Graph Neural Networks for Drug Development

Marinka Zitnik
marinka@hms.harvard.edu

HARVARD
UNIVERSITY

Stanford | ENGINEERING
Computer Science

Drug Development

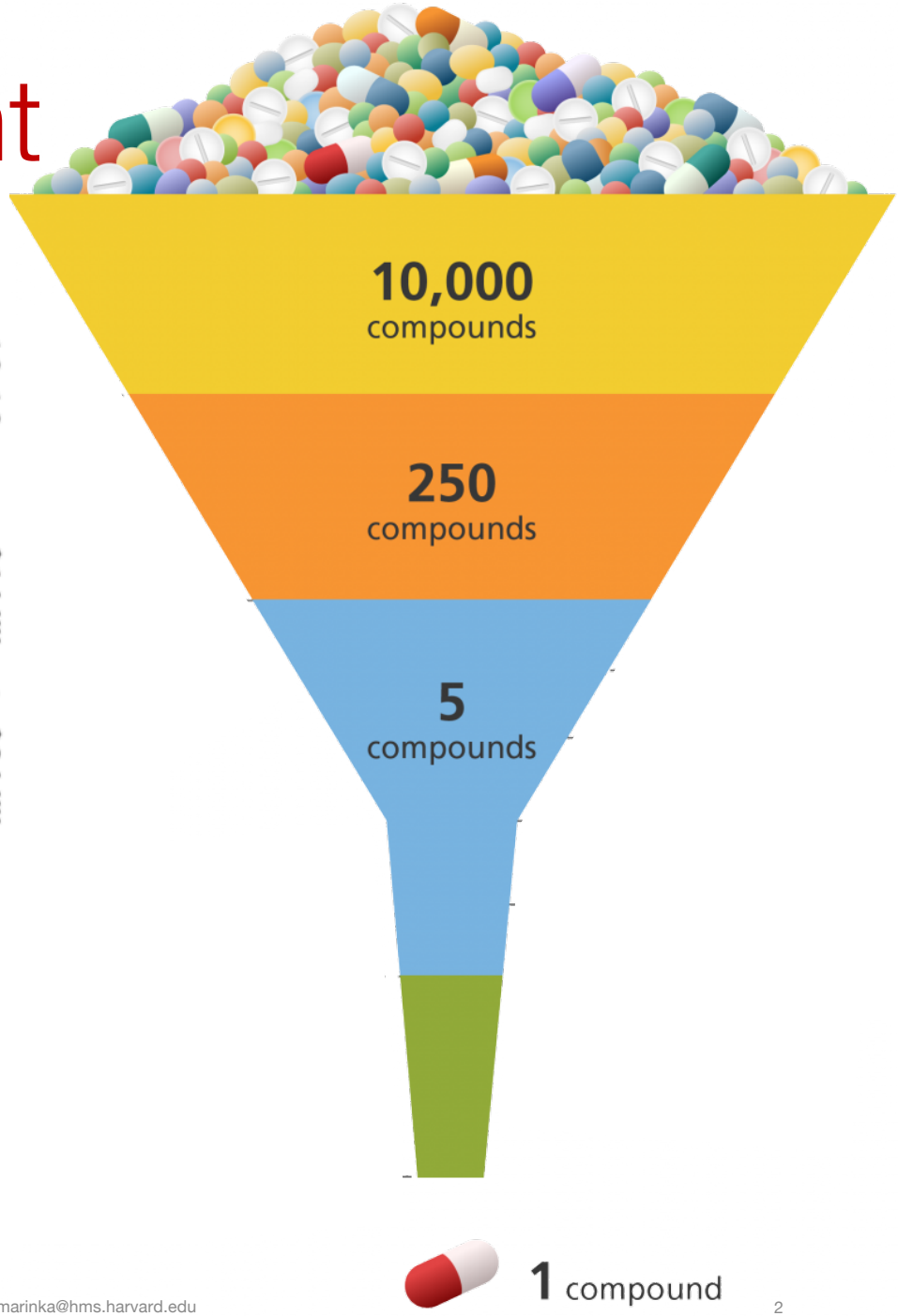
Step 1: Design and Discovery

Step 2: Preclinical Research

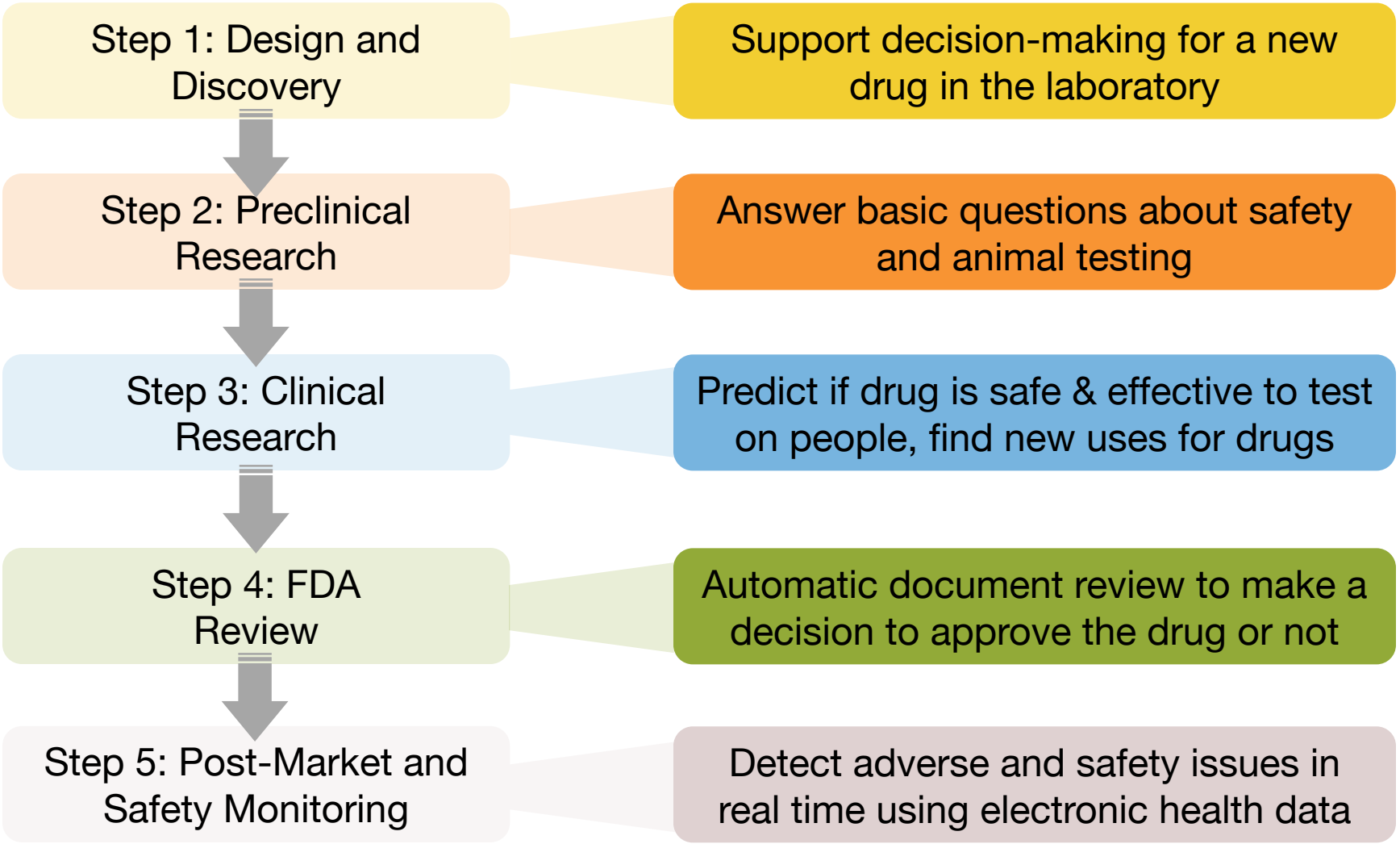
Step 3: Clinical Research

Step 4: FDA Review

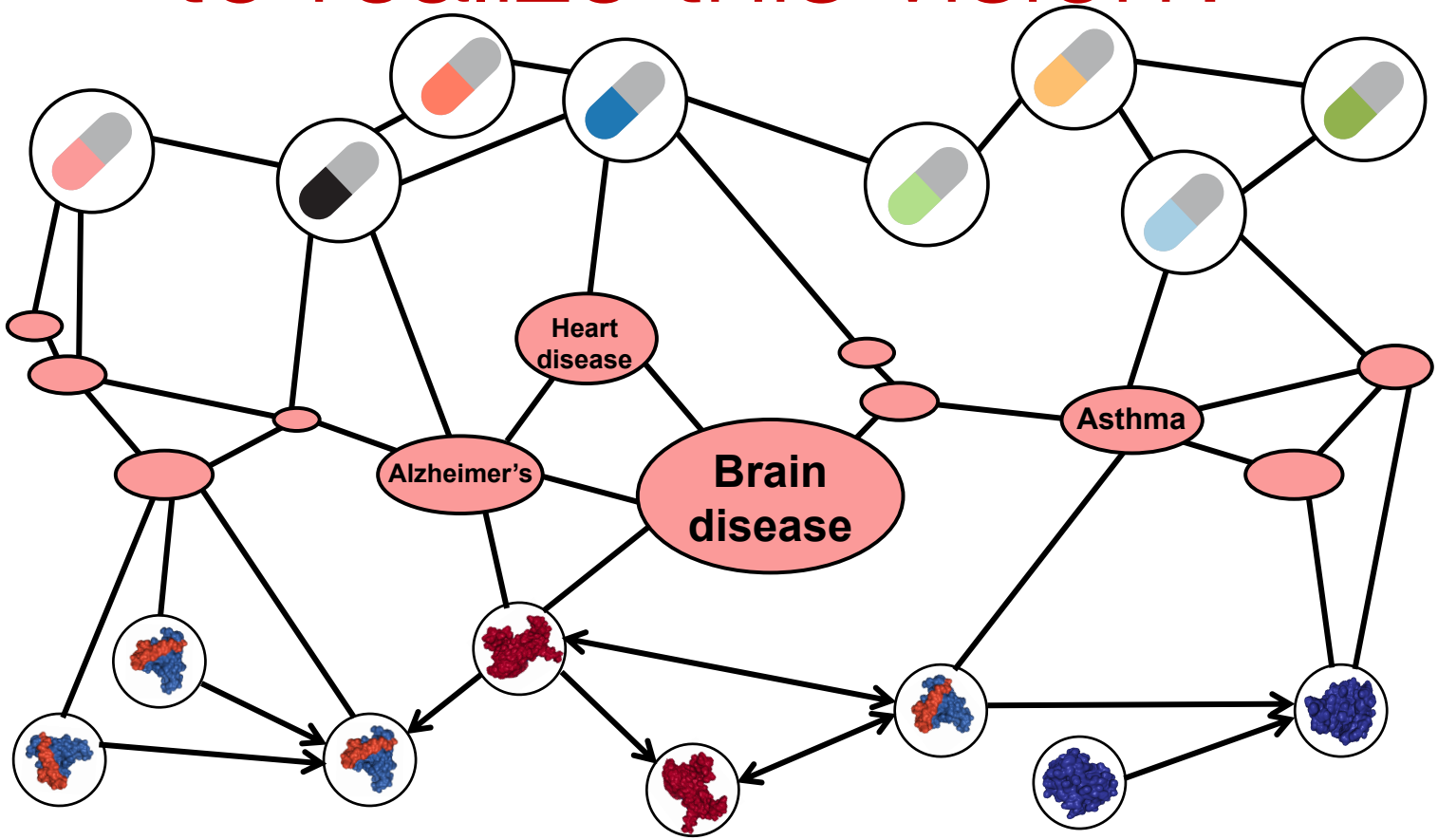
Step 5: Post-Market and Safety Monitoring



Opportunities for AI in Drug Development



Why is it so challenging to realize this vision?



