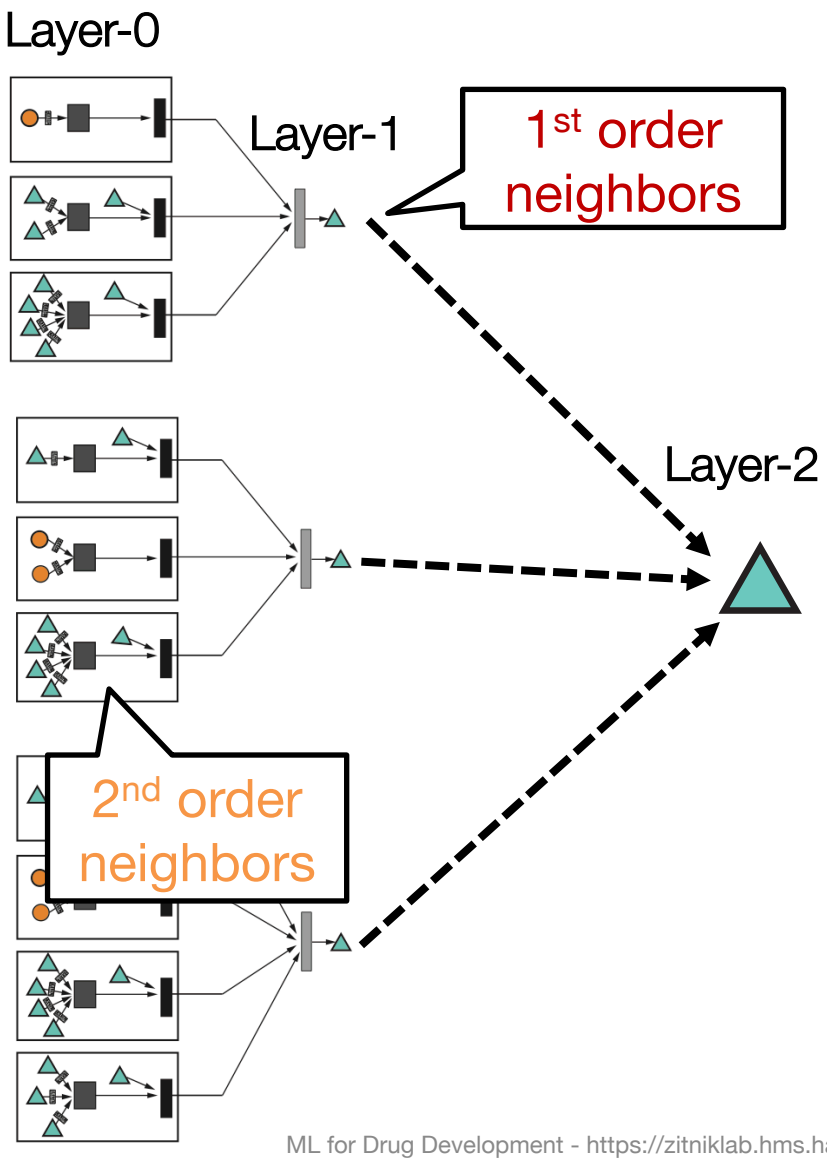


Encoder: Embeddings



A batch of computation graphs

Deep Model: Many Layers



Model can be of arbitrary depth:

- Nodes have embeddings at each layer
- Layer-0 embeddings are nodes' input features

Deep model with K layers:

- Convolves information across K^{th} order neighborhood
- Embedding of a node depends on nodes at most K hops away

Recap: Nodes with **similar network neighborhoods** are embedded **close together**

Graph Neural Encoder

Key element: Each node's computation graph defines a neural network with a different architecture

- Initial 0-th layer embeddings are equal to node features:

$$\mathbf{h}_v^{(0)} = \mathbf{x}_v$$

Aggregate neighbor's previous-layer embeddings, separated by edge type

Ability to integrate side information about nodes

- Per-layer update of n[illegible] embeddings:

$$\mathbf{h}_v^{(k)} = \phi\left(\sum_r \sum_{u \in N_v^r} c_r^{uv} \mathbf{W}_r^{(k-1)} \mathbf{h}_u^{(k-1)} + c_r^v \mathbf{h}_v^{(k-1)}\right) \quad k = 1, \ldots, K$$

