Representation learning: A new approach for biomedical data

Marinka Zitnik (marinka@cs.stanford.edu)

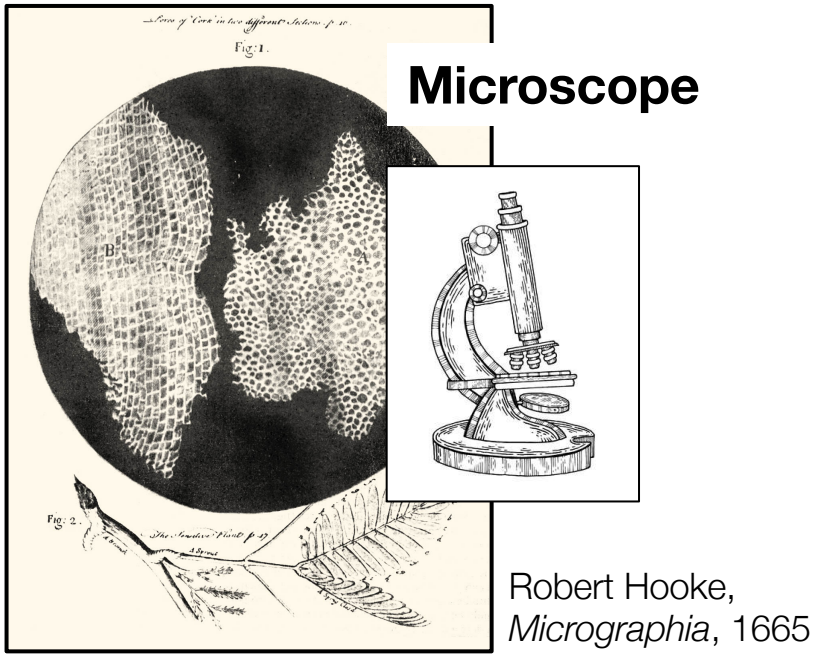
Stanford | ENGINEERING
Computer Science

Stanford MEDICINE

CHAN ZUCKERBERG BIOHUB

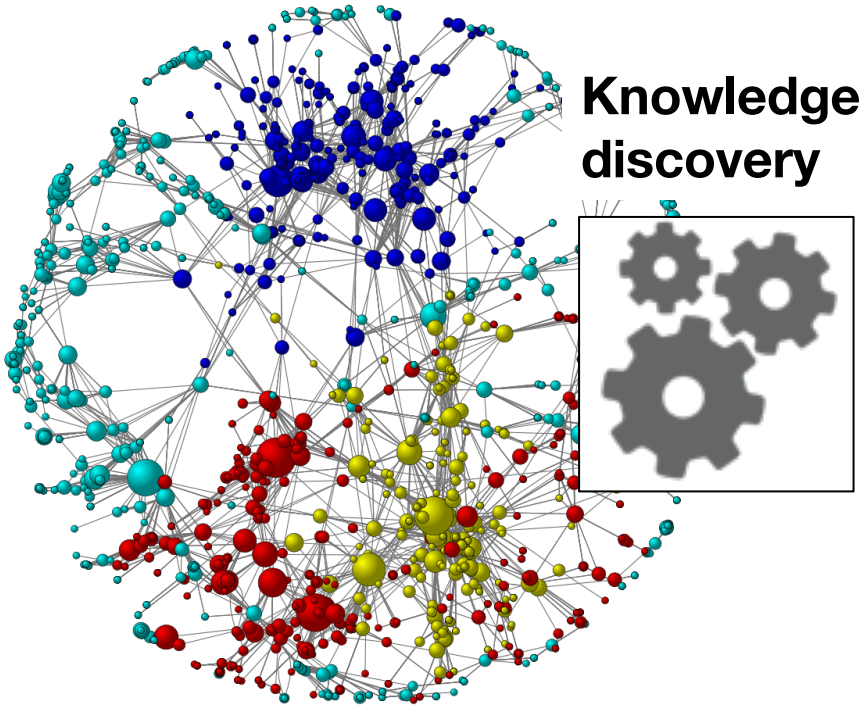
STANFORD INFOLAB

Science crucially depends on scientific instruments



Microscope

Robert Hooke,
Micrographia, 1665

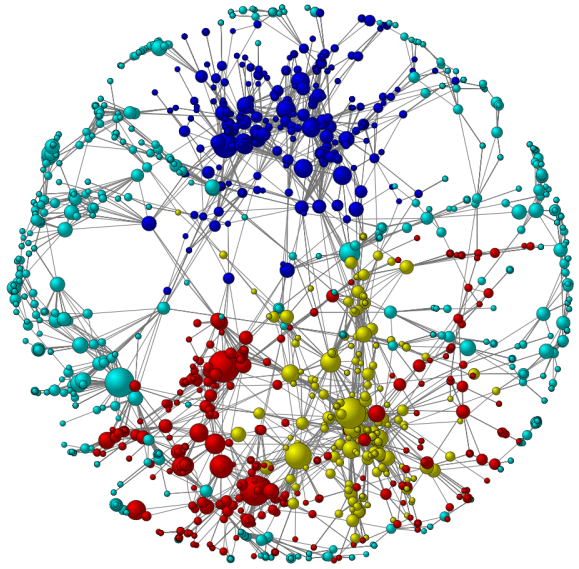


Knowledge discovery

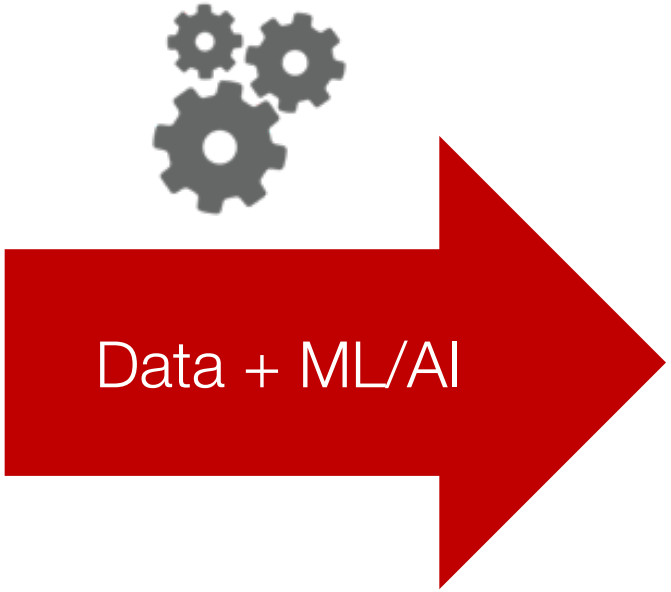
Physical instruments facilitate discoveries

Need instruments for modern, data-intensive sciences

Knowledge Discovery



Data



Predictions
and insights

Opportunities for AI in health & medicine

Preliminary diagnosis, early disease detection, self-care

Automated image diagnosis, language modeling

Clinical trial participation, drug discovery, AI-driven medical devices

Comorbidities, chronic disease treatments

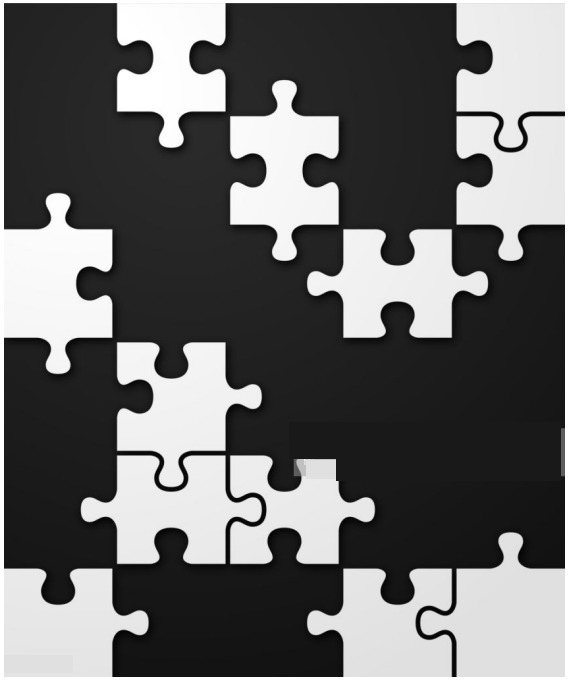
Improve administrative workflows, costly back-office problems

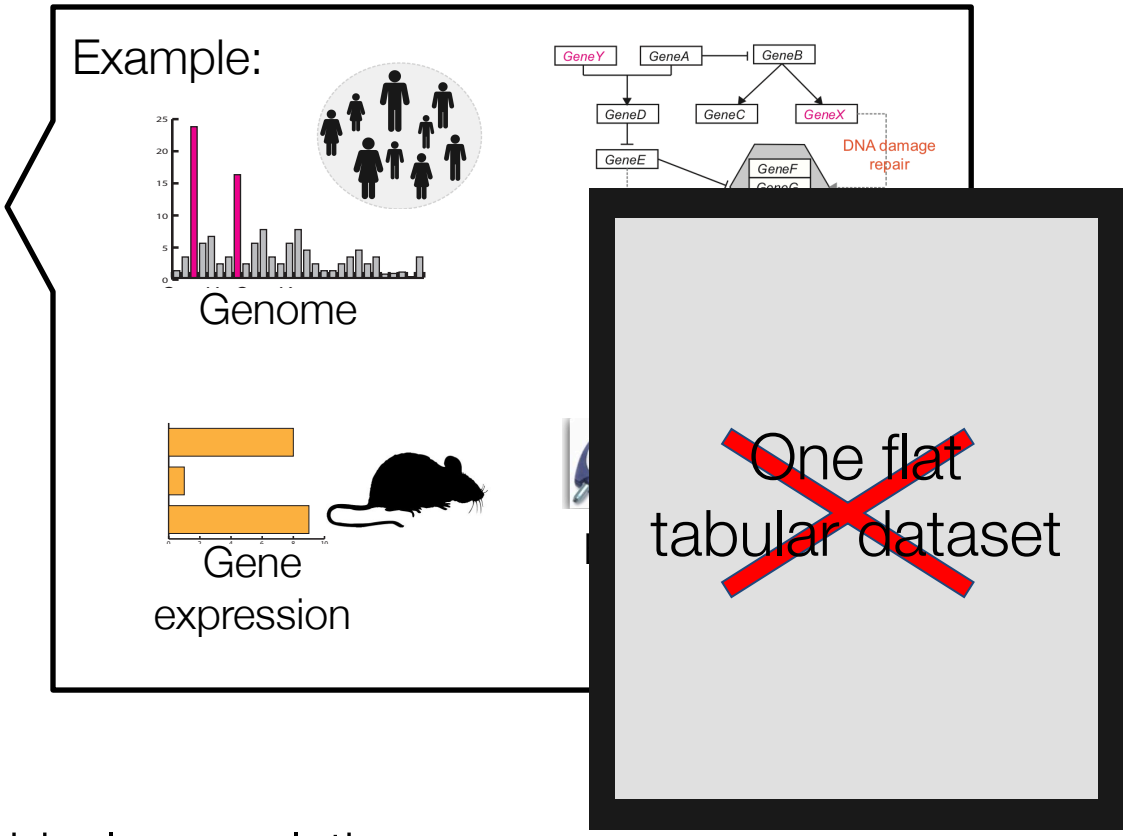
Inpatient & outpatient policies of care

Real-time patient interventions

Help protect health data, avoid medical errors

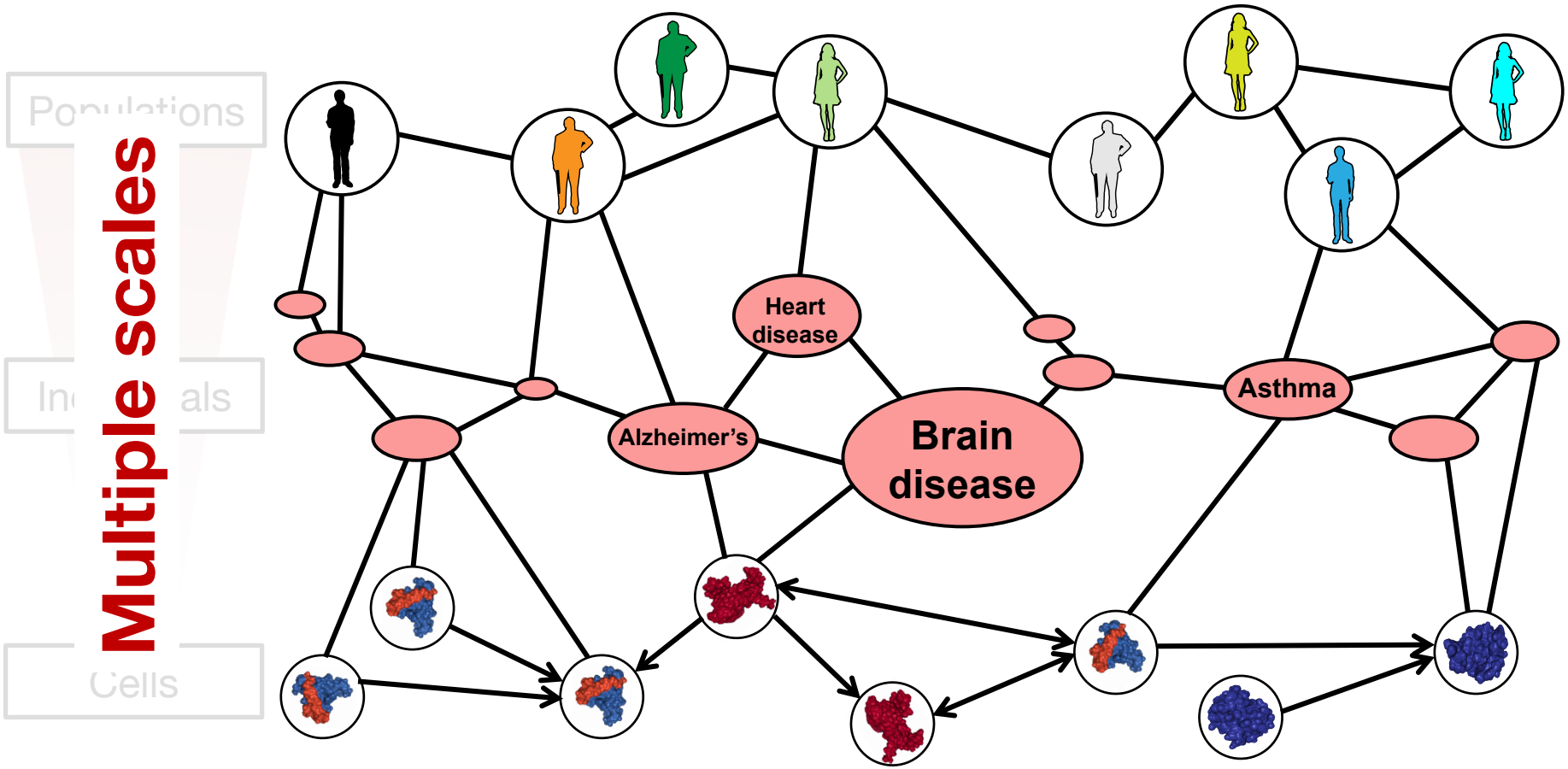
Why is it so challenging to realize this vision?