Finding drugs for disease treatments relies on several types of interactions, *e.g.,* drug-target, protein-protein, drug-drug, drug-disease, disease-protein pairs

Today's Talk

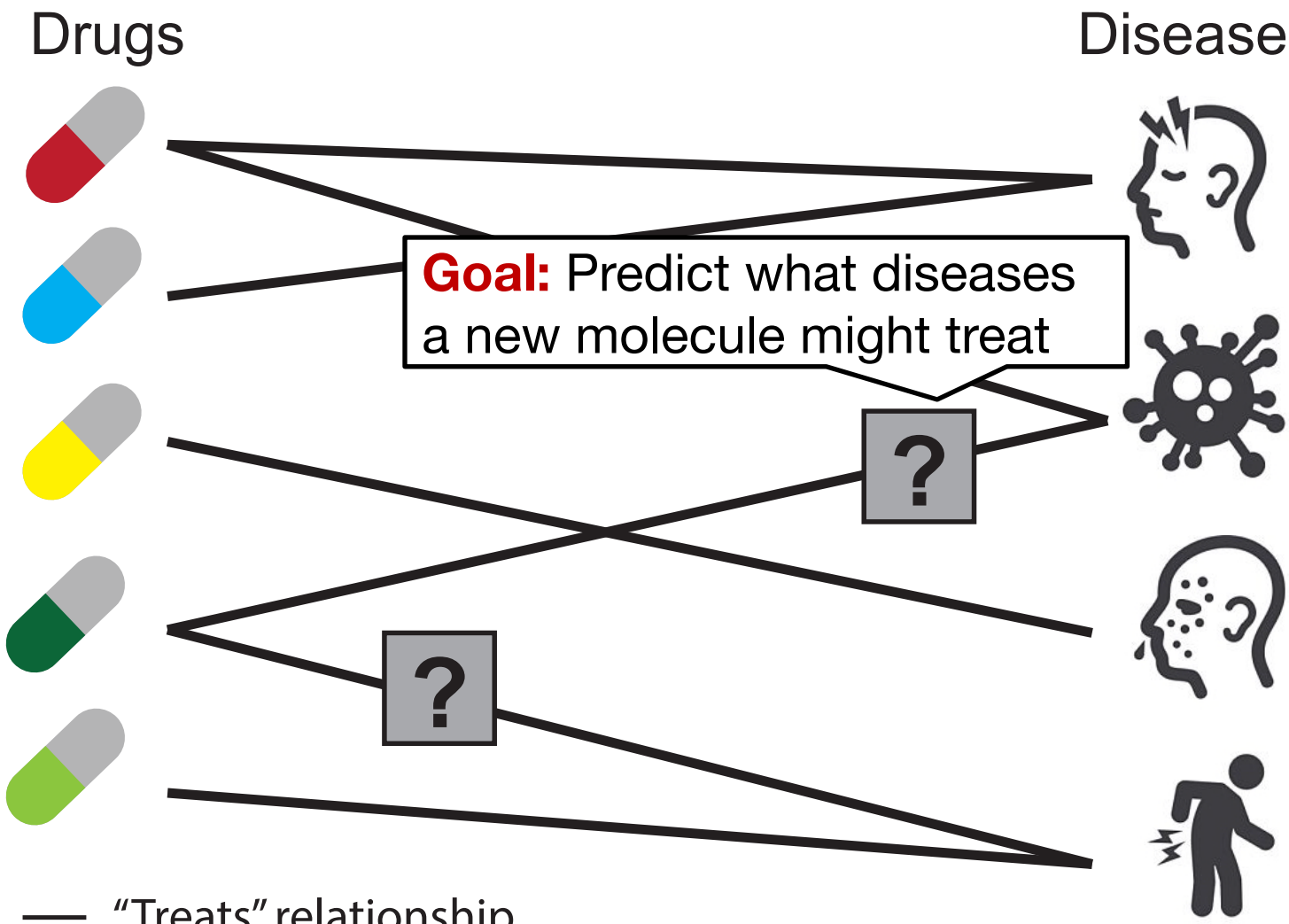
Step 1: Design and Discovery	Support decision-making for a new drug in the laboratory
Step 2: Preclinical Research	Answer basic questions about safety and animal testing
Step 3: Clinical Research	Predict if drug is safe & effective to test on people, find new uses for drugs
Step 4: FDA Review	Automatic document review to make a decision to approve the drug or not
Step 5: Post-Market and Safety Monitoring	Detect adverse and safety issues in real time using electronic health data

New tricks for old drugs

Faced with skyrocketing costs for developing new drugs, researchers are looking at ways to repurpose older ones — and even some that failed in initial trials.

Goal: Find which diseases a drug (new molecule) could treat

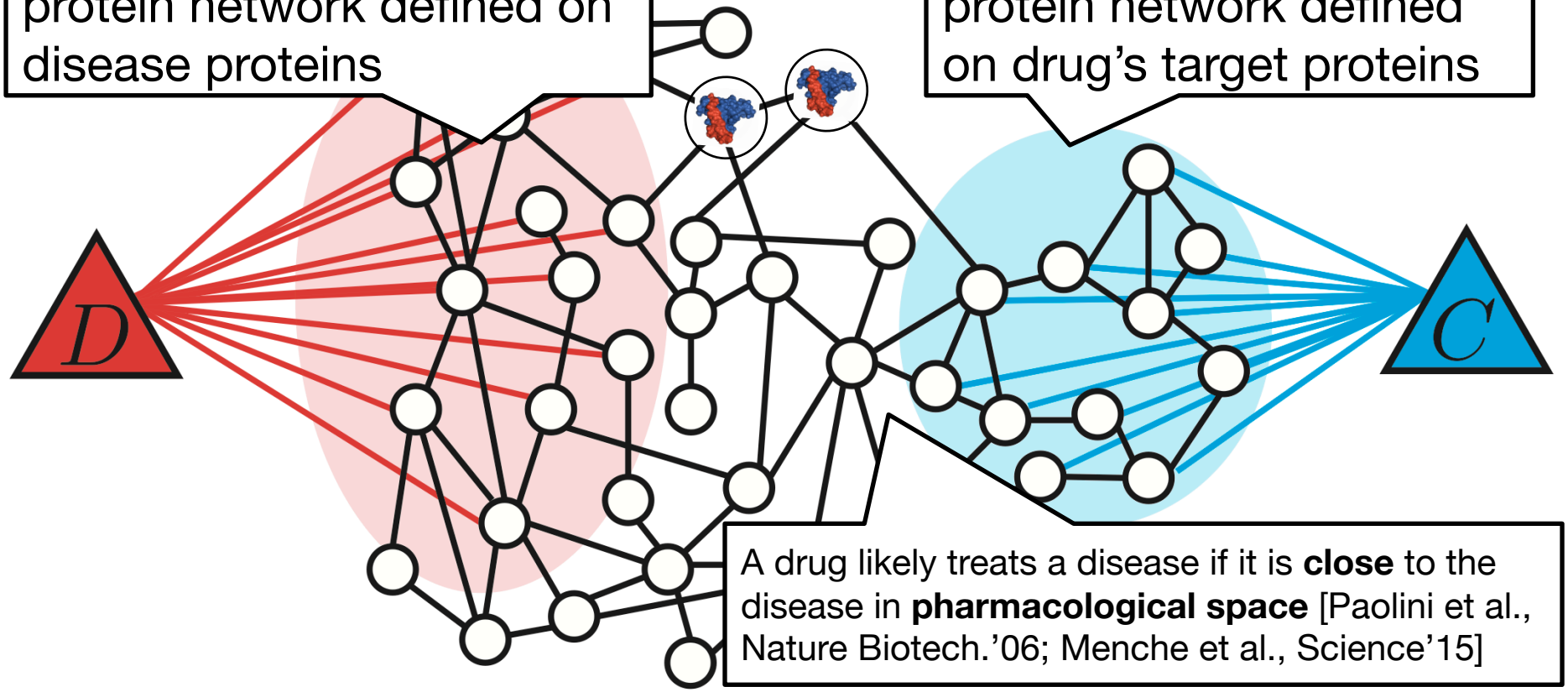
What drug treats what disease?



Key Insight: Subgraphs

Disease: Subgraph of rich protein network defined on disease proteins

Drug: Subgraph of rich protein network defined on drug's target proteins

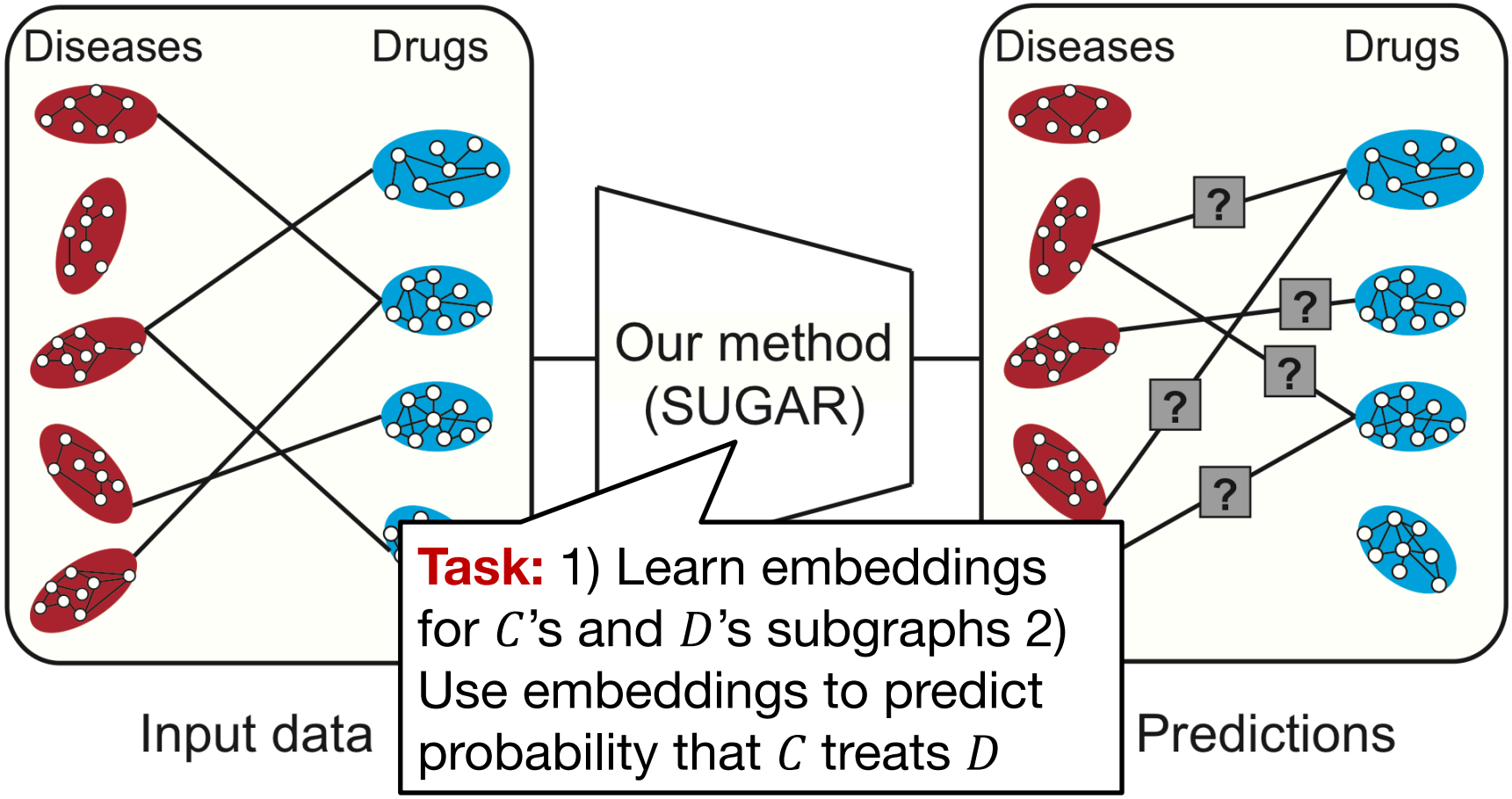


A drug likely treats a disease if it is **close** to the disease in **pharmacological space** [Paolini et al., Nature Biotech.'06; Menche et al., Science'15]

