Modeling Polypharmacy with Graph Convolutional Networks

Marinka Zitnik, Monica Agrawal, and Jure Leskovec

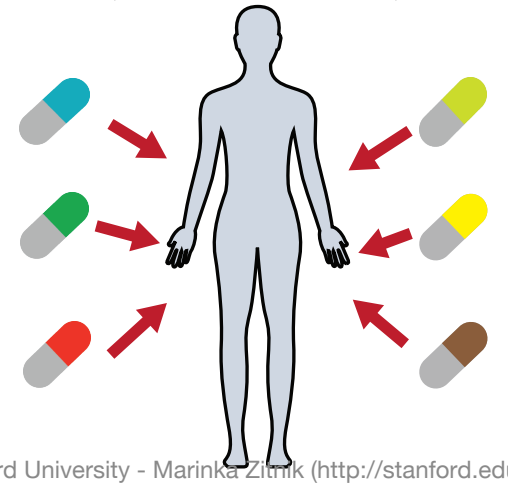
Stanford University

CHAN ZUCKERBERG BIOHUB

Why polypharmacy?

Many patients take multiple drugs to treat complex or co-existing diseases:

- 25% of people ages 65-69 take more than 5 drugs
- 46% of people ages 70-79 take more than 5 drugs
- Many patients take more than 20 drugs to treat heart disease, depression, insomnia, etc.

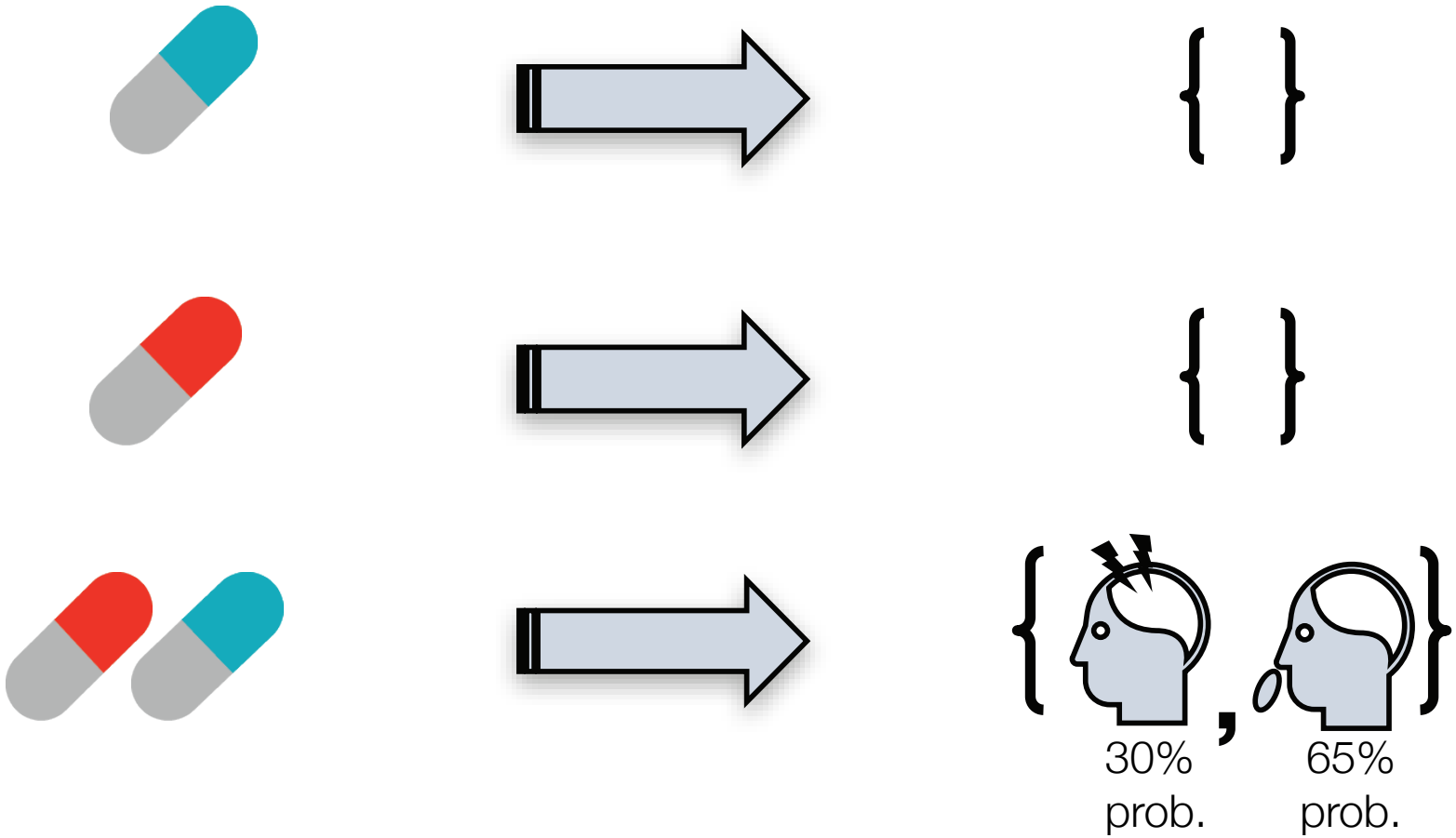


[Charlesworth *et al.*, 2015]

Unwanted Side Effects

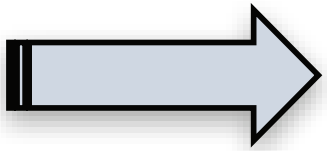
Prescribed drugs

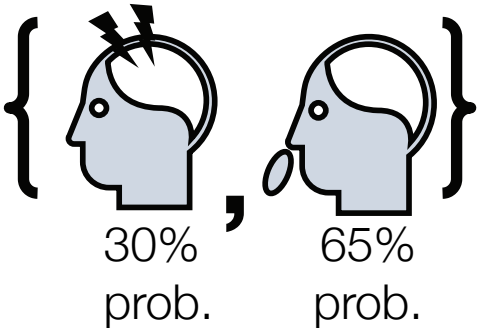
Drug side effect



Unwanted Side Effects

- Side effects due to drug-drug interactions
- Extremely difficult to identify:
  - Impossible to test all combinations of drugs
  - Side effects not observed in controlled trials
- **15%** of the U.S. population affected
- Annual costs exceed **$177 billion**





[Kantor *et al.*, 2015]