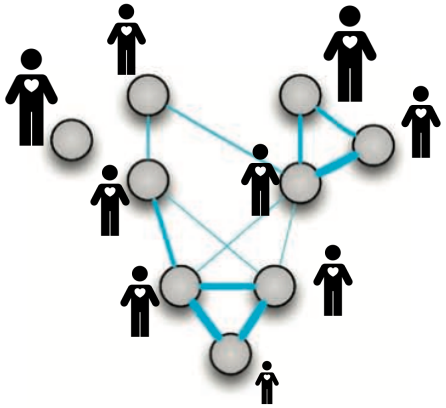
Multi-scale: molecules, individuals, populations
Heterogeneous: experimental readouts, curated annotations, self-reported
Confounded: data from different labs, hospitals, biotech platforms, species

Networks allow for integration of biomedical data

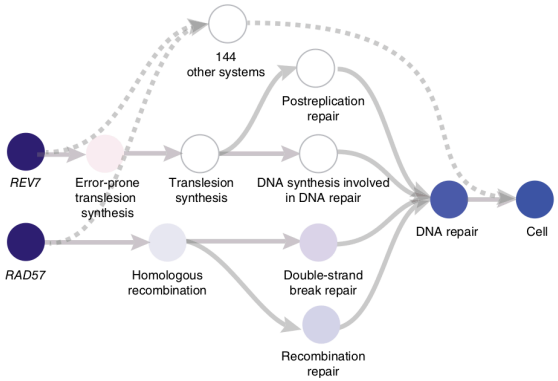


Rich, multimodal data

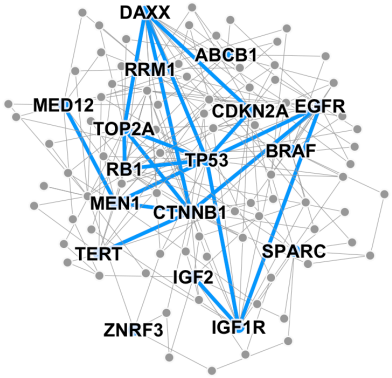
Many biomedical data are networks



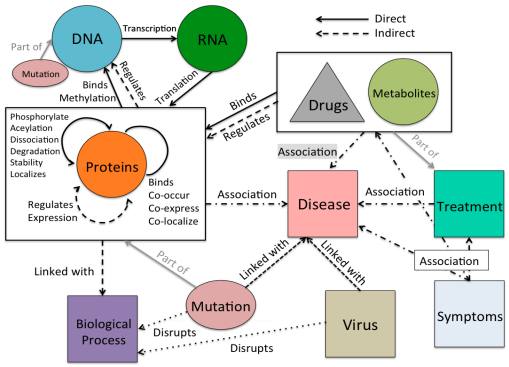
Patient networks



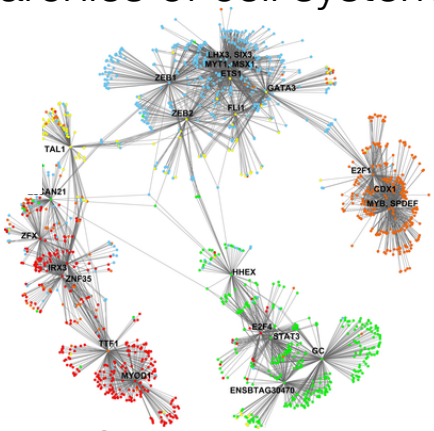
Hierarchies of cell systems



Disease pathways



Biomedical knowledge graphs



Gene interaction networks

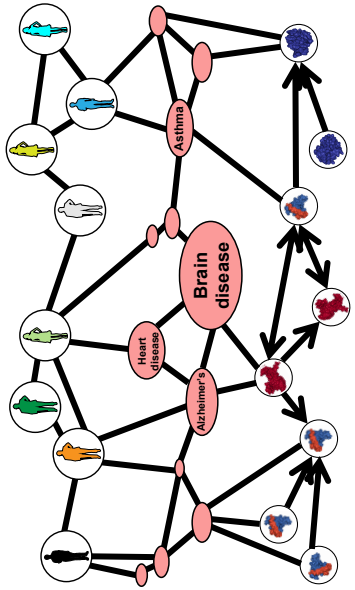
Cell-cell similarity networks

Prioritizing Network Communities, *Nature Communications* 2018

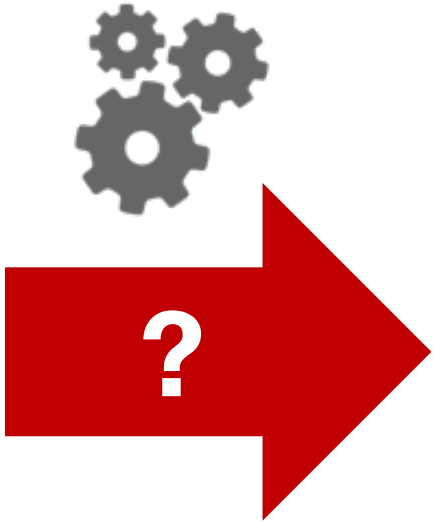
Network Enhancement as a General Method to Denoise Weighted Biological Networks, *Nature Communications* 2018

Evolution of resilience in protein interactomes across the tree of life, *PNAS* 2019

How to do machine learning on biomedical networks?



Networks



Predictions and insights

Biomedical ML opens new avenues for:

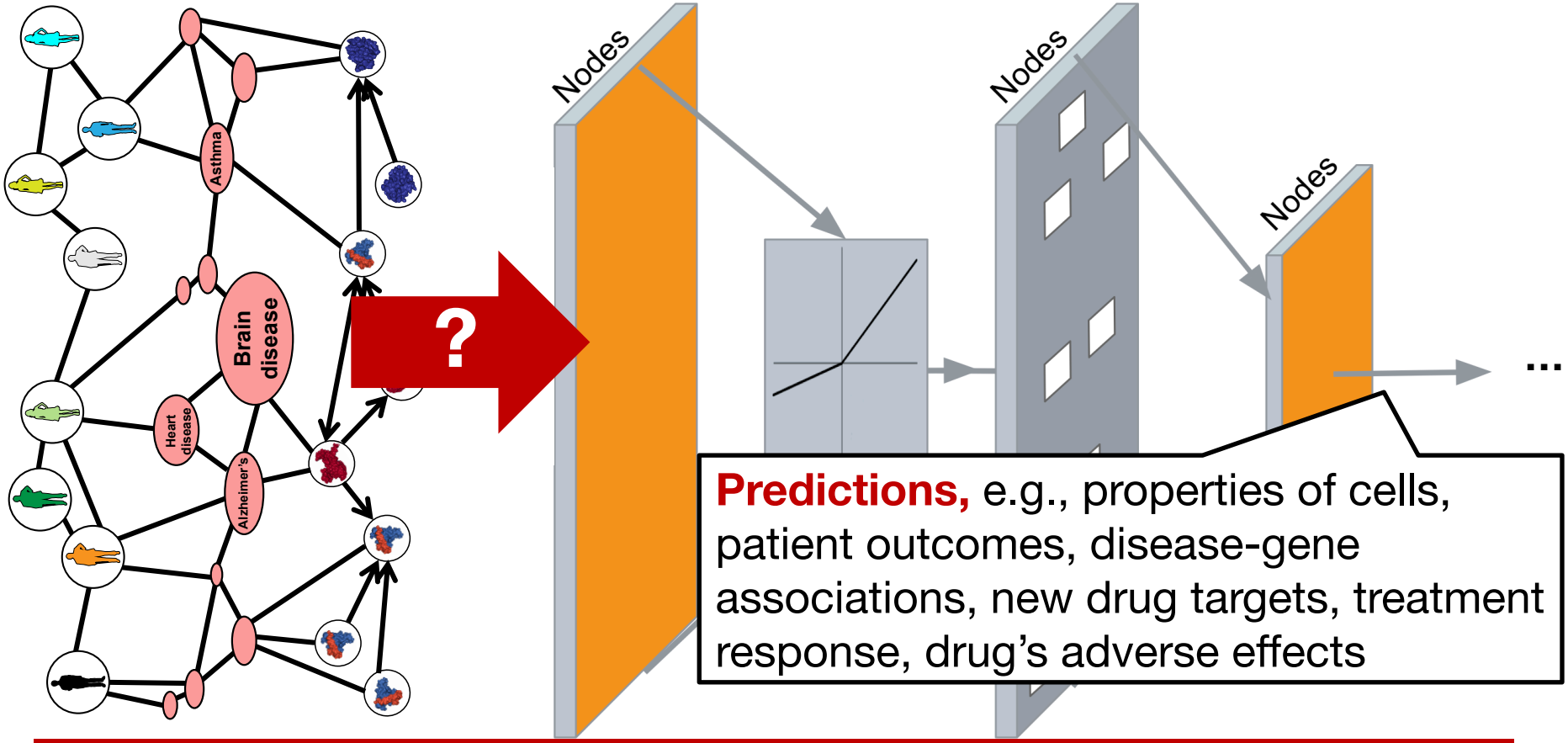
- Understanding nature, analyzing health, and developing medicines
- How predictive modeling is performed today at the fundamental level

Today's Talk

1. **Representation learning** for biomedical data

2. Three research applications:
   - Used new approach to predict **safety** and **side effects of drug combinations**
   - Used new approach to **repurpose old drugs for new diseases**
   - Used new approach to **answer logical queries** on **knowledge graphs**

How to learn deep models on biomedical networks?



Networks are a powerful data representation, but are challenging to work with for prevailing ML

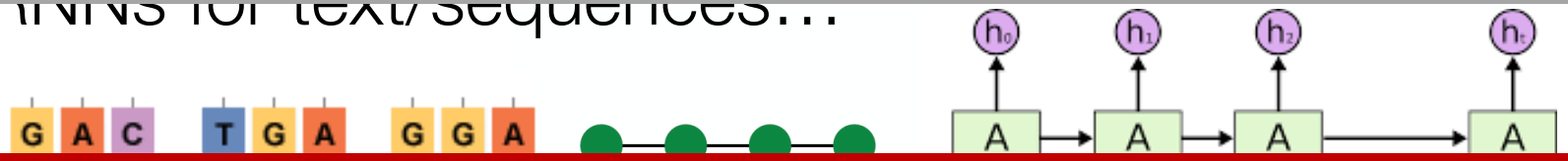
Prevailing Deep Models

Primarily designed for grids or simple sequences:

- CNNs for fixed-size images/grids

These models brought extraordinary gains in **computer vision, natural language processing, speech,** and **robotics**

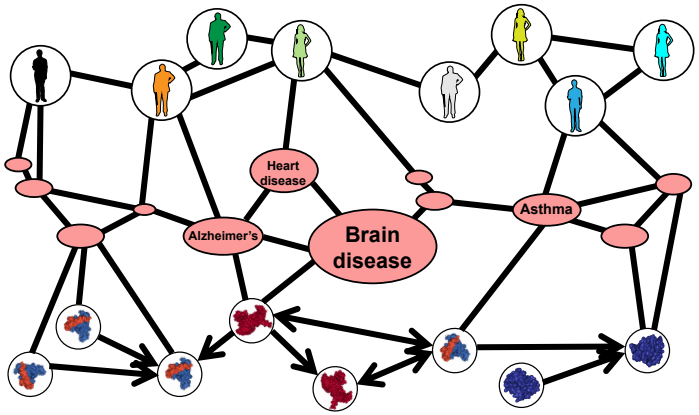
- RNNs for text/sequences…



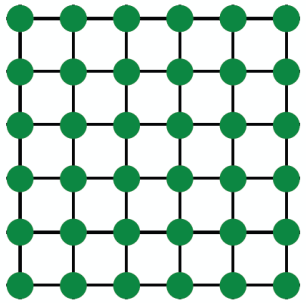
But are **unable to consider interactions**, the essence of networks

Why is deep learning on networks hard?

Biomedical networks are far more complex!



Biomedical networks



Images

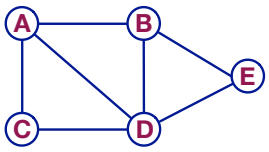
Text

Examples:

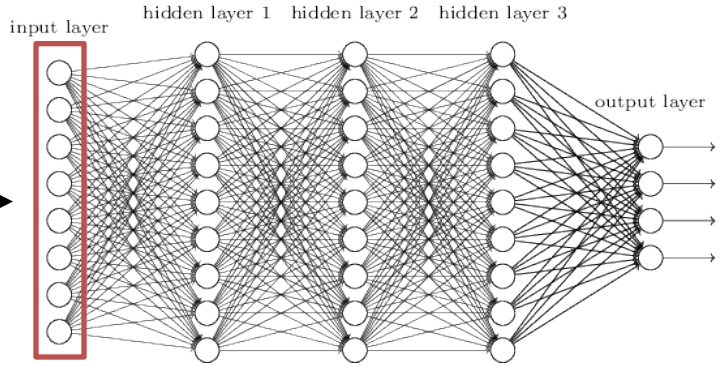
Human contact networks, Disease networks, Patient networks, Cell similarity networks, Medical knowledge graphs

A Naïve Approach

- Join adjacency matrix and features
- Feed them into a deep neural model:



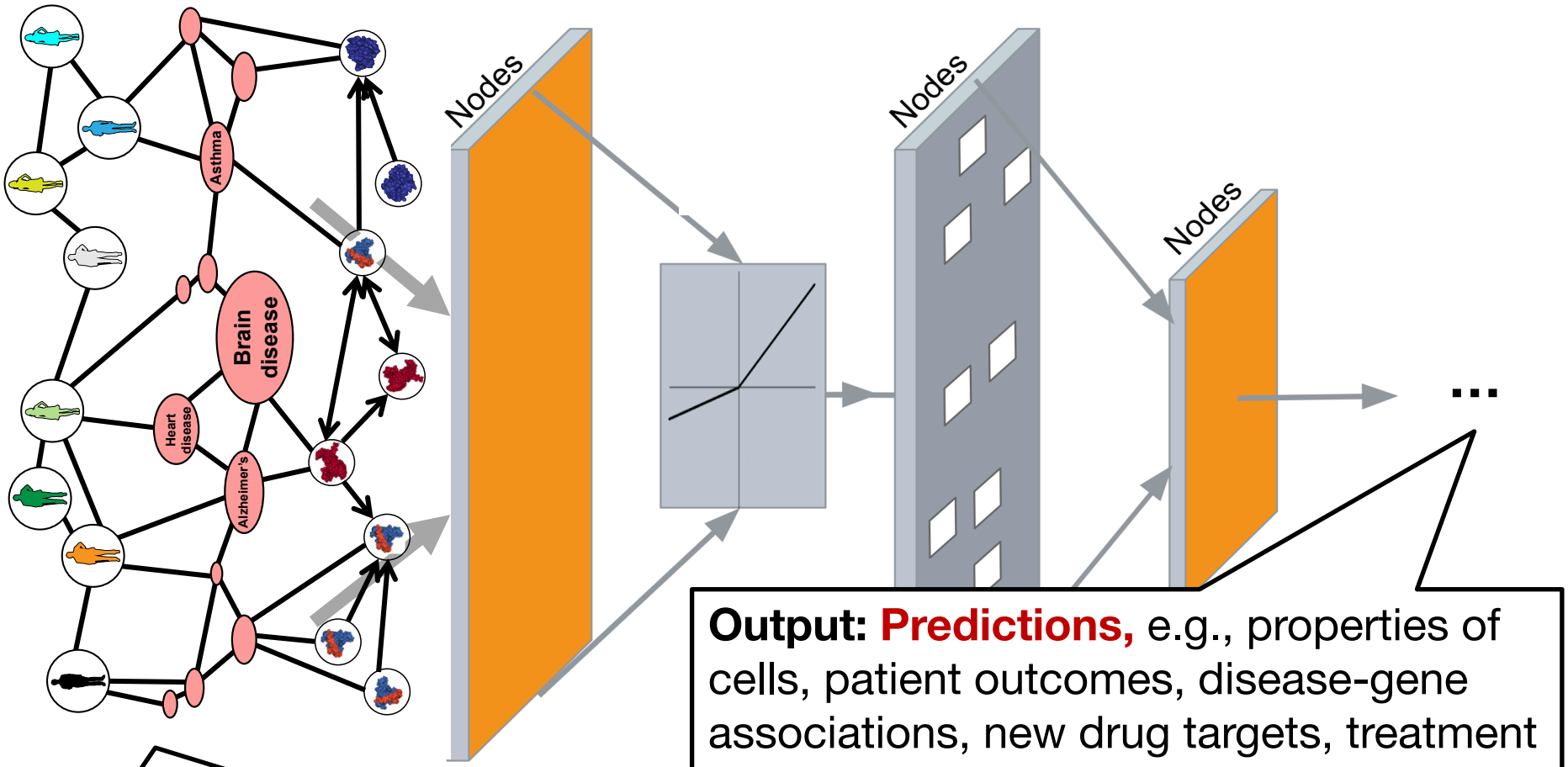
	A	B	C	D	E	Feat	
A	0	1	1	1	0	1	0
B	1	0	0	1	1	0	0
C	1	0	0	1	0	0	1
D	1	1	1	0	1	1	1
E	0	1	0	1	0	1	0



?