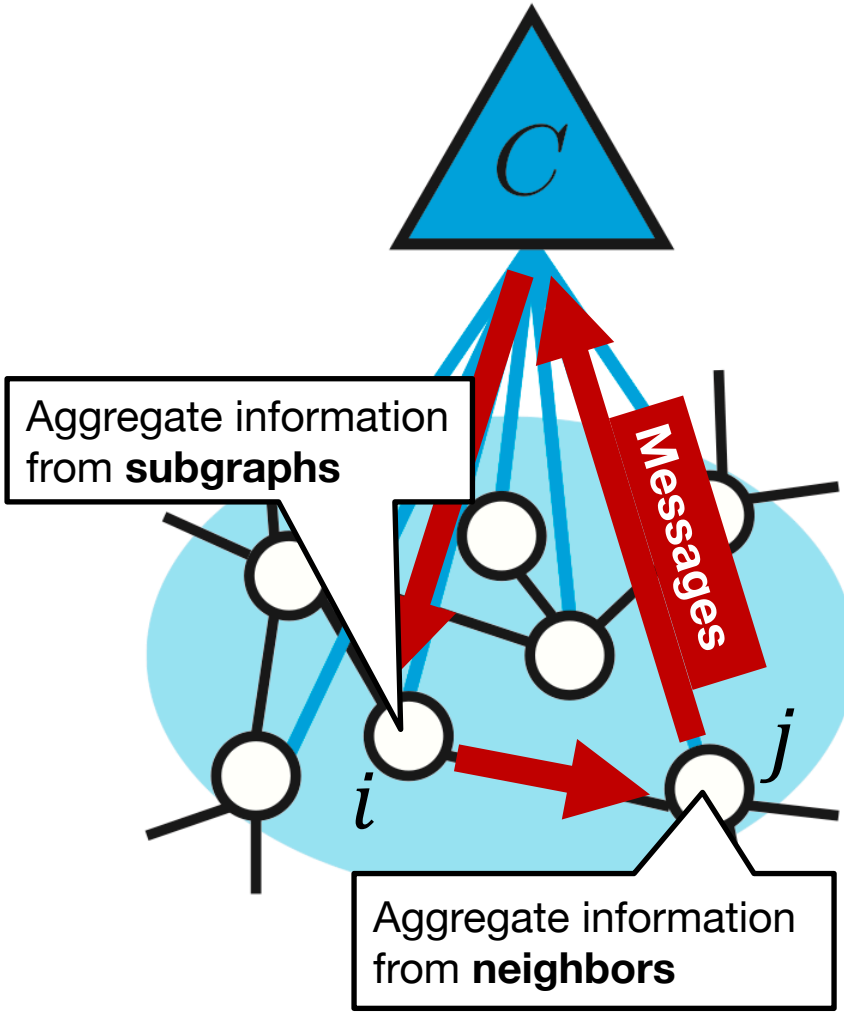
Idea: Use the paradigm of embeddings to operationalize the concept of closeness in pharmacological space

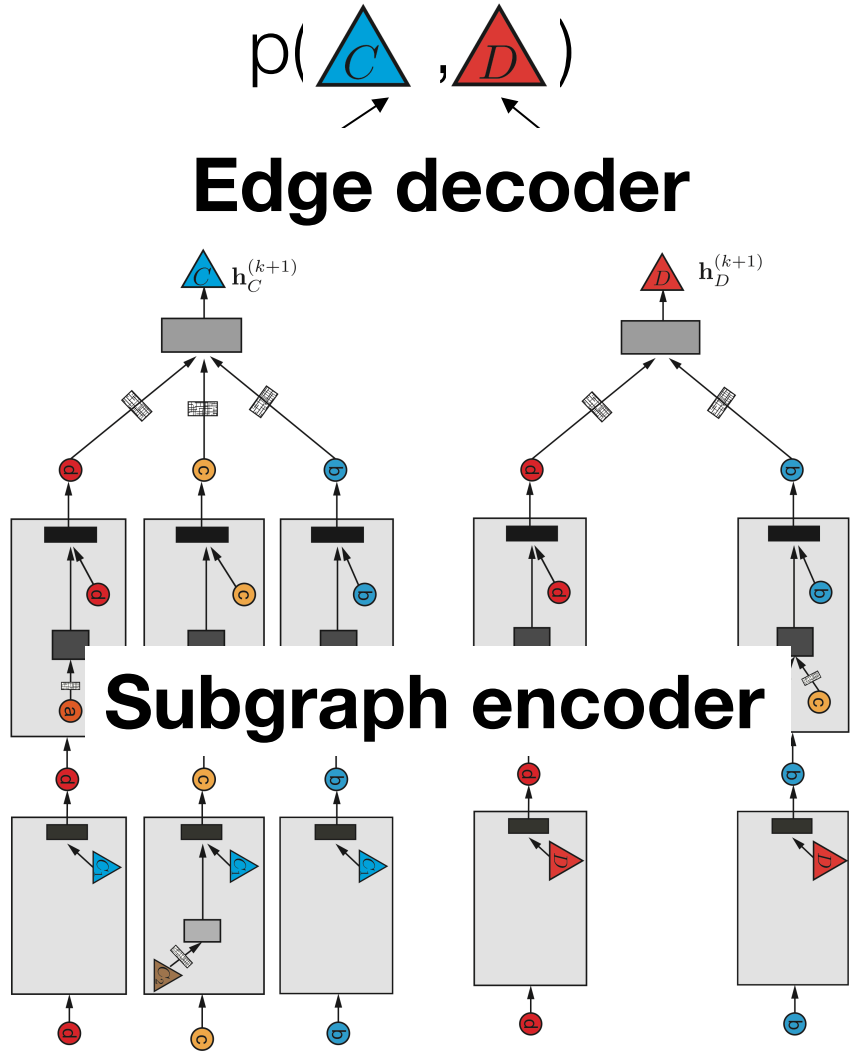
Predicting Links Between Drug and Disease Subgraphs

Task: Given drug C and disease D, predict if C treats D



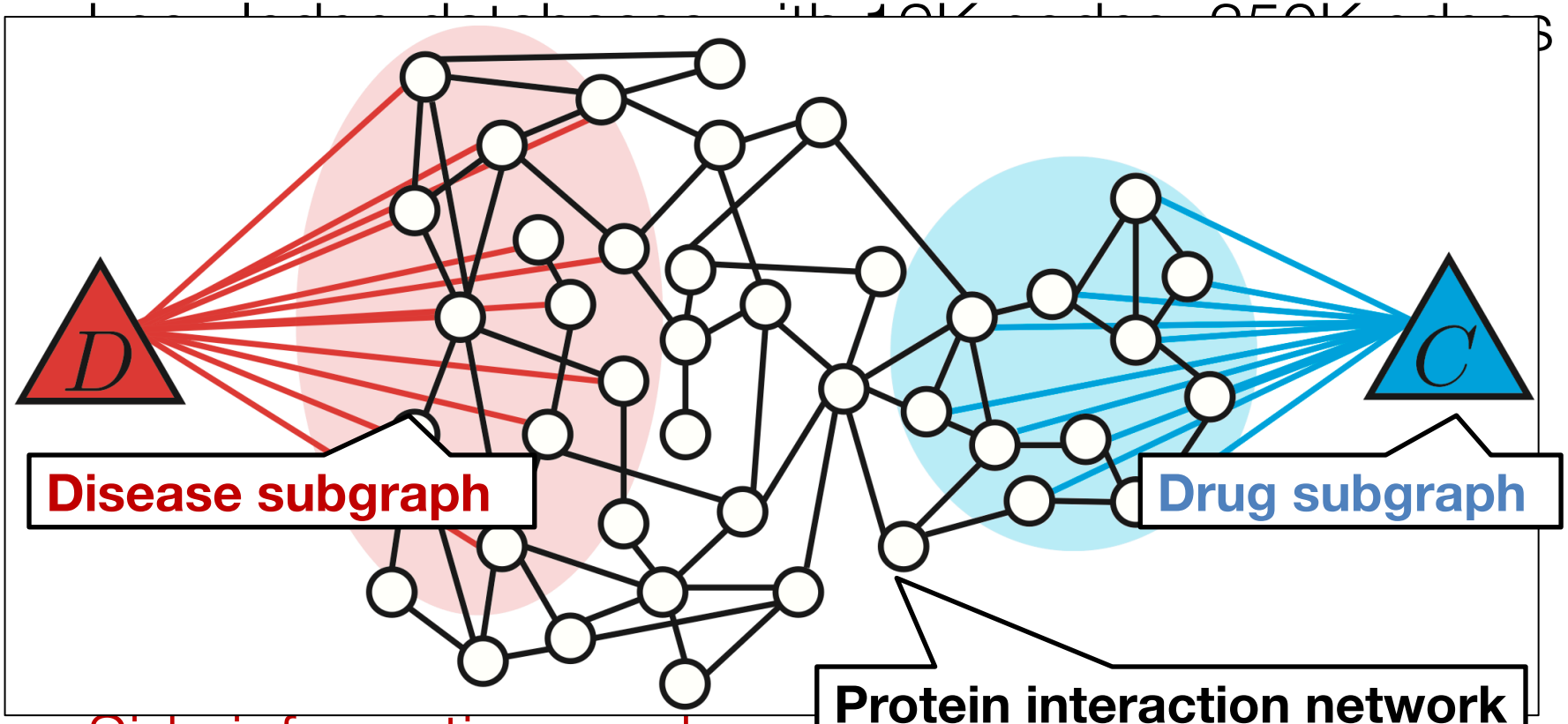
Neural Message Passing





We need drug repurposing dataset

- Protein-protein interaction network culled from 15



- Side information on drugs, diseases, proteins, etc..
  - Molecular pathways, disease symptoms, side effects

Predictive Performance

Task: Given a disease and a drug, predict if the drug could treat the disease

Approach	AUPRC	AUROC
Our method (SUGAR)	0.851	0.888
Graphlets [Bioinformatics'13]		
PREdicting Drug IndiCaTions [Mol. Sys. Biol.'11]		
Bi-directional random walks [Bioinformatics'16]		
Heterogeneous graph inference [Bioinformatics'14]		
Drug-disease closeness [Nat. Commun.'17]		
Drug-disease dispersion [Nat. Commun.'17]		
Gene-based network overlap [Nat. Commun.'17]		

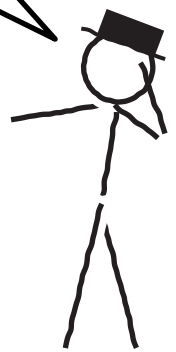
Up to 49% improvement

Up to 172% improvement

Drug Repurposing at Stanford

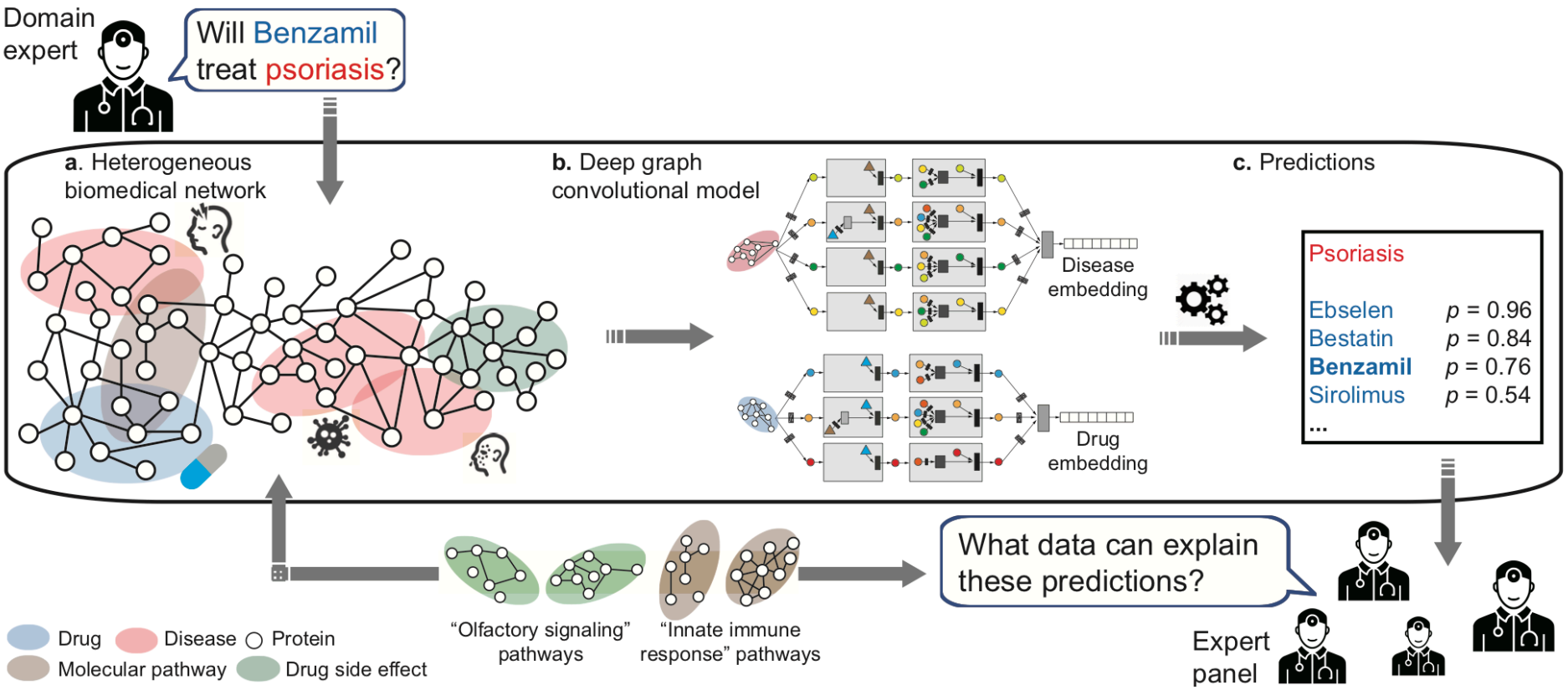
Stanford MEDICINE | SPARK Translational Research Program
From Bench to Bedside

