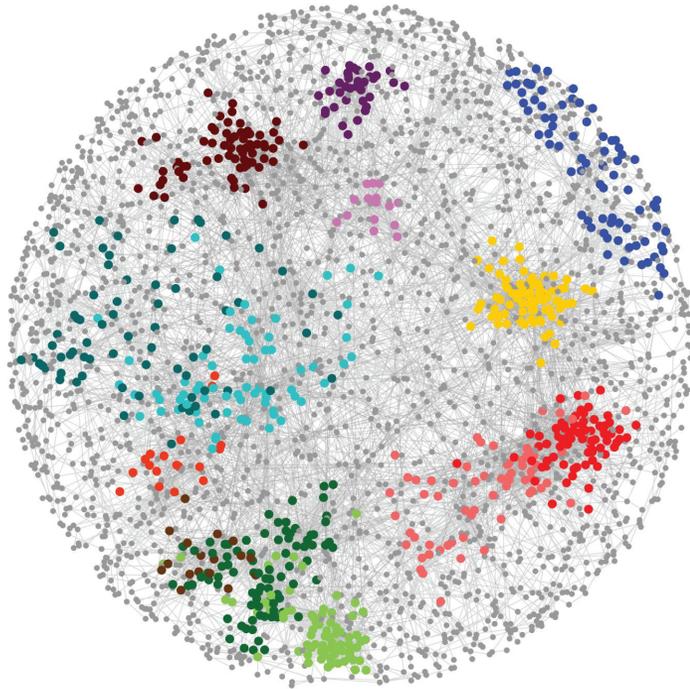# Graph Machine Learning Applications in Biomedicine

Maria Brbić
Leskovec group
Stanford University

Biological systems are naturally represented as networks!
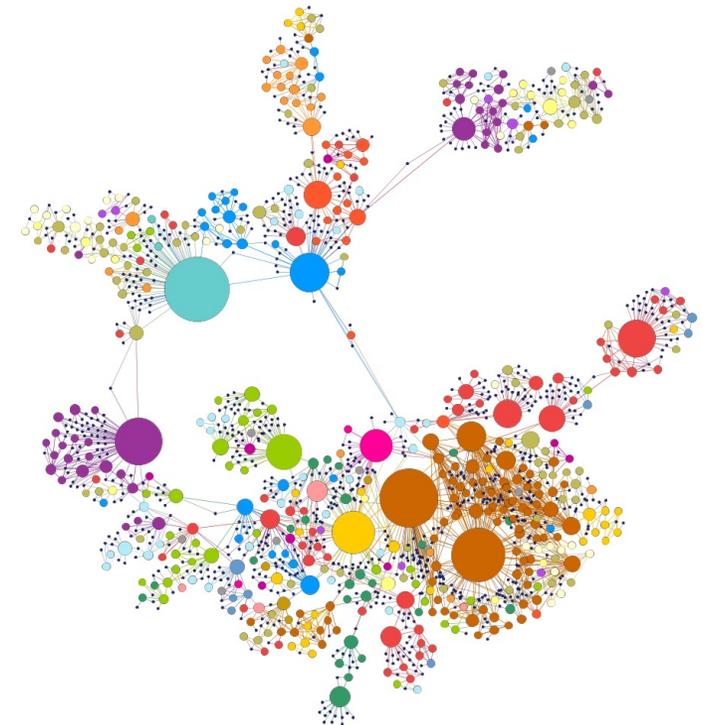
Protein interaction networks   Cell networks   Disease networks

Uncoupled Knowledge

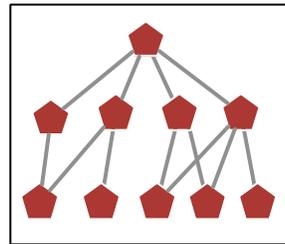Challenge: Many dissociated databases of biological entities

- Drugs: DrugBank, PubChem, ChEBI…
- Disease: MeSH, DiseaseOntology, DDB,…
- Adverse events: MedDRA, ADReCS,…

How can we integrate this knowledge and develop machine learning methods that can reason over it to discover new biology?
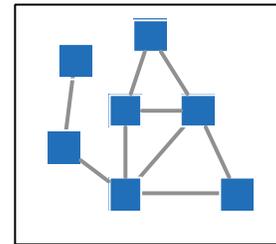
Integrate to Discover
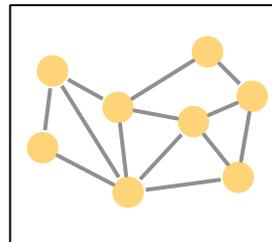
Key: Integrate knowledge to capture complex underlying biological mechanisms
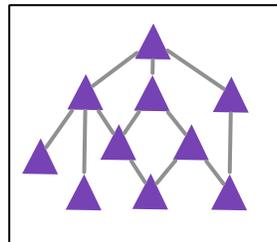
Integrate to Discover