- Issues with this idea:
  - $O(N)$ parameters
  - Not applicable to graphs of different sizes
  - Not invariant to node ordering

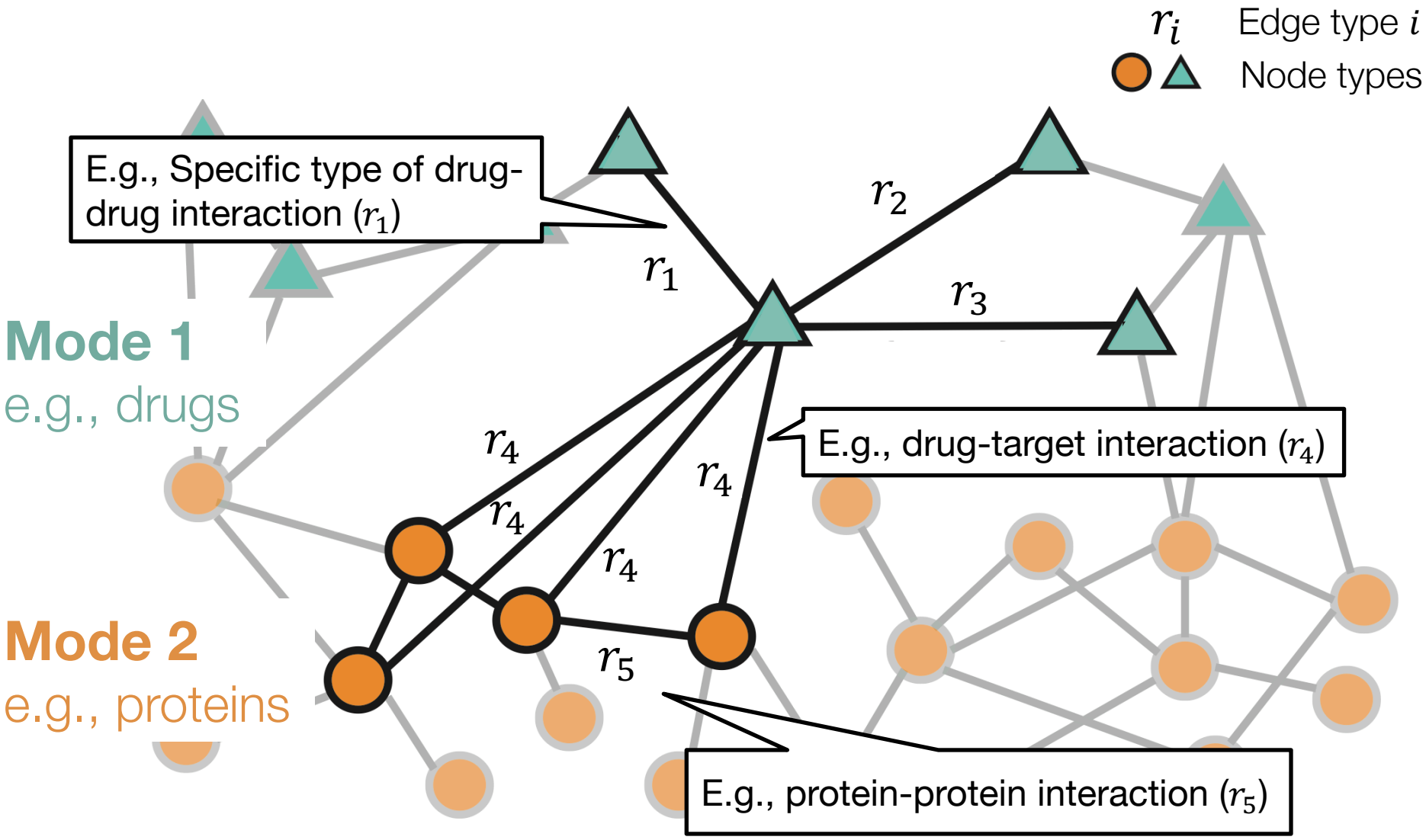
Today's goal: Deep learning for biomedical networks



Input: Knowledge network

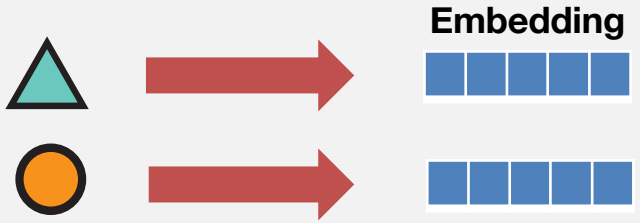
Output: Predictions, e.g., properties of cells, patient outcomes, disease-gene associations, new drug targets, treatment response, drug's adverse effects

Setup: A Multimodal Network

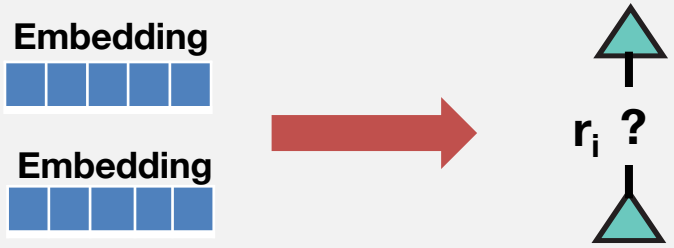


Overview of our deep learning approach for networks

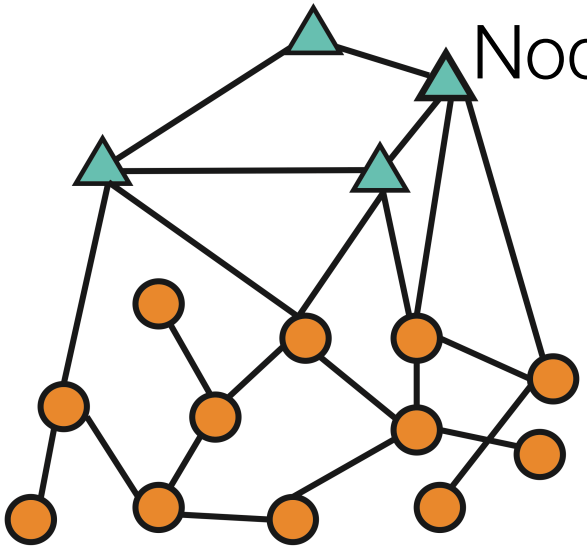
1. **Encoder:** Take a multimodal network and learn an *embedding* for every node



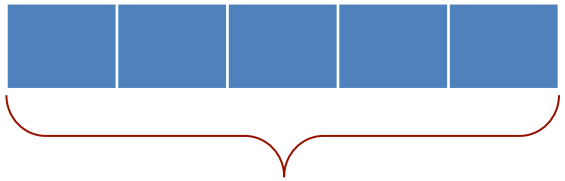
2. **Decoder:** Use the learned embeddings to predict labeled edges between nodes



Embedding Nodes



$$f: v \rightarrow \mathbb{R}^d$$



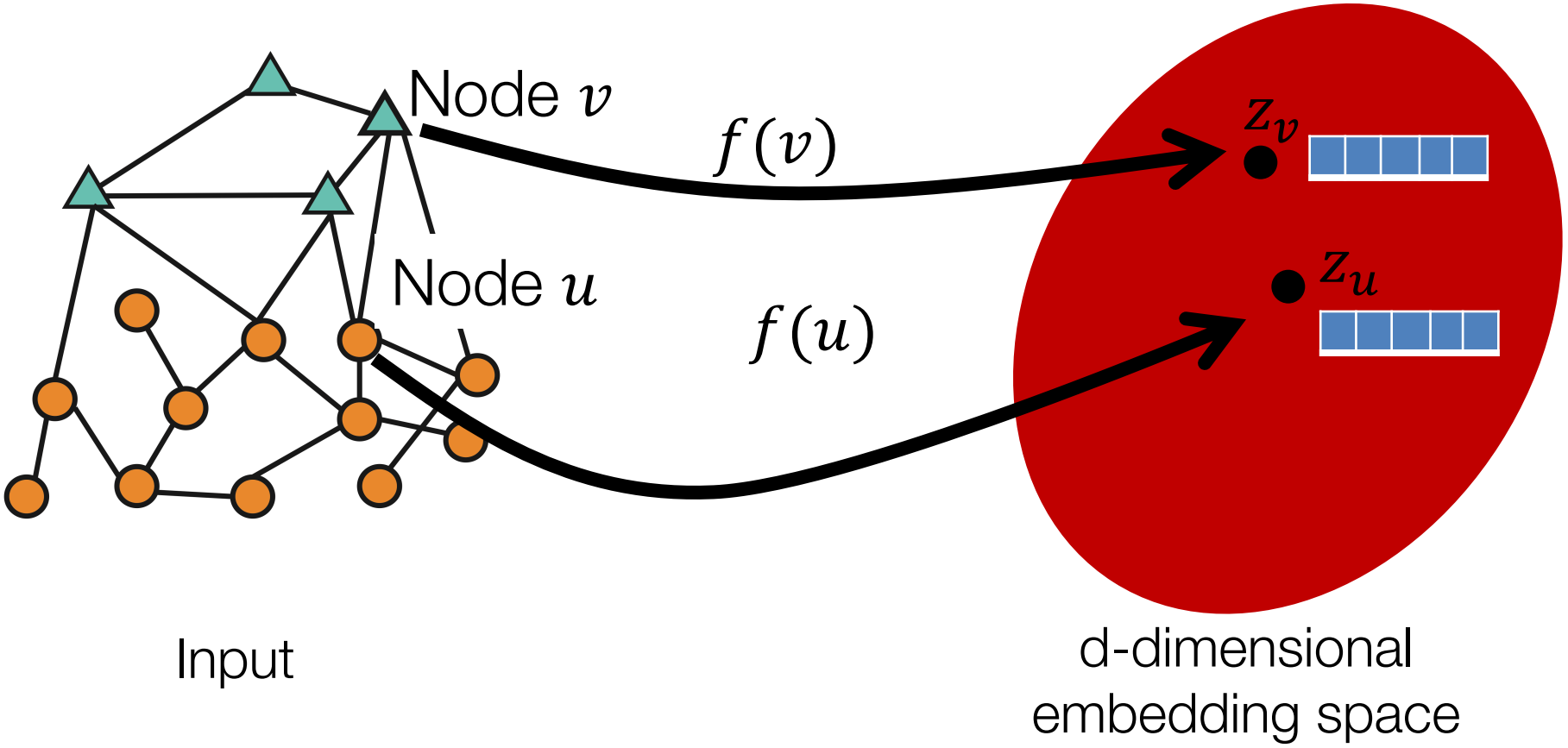
$\mathbb{R}^d$

Feature representation, embedding

Objective: Map nodes to d-dimensional embeddings such that nodes with similar network neighborhoods are embedded close together

Next: How to learn mapping function f?

Embedding Nodes

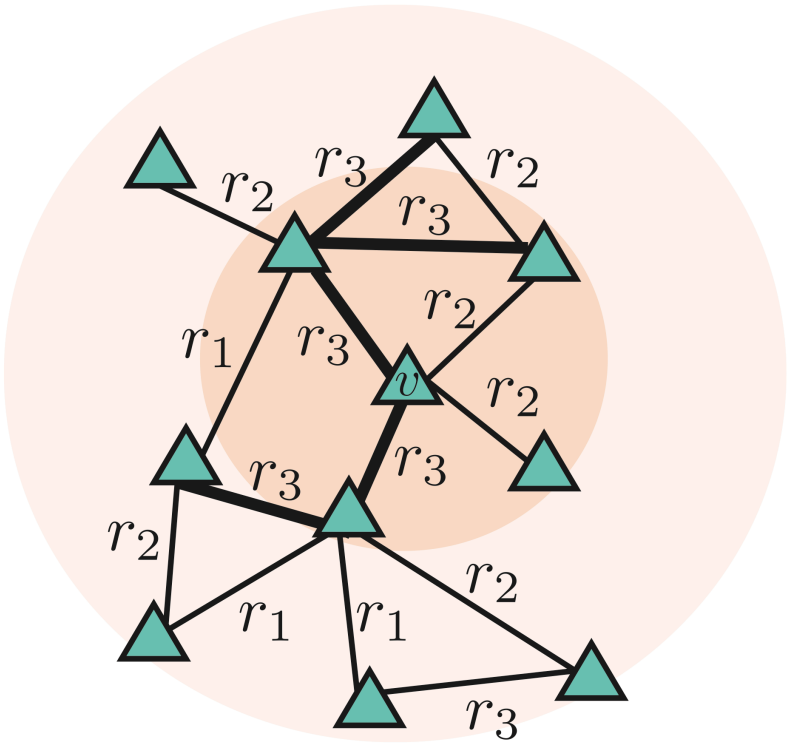


Goal: Map nodes to d-dimensional embeddings such that nodes with similar network neighborhoods are embedded close together