Previous-layer embedding of v

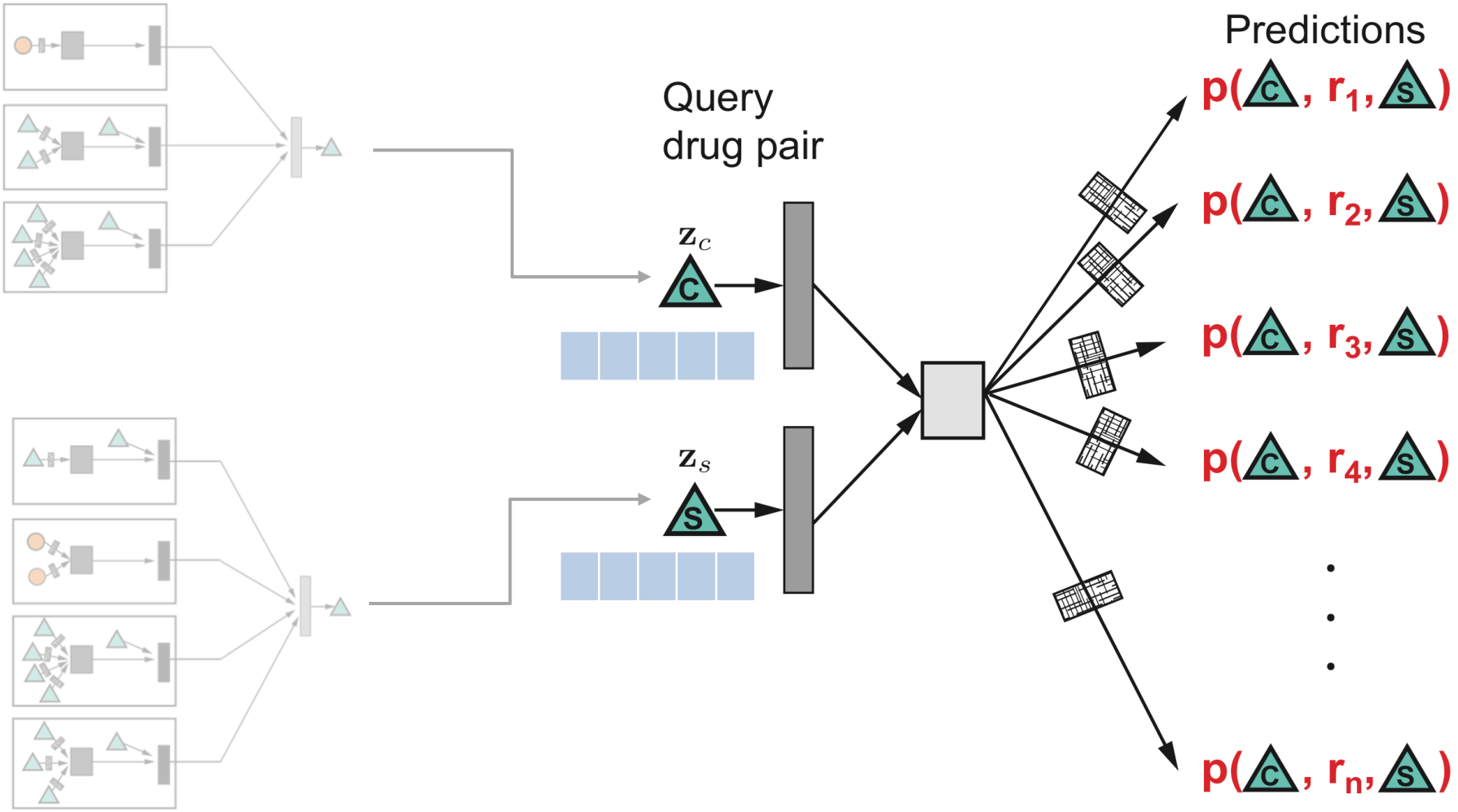
- Embeddings after K layers of neighborhood [illegible]gregation:

$$\mathbf{z}_v = \mathbf{h}_v^{(K)}$$

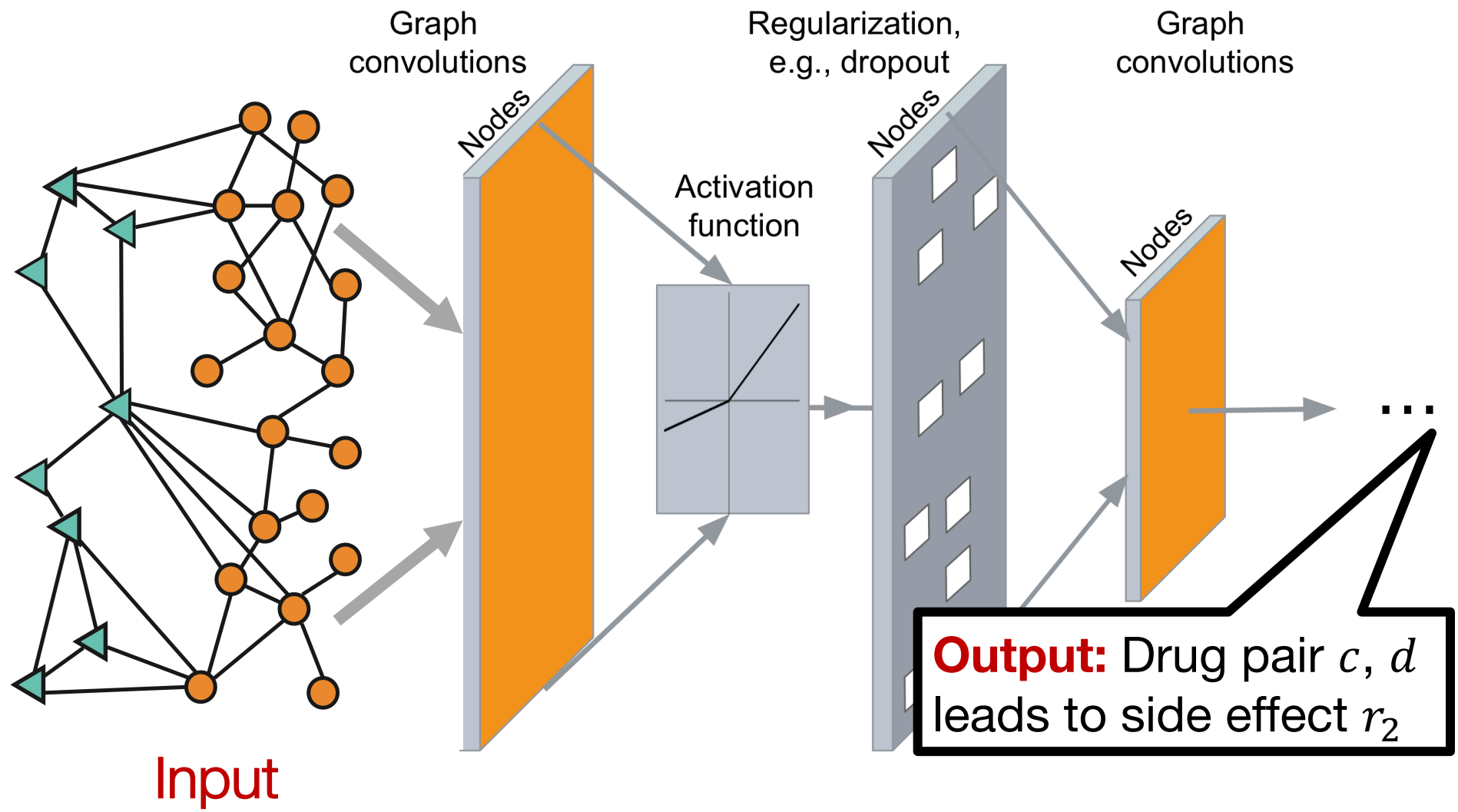
$\mathbf{W}_r^{(k)}$ Para[illegible]