Task: Predict if an existing drug can be repurposed for a new disease

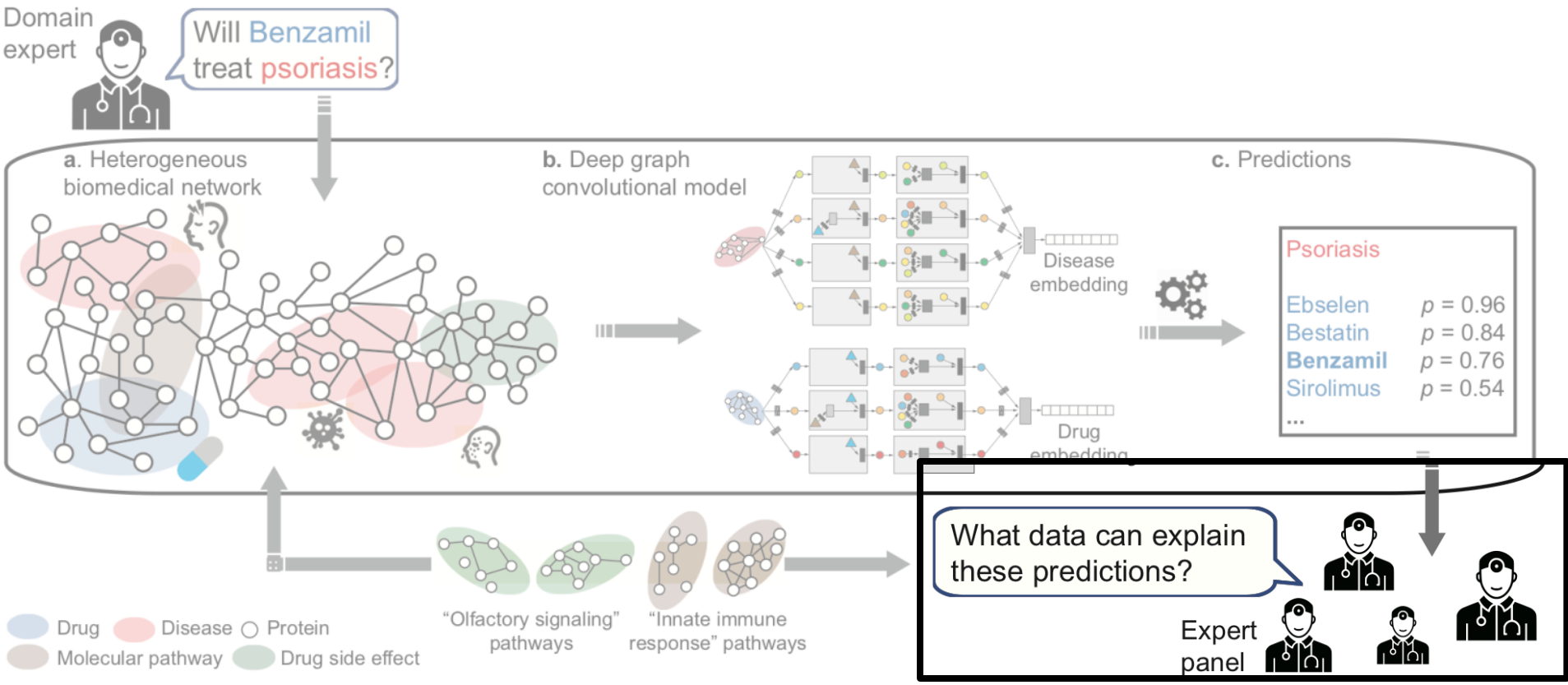


Drug	Disease	
N-acetyl-cysteine	cystic fibrosis	
Xamoterol	neurodegenerat	
Plerixafor	cancer	
Sodium selenite	cancer	Rank: 36/5000
Ebselen	C difficile	Rank: 10/5000
Itraconazole	cancer	Rank: 26/5000
Bestatin	lymphedema	Rank: 11/5000
Bestatin	pulmonary arterial hypertension	Rank: 16/5000
Ketaprofen	lymphedema	Rank: 28/5000
Sildenafil	lymphatic malformation	Rank: 26/5000
Tacrolimus	pulmonary arterial hypertension	Rank: 46/5000
Benzamil	psoriasis	Rank: 114/5000
Carvedilol	Chagas' disease	Rank: 9/5000
Benserazide	BRCA1 cancer	Rank: 41/5000
Pioglitazone	interstitial cystitis	Rank: 13/5000
Sirolimus	dystrophic epidermolysis bullosa	Rank: 46/5000

Feedbacks for the AI Loop



Feedbacks for the AI Loop

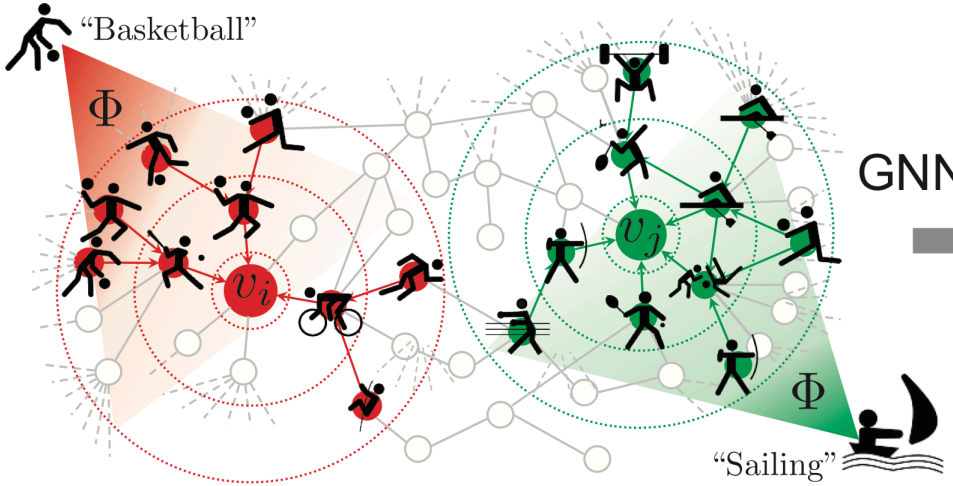


Explaining GNN Predictions

Key idea:

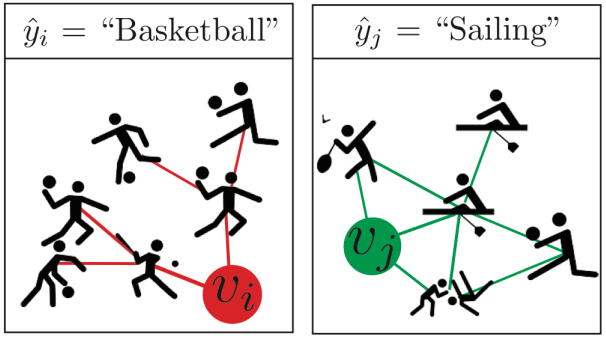
- Summarize where in the data the model "looks" for evidence for its prediction
- Find a small subgraph most influential for the prediction

GNN model training and predictions



GNNExplainer

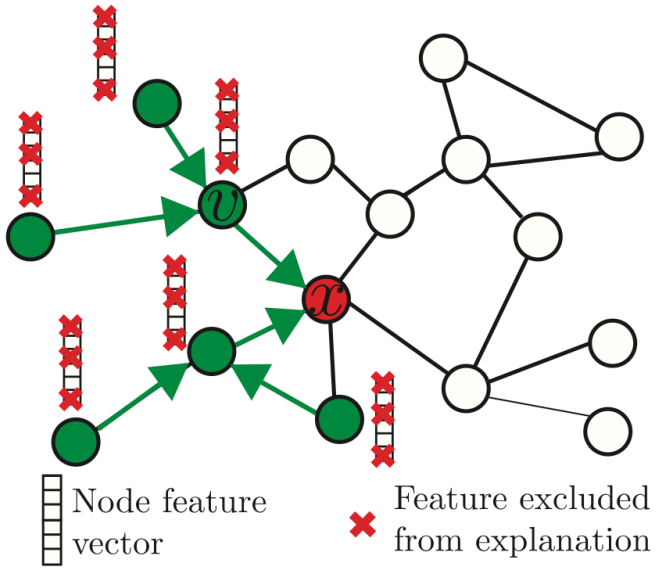
Explaning GNN's predictions



GNN Explainer: Generating Explanations for Graph Neural Networks, *NeurIPS* 2019

GNNExplainer: Key Idea

- Input: Given prediction $f(x)$ for node/link x
- Output: Explanation, a small subgraph M_x together with a small subset of node features:
  - M_x is most influential for prediction $f(x)$
- **Approach:** Learn M_x via counterfactual reasoning
  - Intuition: If removing v from the graph strongly decreases the probability of prediction $\Rightarrow$ v is a good counterfactual explanation for the prediction

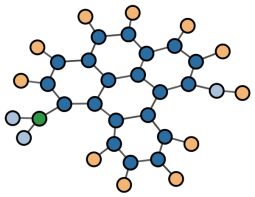
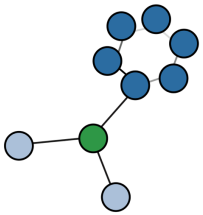
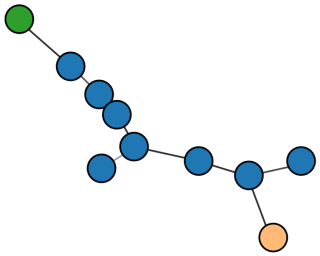
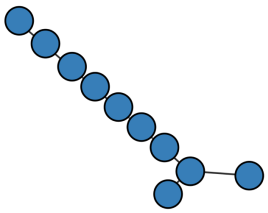
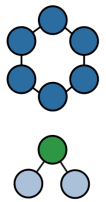
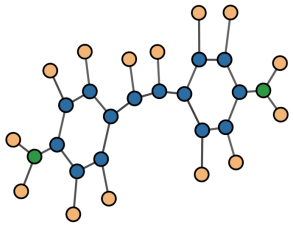
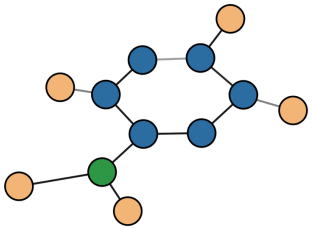
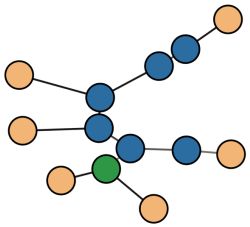
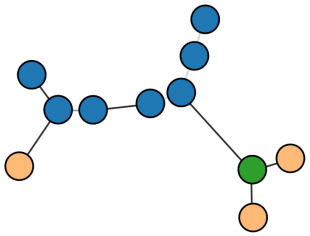
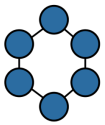