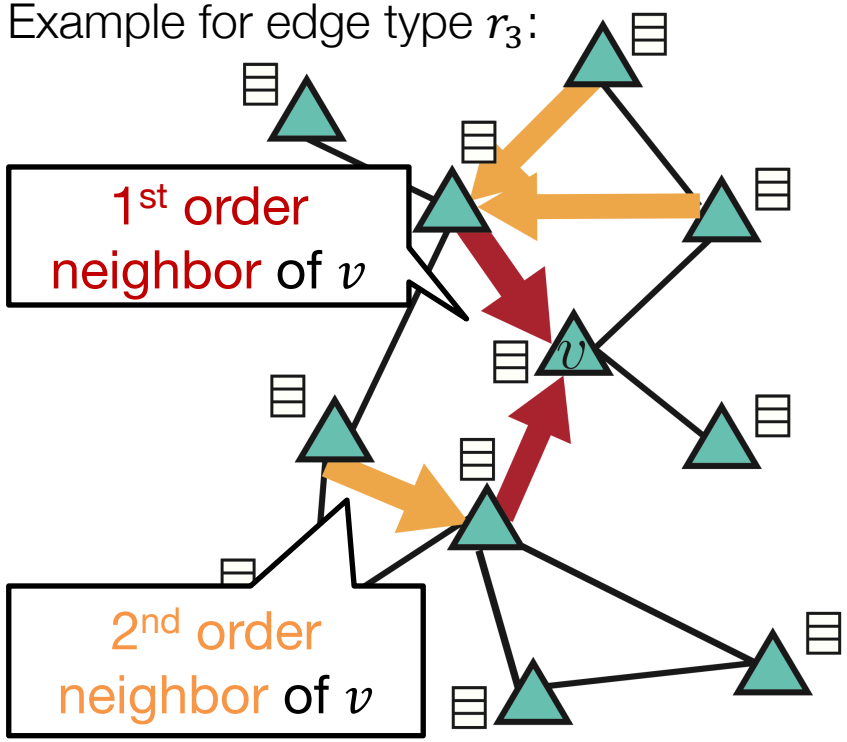
Key Idea: Aggregate Neighbors

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

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

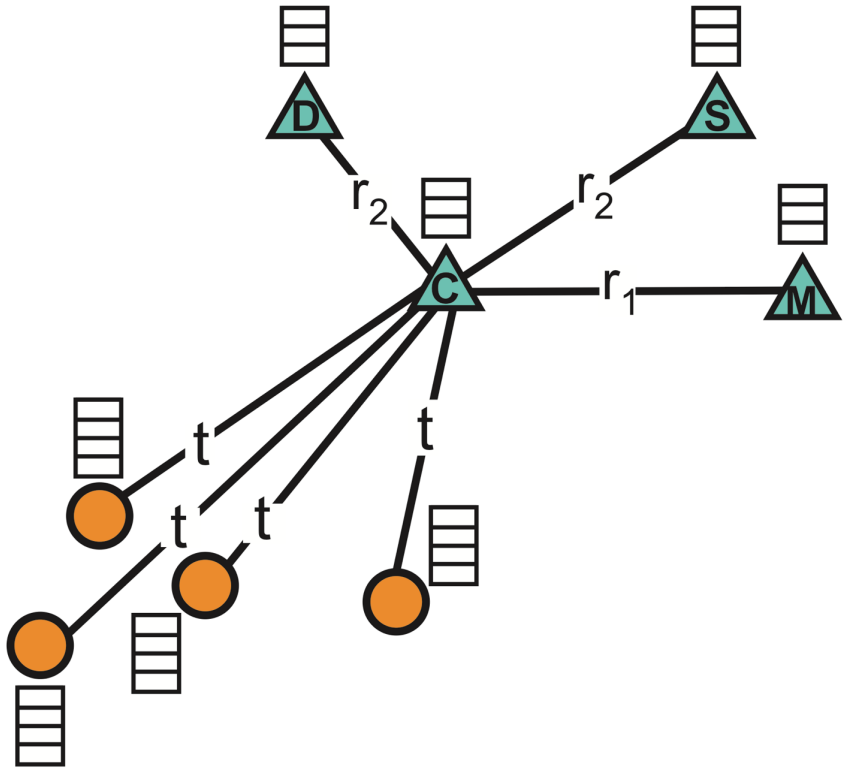


2) Learn how to transform and propagate information across computation graph

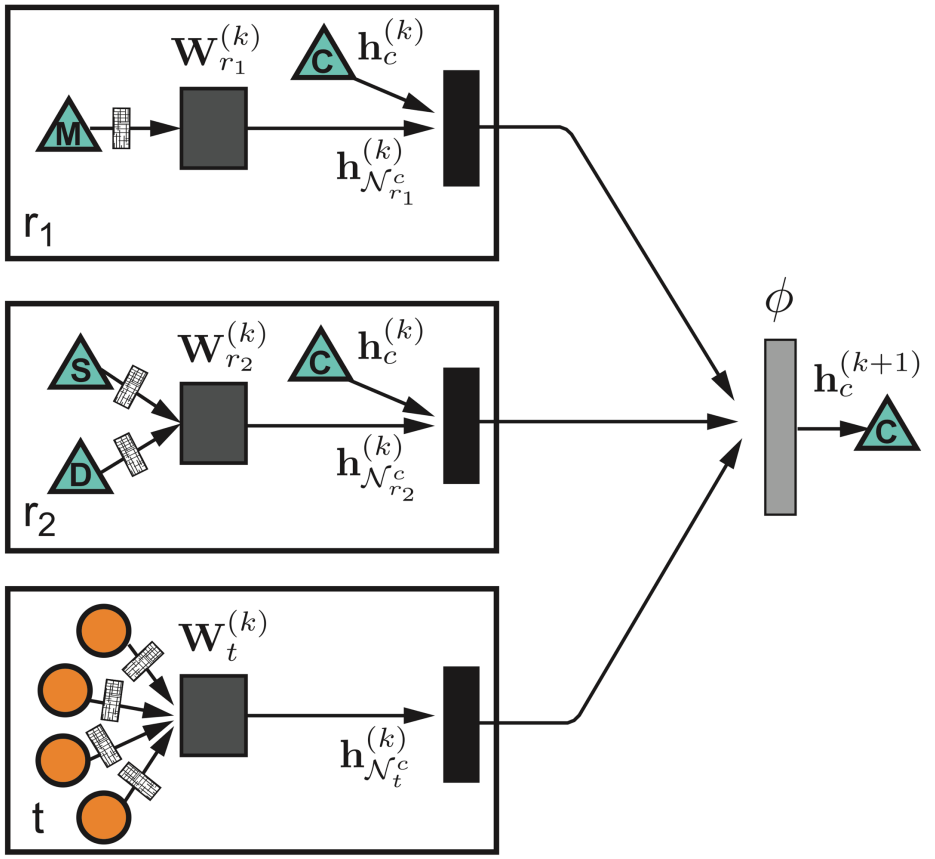


Example: Aggregate Neighbors

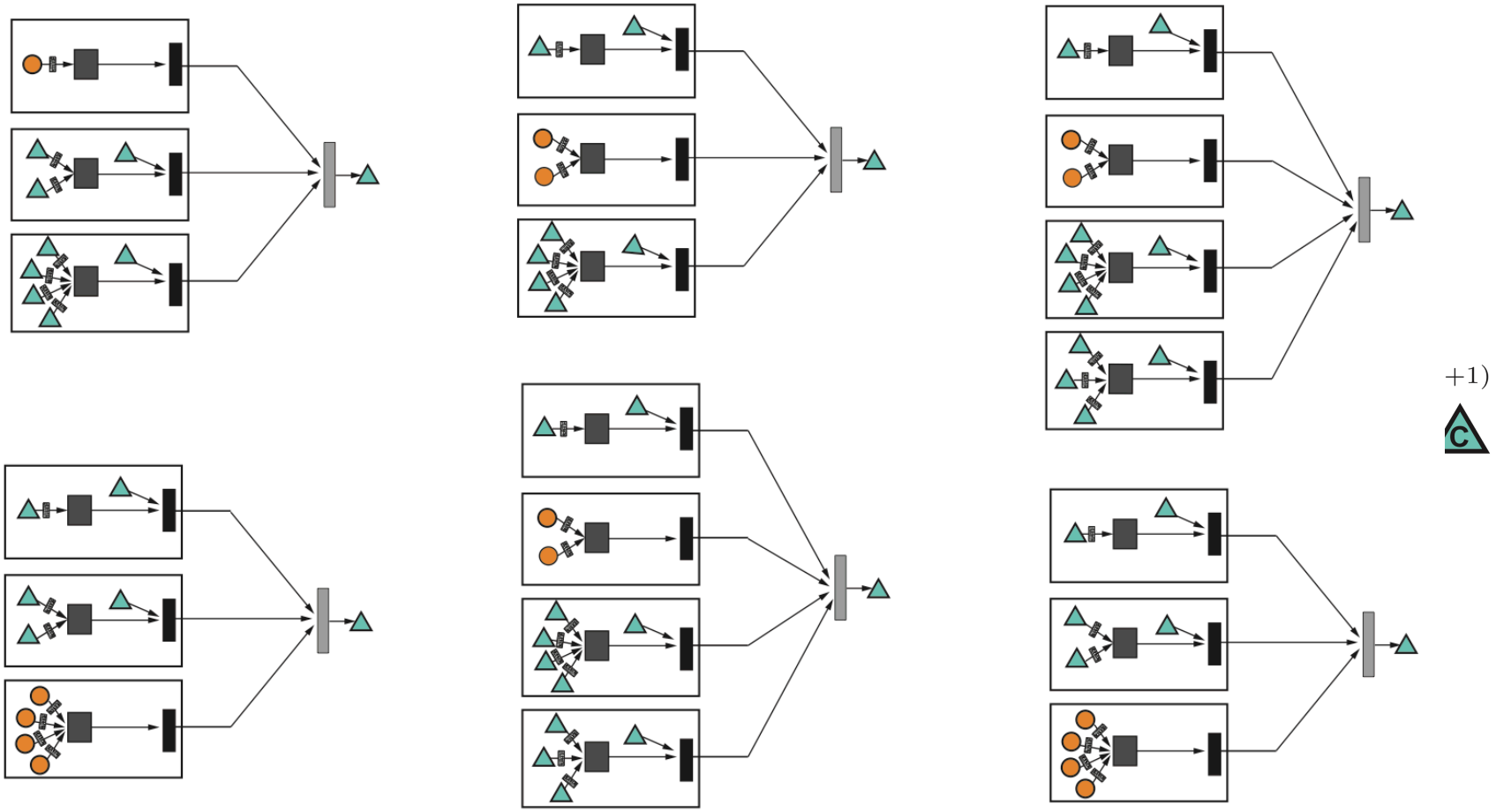
1st order network neighborhood of node C



1st order computation graph of node C

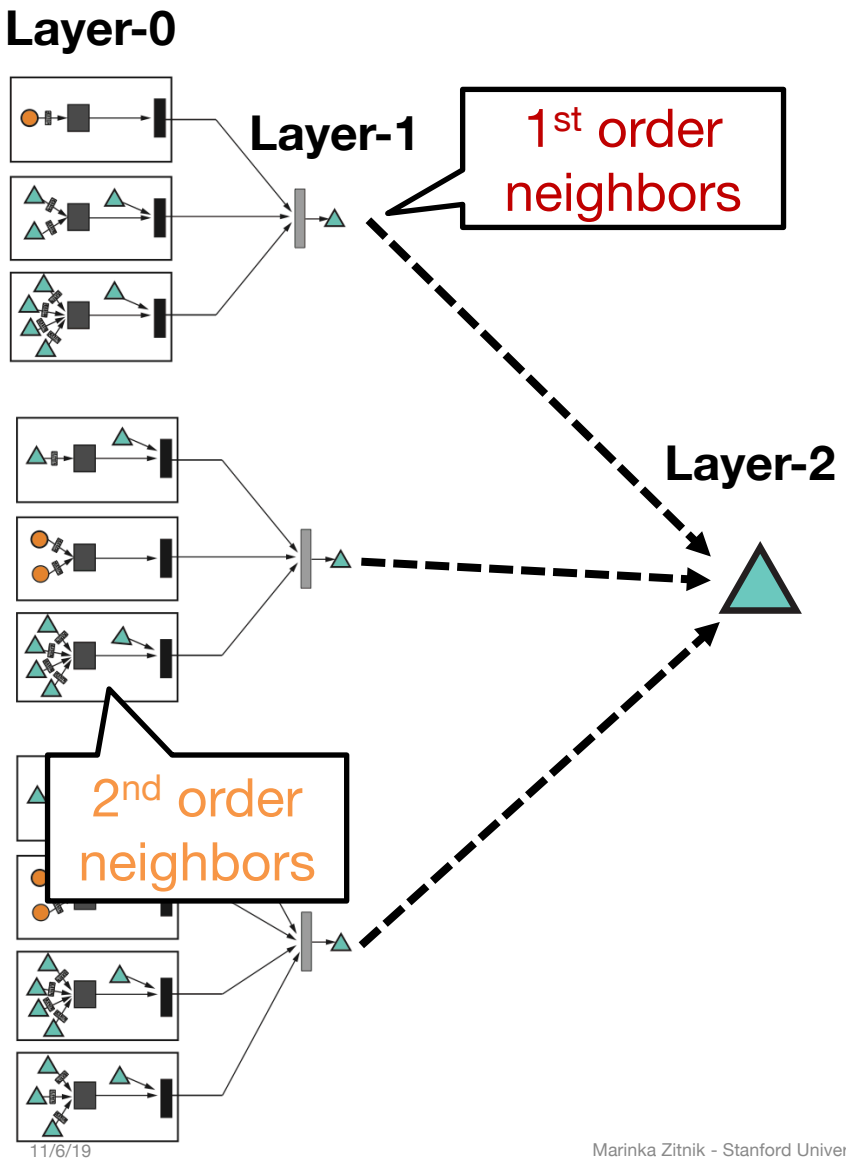


Every node learns how to aggregate its own neighbors



Every node defines a unique computation graph

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**

The Math: Deep Graph 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 node 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 aggregation:**

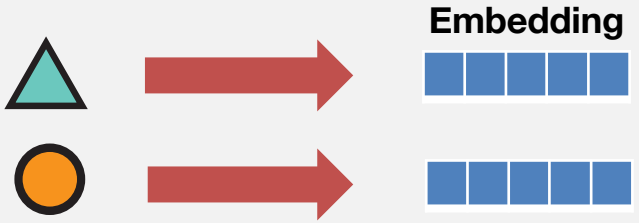
$$\mathbf{z}_v = \mathbf{h}_v^{(K)}$$

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

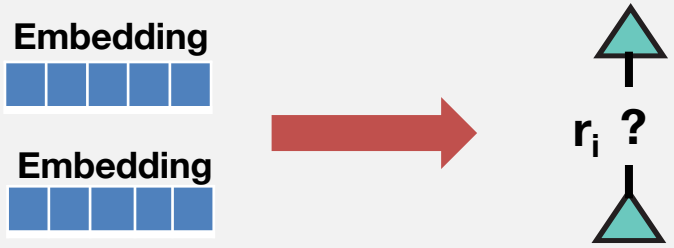
Normalization constant, fixed *e.g.*, $1/|N_v^r|$, or learned

Overview of our deep learning approach for multimodal networks

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



2. **Decoder:** Use the learned embeddings and make predictions



What Can We Predict?

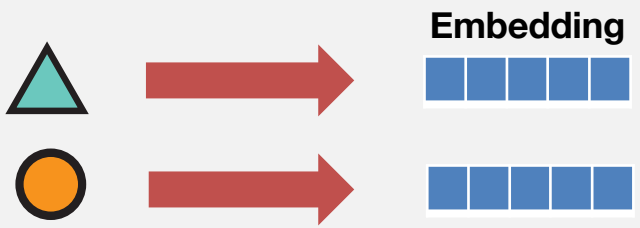
- **Node prediction:** E.g., Predicting protein functions across tissues
- **Pairs of nodes:** E.g., Predicting side-effects and safety of drug combinations
- **Subgraph prediction:** E.g., Predicting what drug treats what disease
- **Graph prediction:** E.g., Predicting properties of molecules

We can now apply deep learning much more broadly, not only to medical images and biological, DNA sequences

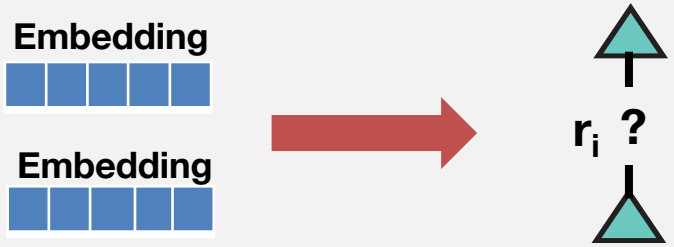
New frontiers for applications in **biology** and **medicine**

Overview of our deep learning approach for multimodal networks

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



2. **Decoder:** Use the learned embeddings and make predictions



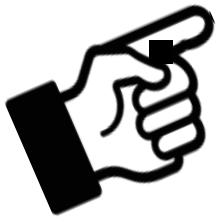
Training the model: Feed embeddings into any loss function and run stochastic gradient descent to train weight parameters:

- Use a loss based on e.g., random walks, node proximity in the graph
- Directly train the model for a supervised task (e.g., node classification)

Today's Talk

1. **Representation learning** for biomedical data

2. Three research applications:



- Used new approach to predict **safety** and **side effects of drug combinations**
- Used new approach to **repurpose old drugs for new diseases**
- Used new approach to **answer logical queries** on **knowledge graphs**

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

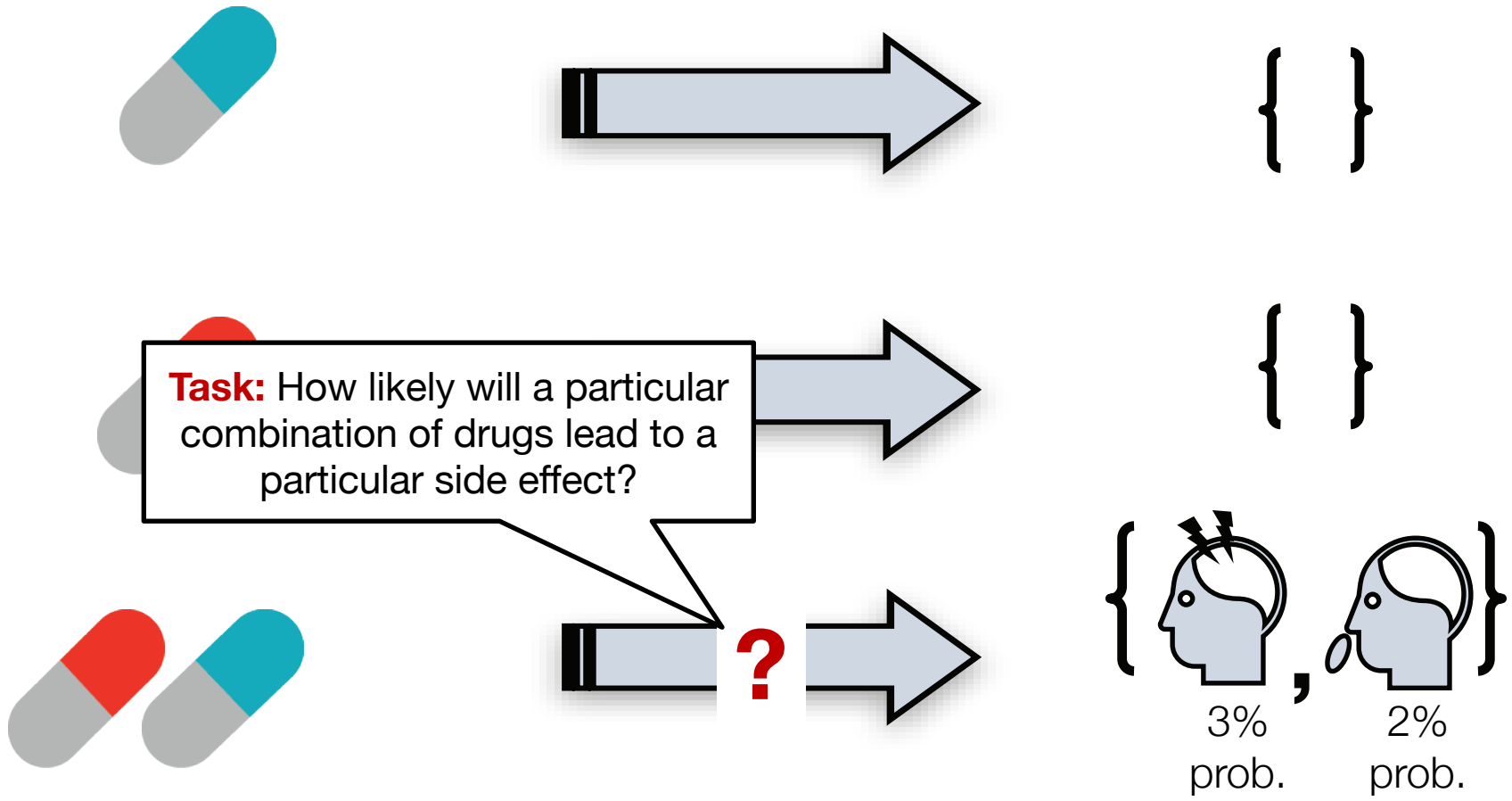
Annual costs in treating side effects exceed **$177** billion in the U.S. alone

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

Unexpected Drug Interactions

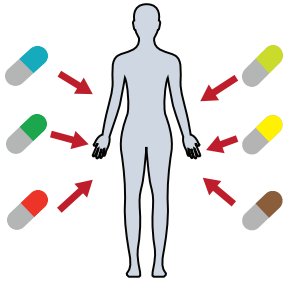
Co-prescribed drugs

Side Effects



Why is modeling polypharmacy hard?