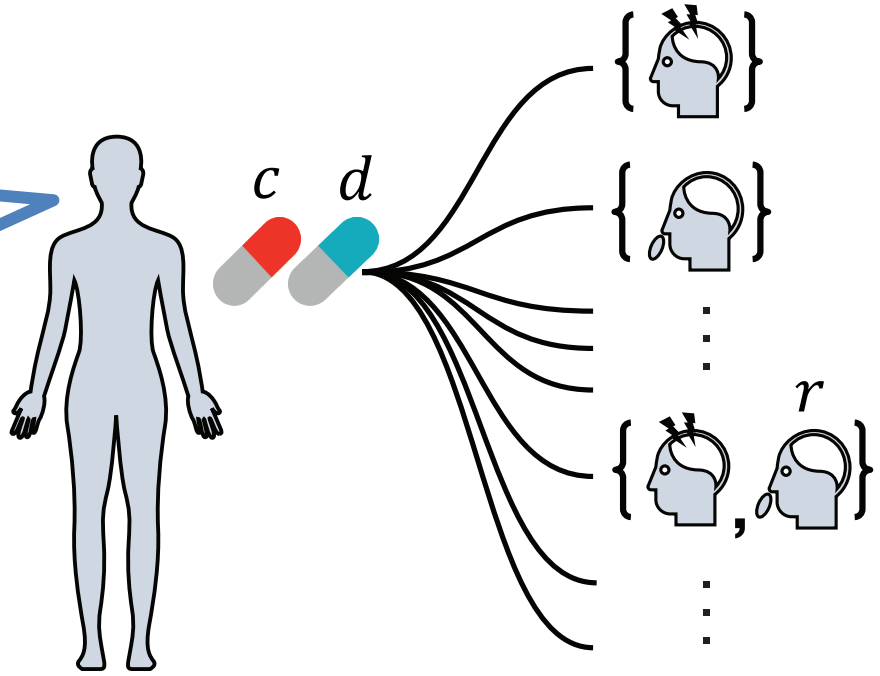
Existing Research

- Experimental screening of drug combs:
  - Expensive, combinatorial explosion
- Computational methods:
  - Supervised methods: Predict probability of a drug-drug interaction [Chen *et al*., 2016; Shi *et al.,* 2017]
  - Similarity-based methods: Similar drugs have similar interactions [Gottlieb *et al.,* 2012; Ferdousi *et al*., 2017; Zhang *et al.,* 2017]

