Conclusion

Jiaxuan You
Assistant Professor at UIUC CDS

CS598: Deep Learning with Graphs, 2024 Fall
https://ulab-uiuc.github.io/CS598/

Review Our Course Objective

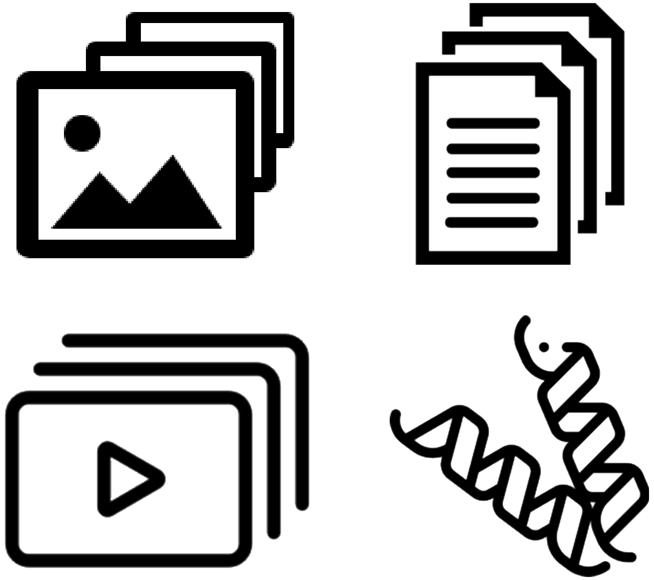
Takeaways from this course:

- **Knowledge** about graph deep learning
  - Core knowledge: Insights, Coding, Math
  - Latest knowledge: Recent research papers
- **Training** for AI research
  - Experience the **full lifecycle as an AI researcher**
    - Read, Ideate, Discuss, Code, Write, Review, Present