Combinatorial explosion

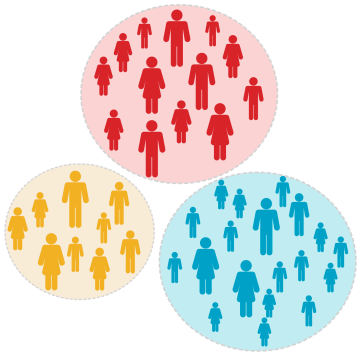


- >13 million possible combinations of 2 drugs
- >20 billion possible combinations of 3 drugs

Non-linear & non-additive interactions

- Different effect than the additive effect of individual drugs

Small subsets of patients



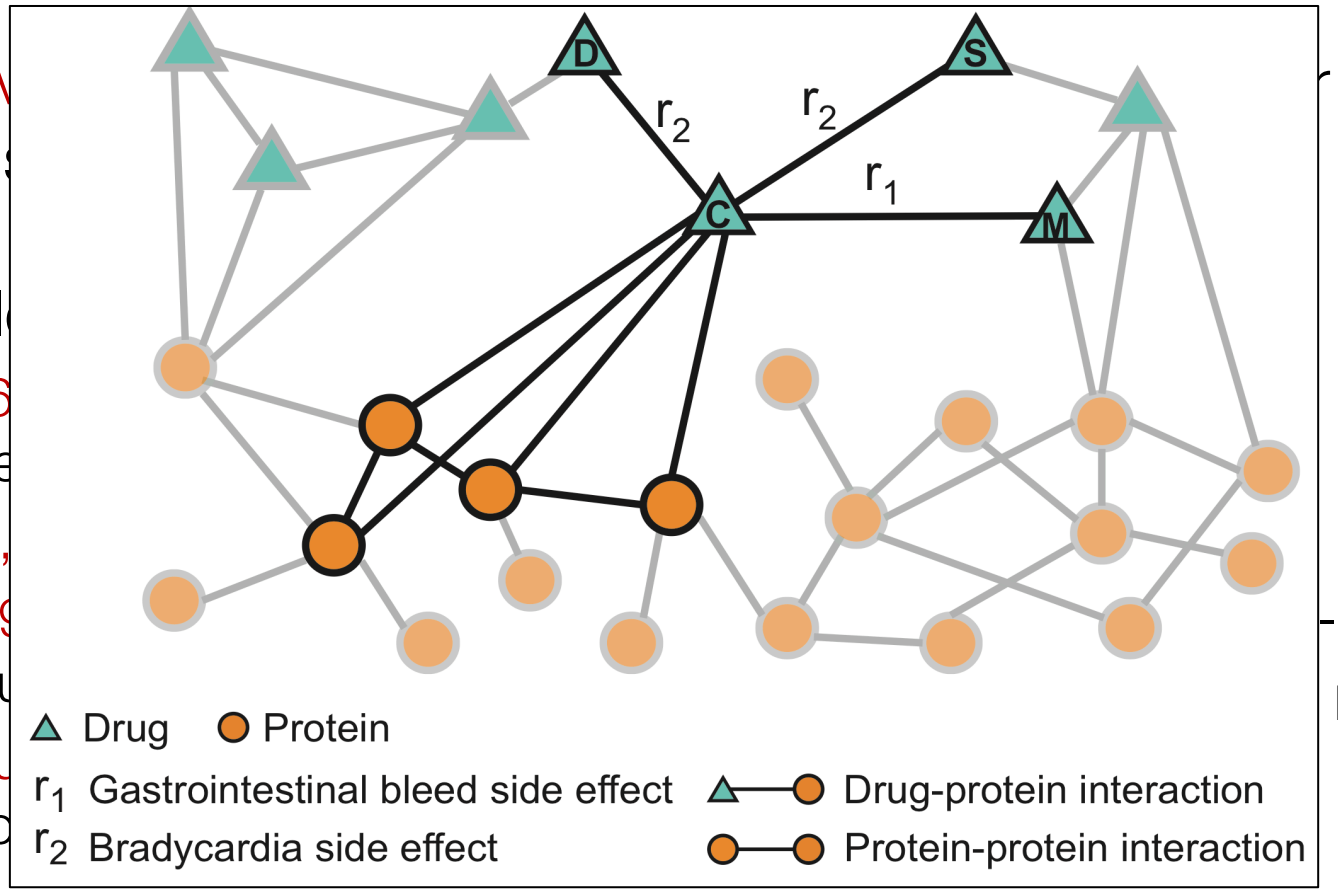
- Side effects are interdependent
- No info on drug combinations not yet used in patients

We need polypharmacy dataset

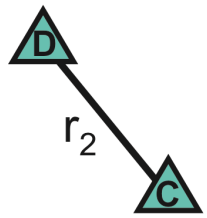
Objectiv
all drug

We buil

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



Drug-drug



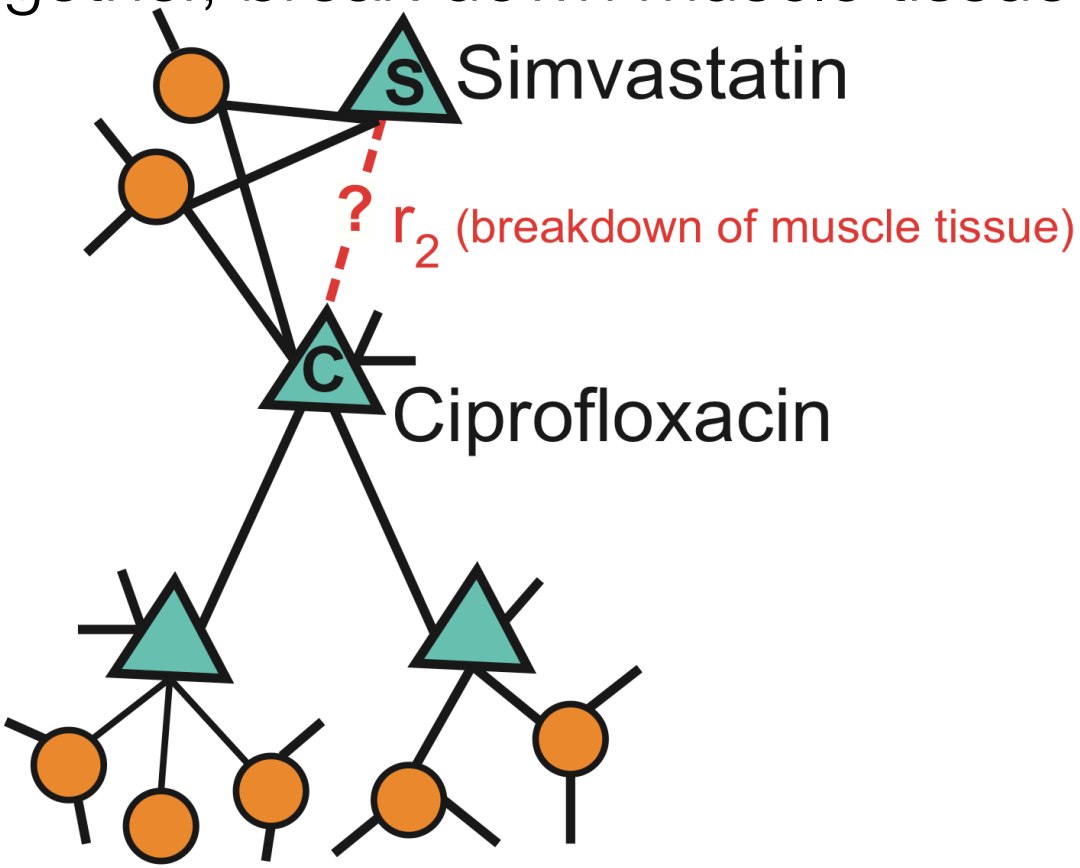
Drug-protein

Protein-protein

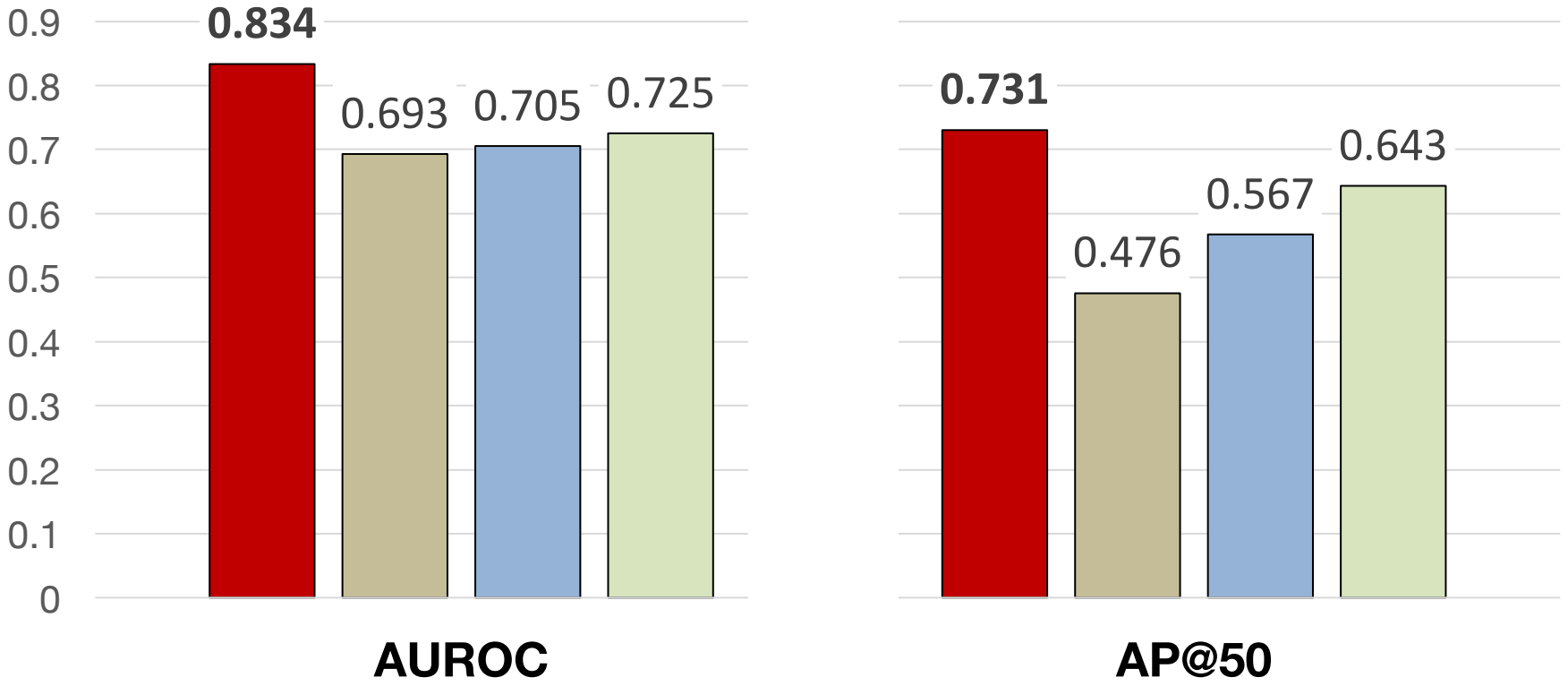
A polypharmacy network with over 5 million edges and over 1,000 different edge types

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]

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:

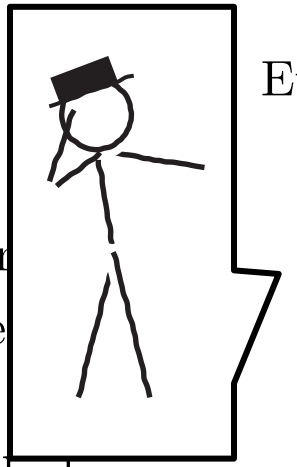
- **Doctors** guidance on whether it is a good idea to prescribe a particular combination of drugs to a particular patient
- **Researchers** guidance on effective wet lab experiments and new drug therapies with fewer side effects

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

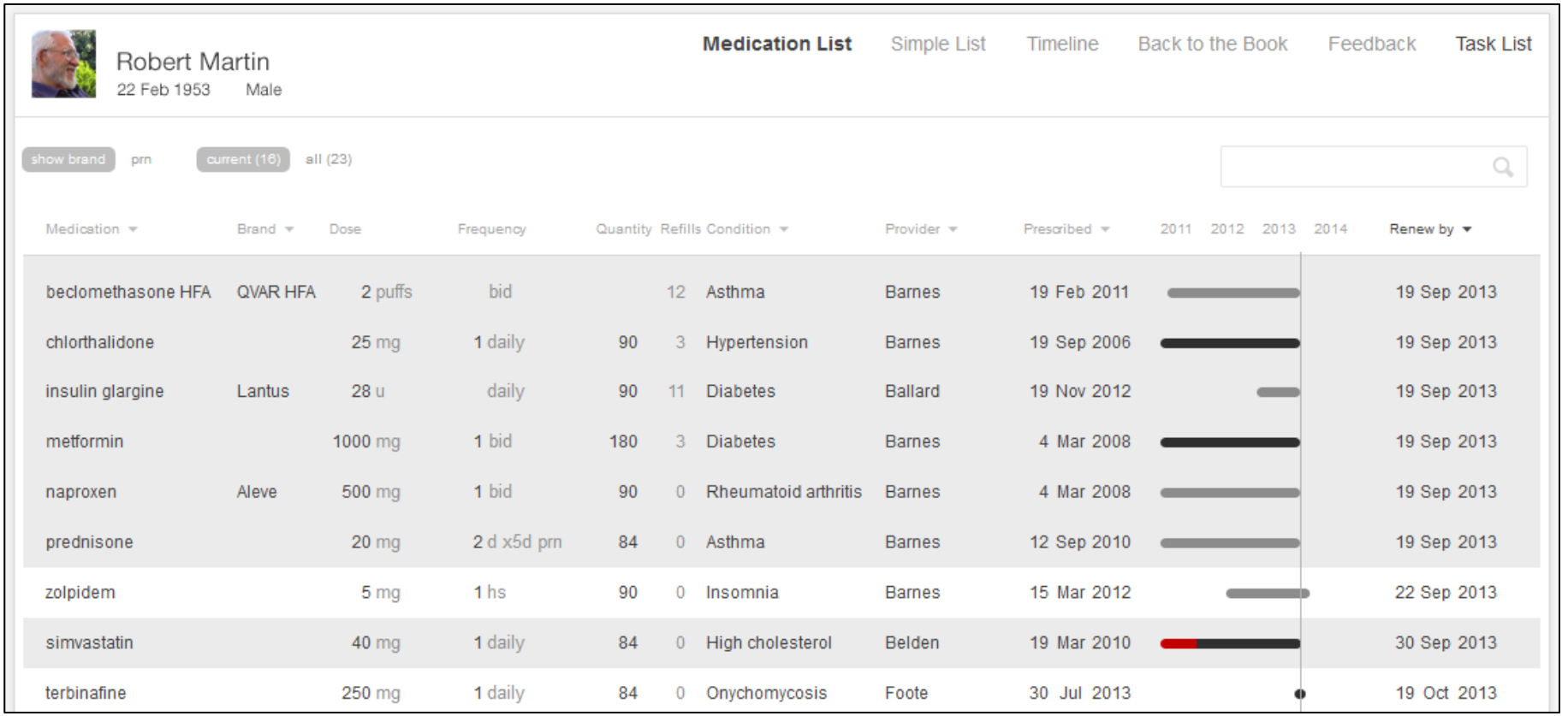
Clinical Validation of New Predictions

Drug interaction markers, lab values, and many other surrogates

NEWTON-WELLESLEY HOSPITAL

MASSACHUSETTS GENERAL HOSPITAL

Stanford MEDICINE



Robert Martin
22 Feb 1953 Male

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

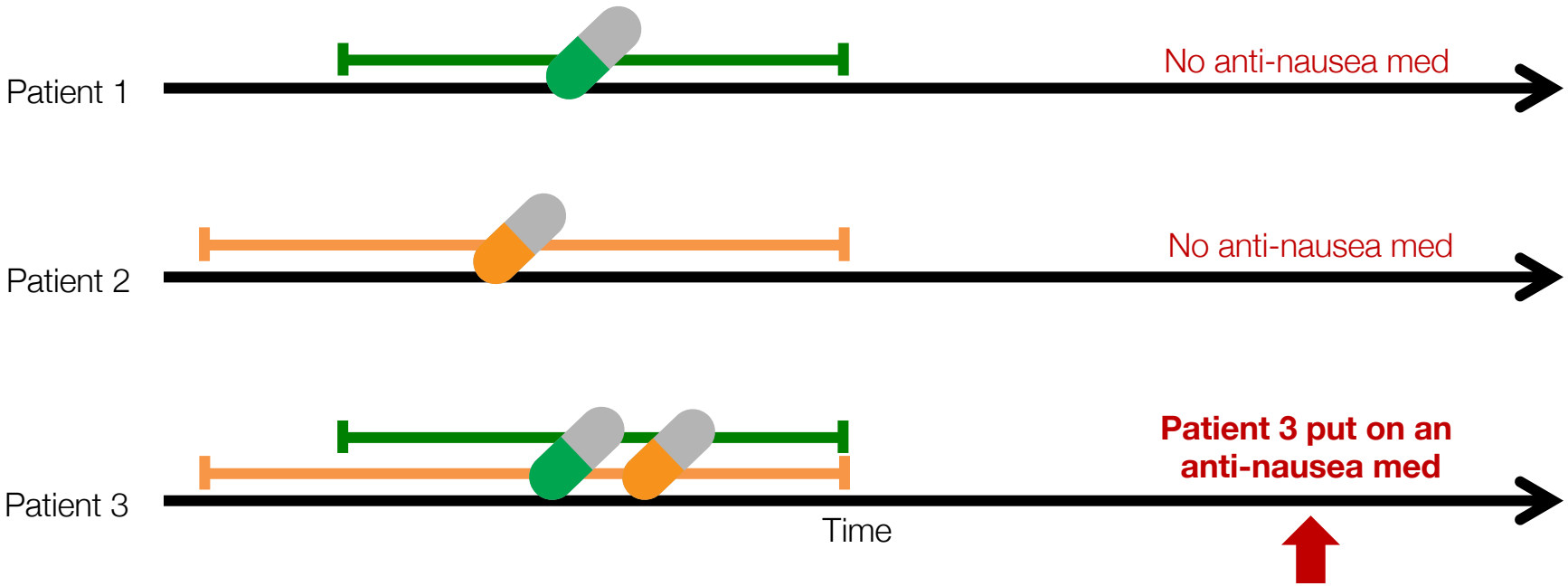
show brand | prn | current (16) | all (23)

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

Clinical Validation: Key Idea

Question: Is it a good idea to prescribe a particular combination of drugs to a particular patient?