GNN Explainer: Generating Explanations for Graph Neural Networks, *NeurIPS* 2019

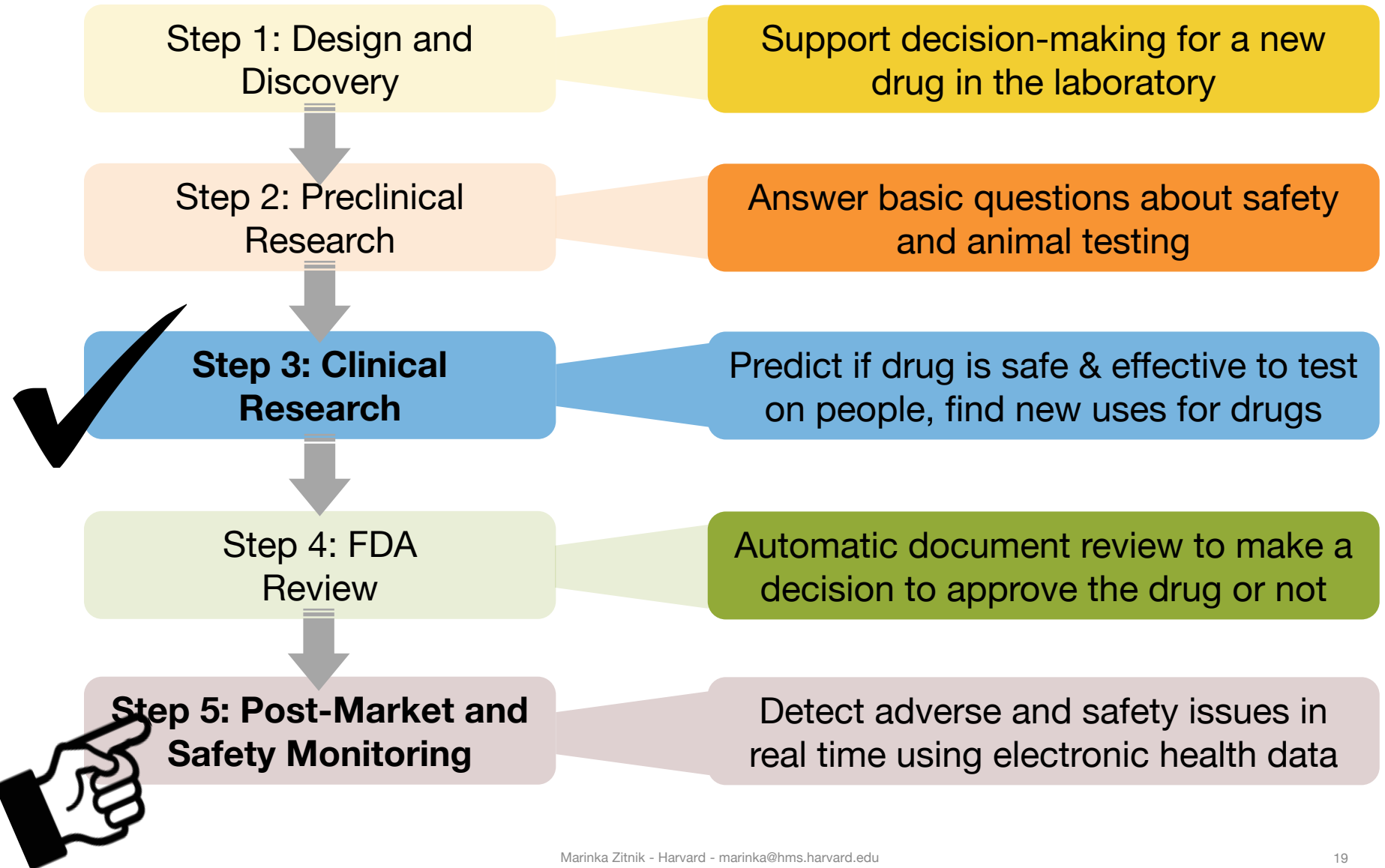
GNNExplainer: Results

"Why did you predict that this molecule will have a mutagenic effect on Gram-negative bacterium *S. typhimurium*?"

Explanation

Input	GNN Explainer	Grad	Att	Ground Truth
				 Cycle NO_2 group
				

Today's Talk


Step 1: Design and Discovery
Support decision-making for a new drug in the laboratory
Step 2: Preclinical Research
Answer basic questions about safety and animal testing
Step 3: Clinical Research
Predict if drug is safe & effective to test on people, find new uses for drugs
Step 4: FDA Review
Automatic document review to make a decision to approve the drug or not
Step 5: Post-Market and Safety Monitoring
Detect adverse and safety issues in real time using electronic health data

Polypharmacy

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

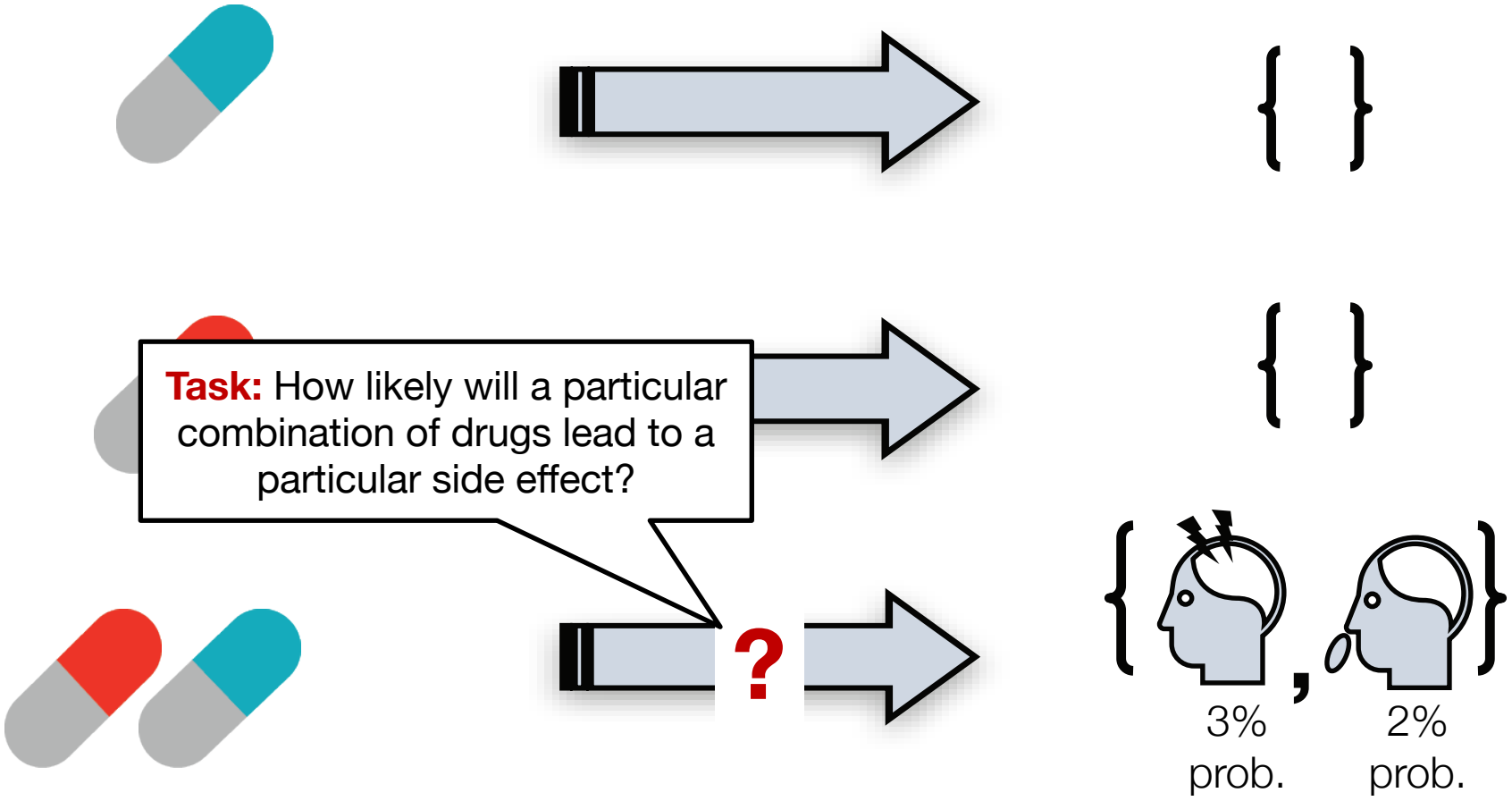
Annual costs in treating side effects exceed **$177** billion in the U.S. alone

[Ernst and Grizzle, JAPA'01; Kantor et al., JAMA'15]

Unexpected Drug Interactions

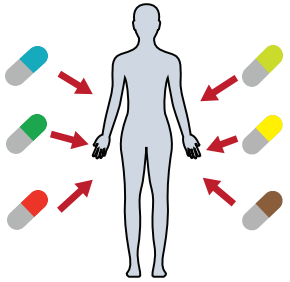
Co-prescribed drugs

Side Effects



Why is modeling polypharmacy hard?

Combinatorial explosion

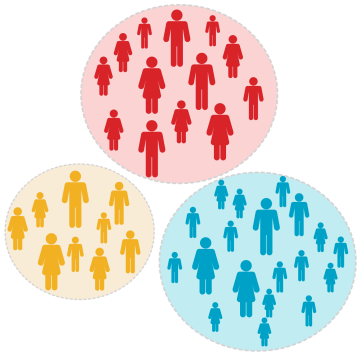


- >13 million possible combinations of 2 drugs
- >20 billion possible combinations of 3 drugs

Non-linear & non-additive interactions

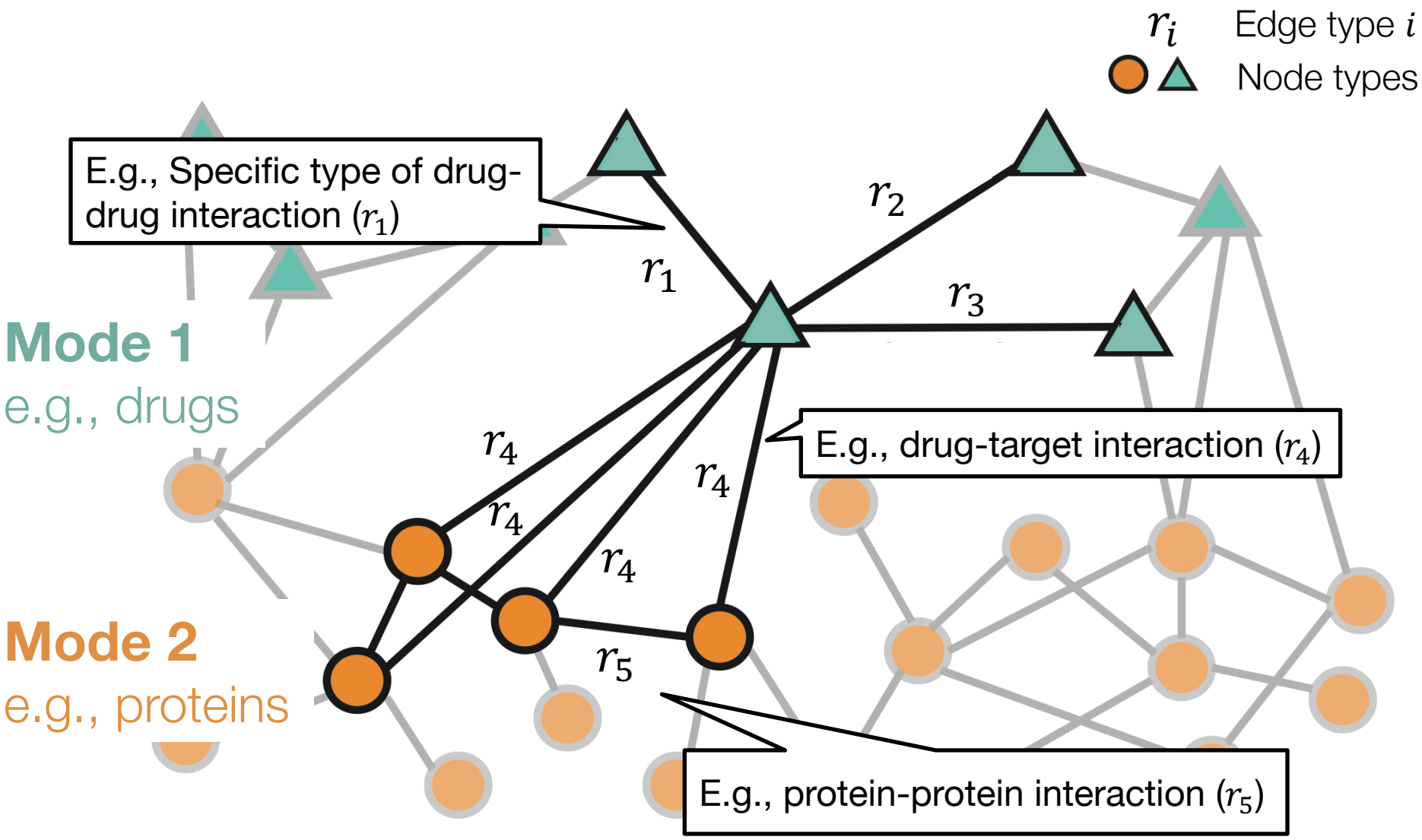
- Different effect than the additive effect of individual drugs