These methods do not predict side effects of drug combinations

This Work

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



Our study: Model and predict side effects of drug pairs

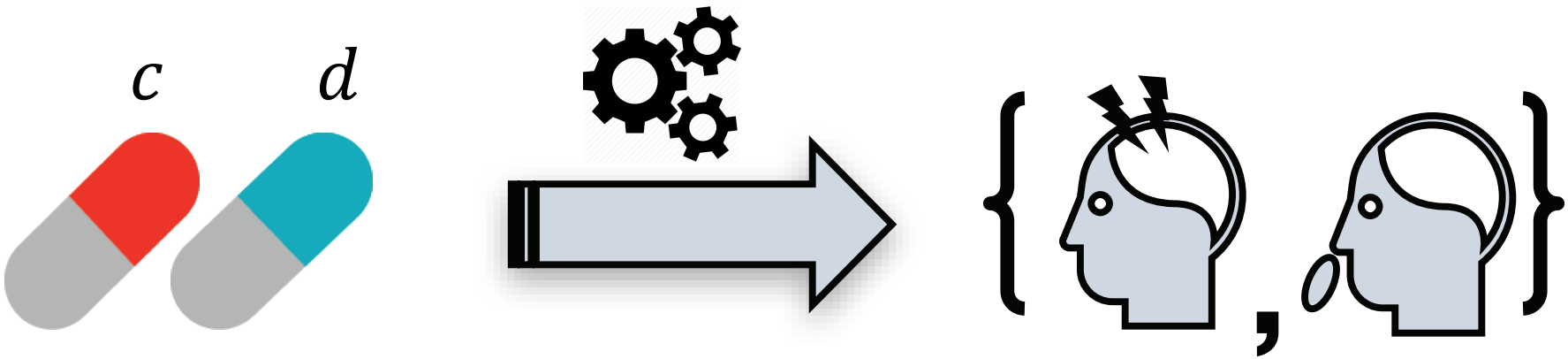
Challenges

- 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

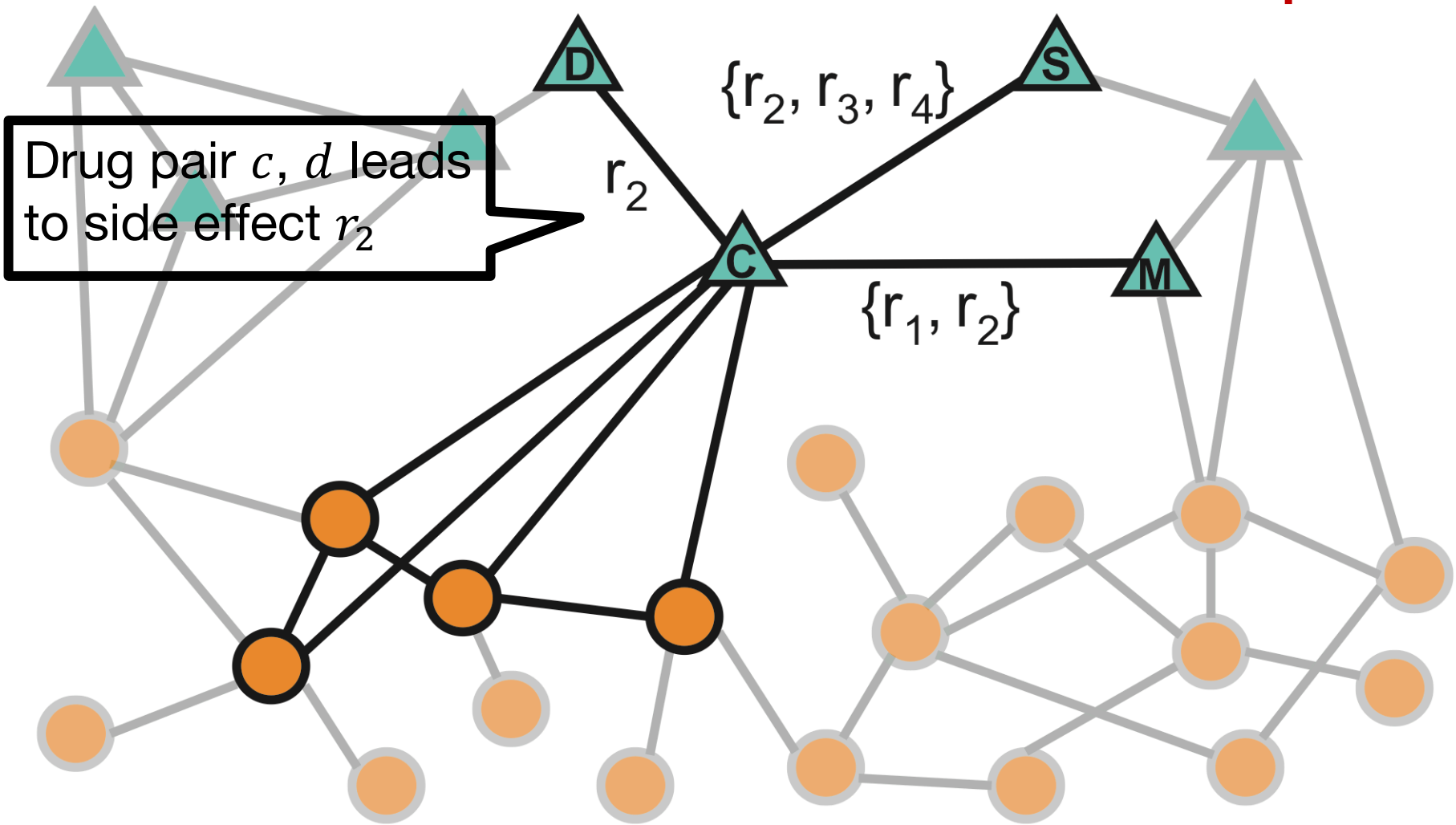
Our Approach

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



Problem Formulation: Graphs

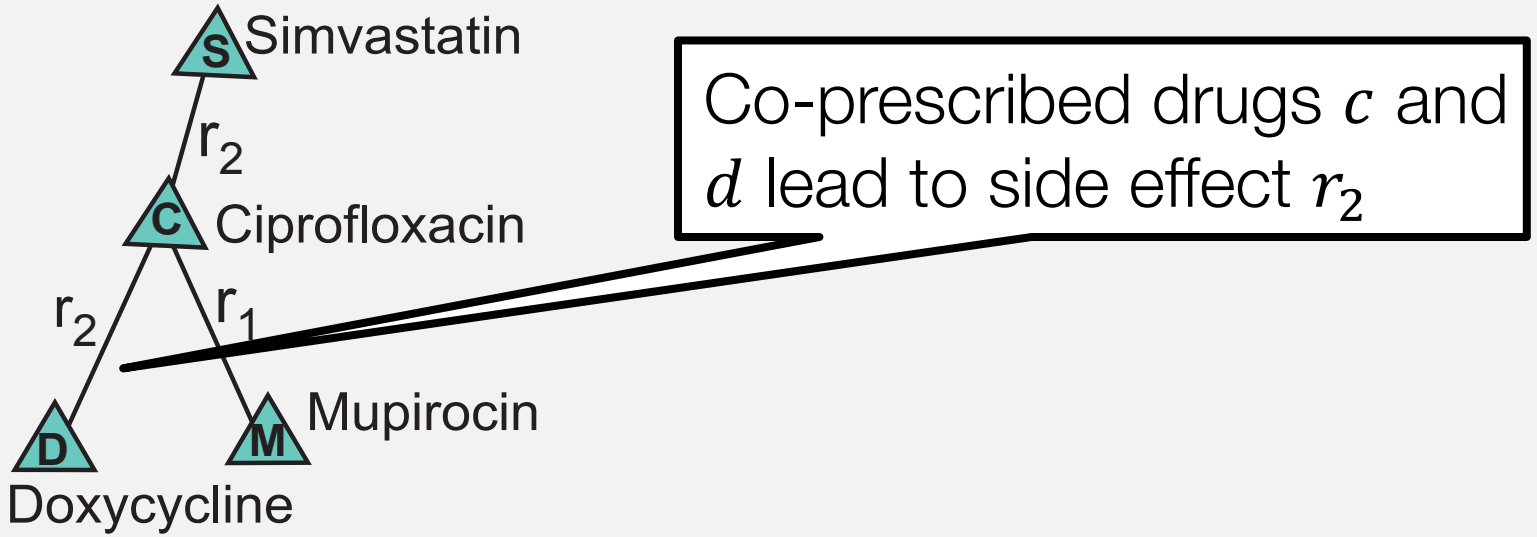


r_1 Gastrointestinal bleed side effect
r_2 Bradycardia side effect
r_3 Nausea side effect
r_4 Mumps side effect
Drug-protein interaction
Protein-protein interaction

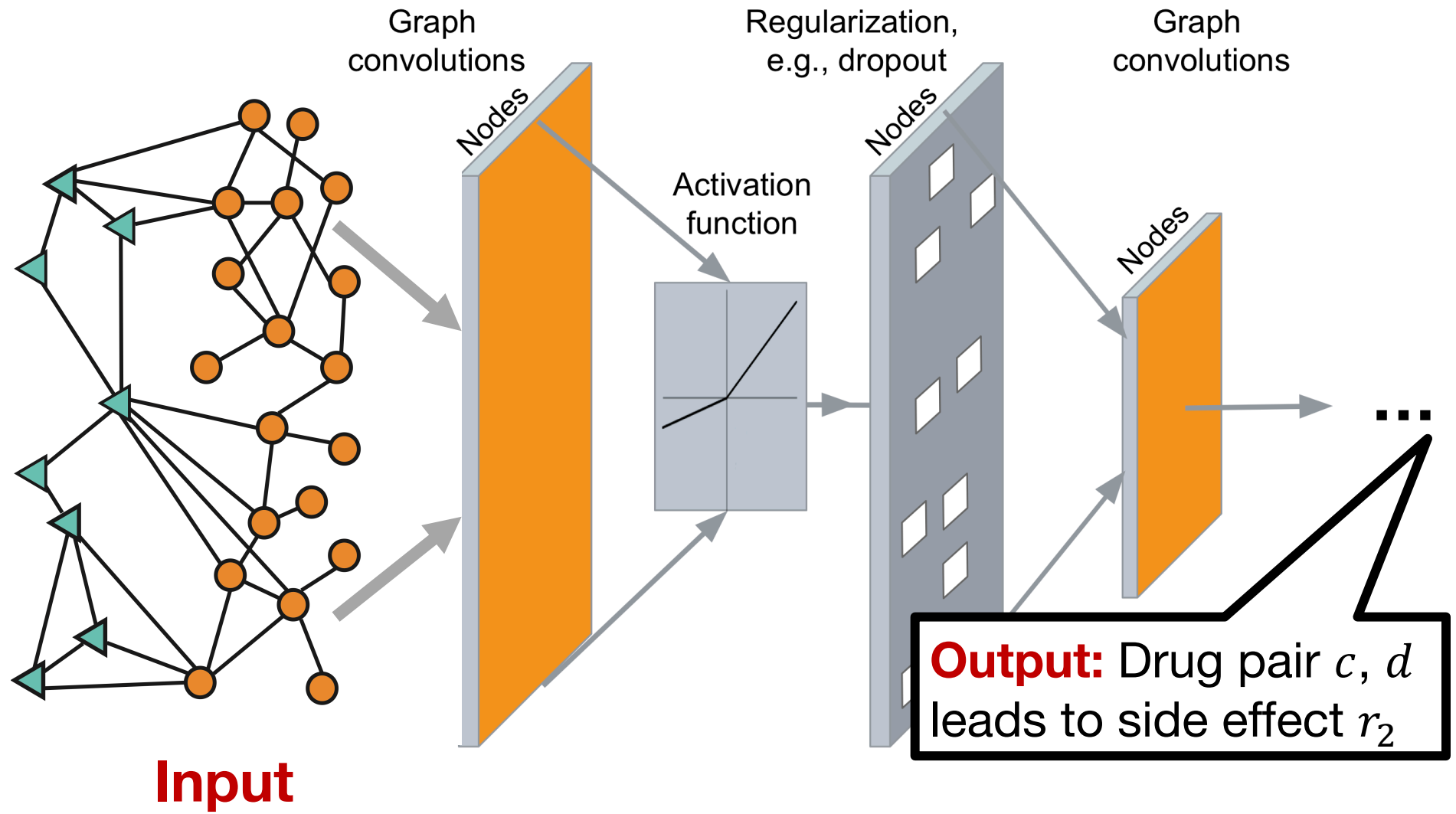
Problem Formulation: Predict

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?