Normalization constant, fixed *e.g.*, $1/|N_v^r|$, or learned

Decoder: Link Prediction

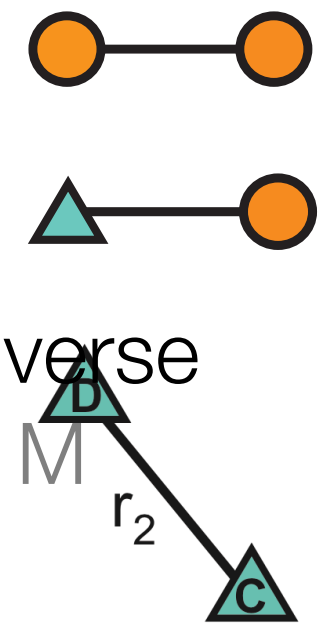


Graph Neural Network



Data: Molecular, Drug & Patient

- Protein-protein interactions: Physical interactions in humans [720 k edges]
- Drug-target relationships [19 k edges]
- Side effects of drug pairs: National adverse event reporting system [4.6 M edges]
- Additional side information



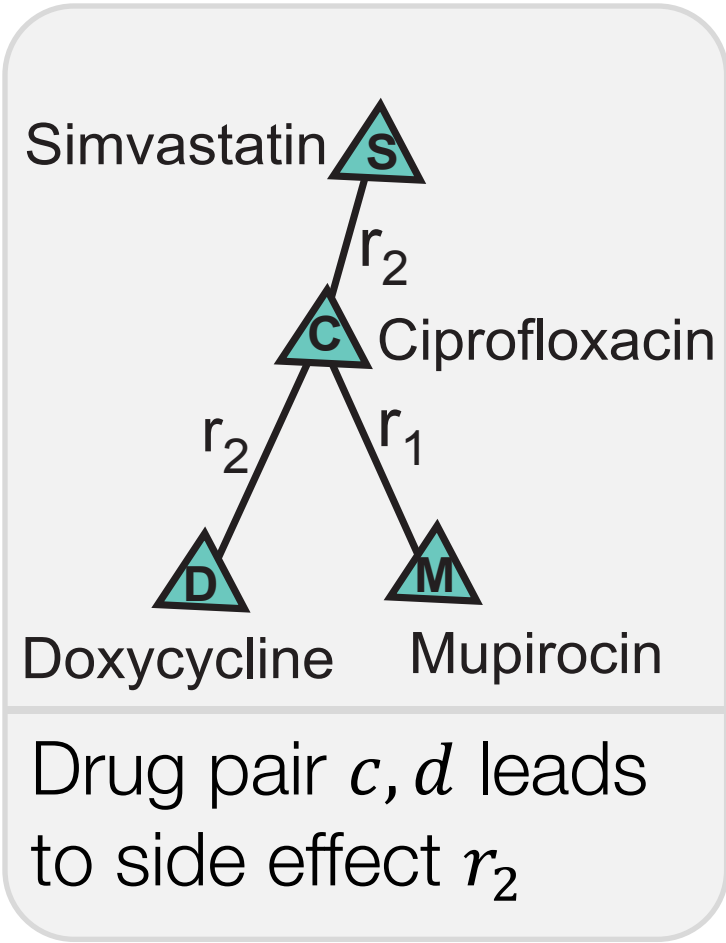
Final graph has 966 different edge types