Key: Integrate knowledge to capture complex underlying biological mechanisms
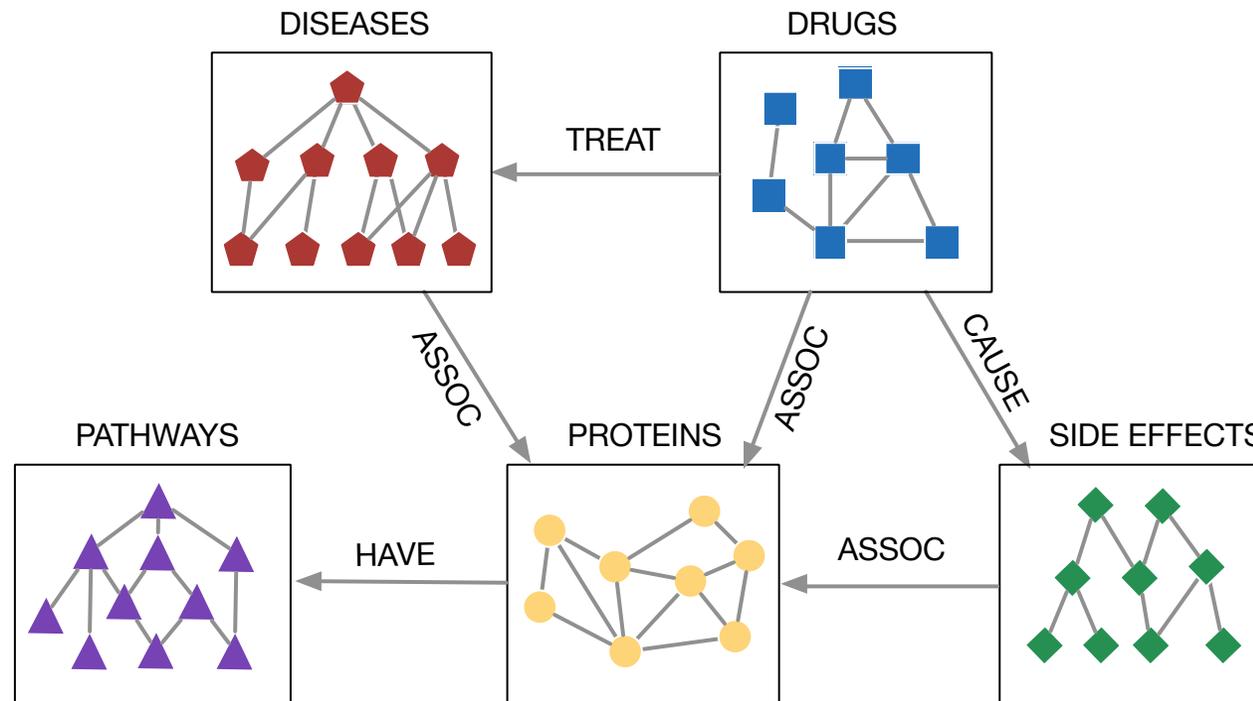
Biological Knowledge Graph
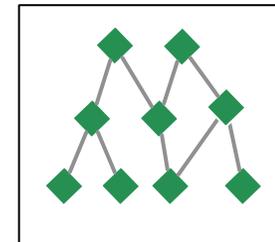
Key idea: Construct knowledge graph that models known biology and learn to reason over it
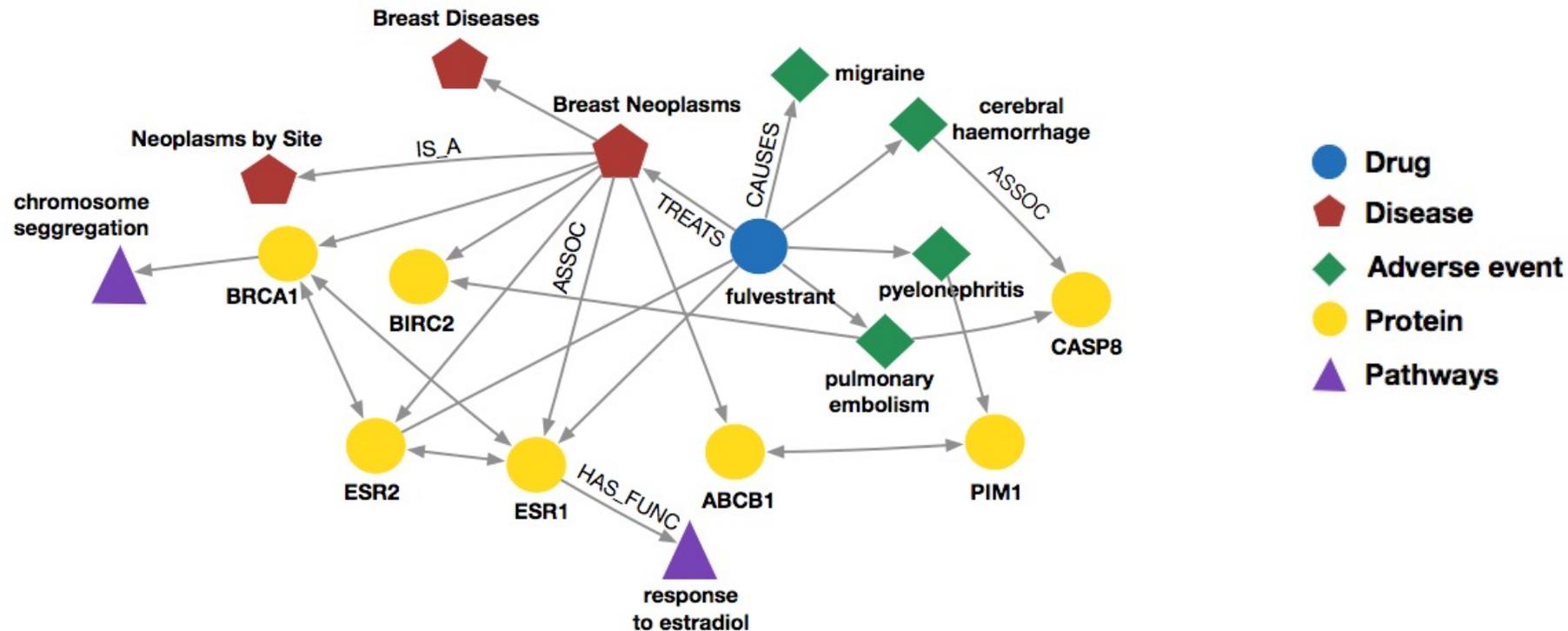
Why Knowledge Graphs?

- We can explicitly store knowledge about underlying biology and chemistry

- Interpretable

- Easy to update and improve
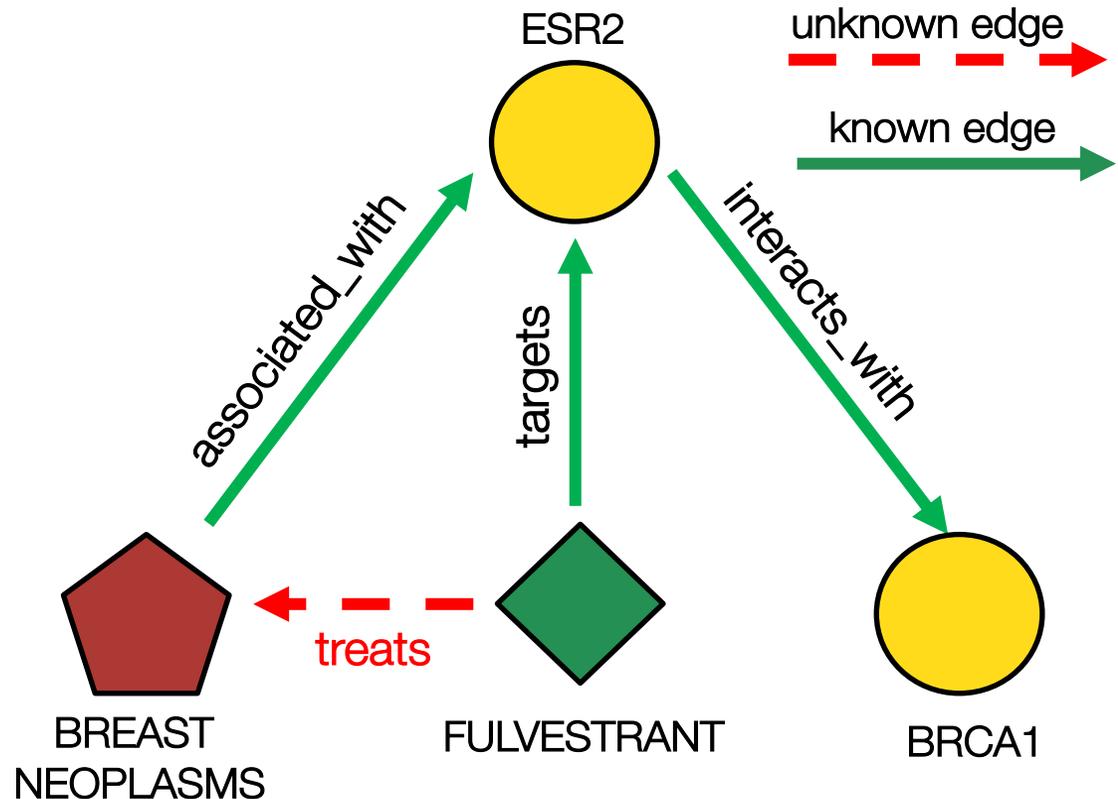
Example: Bio-Knowledge Graph

- Represent facts as triples (h, r, t)
  - ('BRCA1', 'associated_with', 'Breast_Neoplasms')
  - ('Breast_Neoplasms', 'is_a', 'Breast_Disease')
  - …