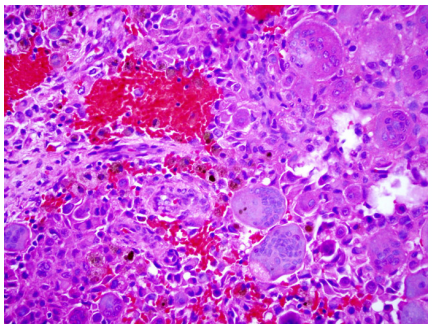


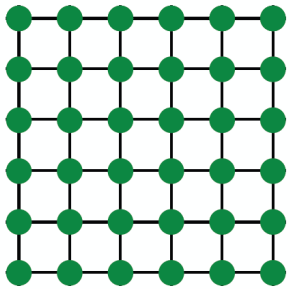
Graph Neural Network

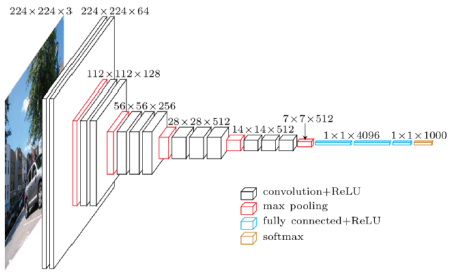


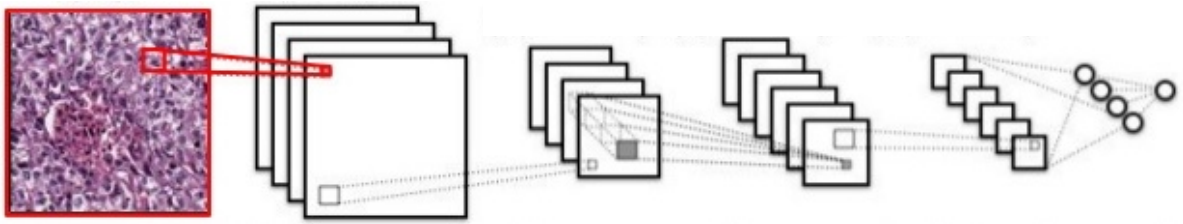
Why Is It Hard?

- Modern deep learning toolbox is designed **for grids** or simple sequences
  - Images have 2D grid structure
  - Can define convolutions (CNN)



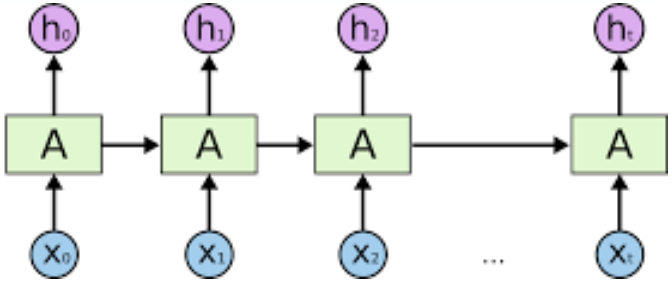






Why Is It Hard?

- Modern deep learning toolbox is designed for grids or **simple sequences**
  - Sequences have linear 1D structure
  - Can define sliding window, RNNs, word2vec, etc.



Why Is It Hard?

- But networks are far more complex!
  - Arbitrary size and complex topological structure (i.e., no spatial locality like grids)

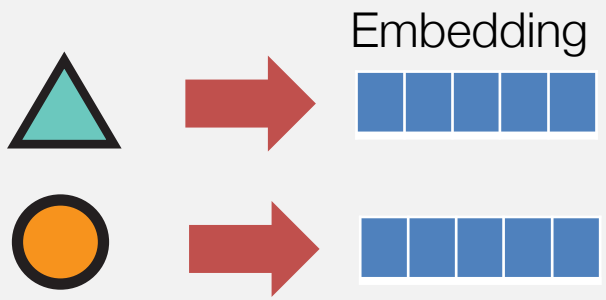
Goal: Generalize convolutions beyond simple lattices

Networks **Images**

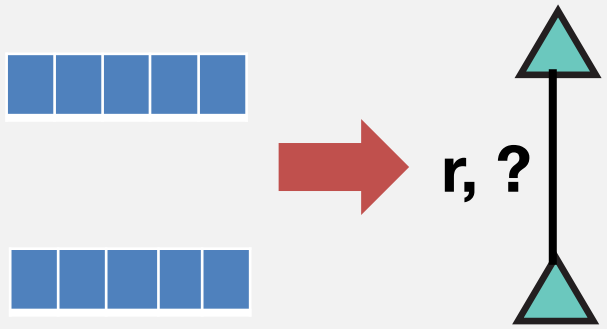
  - No fixed node ordering or reference point
  - Often dynamic and have multimodal features

Decagon: Graph Neural Net

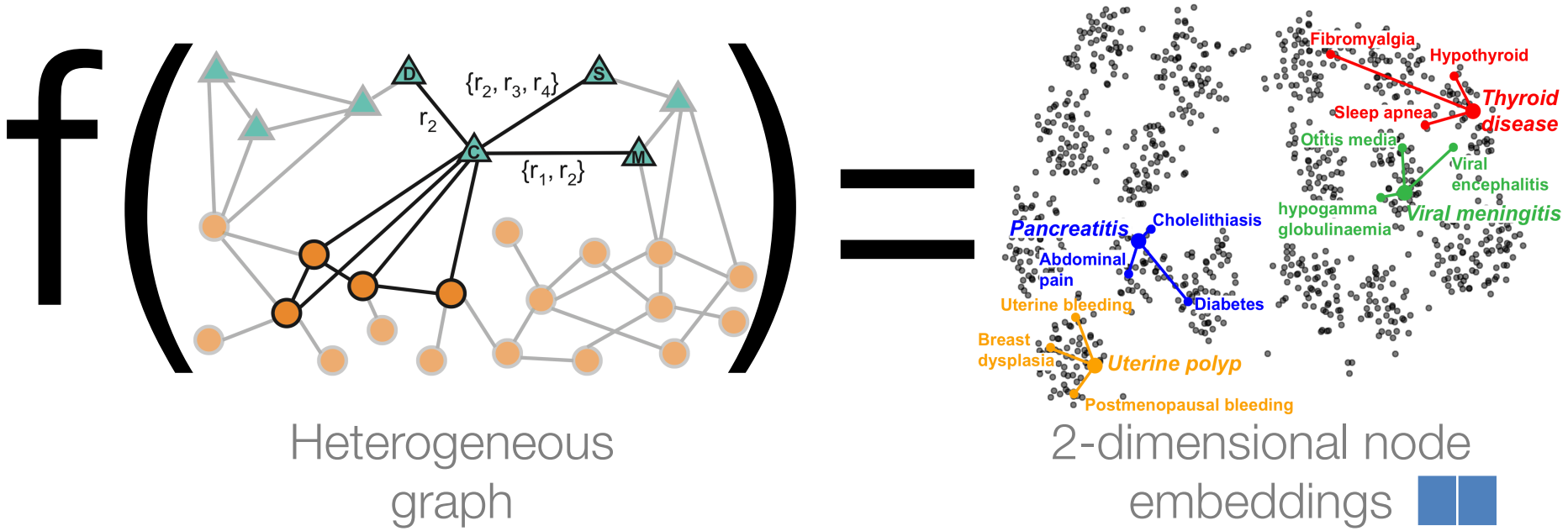
1. **Encoder:** Take the graph and learn an *embedding* for every node