Experimental Setup

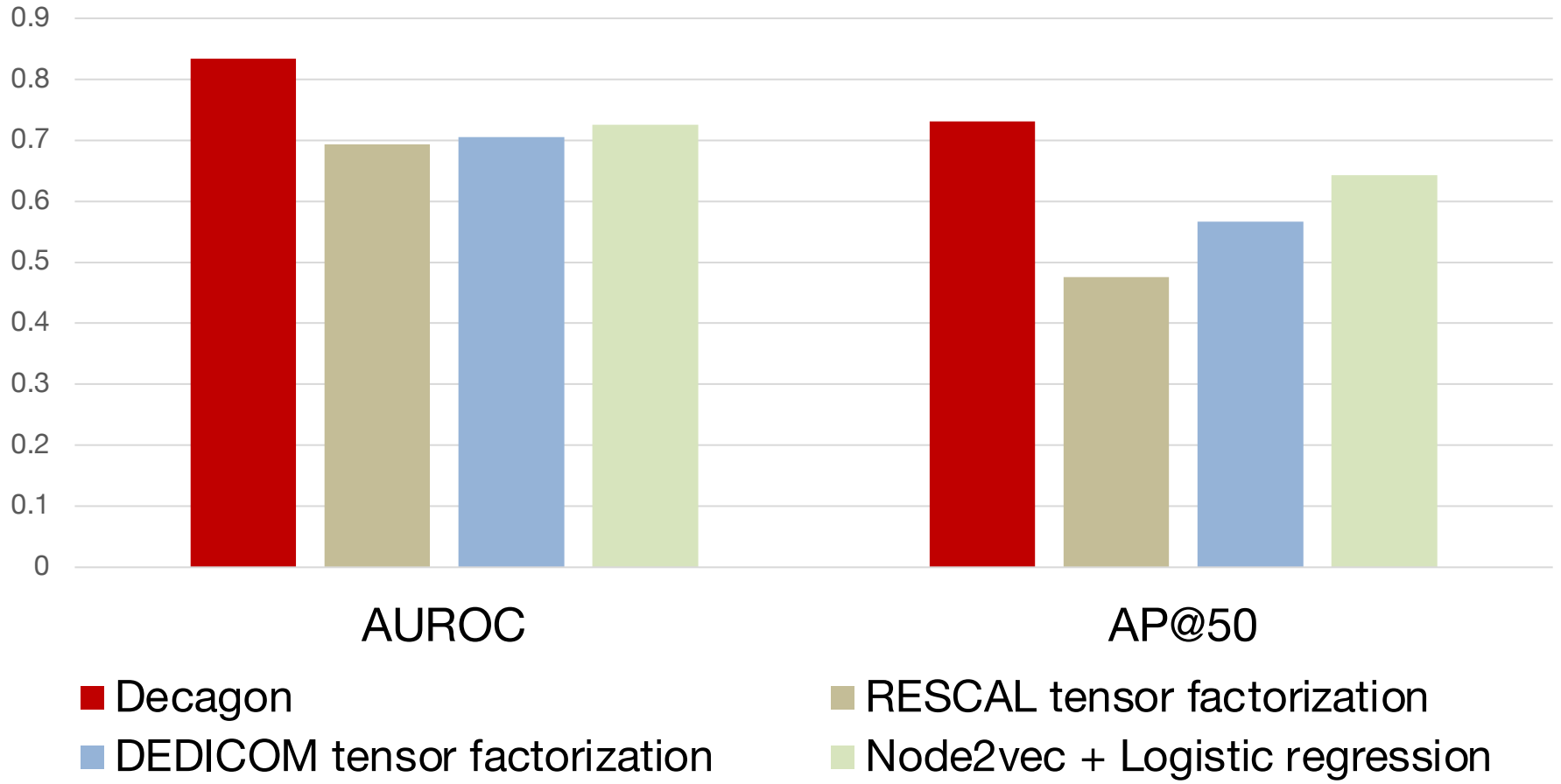
Construct a heterogeneous graph of all the data

Side-effect centric evaluation:

- **Train:** Fit a model on known side effects of drug pairs
- **Test:** Given a query drug pair, predict all types of side effects