Conclusion

Knowledge about graph deep learning

Gap



Interconnected world

Modern ML

Goal of Graph Deep Learning

Enable DL research for the interconnected data

Graph: Ubiquitous across Disciplines

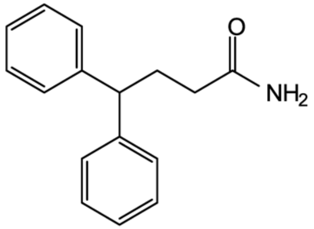


Image credit: MDPI

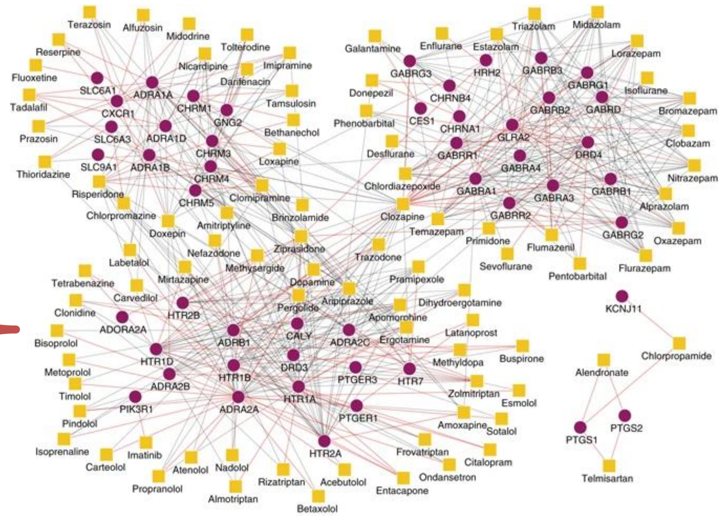


Image credit

Image credit: Medium

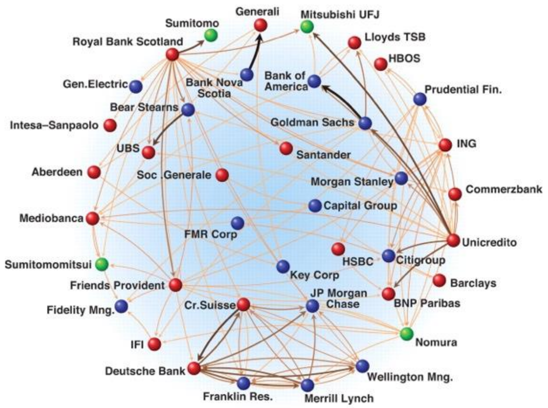


Image credit: Science

Molecule
Molecule design

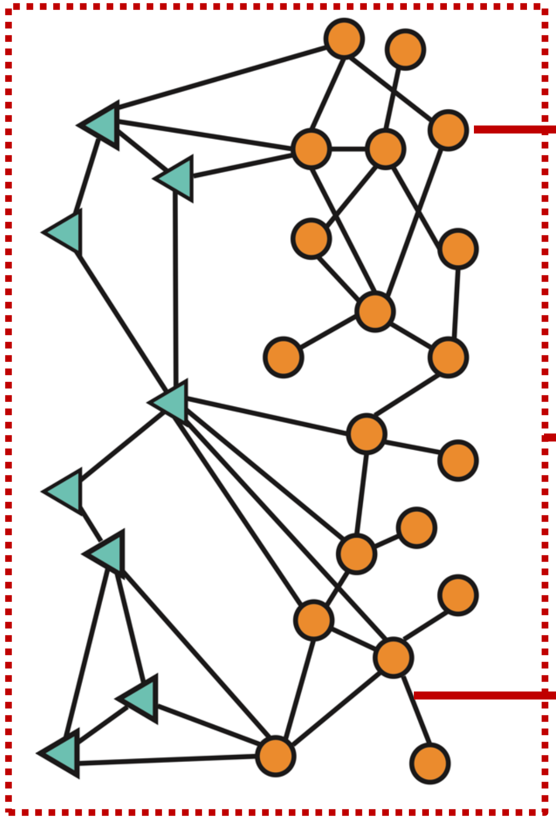
Protein interaction
Drug discovery

Social network
Recommender systems

Economic network
Policy making

- **Graphs:** ***flexible*** and ***expressive***
- **Graphs** can **bridge interdisciplinary data**