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



Embedding Nodes



Heterogeneous graph

2-dimensional node embeddings

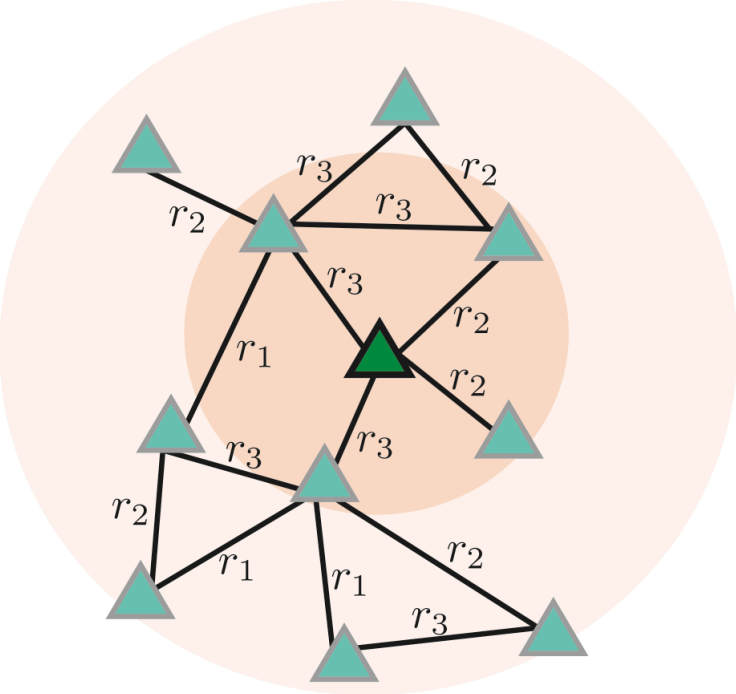
How to learn f?