Small subsets of patients



- Side effects are interdependent
- No info on drug combinations not yet used in patients

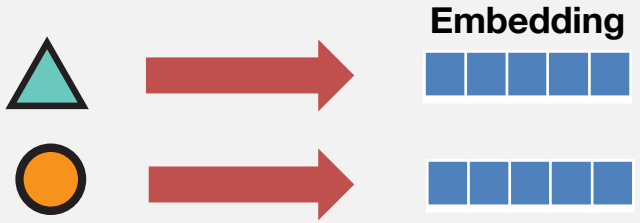
Setup: Multimodal Networks



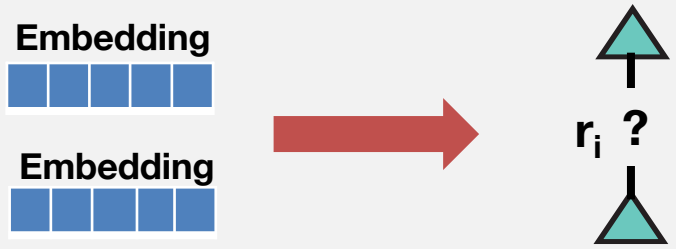
Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

Our Approach: Decagon

1. **Encoder:** Take a multimodal network and learn an *embedding* for every node



2. **Decoder:** Use the learned embeddings to predict typed edges between nodes

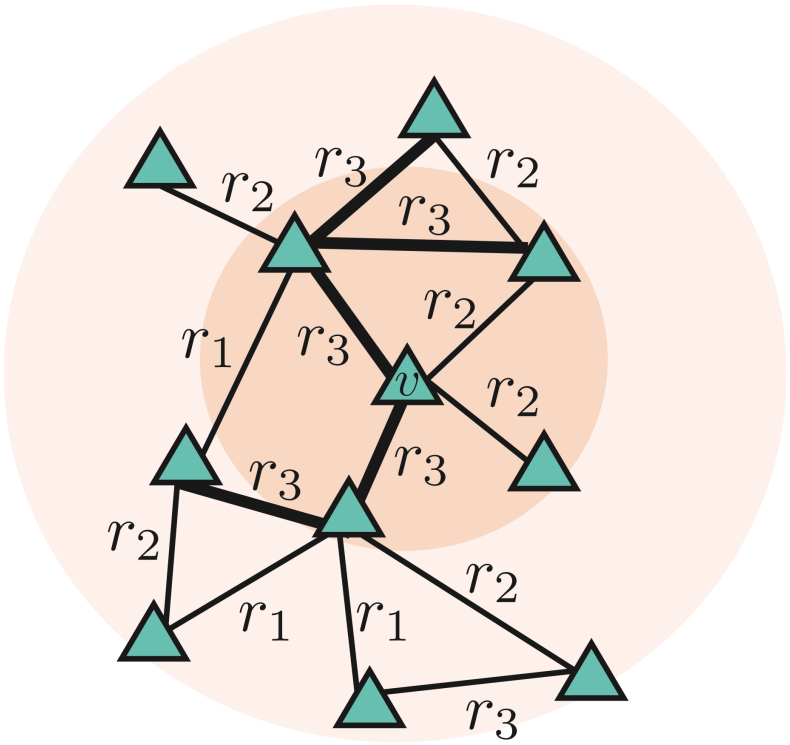


Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

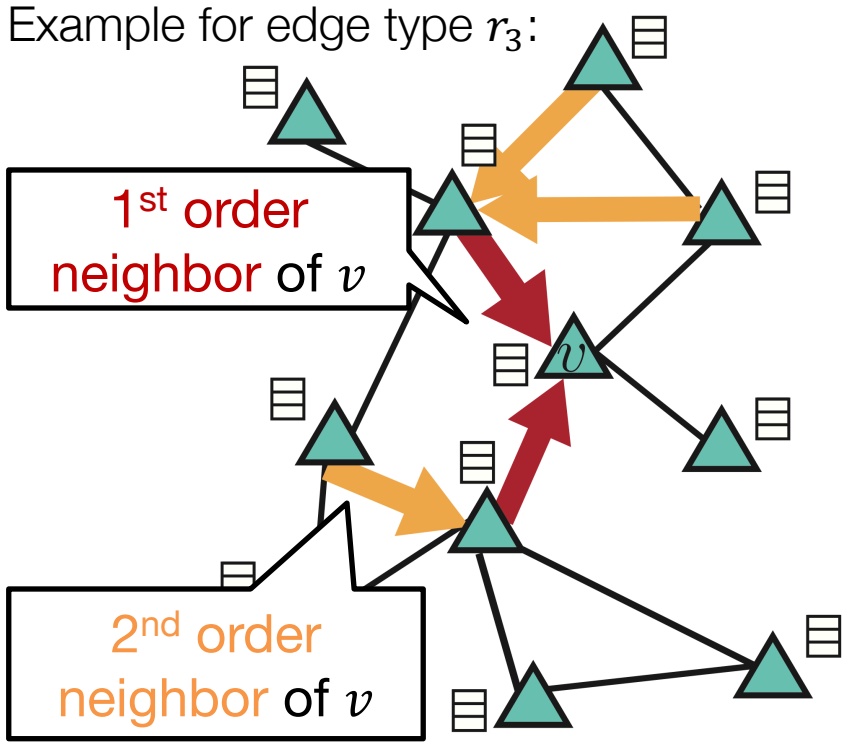
Encoder: Propagate Neighbors

Generate embeddings based on **local network neighborhoods separated by edge type**