- Node types: drug, disease, adverse event, protein, functions, …

- Relation types: causes, assoc, treat, interact, …
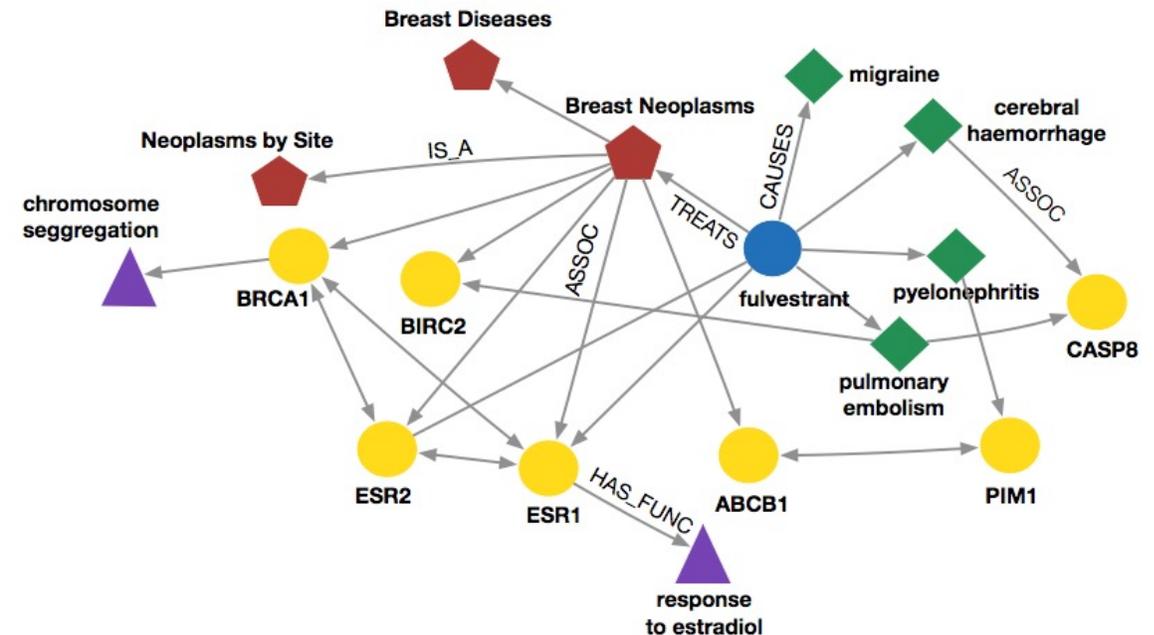
Why Knowledge Graphs?

Discover new biology by predicting missing links in the knowledge graph
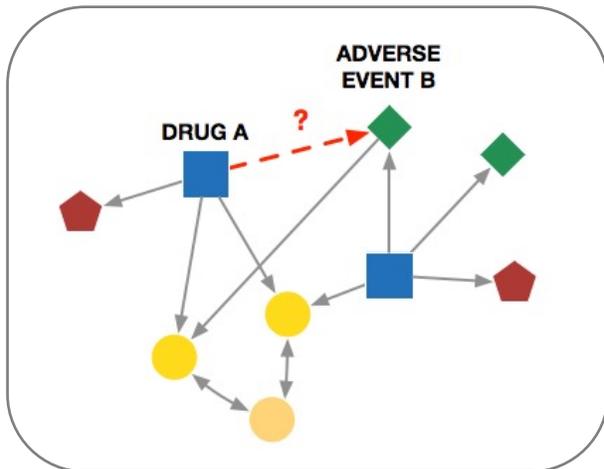
Can drug A cause adverse event B?
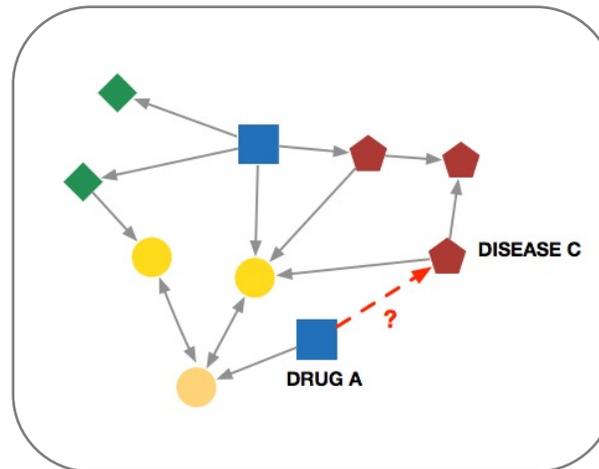
Can drug A treat disease C?
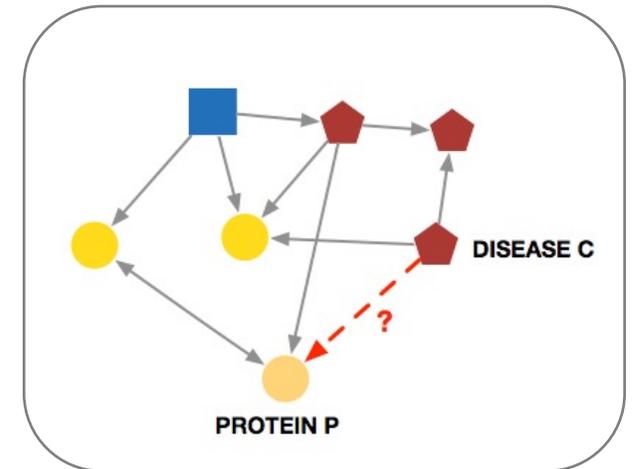
Is disease C associated with protein P?

Building Predictive Models over KGs

How can we leverage knowledge graphs
and build predictive models?

Key idea: Learn to
embed nodes
(e.g., drugs, diseases,
protein)
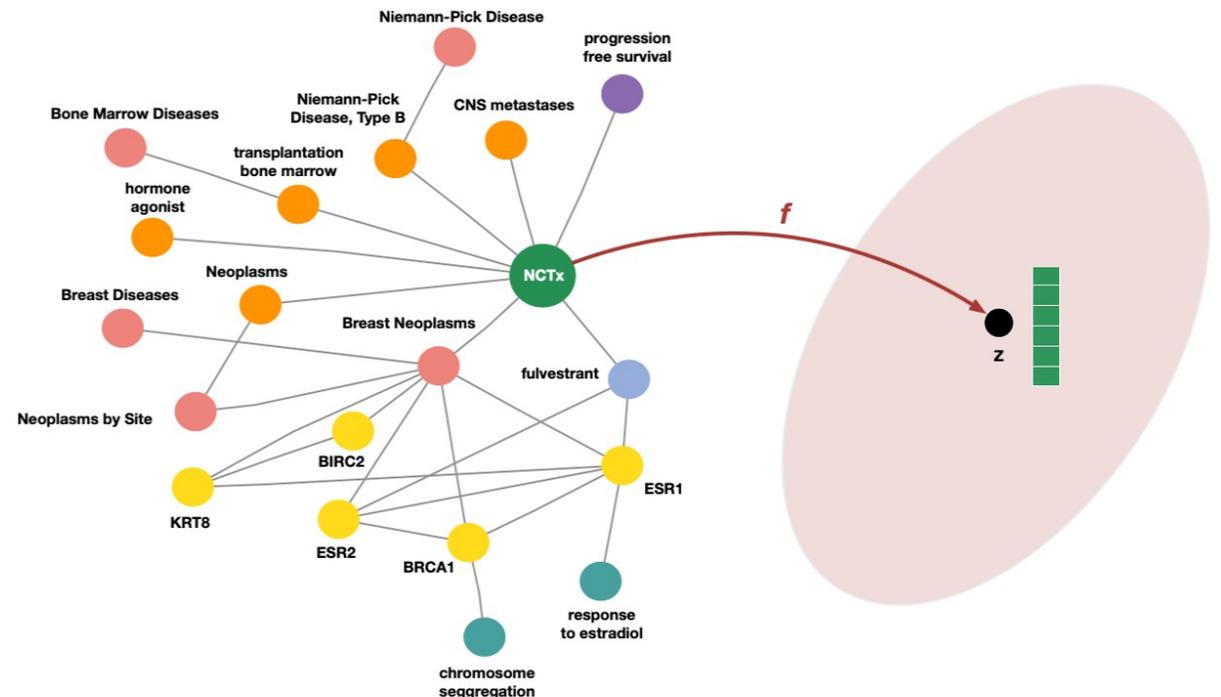
Recap: Graph Convolutional Networks