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

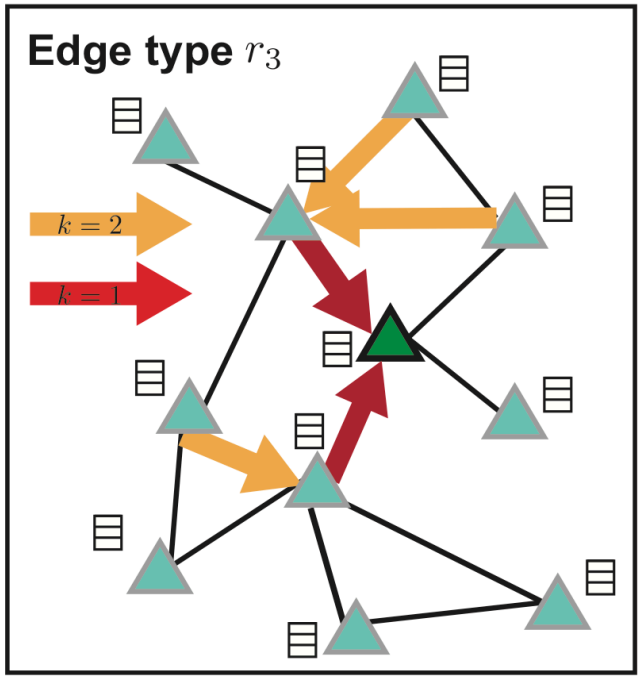
Encoder: Principle

Key idea: Generate node embeddings based on local network neighborhoods

Each edge type is modeled separately

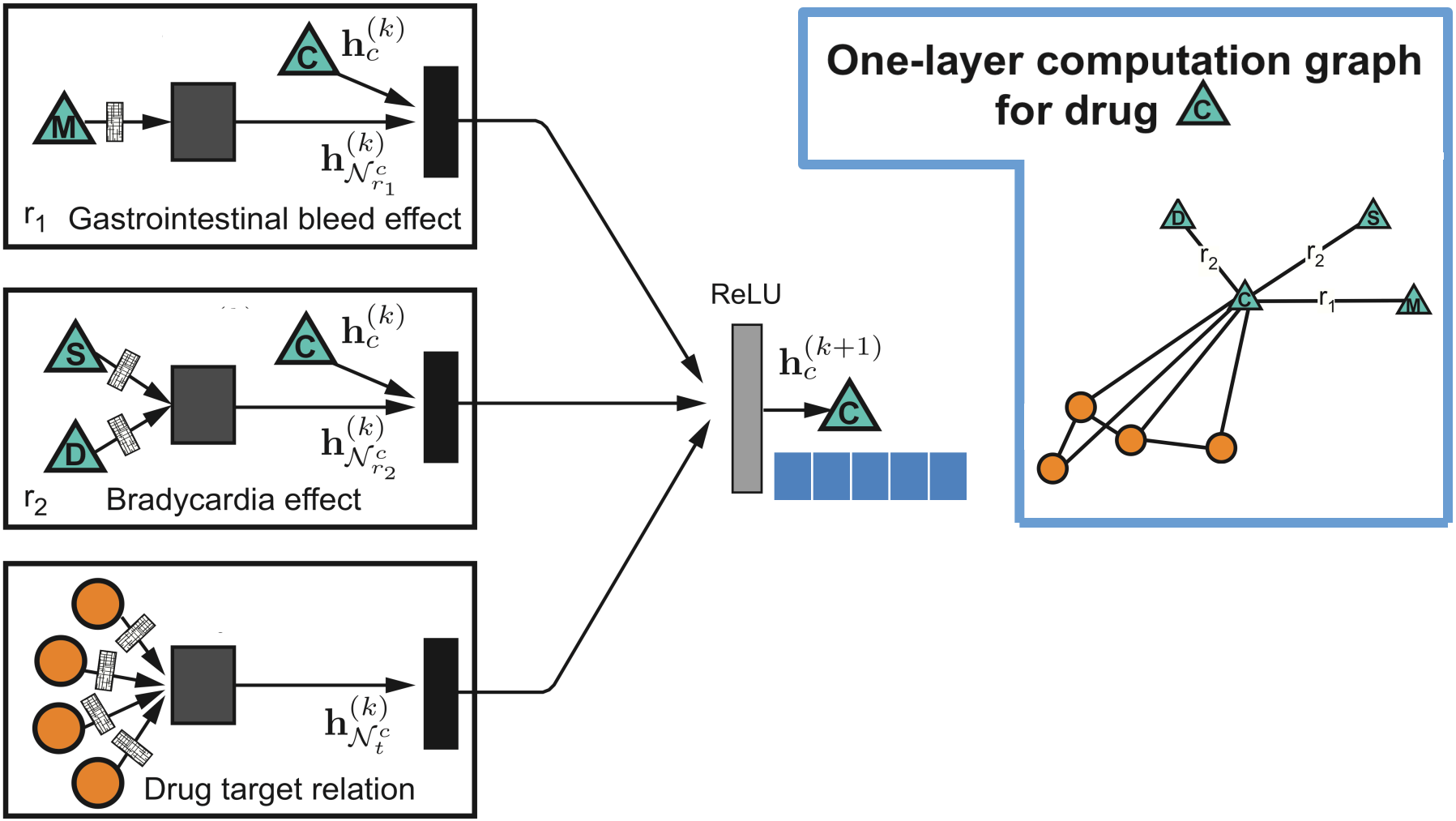


Determine a node's computation graph

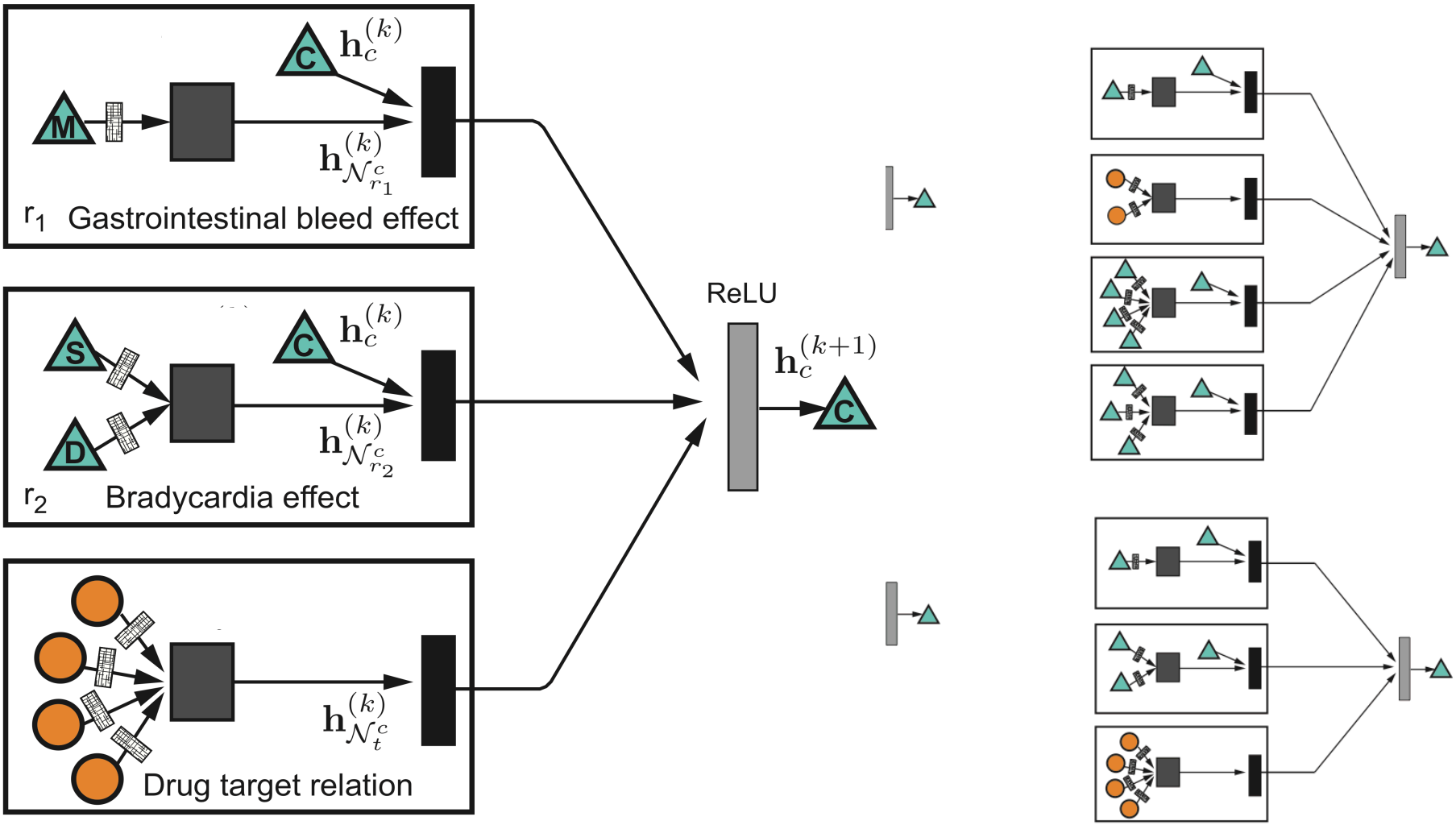


Learn how to transform and propagate information across the graph

Encoder: Embeddings

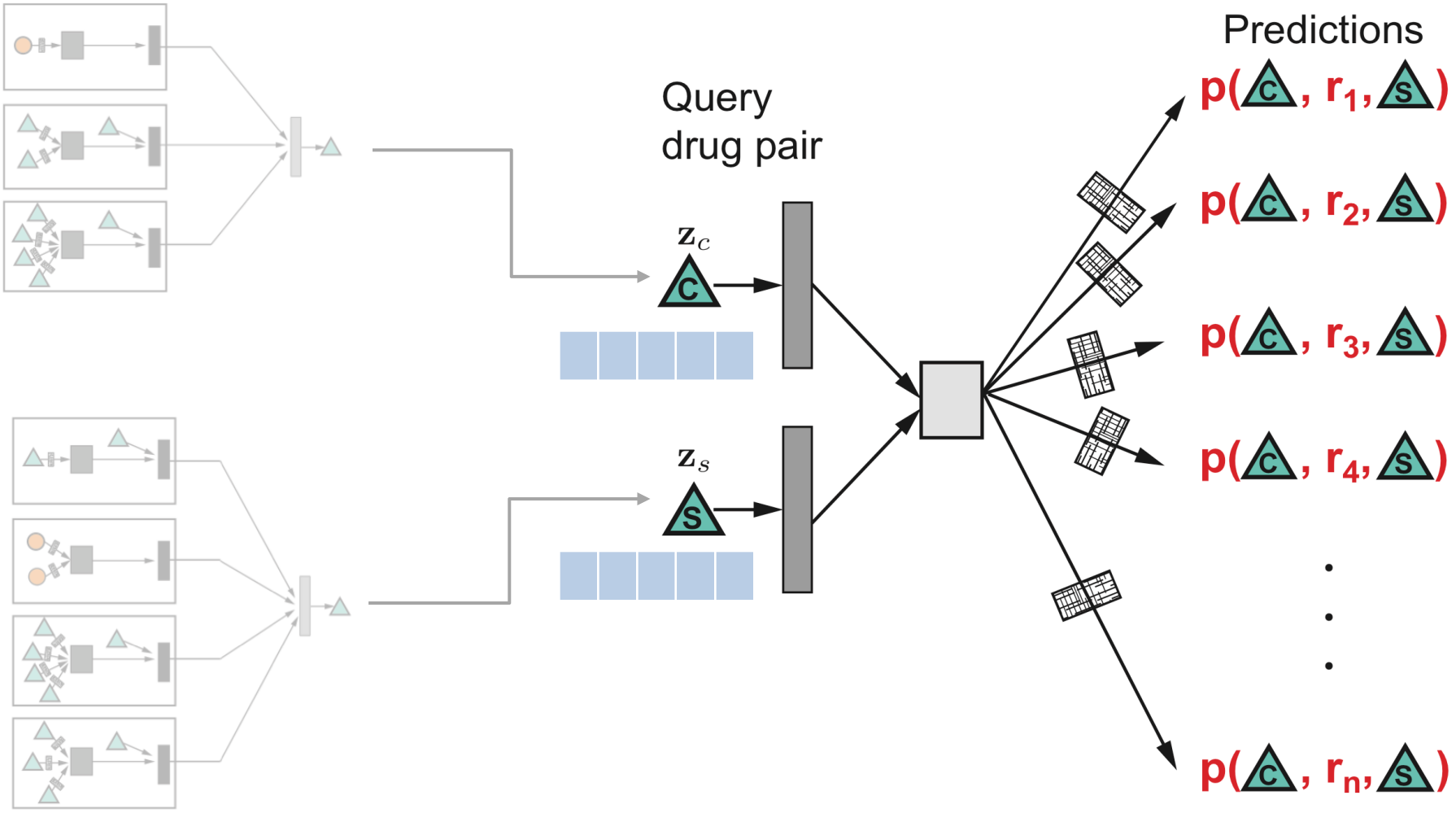


Encoder: Embeddings

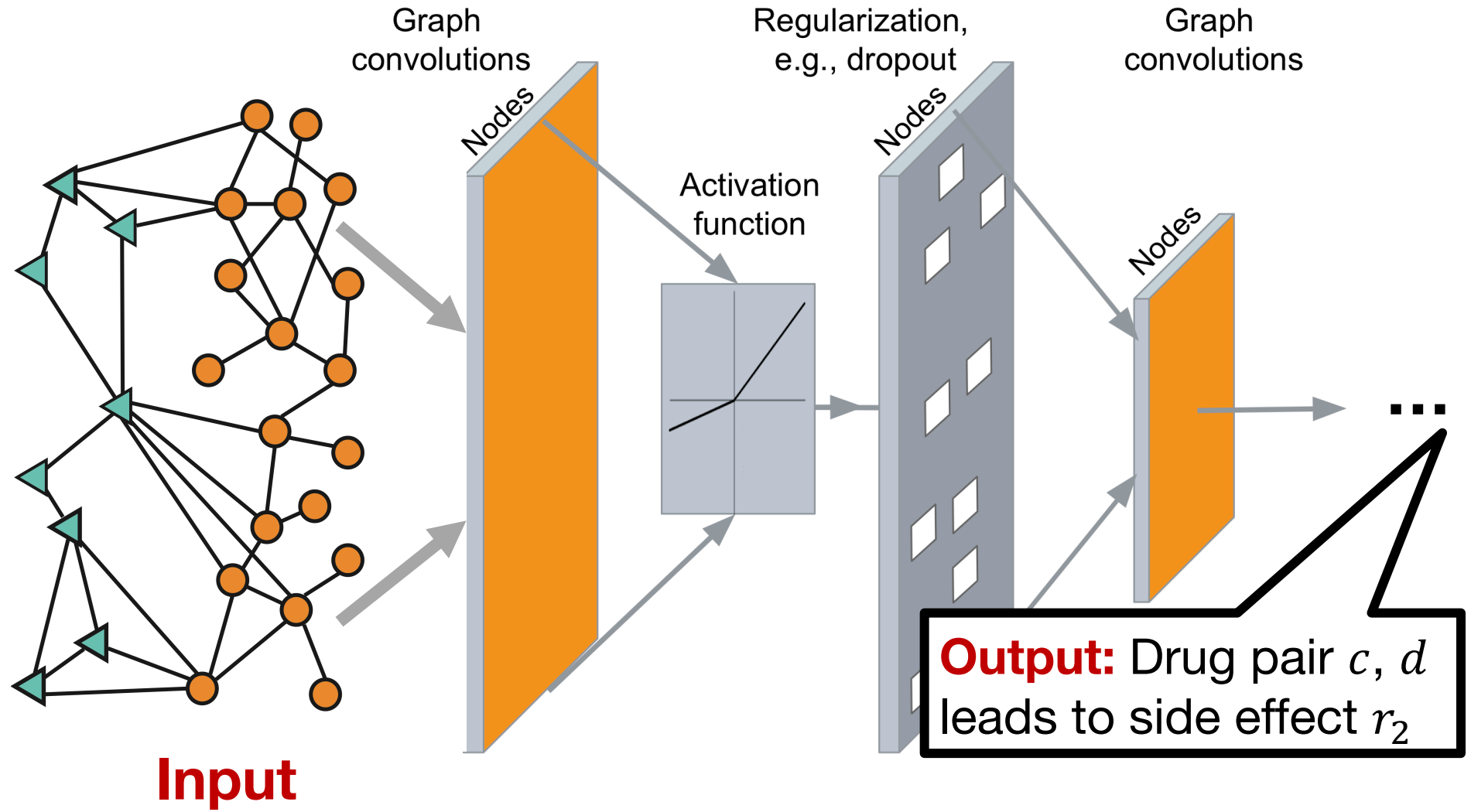


A batch of computation graphs

Decoder: Link Prediction



Graph Neural Network



Deep Learning for Network Biology

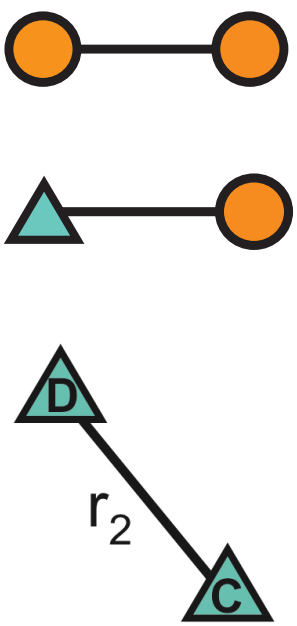
snap.stanford.edu/deepnetbio-ismb

Tutorial at ISMB 2018:

- From basics to state-of-the-art in graph neural nets
- Deep learning **code bases:**
  - End-to-end examples in Tensorflow/PyTorch
  - Popular code bases for graph neural nets
  - Easy to adapt and extend for your application
- Network **analytics tools** and biological **network data**