- E.g., Prediction: {💊,💊} cause nausea as a side effect



NEWTON-WELLESLEY HOSPITAL

MASSACHUSETTS GENERAL HOSPITAL

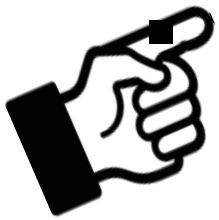
Stanford MEDICINE

Today's Talk

1. **Representation learning** for biomedical data

2. Three research applications:

- Used new approach to predict **safety** and **side effects of drug combinations**



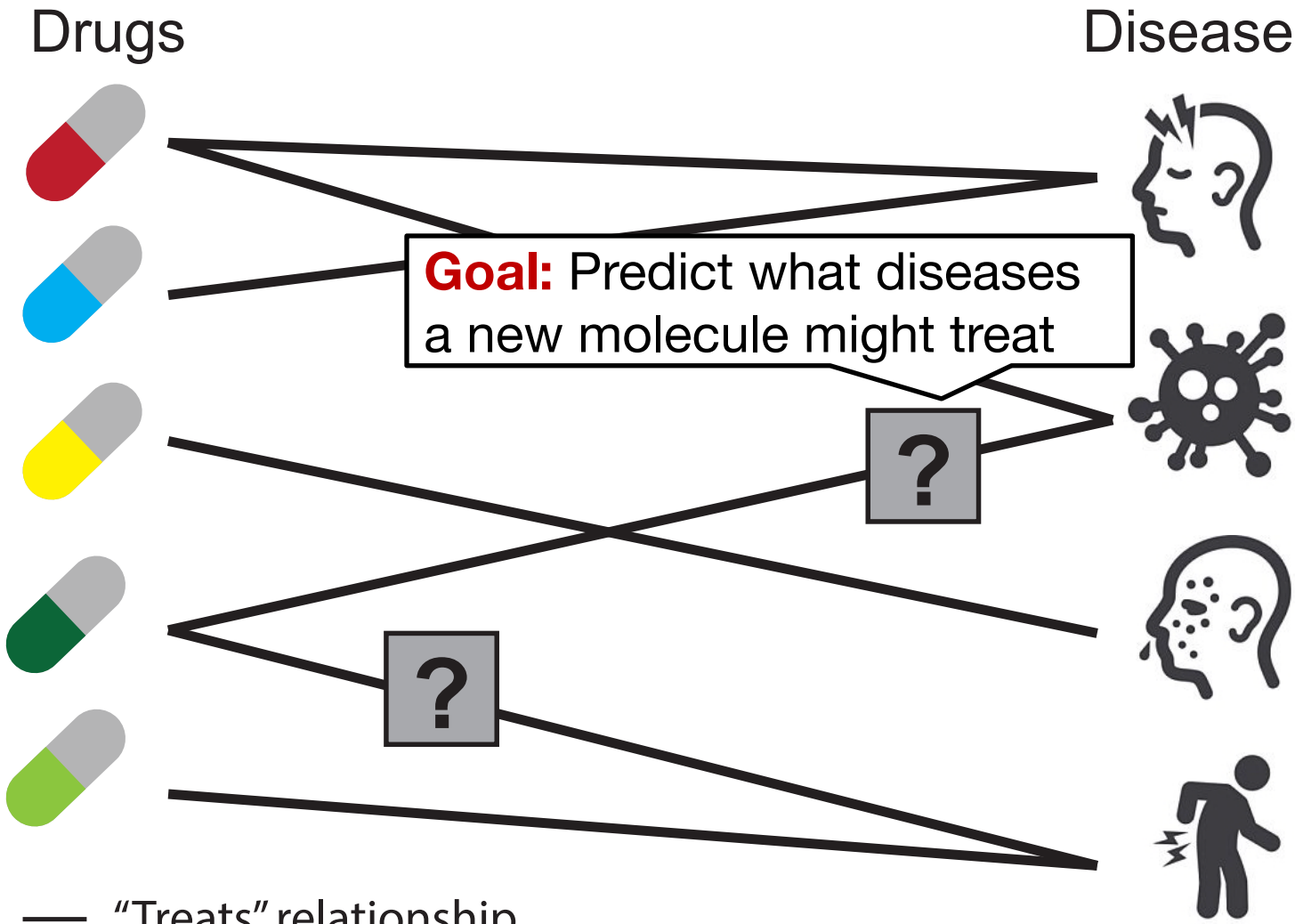
- Used new approach to **repurpose old drugs for new diseases**
- Used new approach to **answer logical queries** on **knowledge graphs**

New tricks for old drugs

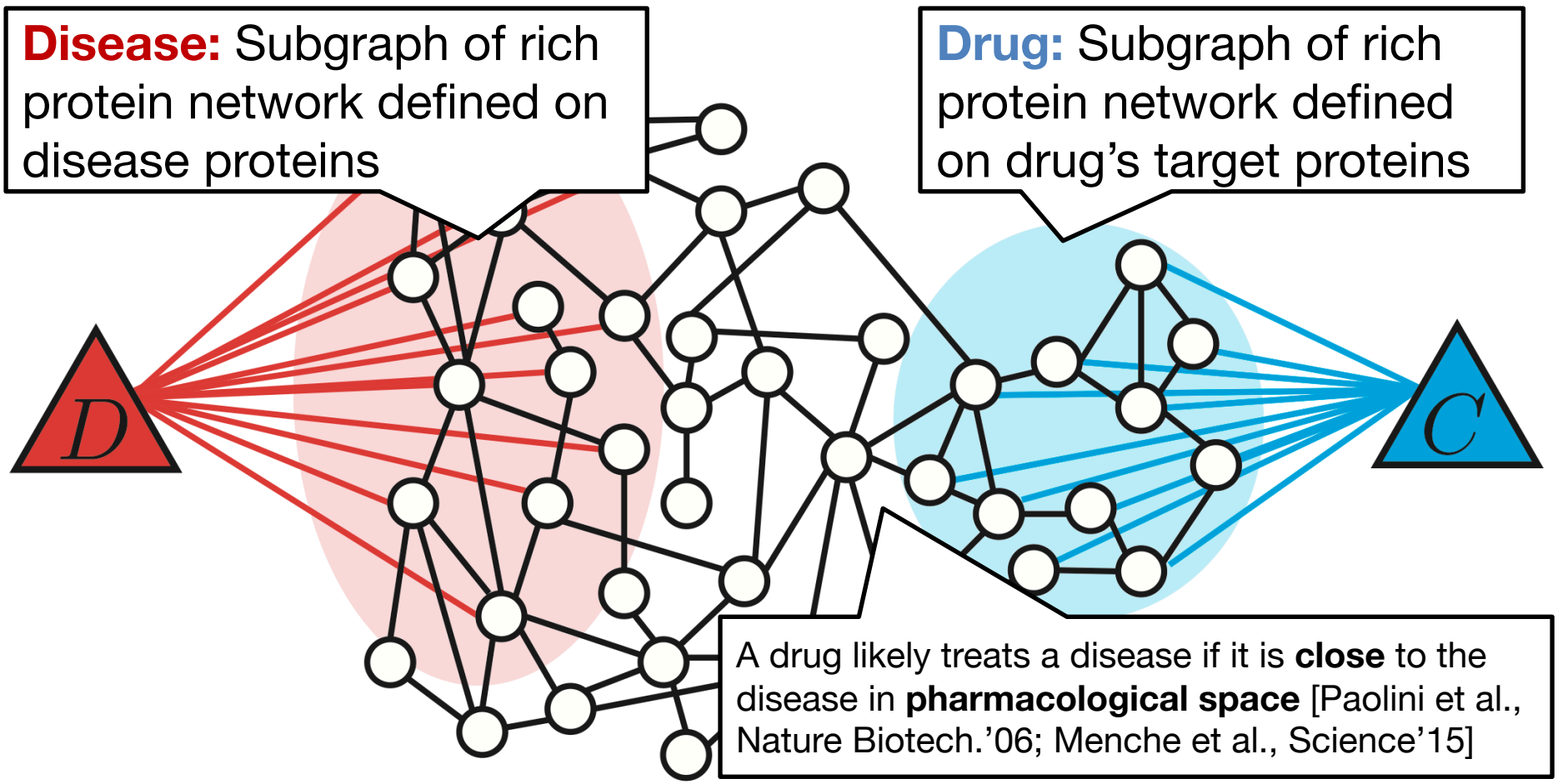
Goal: Find which diseases a drug (new molecule) could treat

Nature 2016

What drug treats what disease?



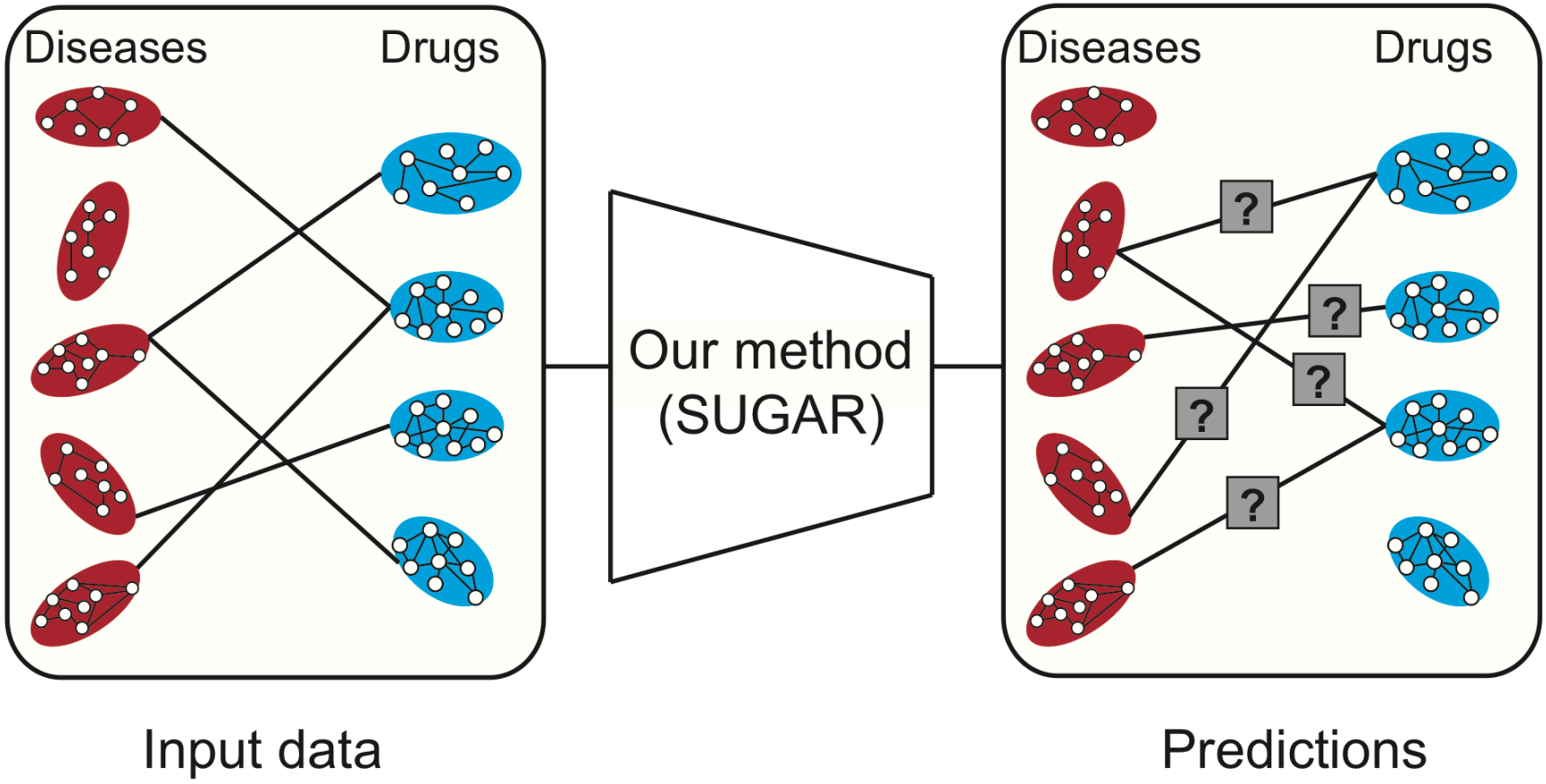
Key Insight: Subgraphs



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

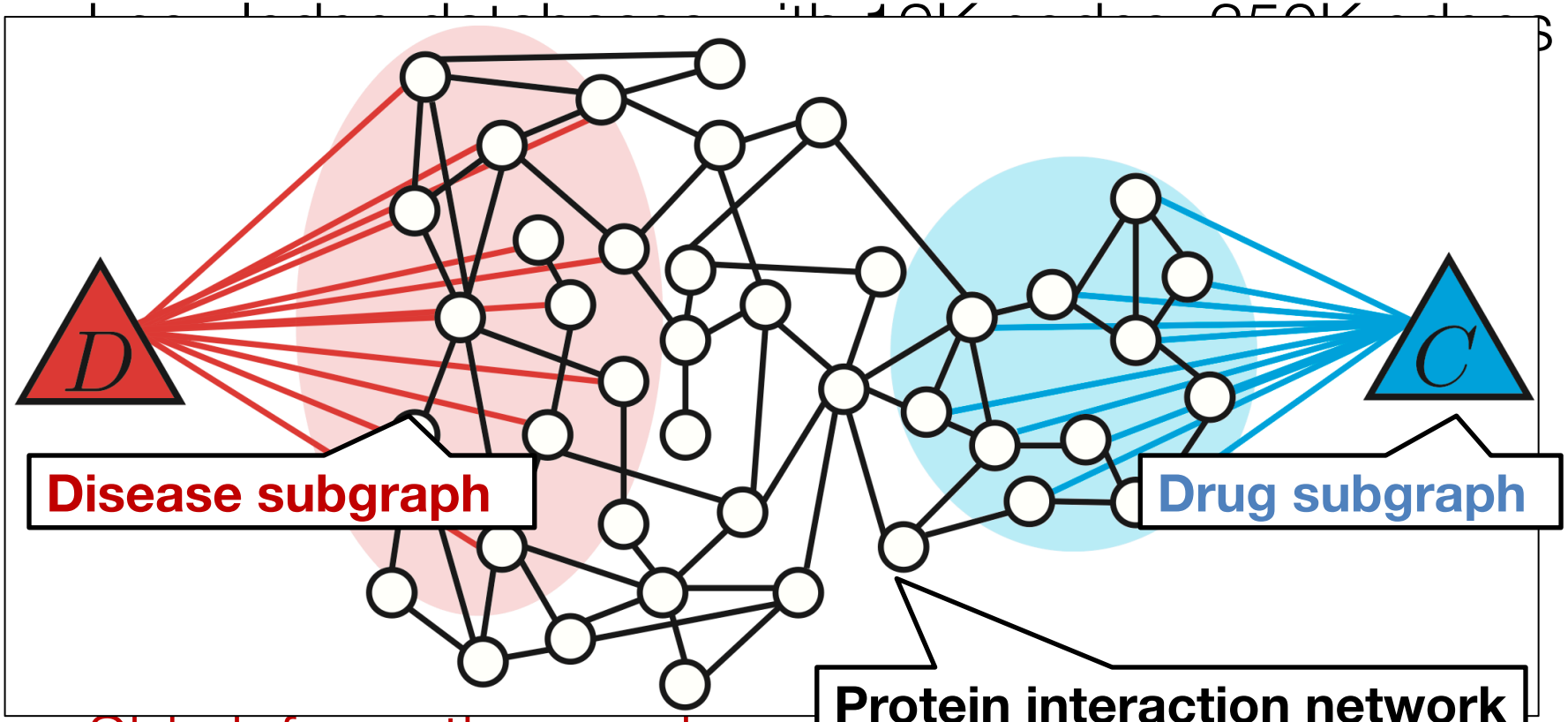
Predicting Links Between Drug and Disease Subgraphs