Idea: Node's neighborhood defines a computation graph
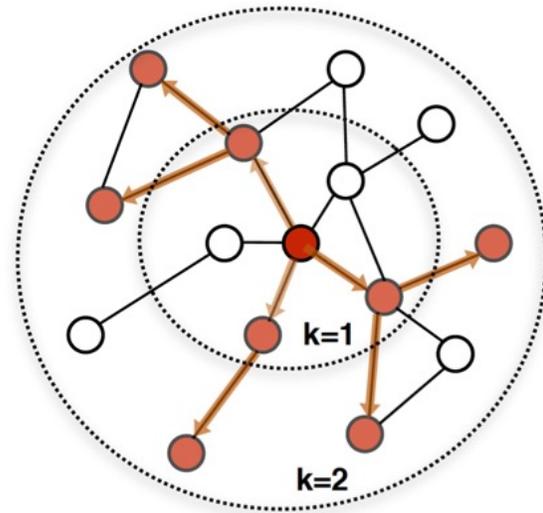
Determine node computation graph
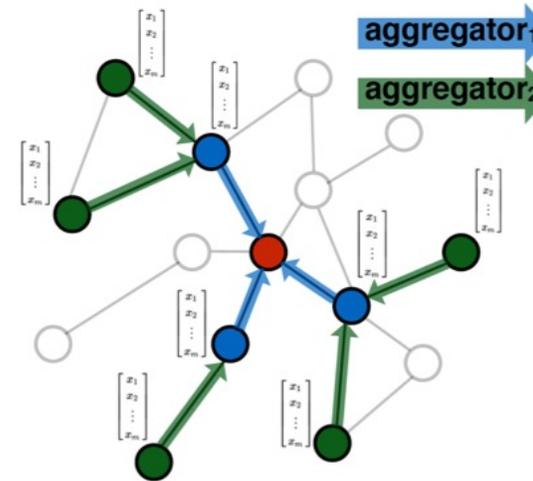
Propagate and transform information

Learn how to propagate information across the graph to compute node features

[Kipf and Welling, *ICLR* 2017]

Relational Graph Convolutional Neural Networks

R-GCN: Learn relation-specific neural network weights

Weights $\mathbf{W}_{r_1}$ for r_1

Weights $\mathbf{W}_{r_2}$ for r_2

Aggregate information from relation-type specific node's neighborhood

[Schlichtkrull, Kipf et al, *Proceedings of ESWC* 2018]
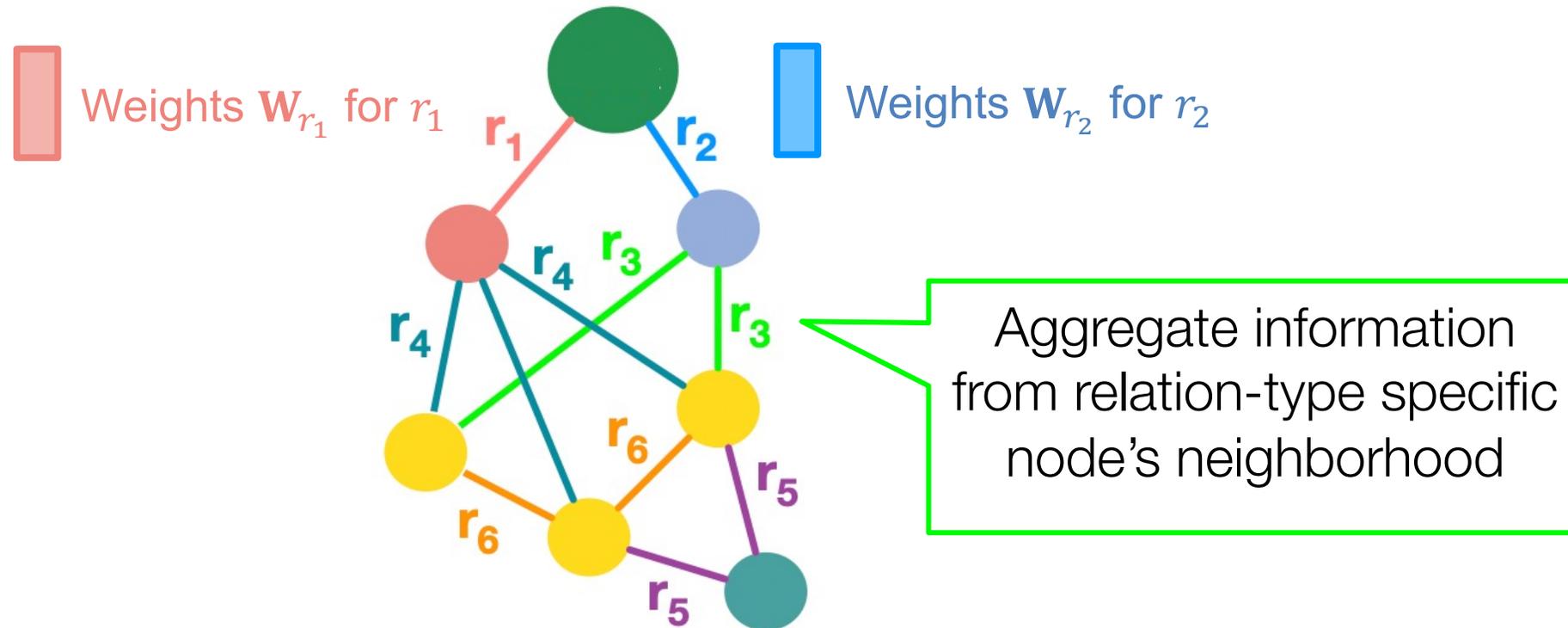
Encoder-Decoder Framework

1) Given a graph, learn a low-dimensional vector (*embedding*) for every node
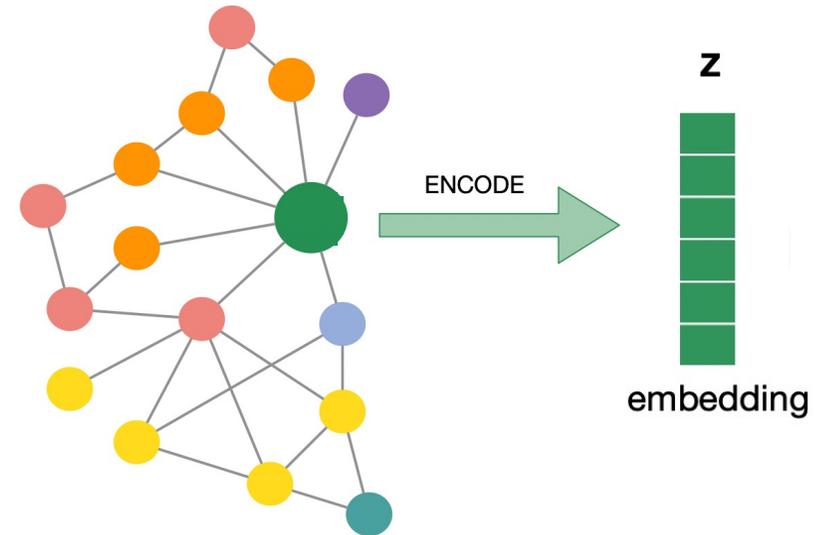
2) Use the learned embeddings for downstream prediction task

z

ENCODE

embedding

z

DOWNSTREAM TASK

embedding
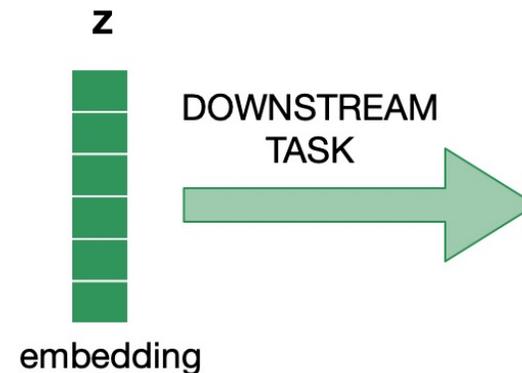
What Tasks Can We Solve?