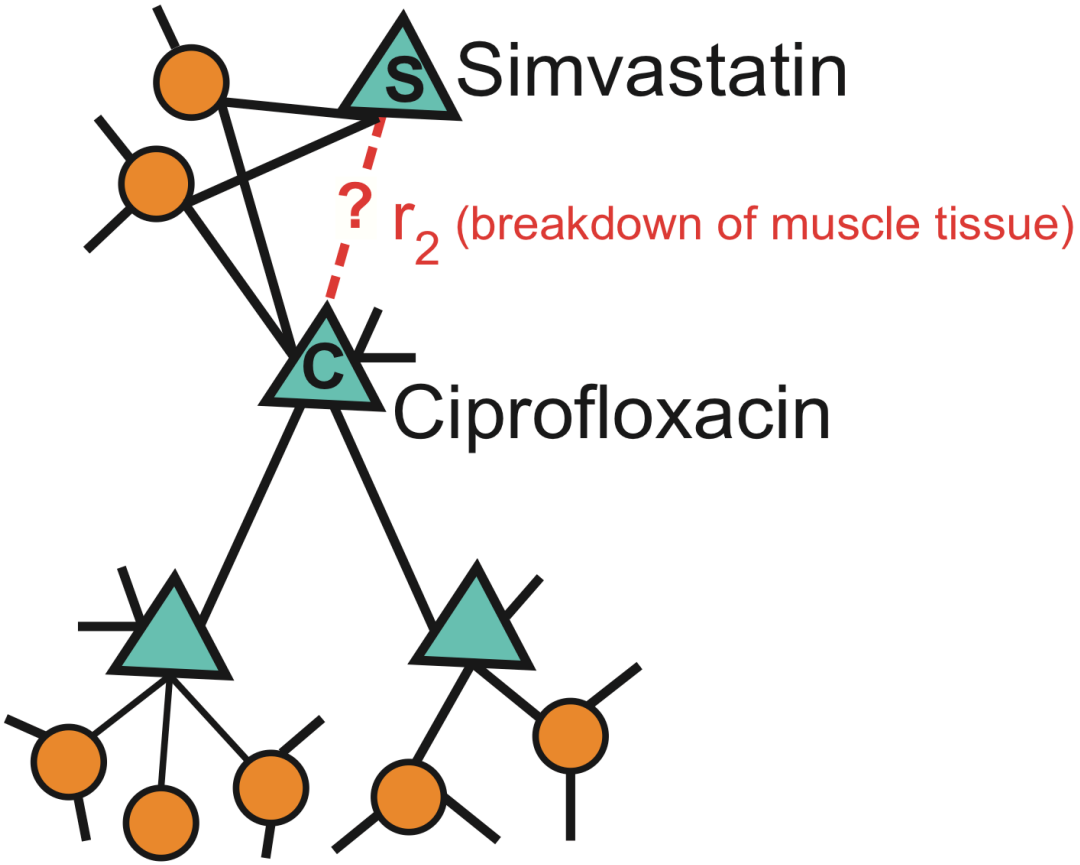
Results: Side Effect Prediction



36% average in AP@50 improvement over baselines

We apply Decagon to the polypharmacy network

E.g.: How likely will Simvastatin and Ciprofloxacin, when taken together, break down muscle tissue?

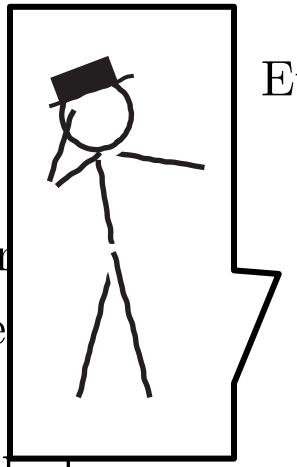


New Predictions

Approach:

1) Train deep model on data generated **prior to 2012**
2) How many **predictions** have been **confirmed after 2012**?

Rank	Drug	Drug	Side effect	Evidence found
1	Pyrimethamine	Aliskiren	Sarcoma	
2	Tigecycline	Bimatoprost	Autonomic r	
3	Telangiectases	Omeprazole	Dacarbazine	
4	Tolcapone	Pyrimethamine	Blood brain	
			eadache	
			ular acidosis	
7	Anagr	Azelaic acid	Cerebral thrombosis	
8	Atorvastatin	Amlodipine	Muscle inflammation	
9	Aliskiren	Tioconazole	Breast inflammation	
10	Estradiol	Nadolol	Endometriosis	



Case Report

Severe Rhabdomyolysis due to Presumed Drug Interactions between Atorvastatin with Amlodipine and Ticagrelor

Conclusions

Decagon predicts side effects of any drug pair:

- Multi-relational Graph Neural Network
- The first AI method for polypharmacy
- Can work even for drug combinations not yet used in patients

Follow-up and Other Work

GNN architectures and chemical structure representations:

- Drug-drug adverse effect prediction with graph co-attention [Deac et al.]
- CASTER: Predicting drug interactions with chemical substructure representation [Huang et al.]
- GENN: Predicting correlated drug-drug interactions with graph energy neural networks [Ma et al.]
- KGNN: Knowledge graph neural network for drug-drug interaction prediction [Lin et al.]
- Bi-level GNNs for drug-drug interaction prediction [Bai et al.]

Drug-drug synergy scoring:

- DeepSynergy: predicting anti-cancer drug synergy with Deep Learning [Preurer et al.]
- Network-based prediction of drug combinations [Cheng et al.]
- MR-GNN: Multi-resolution and dual GNN for predicting structured entity interactions [Xu et al.]
- DeepCCI: End-to-end deep learning for chemical-chemical interaction prediction [Kwon et al.]

Other types of biological relationships:

- Predicting human microbe-drug associations via GCN with conditional random field [Long et al.]
- Deep learning improves prediction of drug-drug and drug-food interactions [Ryu et al.]
- HyperFoods: Machine intelligent mapping of cancer-beating molecules in foods [Veselkov et al.]