Graph ML **Tasks**

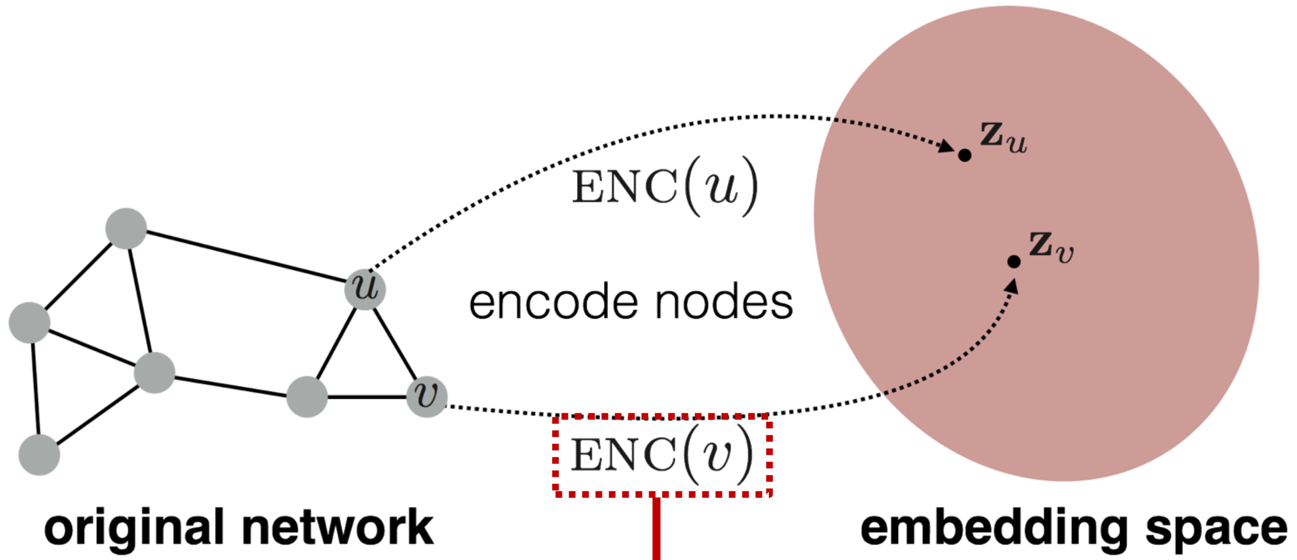


Node-level prediction

Graph-level prediction

Edge-level prediction

Key Idea: **Node Embeddings**

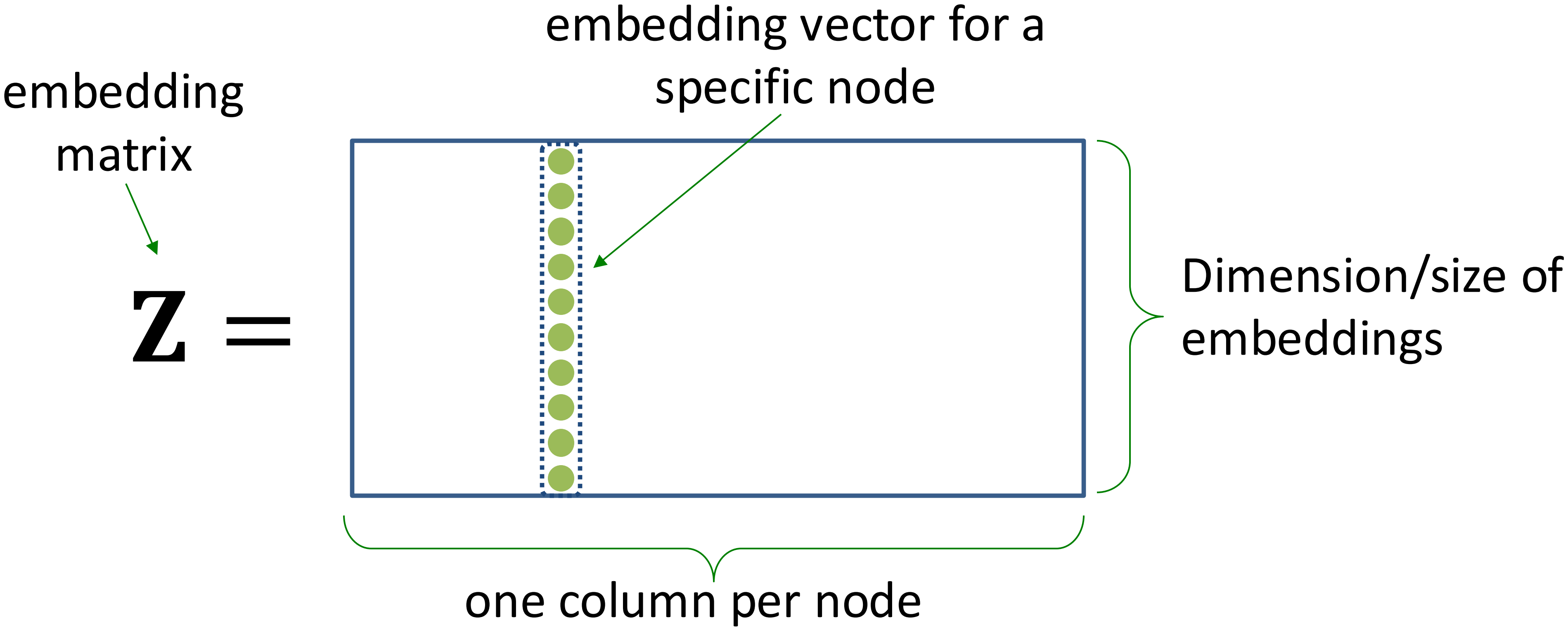


Embedding Matrix,

Graph Neural Networks,

"Shallow" Encoding

- Simplest encoding approach: **encoder is just an embedding-lookup**



A Single GNN Layer

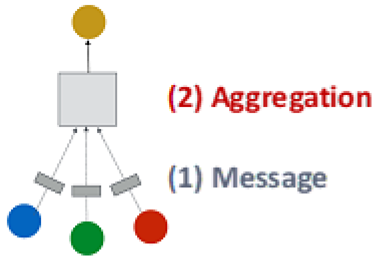
- **Putting things together:**
  - **(1) Message**: each node computes a message

$$\mathbf{m}_u^{(l)} = \mathrm{MSG}^{(l)}\left(\mathbf{h}_u^{(l-1)}\right), u \in \{N(v) \cup v\}$$