Example applications:

1) Predicting polypharmacy side effects

2) Predicting drug-disease treatments

3) Predicting outcomes of clinical trials

Predicting drug polypharmacy side-effects

Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018

Modeling Polypharmacy Side Effects

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

Task: Given a pair of drugs predict adverse side effects

[Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018]
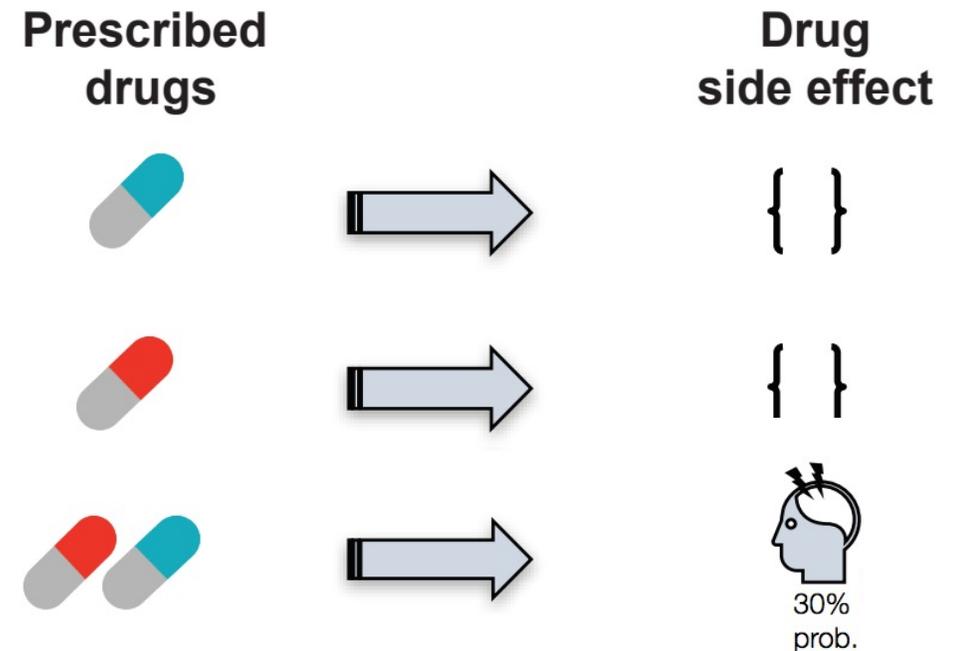
Approach: Build a Graph

$\mathbf{r_i}$ Edge type i

△ Drug node

● Protein node

Drug-drug interaction of type r_2, e.g., nausea

$\{r_2, r_3, r_4\}$

r_2

$\{r_1, r_2\}$

Drug-protein interaction

Protein-protein interaction
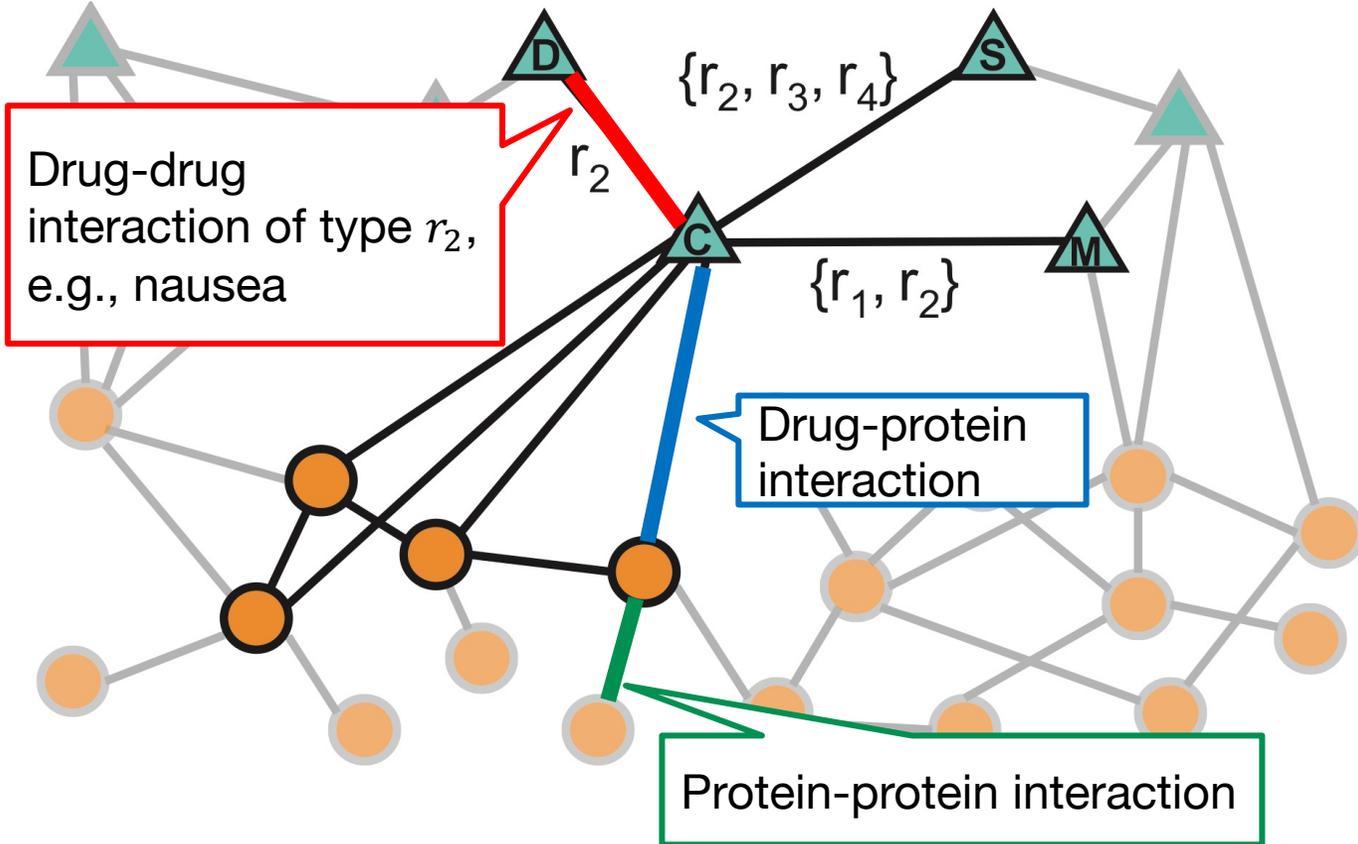
Model:

- Polypharmacy side effects
- Drug-protein interactions
- Protein-protein interactions

[Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018]

Task: Link Prediction

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

Example query: Given drugs c, d, how likely is an edge (c, r_2, d)?

S Simvastatin

r_2

C Ciprofloxacin

r_2 r_1
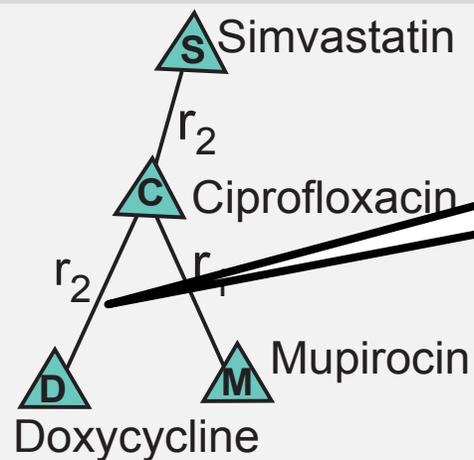
Co-prescribed drugs c and d lead to side effect r_2

M Mupirocin

D Doxycycline

[Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018]

Approach: Drug Side Effects

1) Take the graph and learn a d-dimensional vector (*embedding*) for every node
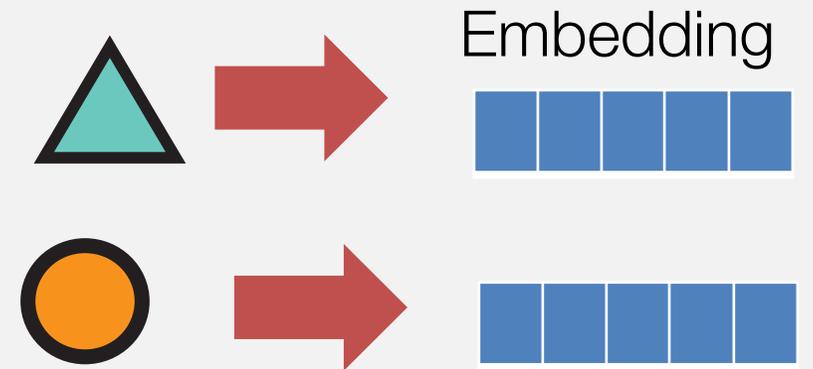
Embedding

2) Use the learned embeddings to predict side effects of drug pairs

r, ?

[Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018]
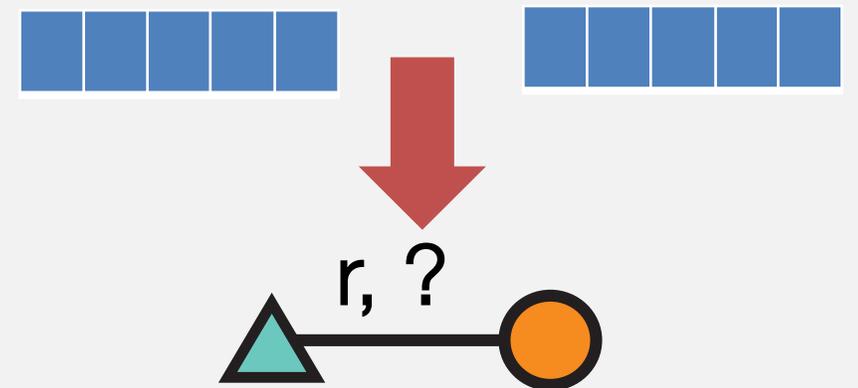
De novo Predictions

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	Autonomic neuropathy	
3	Omeprazole	Dacarbazine	Telangiectases	
4	Tolcapone	Pyrimethamine	Breast disorder	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

Predicting drug-disease treatments

Ruiz, Zitnik, Leskovec. *Nature Communications* 2021

Disease-Drug Treatments

Goal: Understanding drug-disease mechanisms of action

Will the drug treat a disease?

What genes alter drug efficacy?

[Ruiz, Zitnik, Leskovec. *Nature Communications* 2021]

Multi-Scale Interactome

[Ruiz, Zitnik, Leskovec. *Nature Communications* 2021]
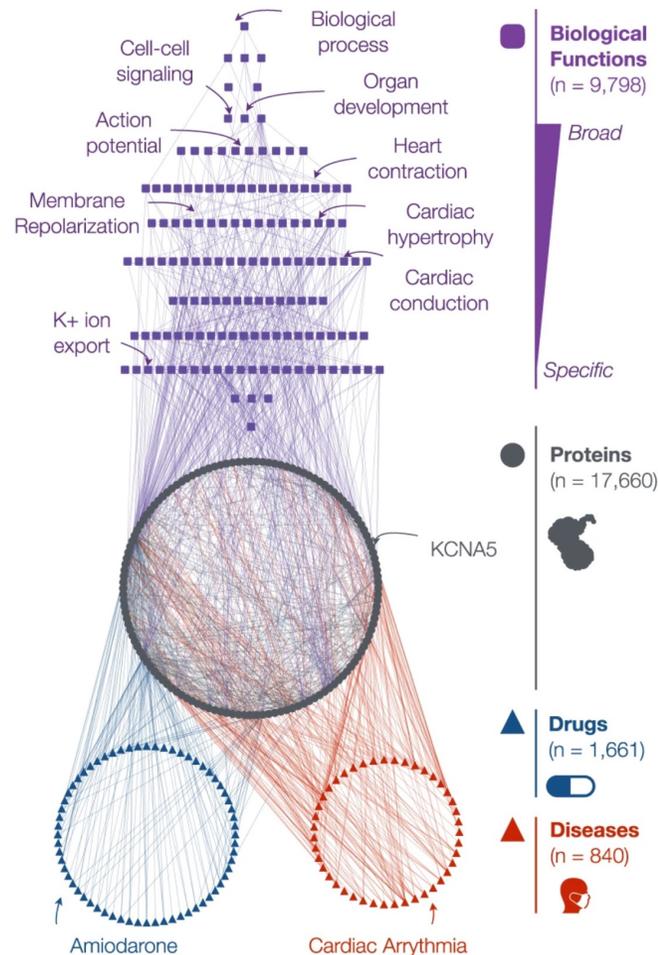
Model drug-disease treatment by integrating proteins and a hierarchy of biological functions

Covers:

- 1,661 drugs

- 840 diseases

- 17,660 proteins

- 9,798 biological functions

Explaining Drug Disease Mechanisms

Approach: Compute network diffusion profiles by biased random walks to explain how drug and disease effects propagate in a biological network

[Ruiz, Zitnik, Leskovec. *Nature Communications* 2021]

Multi-Scale Interactome: Results

Significant improvement over baselines

Interpretability: Can reveal the proteins and biological functions relevant to treatment

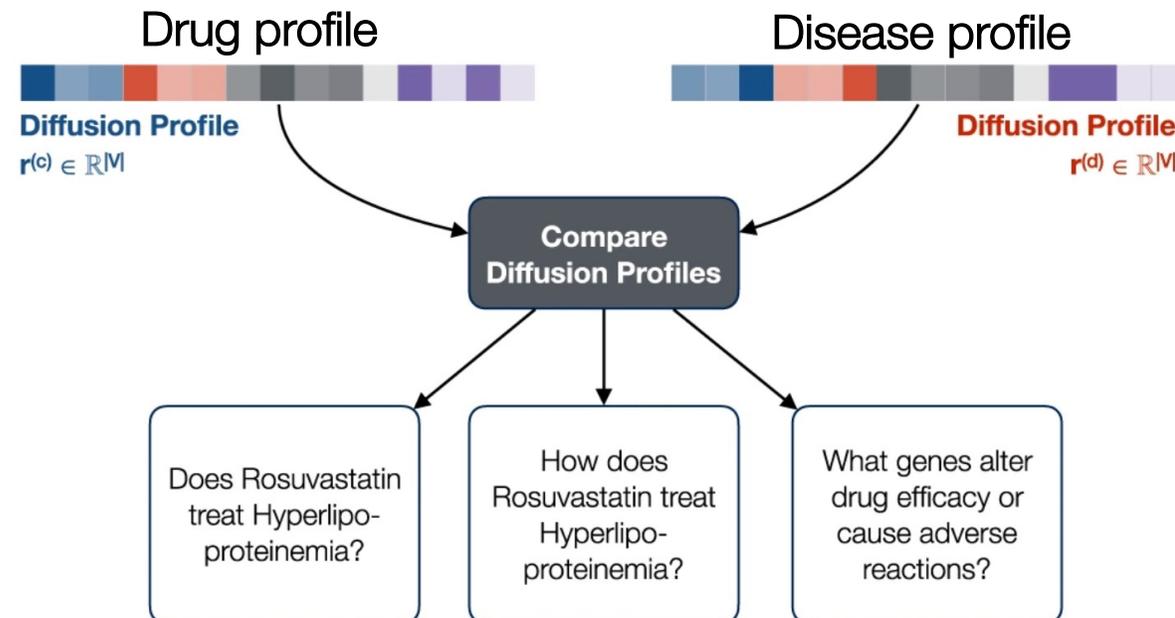| Method	Graph	AUROC	Avg Prec	Rec@ 50
Protein Overlap		0.499	0.064	0.298
Functional Overlap		0.558	0.050	0.237
Molecular-scale Interactome		0.620	0.065	0.264
Multiscale Interactome		**0.705**	**0.091**	**0.347**
Multiscale vs Molecular-scale Interactome		+13.7%	+40.0%	+31.4%