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



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

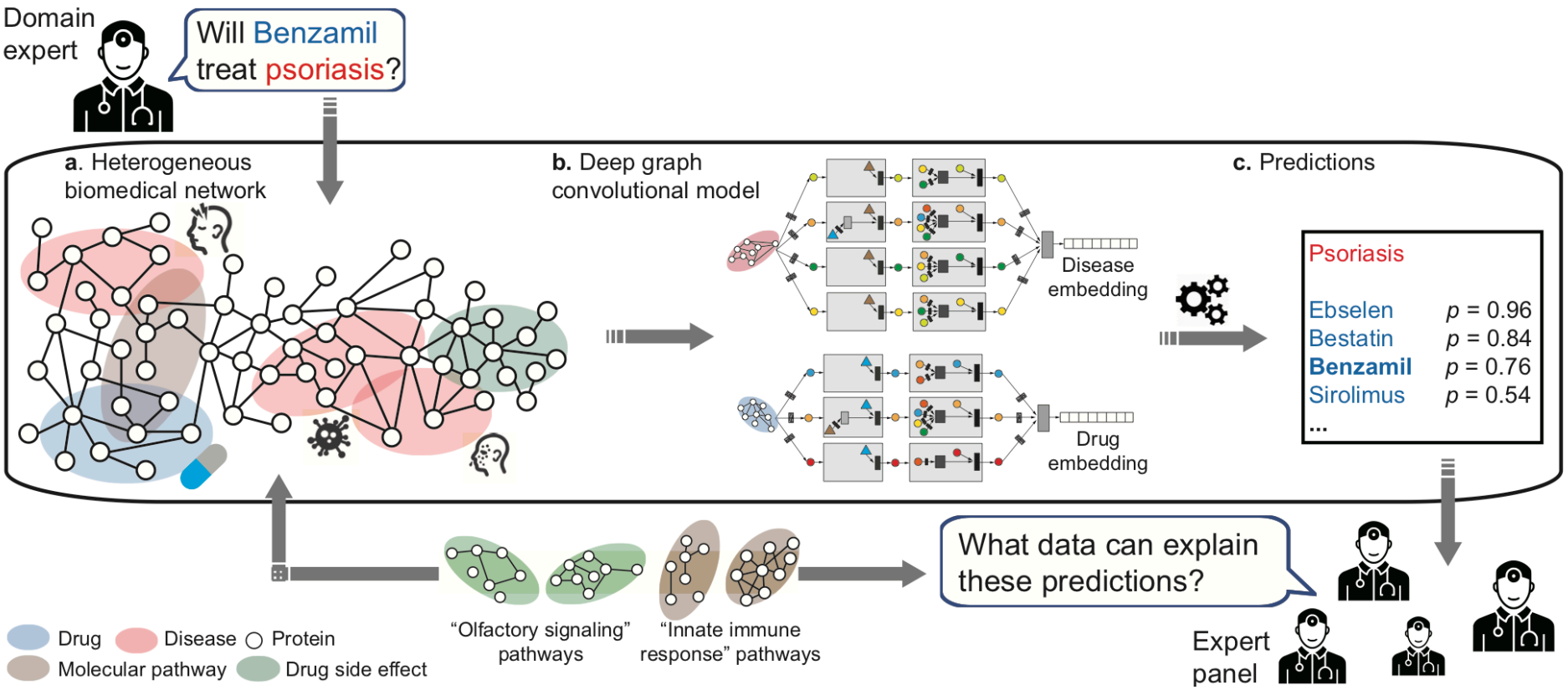
Task: Given a disease and a drug, predict if the drug could treat the disease

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]		

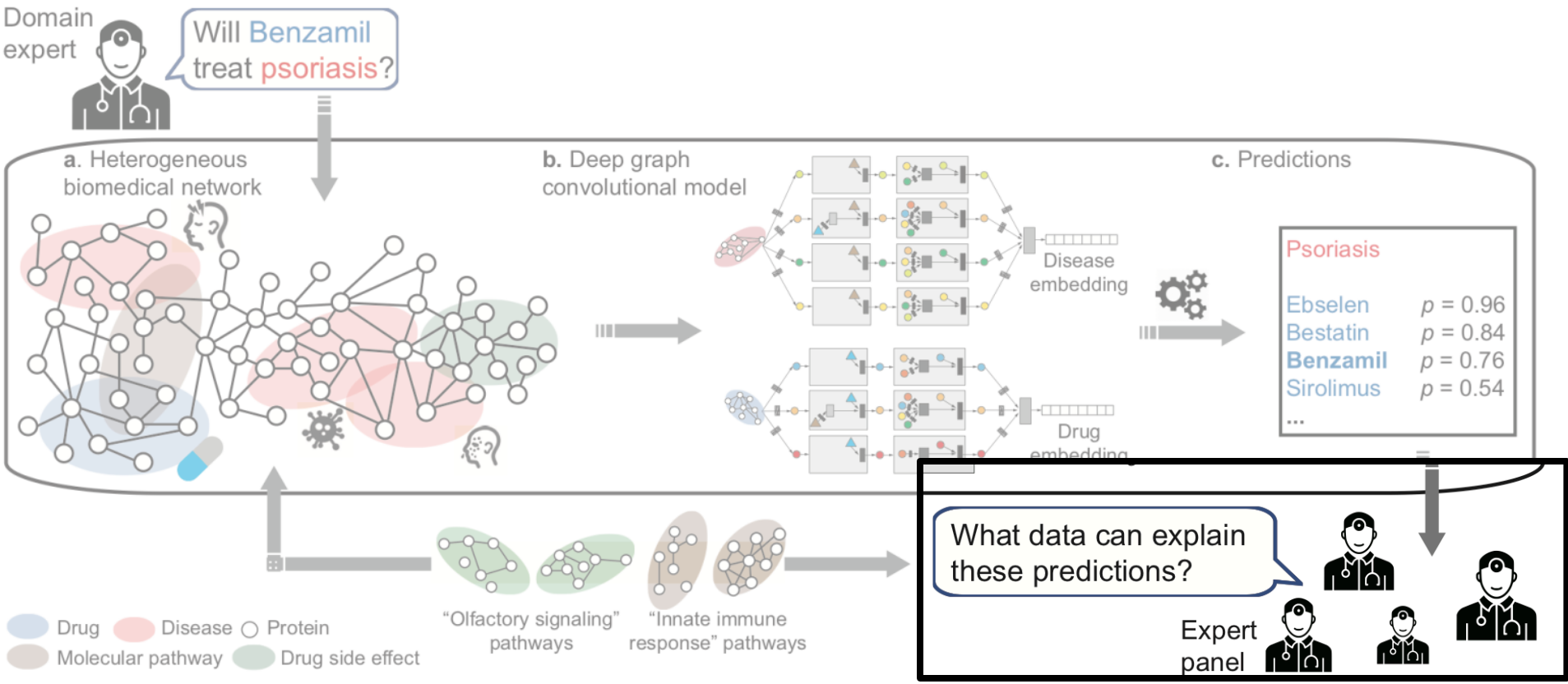
Up to 49% improvement

Up to 172% improvement

Feedbacks for the AI Loop



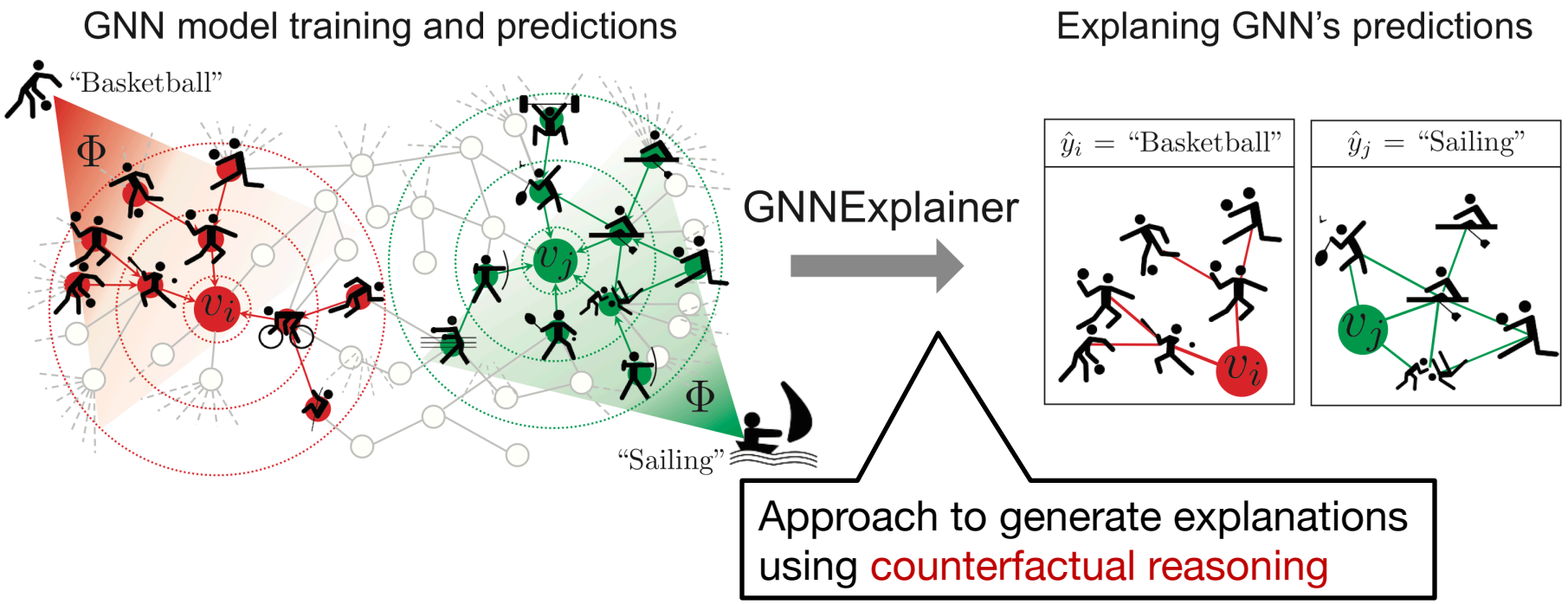
Feedbacks for the AI Loop



Explaining Machine Predictions

Key idea:

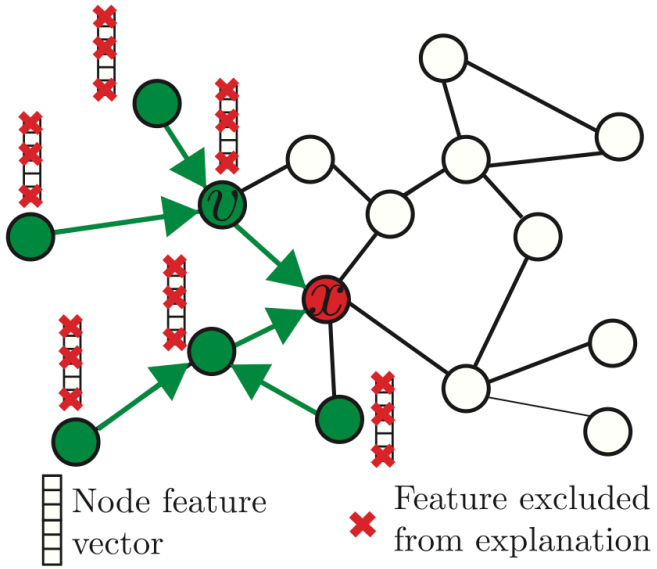
- Summarize where in the data the model "looks" for evidence for its prediction
- Find a small subgraph most influential for the prediction



GNN Explainer: Generating Explanations for Graph Neural Networks, *NeurIPS* 2019 (to appear)

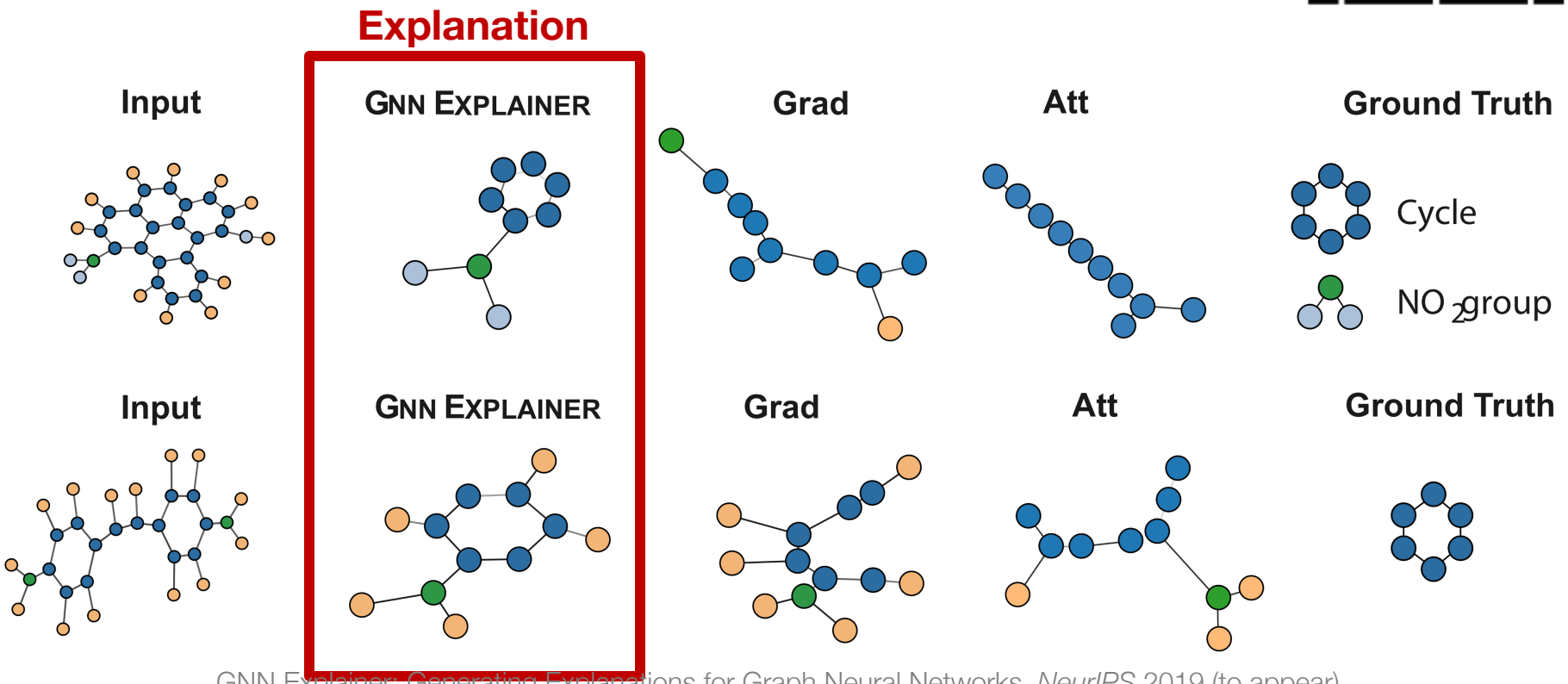
GNNExplainer: Key Idea

- 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



GNNExplainer: Results

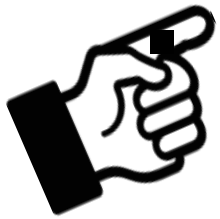
"Why did you predict that this molecule will have a mutagenic effect on Gram-negative bacterium *S. typhimurium*?"