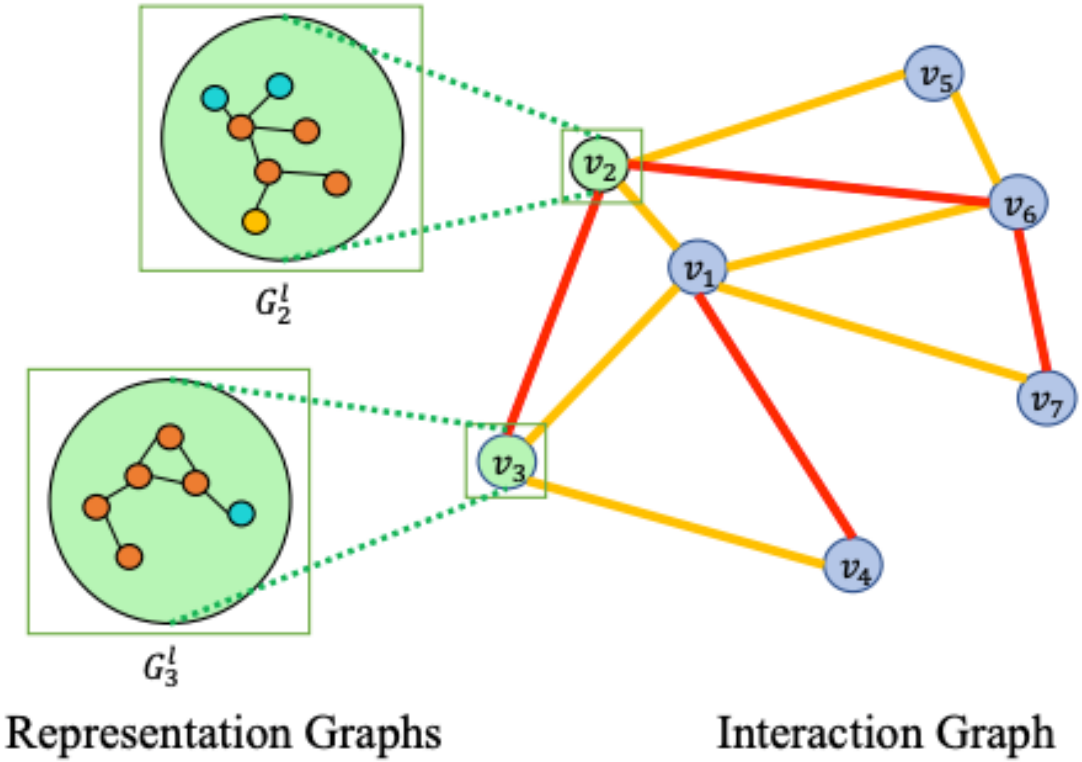
Drug-drug interactions

Paper:

Yunsheng Bai, Ken Gu, Yizhou Sun, Wei Wang. Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction, *arXiv:2006.14002*

Approach

- **Bi-level graph view** of DDI data for multi-scale prediction
- Typically, GNN methods operate only on either the representation graphs or a single interaction graph without



Node colors in the representation graphs denote molecular level element types. Edge colors in the interaction graph denote drug interactions types

Results

Performance of all methods on DRUGBANK under different training data ratios (TR). Further breakdown of performance under different node degree splits are shown.