$r^{(c)}$ $r^{(d)}$

k most visited nodes
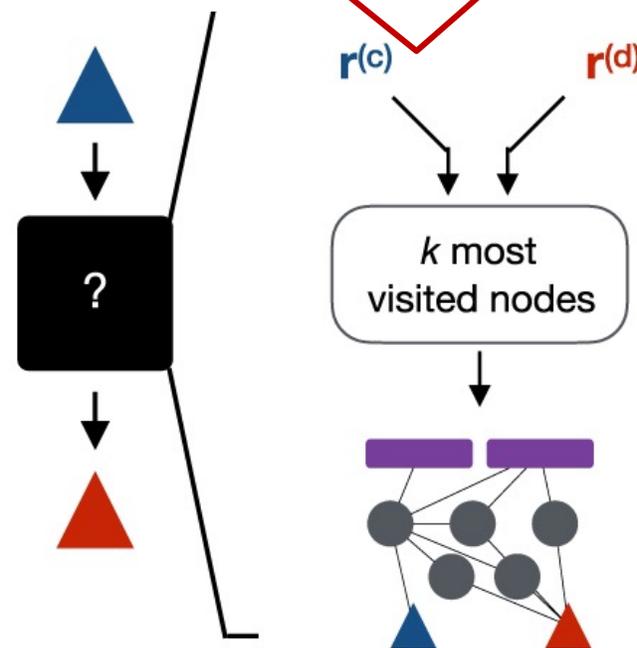
Predicting outcome of clinical trials

Joint work with:
Prabhat Agarwal, Michihiro Yasunaga, Jure Leskovec

Machine Learning for Clinical Trials

Can we use machine learning to guide clinical trials design?

Why is it hard?

- Unstructured and incomplete text
- Challenging to define success
- Multiple patients sub-groups and multiple phases
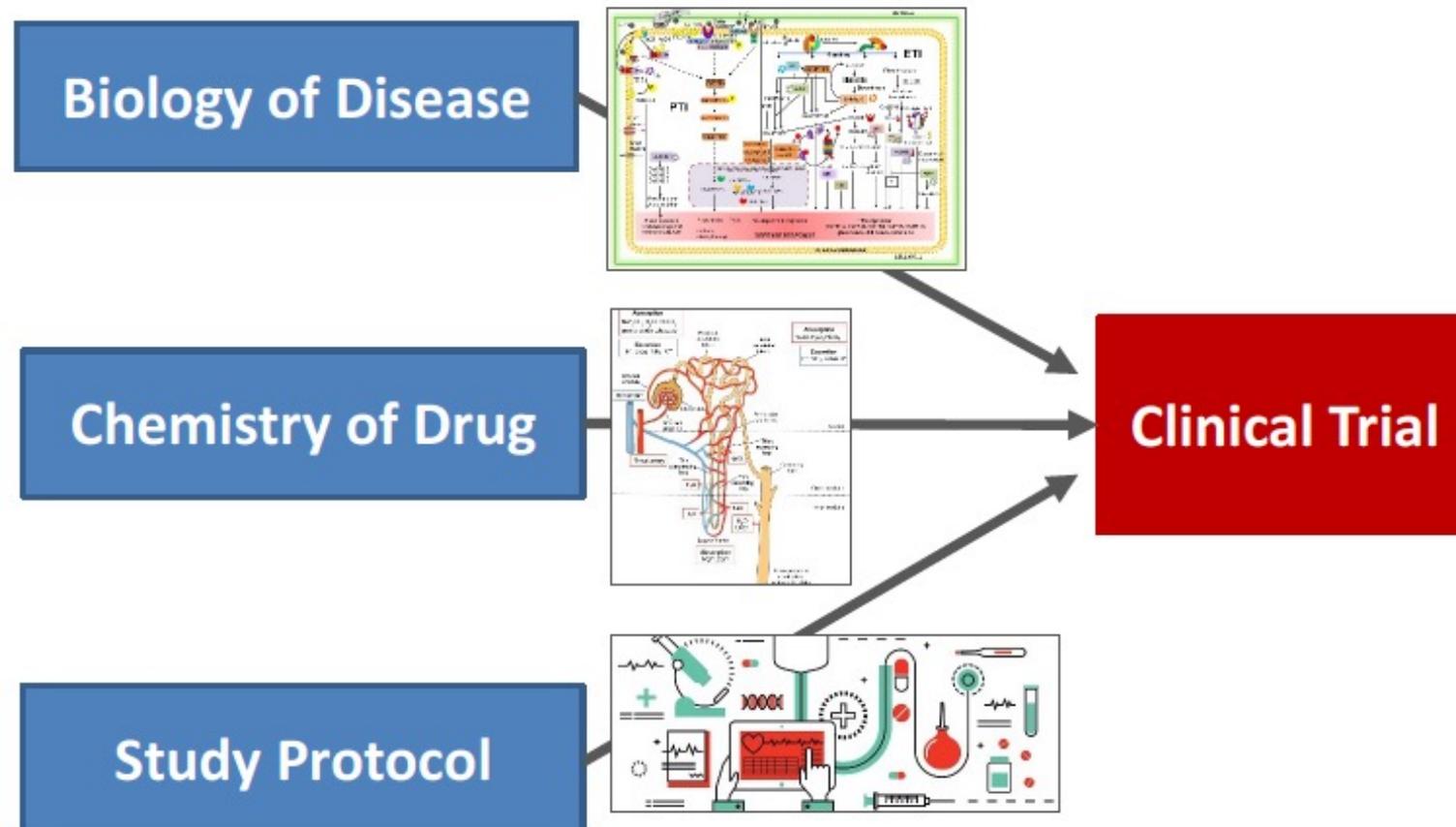- How to model complex biological and chemical mechanisms?

Predicting Outcome of Clinical Trials

- Model and disentangle 3 key factors:

Why It is Hard?

- **Challenge:** Clinical trials database consists of highly unstructured, unlabeled data

Intervention/treatment ⓘ

Drug: LY03003 (Rotigotine, extended-release microspheres)

Patients to be enrolled to 70 mg dose group will receive 14 mg in the first week, 28 mg in the second week, 42 mg in the third week,56 mg in the fourth week and then 70 mg in the next 5 weeks.

Inclusion Criteria:

1. Patient had Parkinson's Disease that meet the clinical diagnostic criteria of the brain bank of the Parkinson's Disease Association of the United Kingdom.
2. Patient was Hoehn & Yahr stage ≤3 (excluding stage 0) ;
3. Patient was male or female aged 18 to 75 years;
4. Patient had a Mini Mental State Examination (MMSE) score of ≥25;
5. Patient had a Unified Parkinson's Disease Rating Scale (UPDRS) motor score (Part III) of ≥10 but ≤30 at Screening.
6. Patient who signed the informed consent form volunteered to participate in this clinical trial and could cooperate with the prescribed inspections.