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



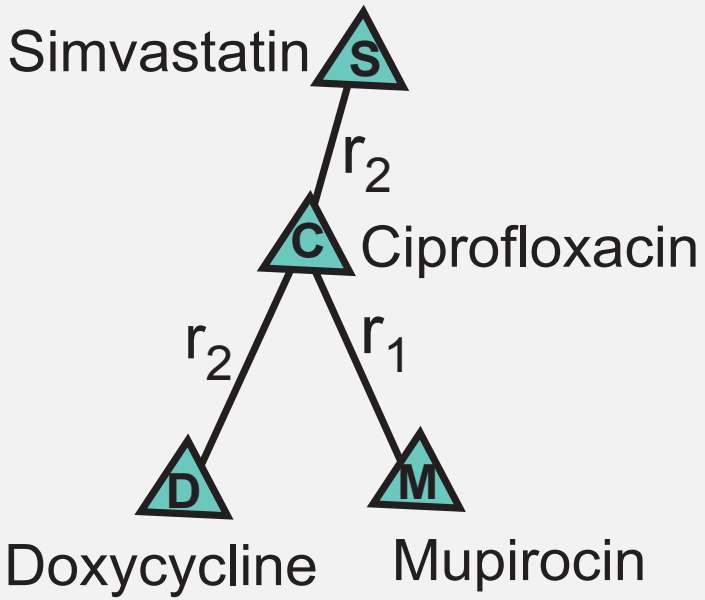
Final graph has **966 different edge types**

Experimental Setup

Construct a heterogeneous graph of all the data

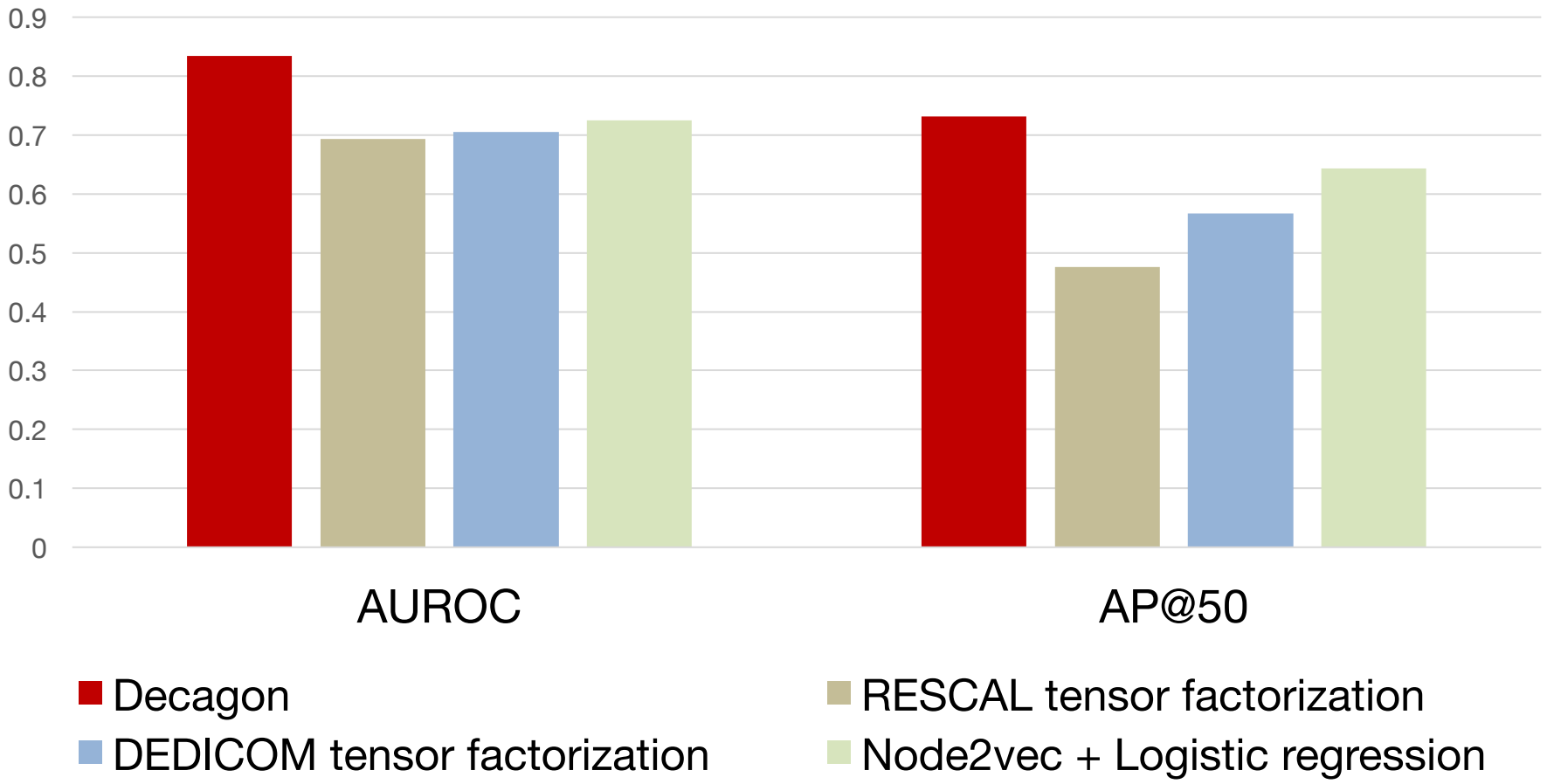
Side-effect centric evaluation:

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



Drug pair c, d leads to side effect r_2

Results: Side Effect Prediction



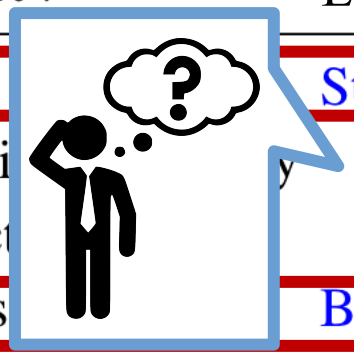
36% average in AP@50 improvement over baselines

De novo Predictions

Rank	Drug c	Drug d	Side effect r
1	Pyrimethamine	Aliskiren	Sarcoma
2	Tigecycline	Bimatoprost	Autonomic neuropathy
3	Omeprazole	Dacarbazine	Telangiectases
4	Tolcapone	Pyrimethamine	Breast disorder
5	Minoxidil	Paricalcitol	Cluster headache
6	Omeprazole	Amoxicillin	Renal tubular acidosis
7	Anagrelide	Azelaic acid	Cerebral thrombosis
8	Atorvastatin	Amlodipine	Muscle inflammation
9	Aliskiren	Tioconazole	Breast inflammation
10	Estradiol	Nadolol	Endometriosis

De novo Predictions

Rank	Drug c	Drug d	Side effect r	Evidence found
1	Pyrimethamine	Aliskiren	Sarcoma	Stage *et al.* 2015
2	Tigecycline	Bimatoprost	Autonomi	
3	Omeprazole	Dacarbazine	Telangiect	
4	Tolcapone	Pyrimethamine	Breast dis	Bicker *et al.* 2017
5	Minoxidil	Paricalcitol	Cluster headache	
6	Omeprazole	Amoxicillin	Renal tubular acidosis	Russo *et al.* 2016
7	Anagrelide	Azelaic acid	Cerebral thrombosis	
8	Atorvastatin	Amlodipine	Muscle inflammation	Banakh *et al.* 2017
9	Aliskiren	Tioconazole	Breast inflammation	Parving *et al.* 2012
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:

- The first method to do that
- Even for drug combinations not yet used in patients

Project website with data & code:
snap.stanford.edu/decagon

Deep learning for network biology:
snap.stanford.edu/deepnetbio-ismb