(a) LL-GNN

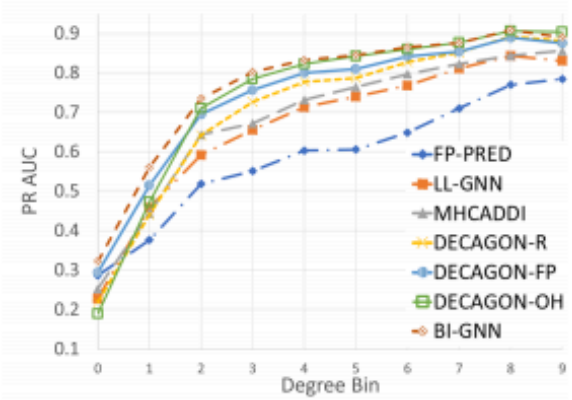
(b) LL-GNN

(c) DECAGON-OH

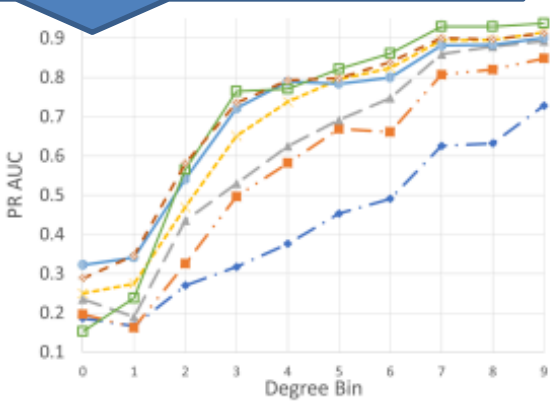
(d) DECAGON-OH

(f) BI-GNN

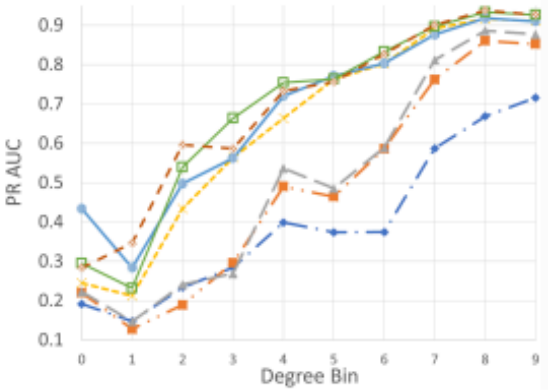
Bi-level GNN



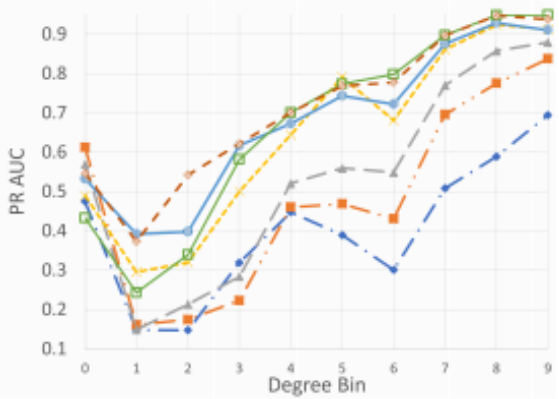
(a) TR=0.1



(b) TR=0.3



(c) TR=0.5



(d) TR=0.7

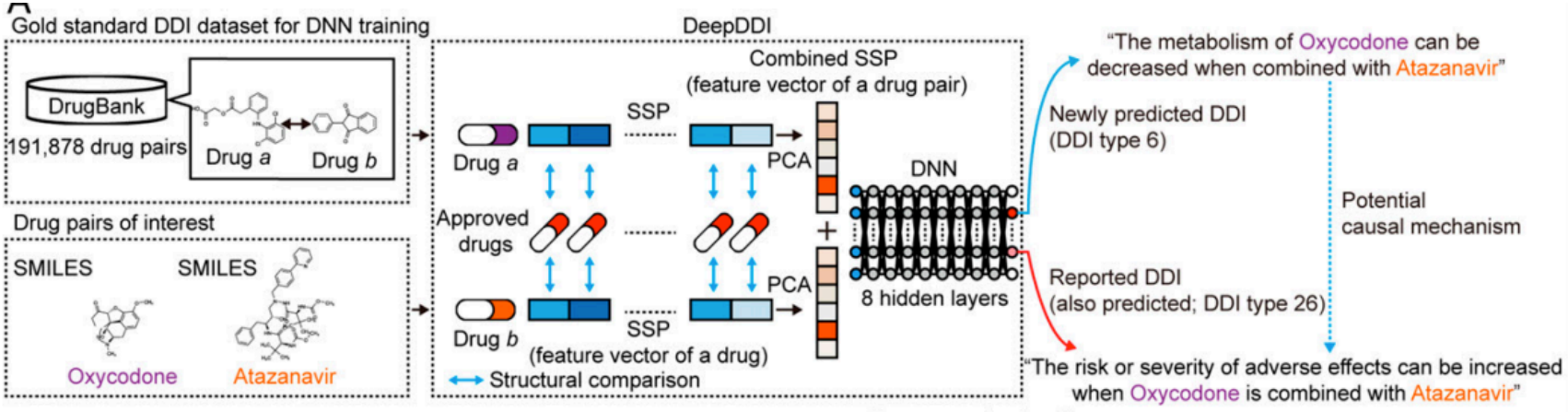
Drug-drug interactions and food-drug interactions

Paper:

Jae Yong Ryu, Hyun Uk Kim, and View ORCID ProfileSang Yup Lee. Deep learning improves prediction of drug-drug and drug-food interactions, *PNAS* 2018

Approach

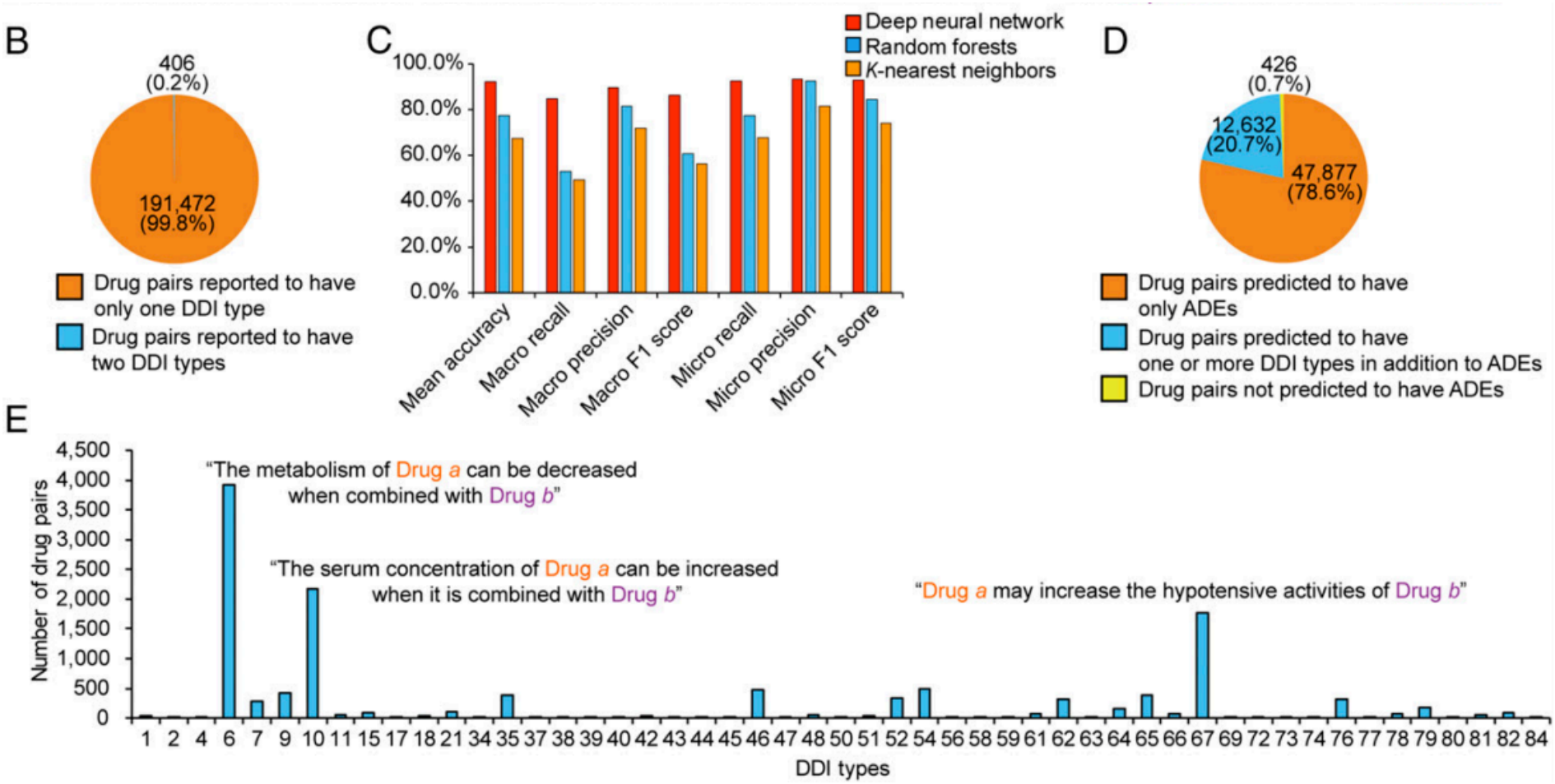
DeepDDI designs a feature called structural similarity profile (SSP) combined with MLP for prediction