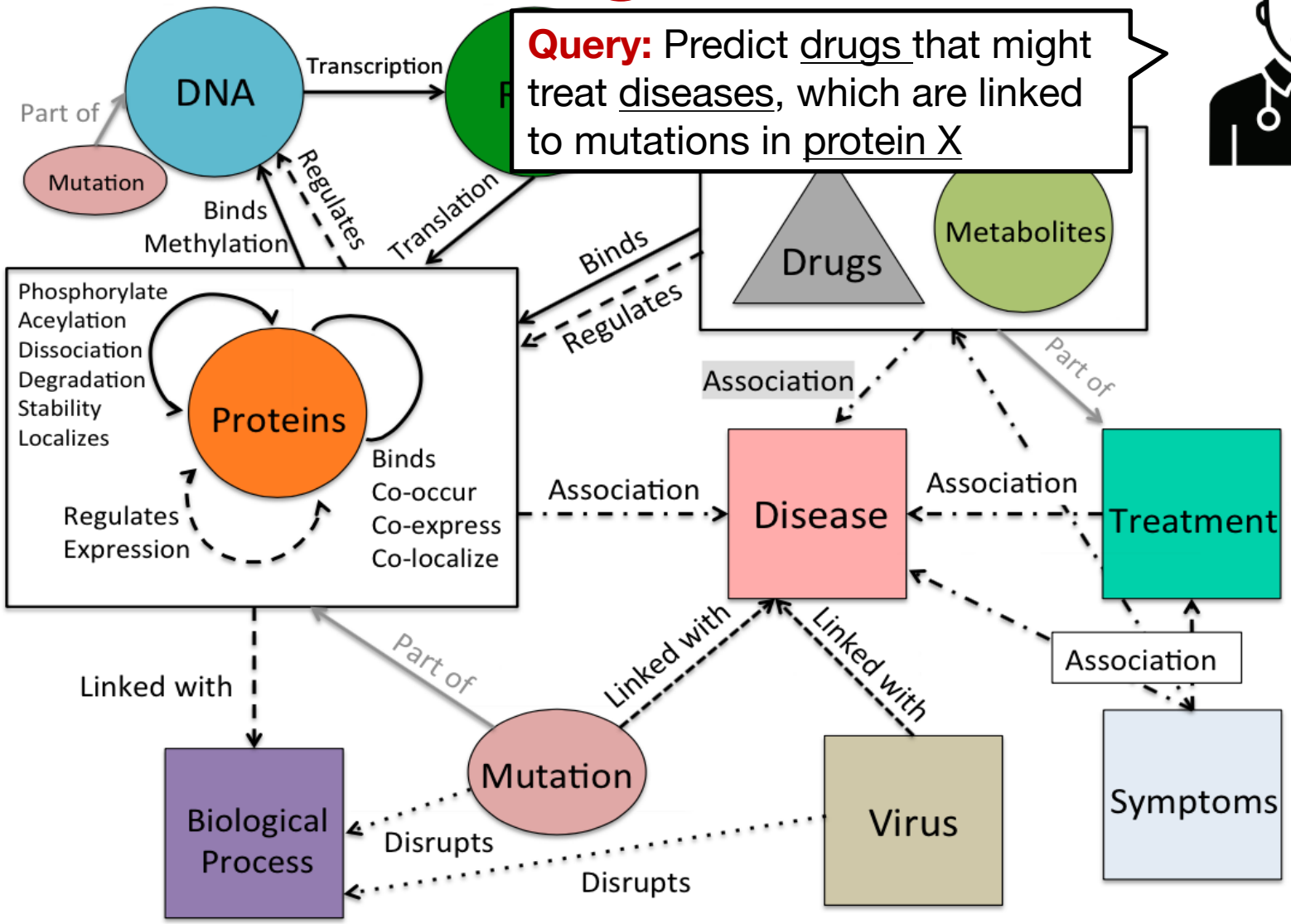
Today's Talk

1. **Representation learning** for biomedical data

2. Three research applications:
   - Used new approach to predict **safety** and **side effects of drug combinations**
   - Used new approach to **repurpose old drugs for new diseases**
   - Used new approach to **answer logical queries** on **knowledge graphs**



Knowledge Graphs

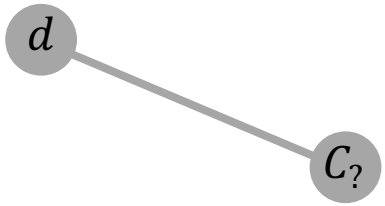


Learn over Knowledge Graphs

Simple edge prediction

Predict drugs $C_?$ that treat disease d

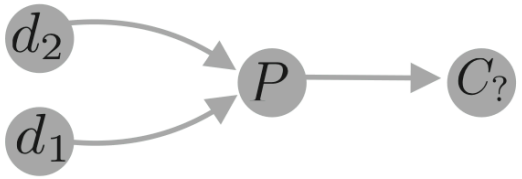
$$C_? : \text{TREAT}(C_?, d)$$

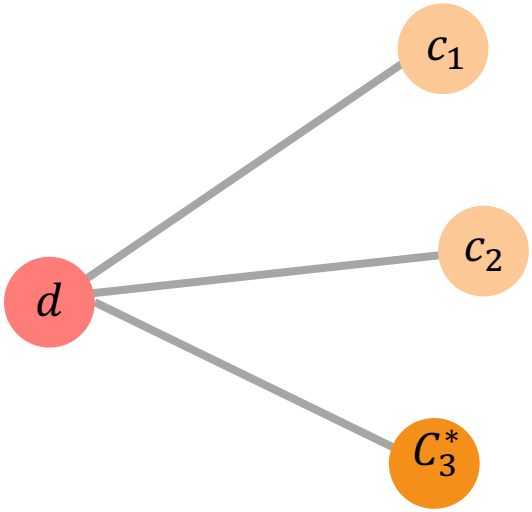


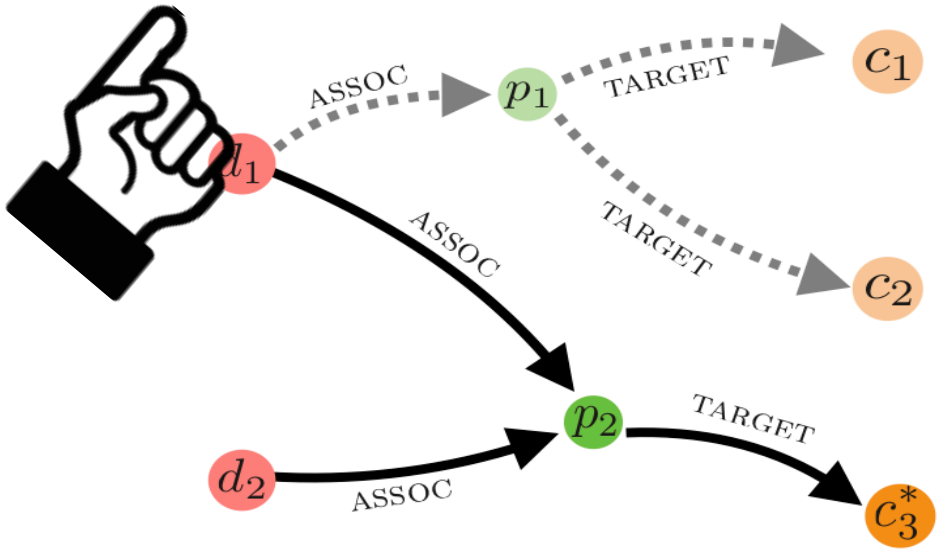
Answering logical queries

Predict drugs $C_?$ that might TARGET proteins, which are in turn ASSOCiated with diseases d_1 and d_2

$$C_?.\exists P : \text{ASSOC}(d_1, P) \wedge \text{ASSOC}(d_2, P) \wedge \text{TARGET}(P, C_?)$$







Embedding Logical Queries on Knowledge Graphs, *NeurIPS* 2018

Why is query prediction on knowledge graphs a hard problem?

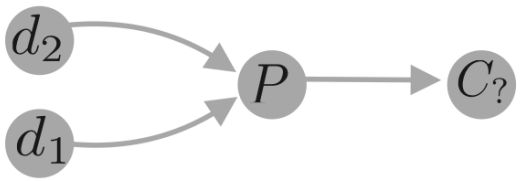
1) Massive enumerations
   - E.g., the protein node is an existentially quantified variable
   - Need to enumerate over all possible protein nodes

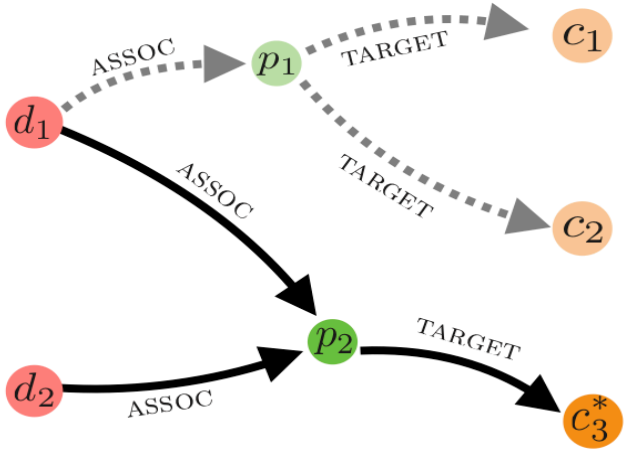
1) Exponential computations
   - Combinatorial number of possible answers to the query
   - Naive enumeration approach has exponential time complexity in the number of query variables

Logical query

Predict drugs $C_?$ that might TARGET proteins, which are in turn ASSOCiated with diseases d_1 and d_2

$$C_?.\exists P : \text{ASSOC}(d_1, P) \wedge \text{ASSOC}(d_2, P) \wedge \text{TARGET}(P, C_?)$$



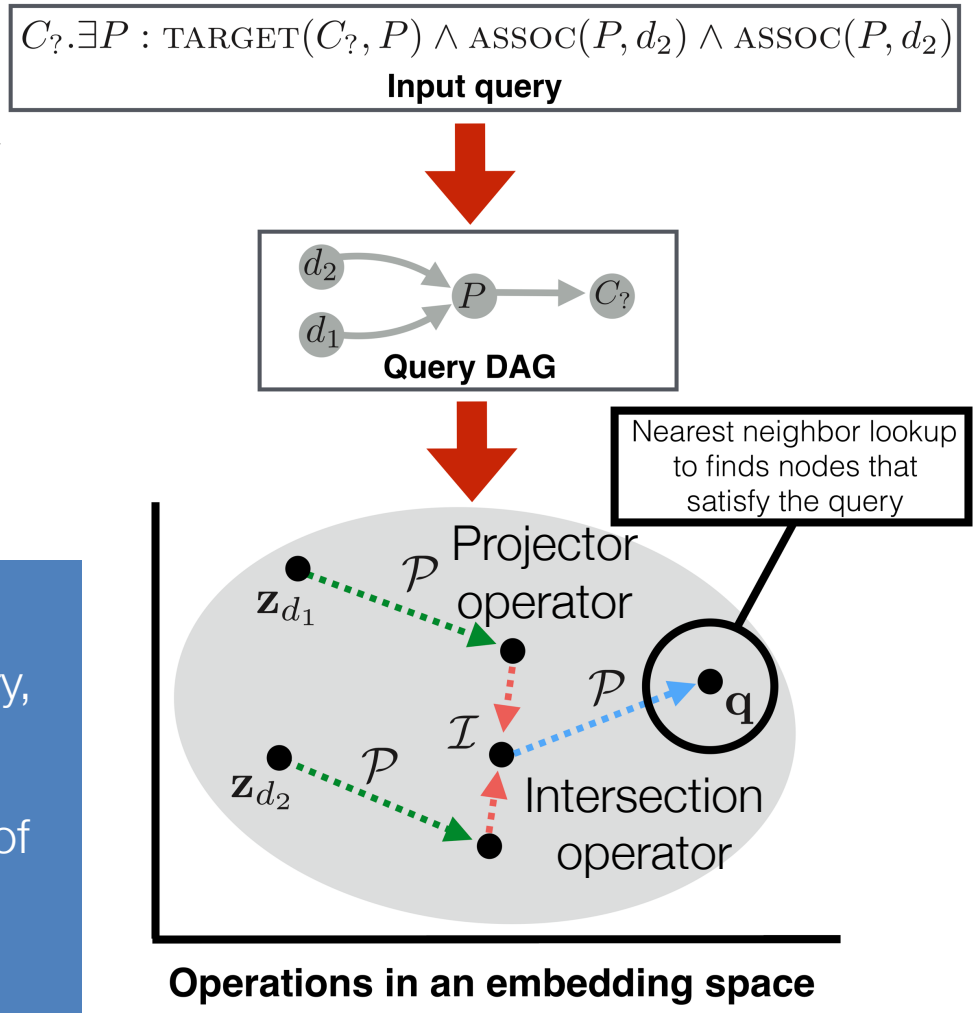


Approach: Query Embeddings

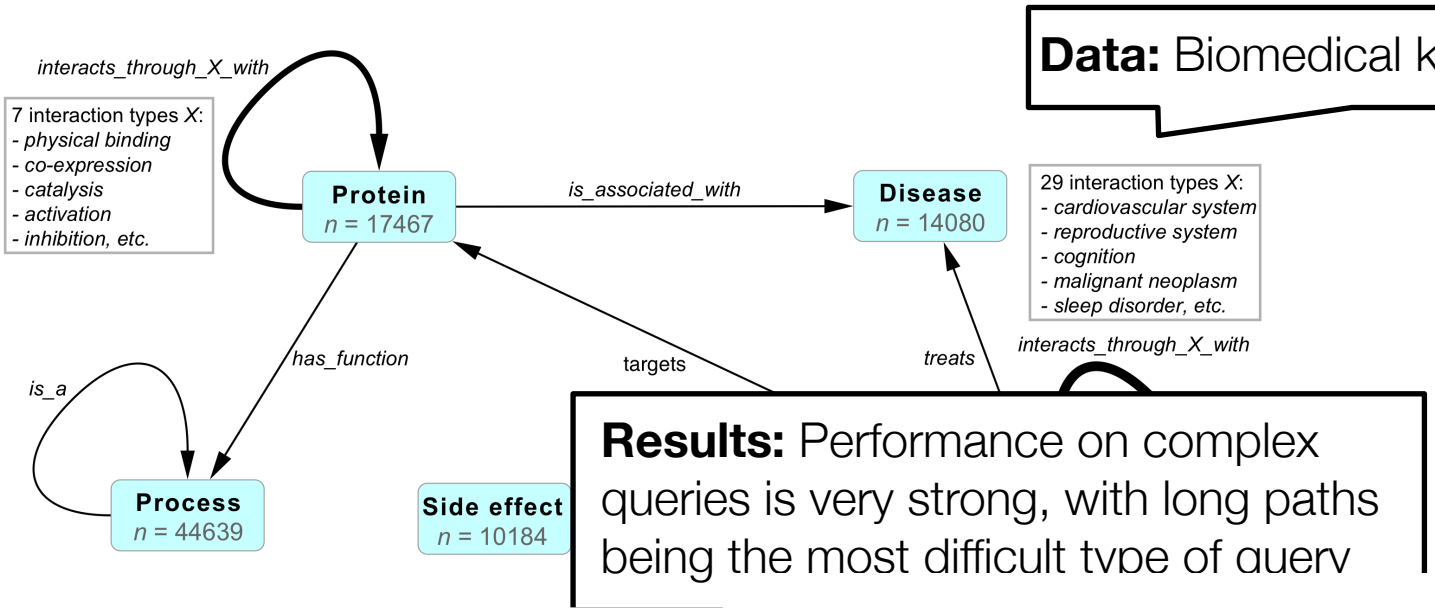
Two key steps:

1) Generate an embedding for every node in the graph
2) Represent logical operators as learned geometric operations (e.g., translation, rotation) in this embedding space

- **Any conjunctive query:** Can predict which nodes are likely to satisfy any query, even if it involves unobserved edges
- **Efficient:** Linear time complexity in size of the query and constant in size of the knowledge graph



Query Embeddings: Results



Data: Biomedical knowledge graph

Results: Performance on complex queries is very strong, with long paths being the most difficult type of query

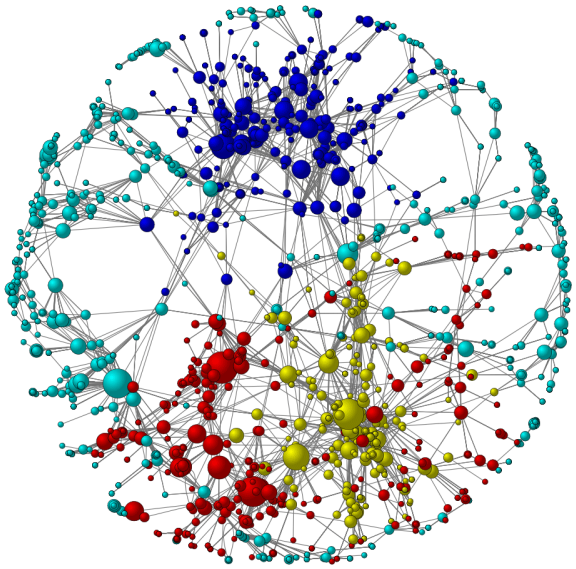
Results: Ablated models that are only trained on edge prediction perform much worse than query embeddings

Summary of Results

1. Used new approach to **predict safety and side effects of drug combinations:**
   - First-ever systematic and predictive study of drug combinations
   - Follow-up research on prostate cancer and validations in the clinic
2. Used new approach to **repurpose old drugs for new diseases:**
   - Outperforms baselines by up to 172%
   - Correctly predicted drugs repurposed at Stanford SPARK
3. Used new approach to **answer logical queries on knowledge graphs:**
   - *Predict drugs that might treat diseases linked to mutations in protein X*
   - Ability to answer logical queries in a linear instead of exponential time

Large datasets are transforming science and medicine

New machine learning methods can unlock these datasets and open doors for scientific discoveries



Data

Data + ML/AI

Predictions and insights

Thank you!

Papers, tutorials, data & code
ai.stanford.edu/~marinka

HARVARD
UNIVERSITY

I am hiring outstanding postdocs for projects in machine learning and biomedical data!