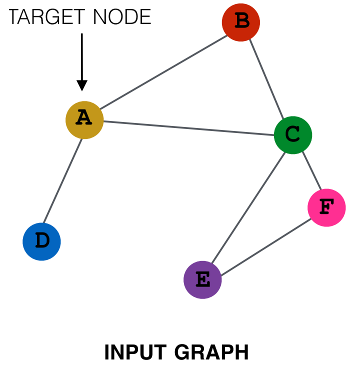
  - **(2) Aggregation**: aggregate messages from neighbors

$$\mathbf{h}_v^{(l)} = \mathrm{AGG}^{(l)}\left(\left\{\mathbf{m}_u^{(l)}, u \in N(v)\right\}, \mathbf{m}_v^{(l)}\right)$$

  - **Nonlinearity (activation):** Adds expressiveness
    - Often written as $\sigma(\cdot)$: $\mathrm{ReLU}(\cdot)$, $\mathrm{Sigmoid}(\cdot)$, …
    - Can be added to **message or aggregation**

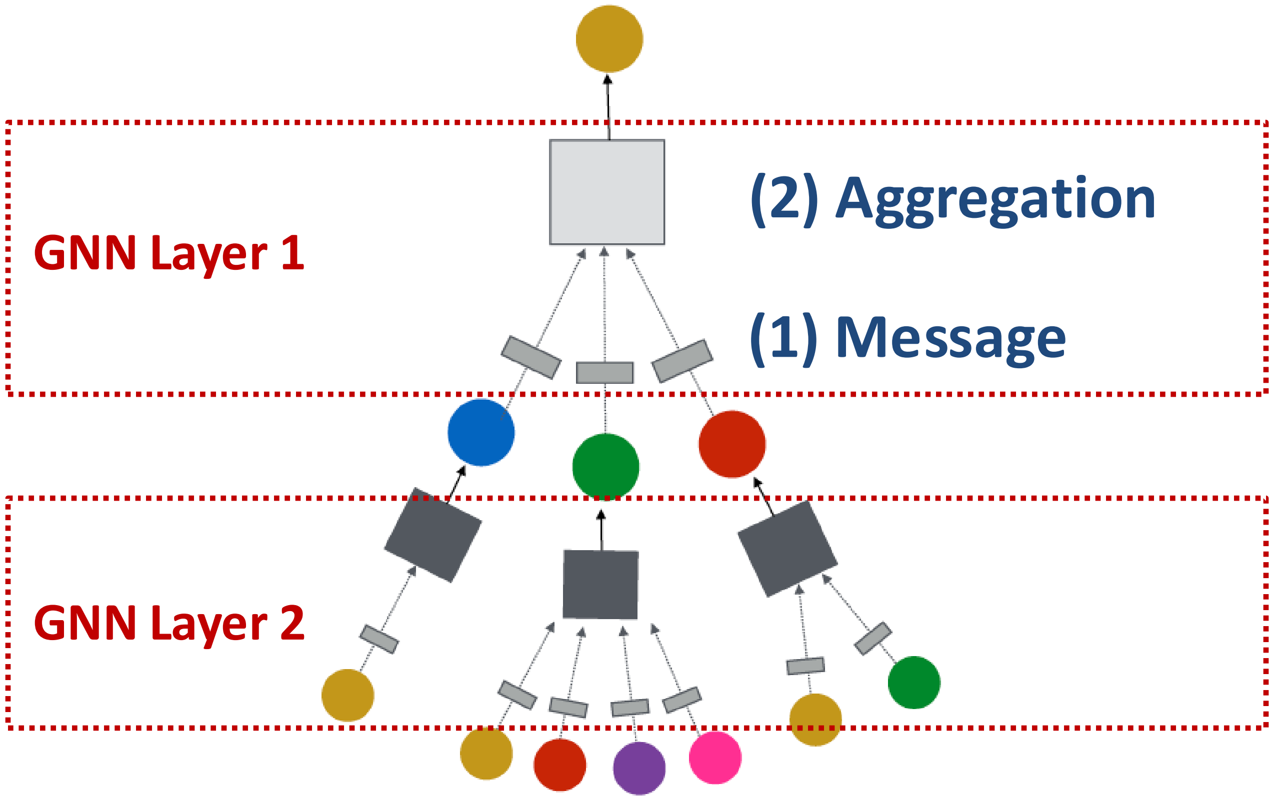


Recap: A General GNN Framework



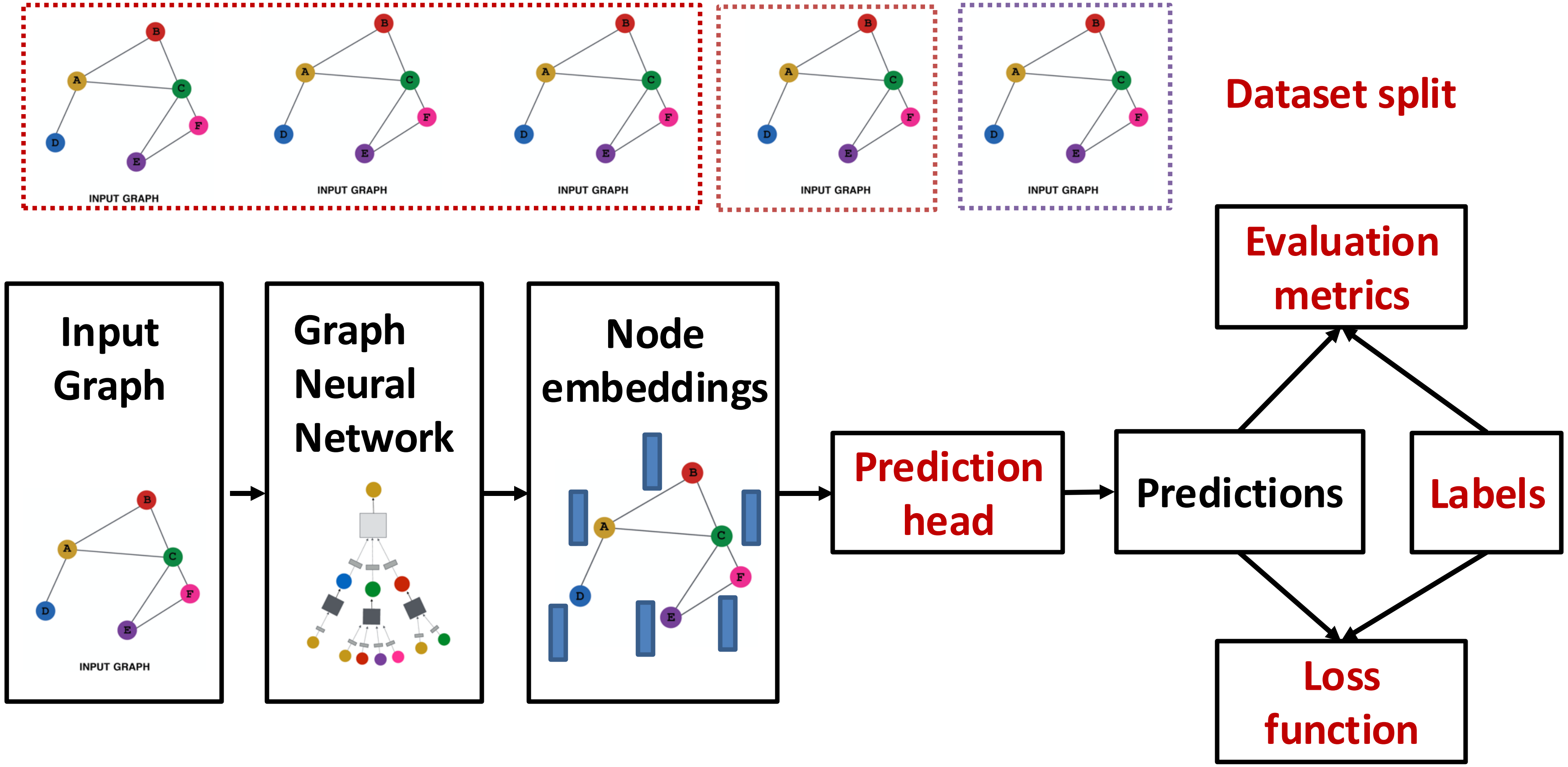
(5) Learning objective

(3) Layer connectivity



(4) Graph augmentation

GNN Training Pipeline



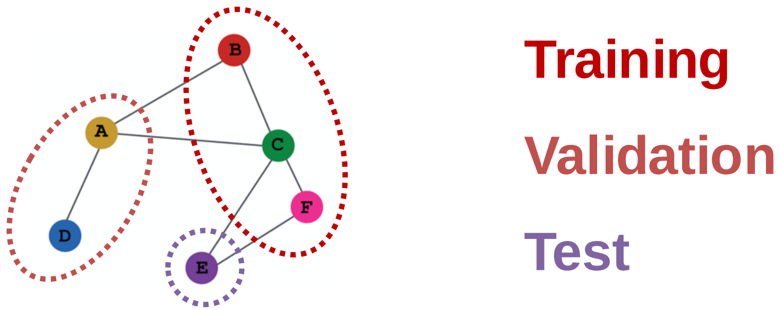
Implementation resources:

[DeepSNAP] provides core modules for this pipeline

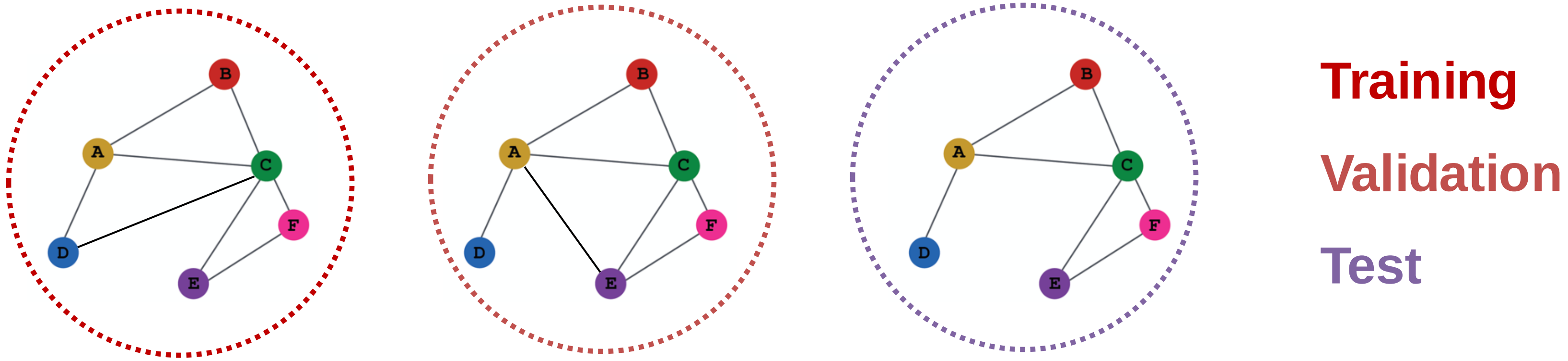
[GraphGym] further implements the full pipeline to facilitate GNN design

Example: Node Classification

- **Transductive** node classification
  - **All the splits can observe the entire graph structure**, but can only observe the labels of their respective nodes