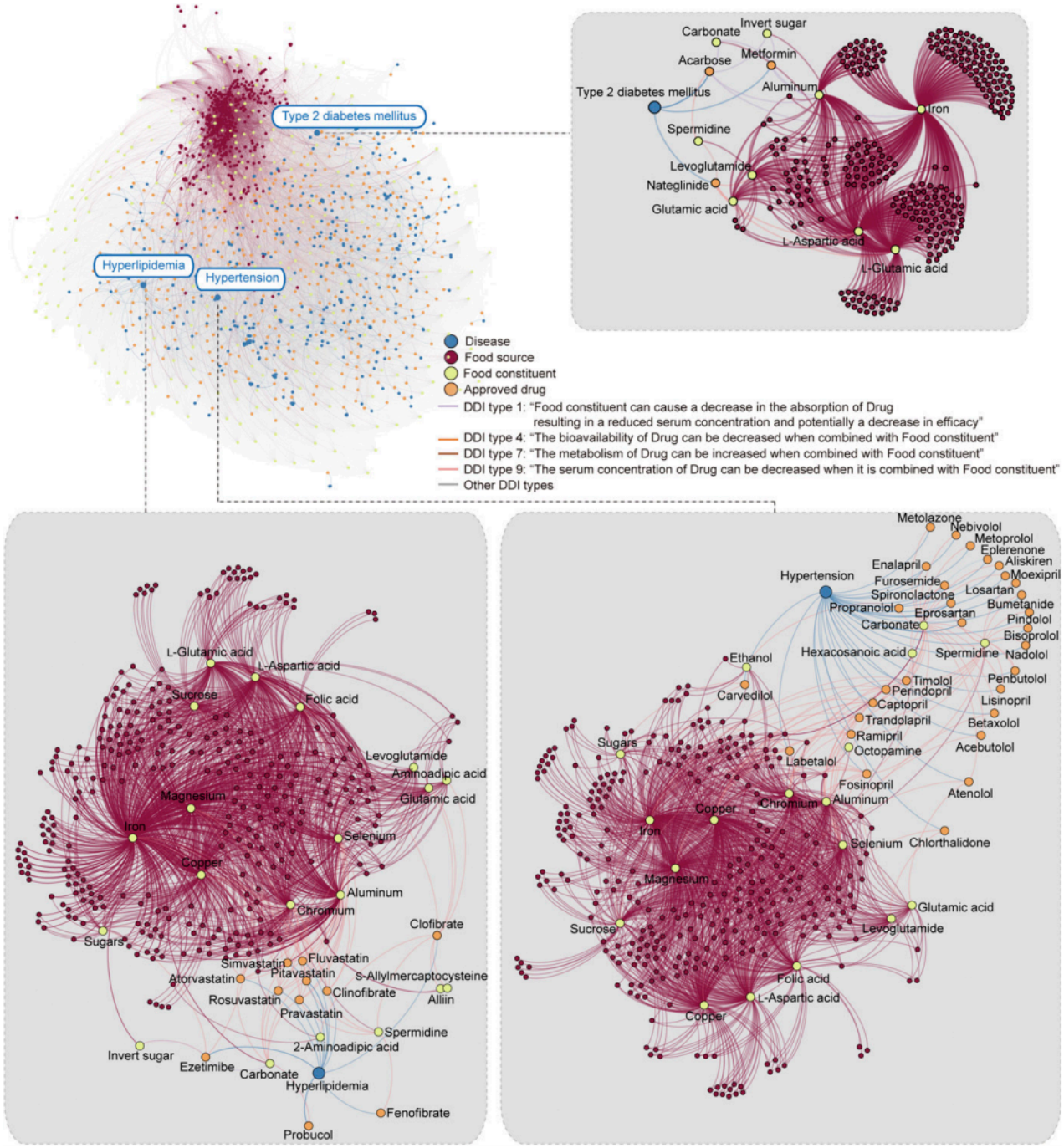


Input: DeepDDI accepts chemical structures (in SMILES describing the structure of a chemical compound) and names of drugs in pairs as inputs

Output: It predicts their potential drug–drug interaction (DDI) types as outputs in human-readable sentences having the input drug names

Results: DDI Prediction





Prediction of food constituents that reduce the in vivo concentration of approved drugs. A network showing relationships among 357 diseases, 430 approved drugs, 274 food constituents, and 356 food sources was created using the DeepDDI output sentences obtained from 358,995 drug-food constituent pairs.

Outline

- ✓ Overview and introduction
- ✓ Part 1: Virtual drug screening and drug repurposing
- ✓ Part 2: Adverse drug effects, drug-drug interactions
- Part 3: Clinical trial site identification, patient recruitment
- Part 4: Molecule optimization, molecular graph generation, multimodal graph-to-graph translation
- Part 5: Molecular property prediction and transformers
- Demos, resources, wrap-up & future directions