Primary Outcome Measures ⓘ :

1. Frequency of adverse events [Time Frame: From screening up to day 50]

    Adverse events to evaluate the safety and tolerability of LY03003
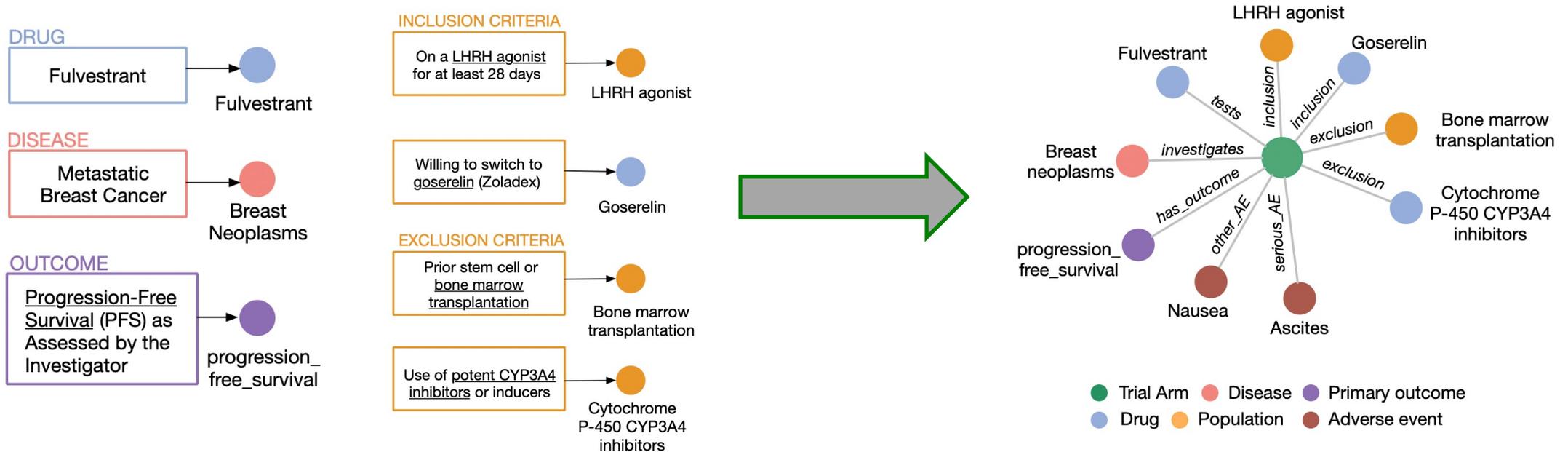
How is drug applied?

Who is eligible to apply?

How is success measured?

How can we structure this data?

Clinical Trials Knowledge Graph

- **Node types:** trial arm, drug, disease, primary outcome, population, adverse events
- **Relation types:** tests, investigates, inclusion, exclusion, has_outcome, has_AE, has_serious_AE

Grounding Clinical Trials KG

Approach: Ground clinical trials knowledge graph in underlying biology and chemistry

- Biology of disease:
  - Disease-associated genes, protein interactions, protein pathways,…

- Chemistry of drug:
  - Drug target and off-target proteins, chemical structure, side effects, …

TrialNet Knowledge Graph

Legend: Trial Arm, Disease, Primary outcome, Drug class, Protein, Drug, Population, Adverse event, Function

- Covers 70k interventional clinical trials

- 300k nodes (clinical trials protocol entities, biological and chemical entities)

- 10.8 millions edges across 18 relation types

Prediction Tasks

- Efficacy: How likely will a new drug surpass existing treatments?

$$f(\overset{z_1}{\blacksquare}, \overset{z_2}{\blacksquare}) = \{1,2\}$$

Predict which treatment will result in higher progression-free-survival

- Safety: Is the drug safe for use?

$$f(\blacksquare) = \{0,1\}$$

Predict which adverse events will happen in a trial

TrialNet: EfficacyResults

- Our model significantly outperforms PubMedBERT
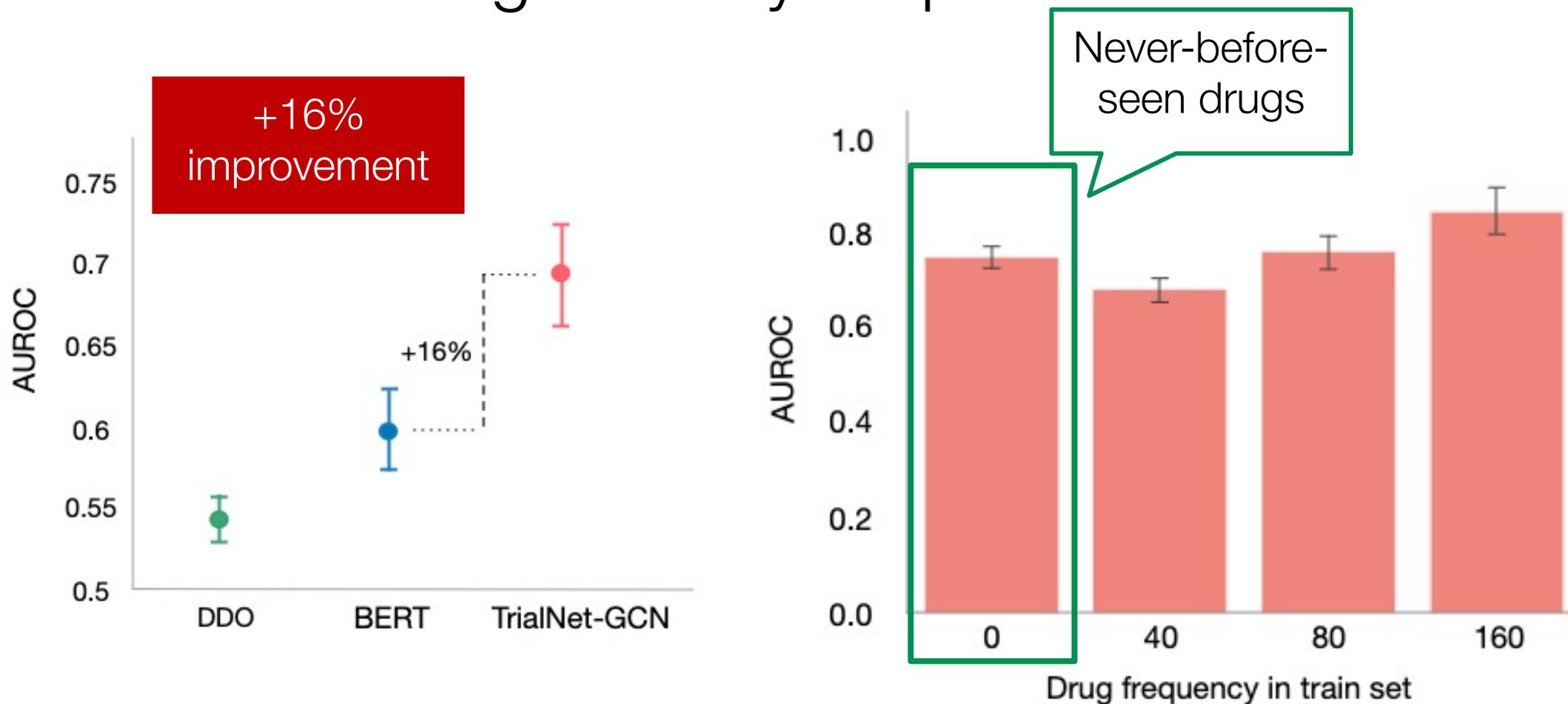
Takeaways

- Knowledge graphs are powerful representation for storing and capturing known biology

- Graph neural networks that operate on heterogneous graphs can be used to build predictive models

- We can predict drug safety, drug efficacy, outcome of clinical trials

Acknowledgements

SNAP group

Jure Leskovec (PI)
Michihiro Yasunaga
Prabhat Agarwal
Camilo Ruiz
Antoine Bosselut
Adrijan Bradaschia
Kaidi Cao
Serina Chang
Weihua Hu
Hamed Nilforoshan
Alex Porter
Hongyu Ren
Yusuf Roohani
Rok Sosic
Tailin Wu
Rex Ying
Jiaxuan You

Jure Leskovec
Stanford CS,
CZ Biohub

Collaborators:

Marinka Zitnik, Harvard
Russ Altman, Stanford Bioengineering
Hongjie Li, Baylor College of Medicine
Liqun Luo, Stanford Biology
Colleen McLaughlin, Stanford Biology
Angela Pisco, CZ Biohub
Stephen Quake, Stanford Biology
ChiChi Xie, Stanford Biology