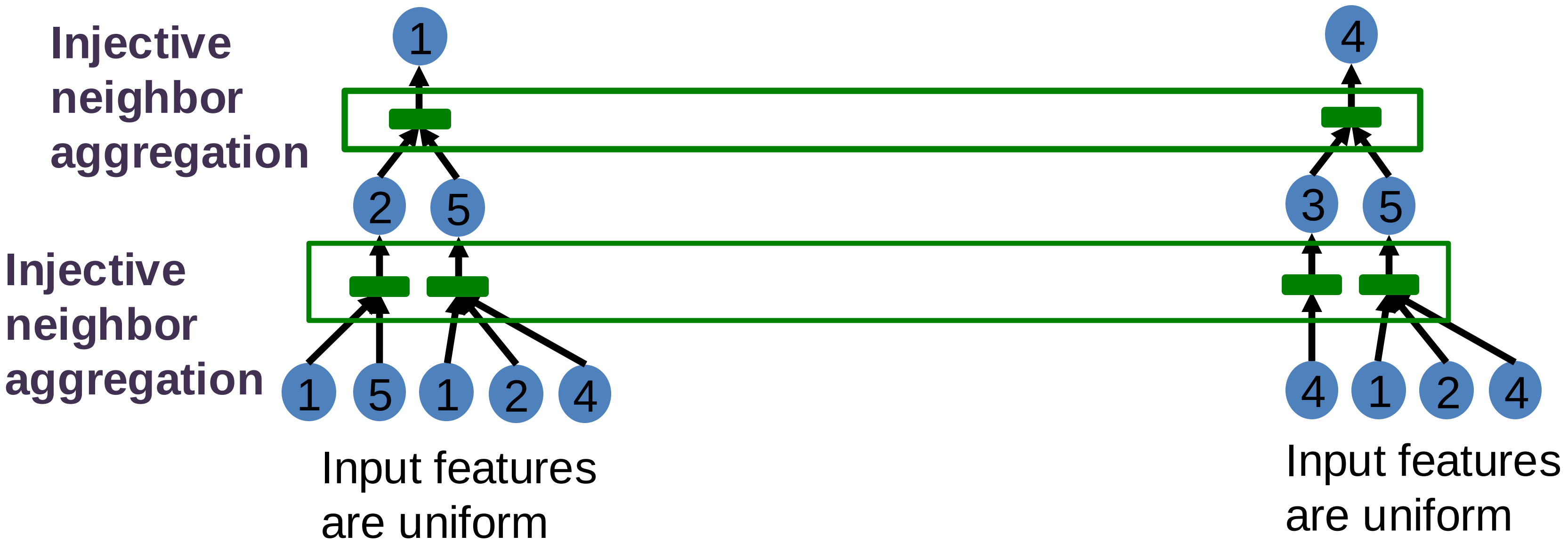


- **Inductive** node classification
  - Suppose we have a dataset of 3 graphs
  - **Each split contains an independent graph**



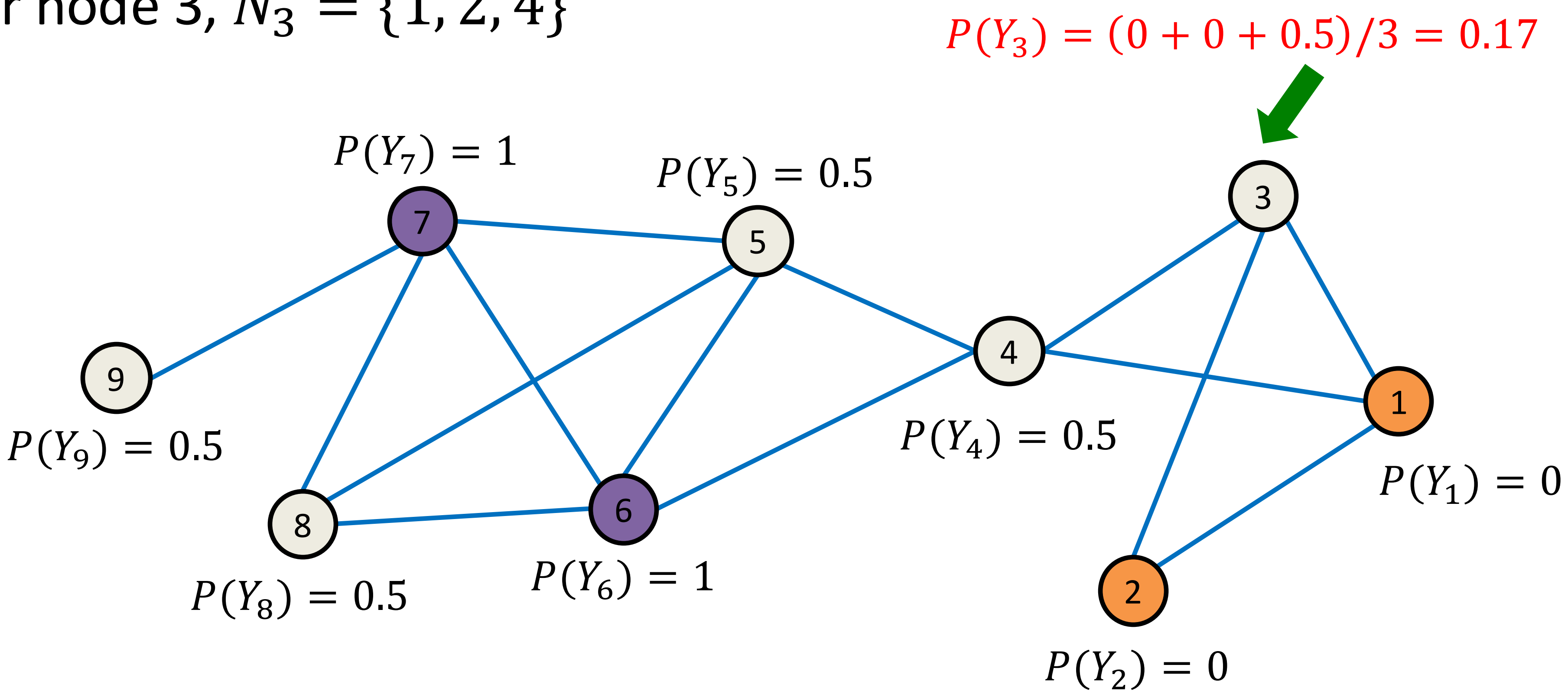
How Expressive is a GNN?