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

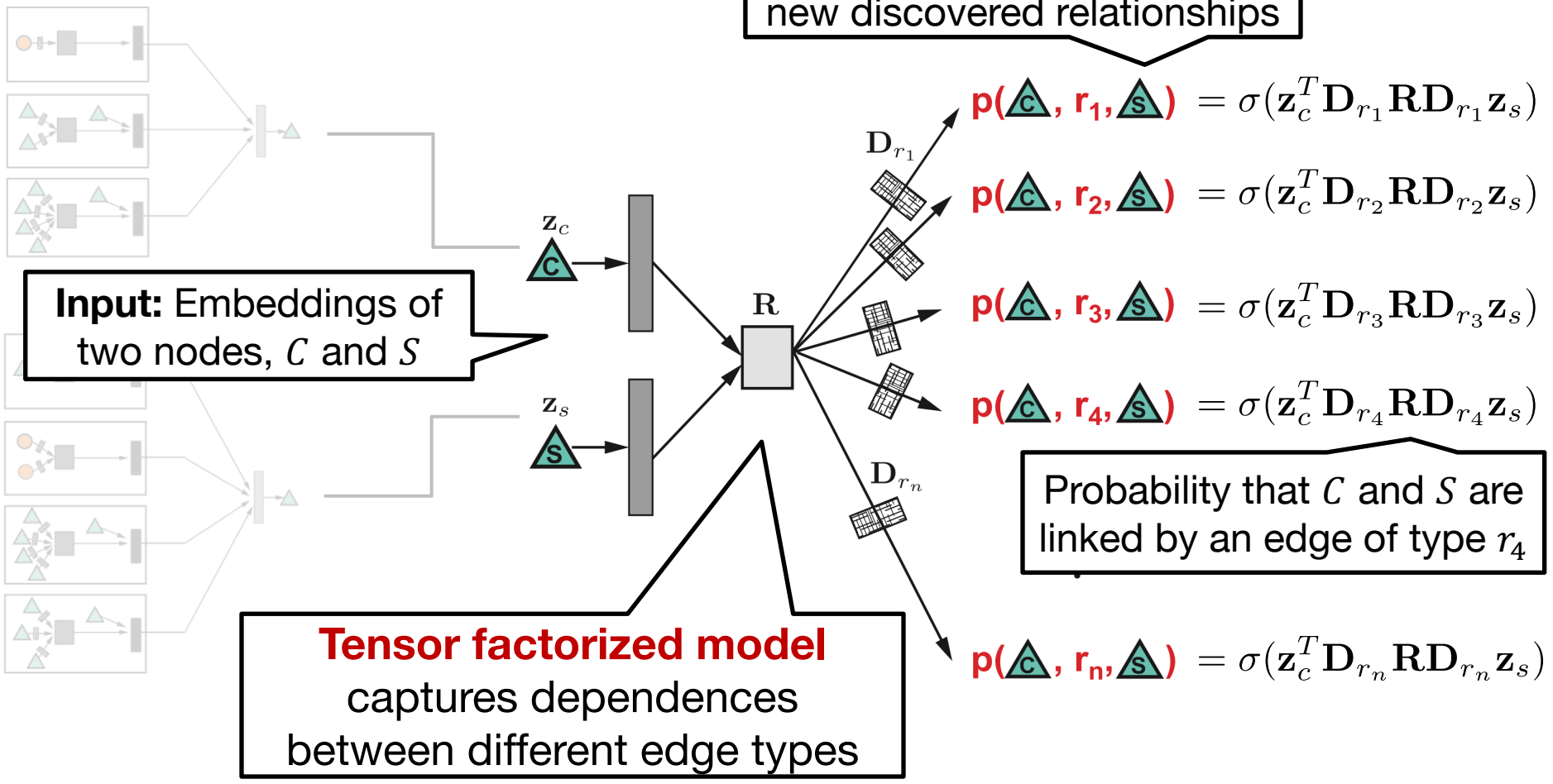


2) Learn how to transform and propagate information across computation graph



Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

Decoder: Weighted, Typed Edges



$\mathbf{R}, \mathbf{D}_{r_i}$ Parameter weight matrices

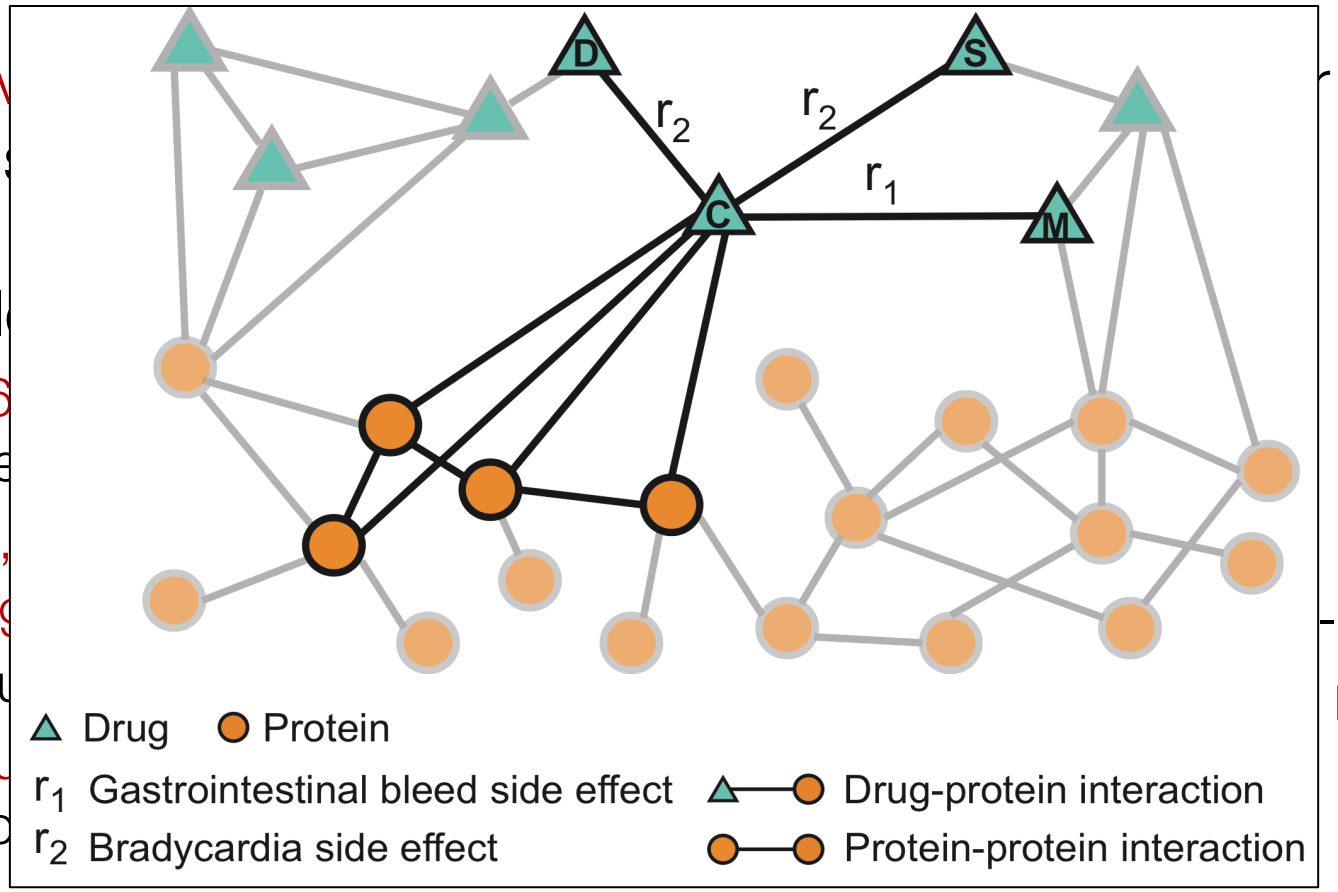
Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

We need polypharmacy dataset

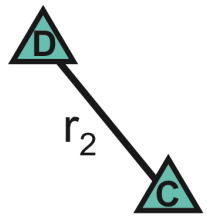
Objectiv
all drug

We buil

- 4,6
  eve
- 18,
- 719
  cou
- Dru
  pro



Drug-drug



Drug-protein

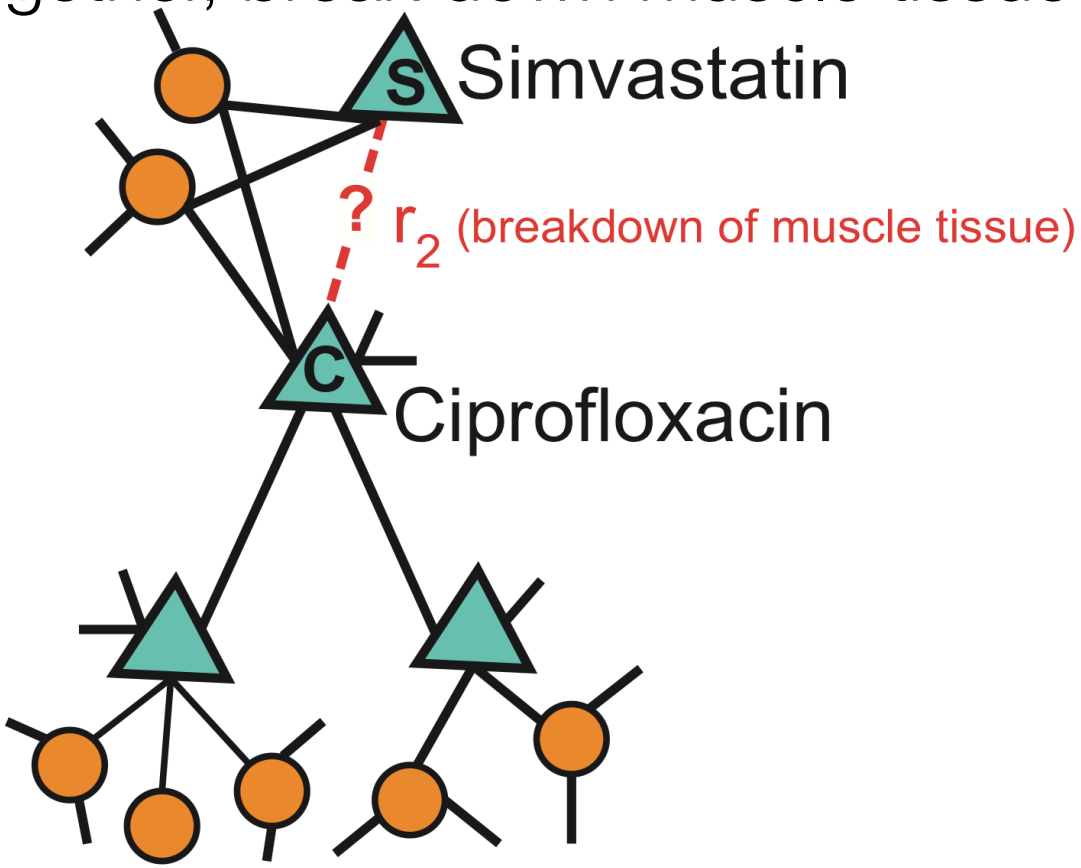
Protein-protein

A polypharmacy network with over 5 million edges and over 1,000 different edge types

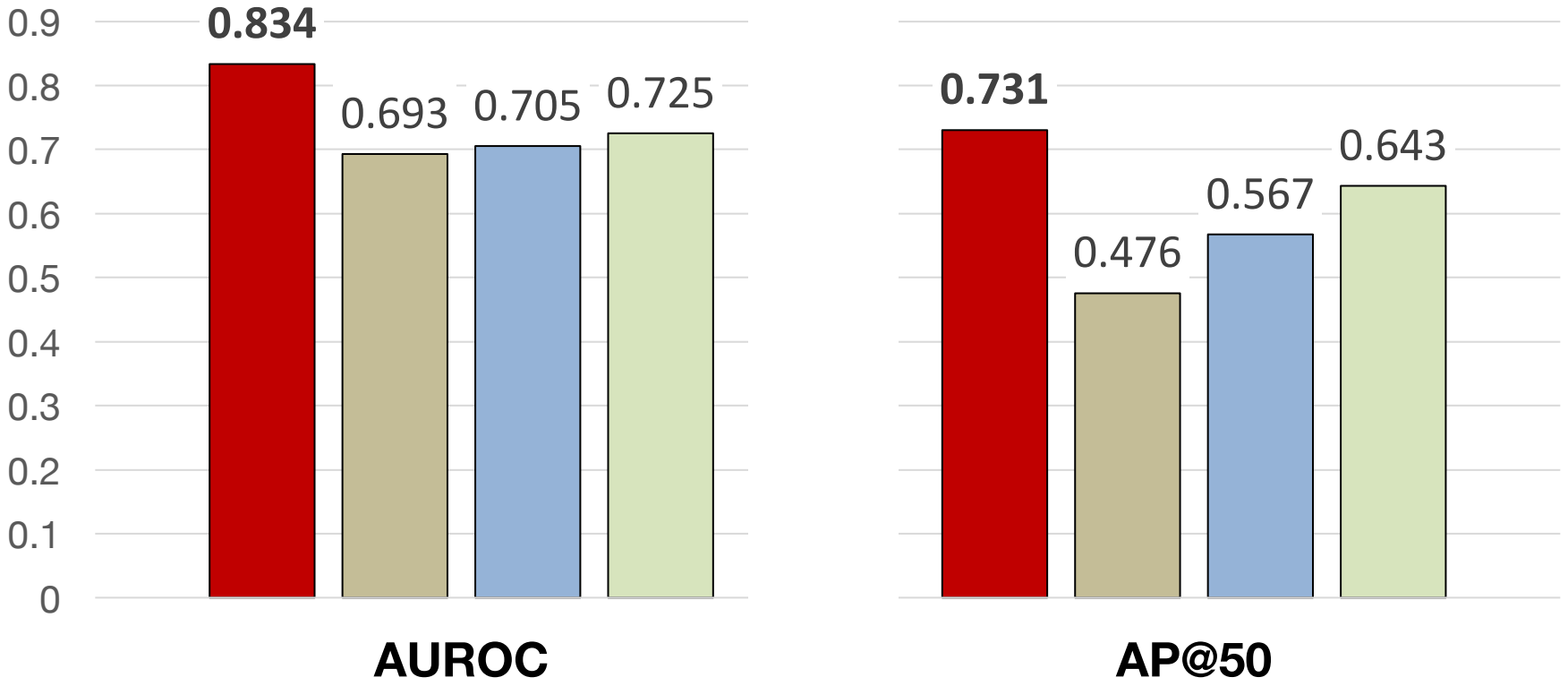
Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

We apply our deep approach to the polypharmacy network

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



Results: Side Effect Prediction



- Our method (Decagon)
- RESCAL Tensor Factorization [Nickel et al., ICML'11]
- Multi-relational Factorization [Perros, Papalexakis et al., KDD'17]
- Shallow Network Embedding [Zong et al., Bioinformatics'17]

Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

New Predictions

First AI method to **predict side effects of drug combinations,** even for combinations not yet used in patients

Next: Can the method generate hypotheses and give:

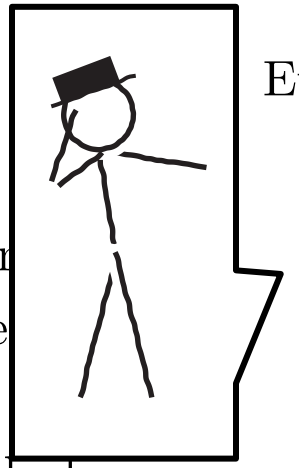
- **Doctors** guidance on whether it is a good idea to prescribe a particular combination of drugs to a particular patient
- **Researchers** guidance on effective wet lab experiments and new drug therapies with fewer side effects

Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

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

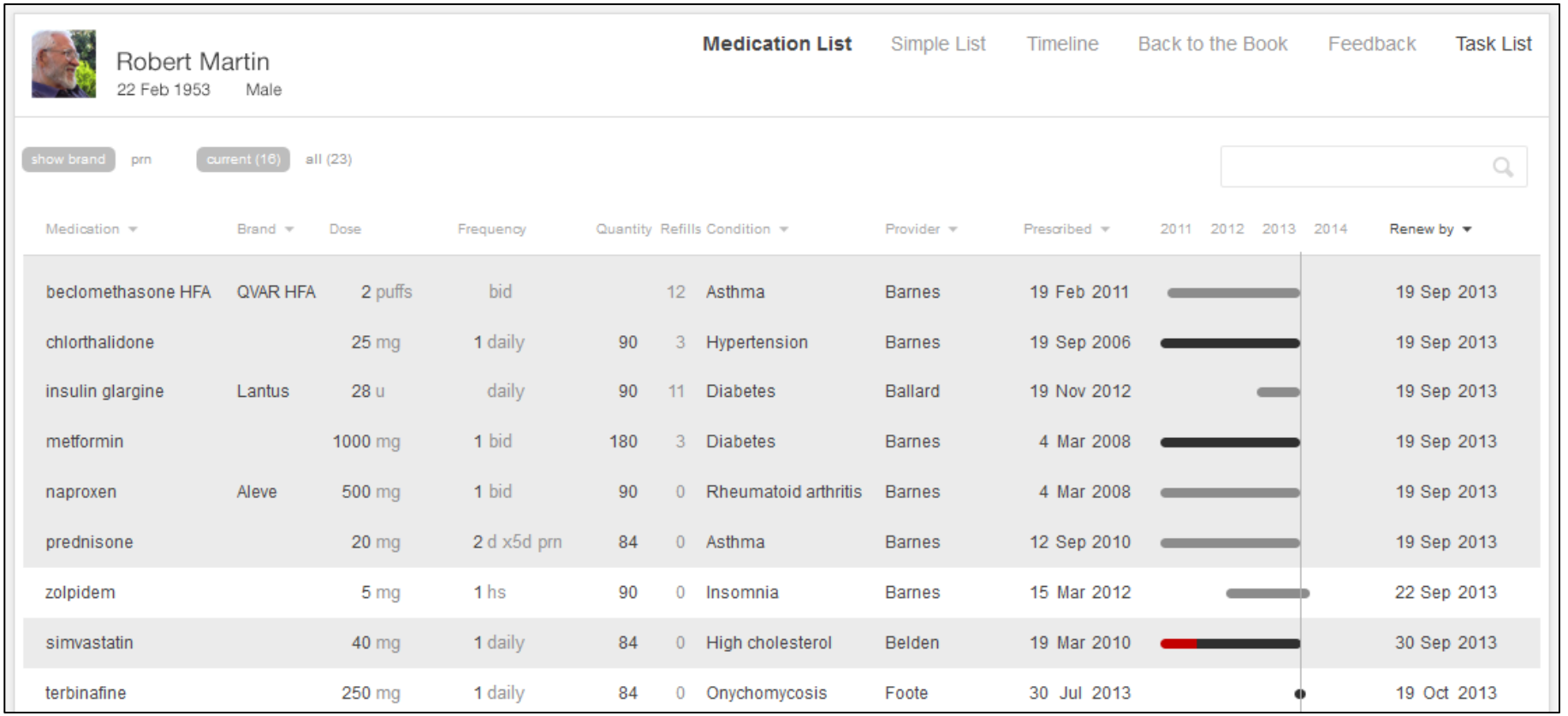
Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 2018

Clinical Validation of New Predictions

Drug interaction markers, lab values, and many other surrogates

NEWTON-WELLESLEY HOSPITAL

MASSACHUSETTS GENERAL HOSPITAL



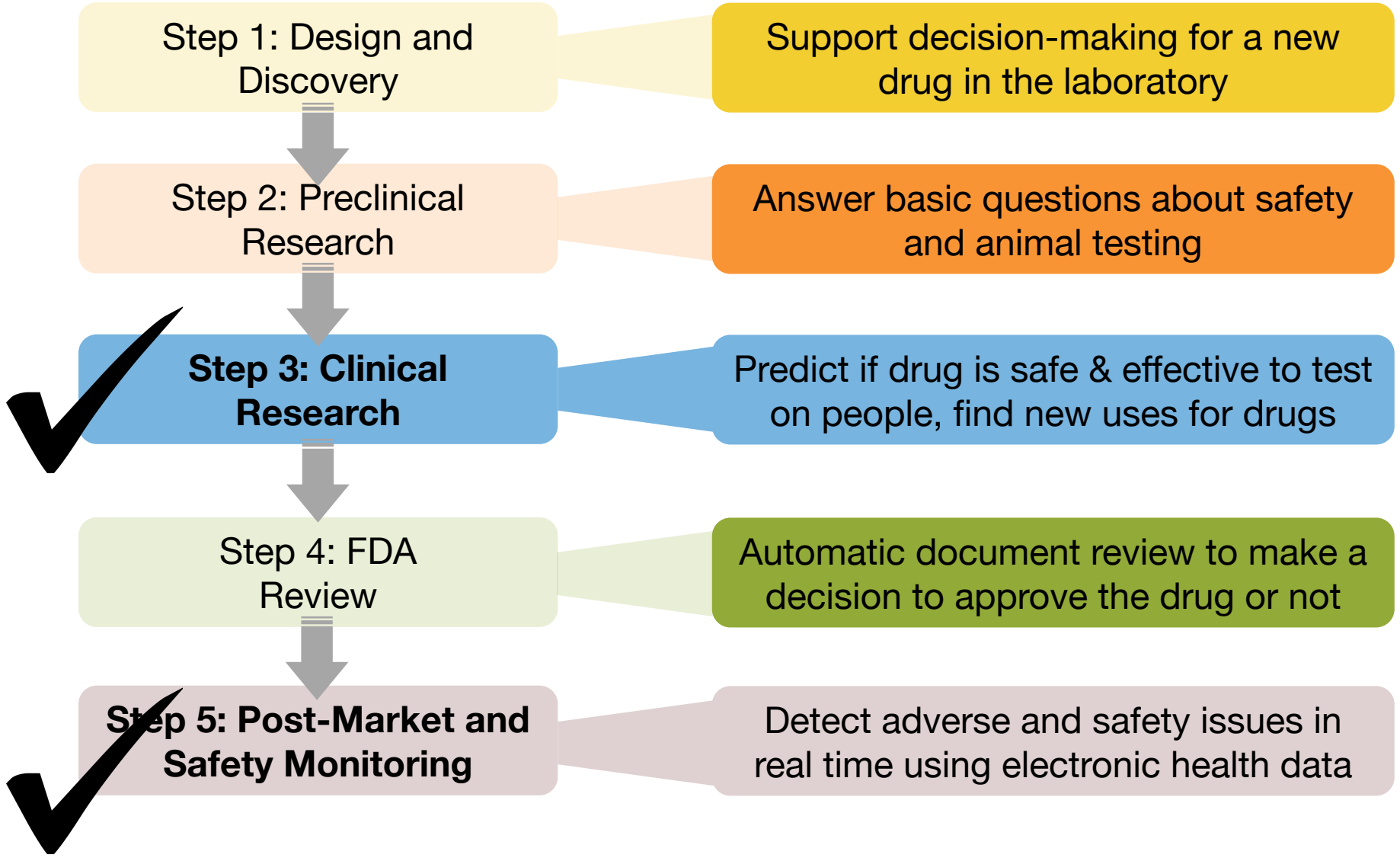
Robert Martin
22 Feb 1953 Male

Medication List | Simple List | Timeline | Back to the Book | Feedback | Task List

show brand | prn | current (16) | all (23)

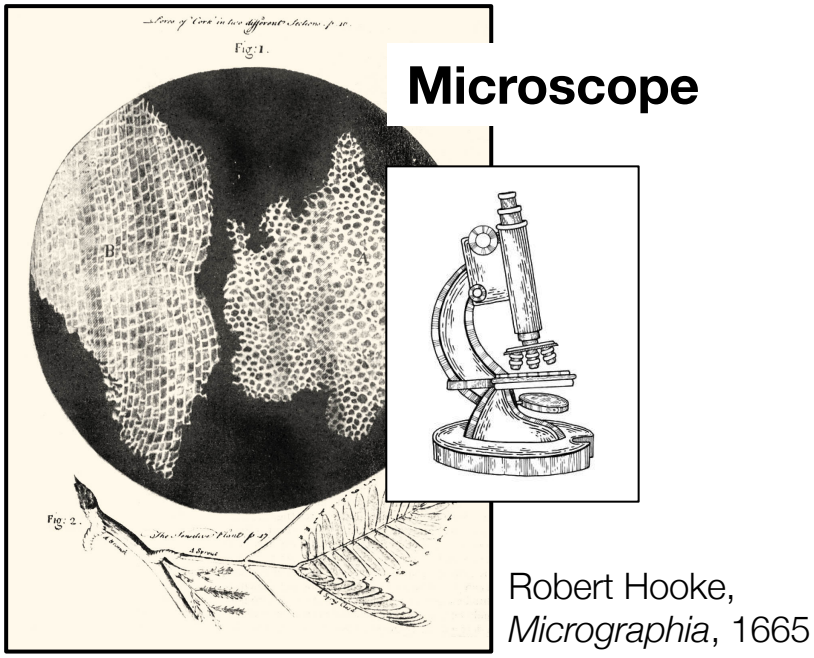
Medication	Brand	Dose	Frequency	Quantity	Refills	Condition	Provider	Prescribed	Renew by
beclomethasone HFA	QVAR HFA	2 puffs	bid		12	Asthma	Barnes	19 Feb 2011	19 Sep 2013
chlorthalidone		25 mg	1 daily	90	3	Hypertension	Barnes	19 Sep 2006	19 Sep 2013
insulin glargine	Lantus	28 u	daily	90	11	Diabetes	Ballard	19 Nov 2012	19 Sep 2013
metformin		1000 mg	1 bid	180	3	Diabetes	Barnes	4 Mar 2008	19 Sep 2013
naproxen	Aleve	500 mg	1 bid	90	0	Rheumatoid arthritis	Barnes	4 Mar 2008	19 Sep 2013
prednisone		20 mg	2 d x5d prn	84	0	Asthma	Barnes	12 Sep 2010	19 Sep 2013
zolpidem		5 mg	1 hs	90	0	Insomnia	Barnes	15 Mar 2012	22 Sep 2013
simvastatin		40 mg	1 daily	84	0	High cholesterol	Belden	19 Mar 2010	30 Sep 2013
terbinafine		250 mg	1 daily	84	0	Onychomycosis	Foote	30 Jul 2013	19 Oct 2013

Today's Talk


Step 1: Design and Discovery
Support decision-making for a new drug in the laboratory
Step 2: Preclinical Research
Answer basic questions about safety and animal testing
Step 3: Clinical Research
Predict if drug is safe & effective to test on people, find new uses for drugs
Step 4: FDA Review
Automatic document review to make a decision to approve the drug or not
Step 5: Post-Market and Safety Monitoring
Detect adverse and safety issues in real time using electronic health data

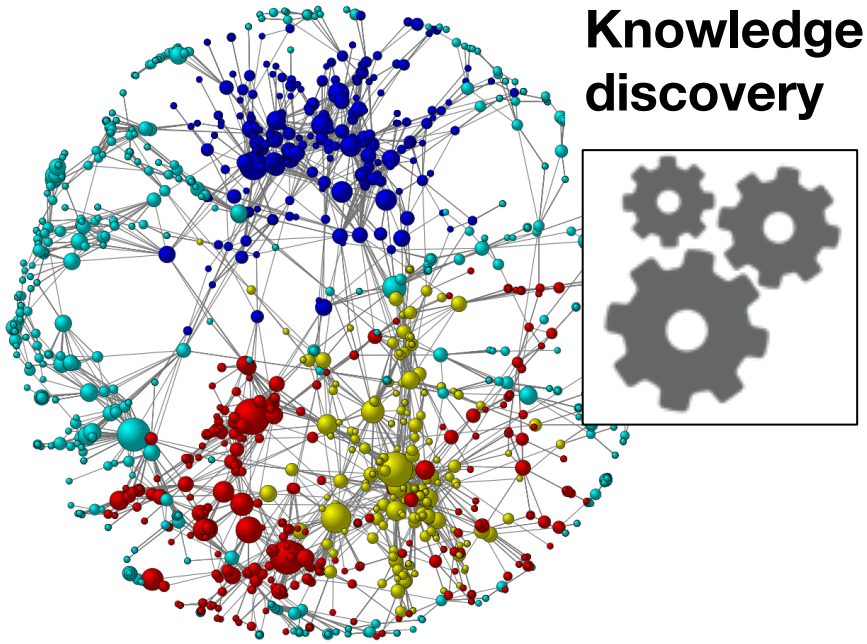
Complex, interconnected datasets are transforming science and medicine

Graph ML can unlock these datasets



Robert Hooke, *Micrographia*, 1665

Physical instruments facilitate discoveries



Instruments for modern, data-intensive sciences

Thank you!

And thanks to my collaborators:

Jure Leskovec, Russ B. Altman, Will Hamilton, Rex Ying, Monica Agrawal, Dylan Bourgeois, Jiaxuan You, Evan Sabri Eyuboglu

Papers, data & code

cs.stanford.edu/~marinka

snap.stanford.edu/biodata

HARVARD UNIVERSITY

Students and postdocs for projects in machine learning on biomedical data

marinka@hms.harvard.edu