- In other words, most expressive GNN would use an **injective neighbor aggregation** function at each step.
  - Maps different neighbors to different embeddings.



Example: 1st Iteration, Update Node 3

- Update for the 1st Iteration:
  - For node 3, $N_3 = \{1, 2, 4\}$



GNN Design Space

- **Overall: A GNN design space**
  - Intra-layer design

Batch Normalization	Dropout	Activation	Aggregation
True, False	False, 0.3, 0.6	RELU, PRELU, SWISH	MEAN, MAX, SUM

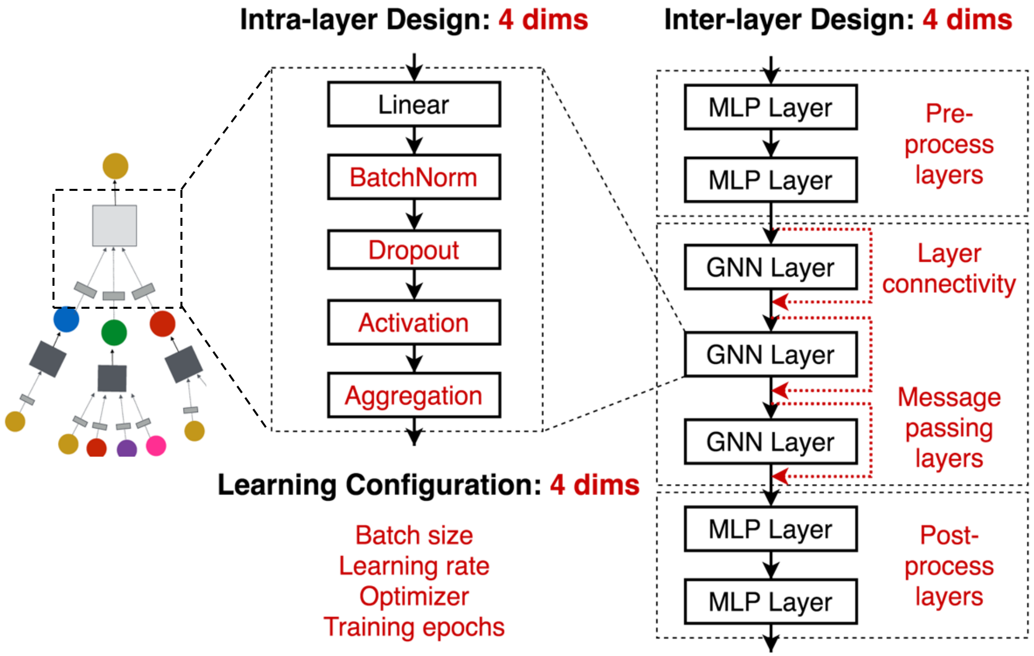
  - Inter-layer design

Layer connectivity	Pre-process layers	Message passing layers	Post-precess layers
STACK, SKIP-SUM, SKIP-CAT	1, 2, 3	2, 4, 6, 8	1, 2, 3

  - Learning configuration

Batch size	Learning rate	Optimizer	Training epochs
16, 32, 64	0.1, 0.01, 0.001	SGD, ADAM	100, 200, 400

  - **In total:** 315K possible designs



- **Purpose:**
  - We don't want to (and we cannot) cover all the possible designs
  - **A mindset transition:** We want to demonstrate that **studying a design space is more effective than studying individual GNN designs**

Heterogeneous Graphs: Motivation

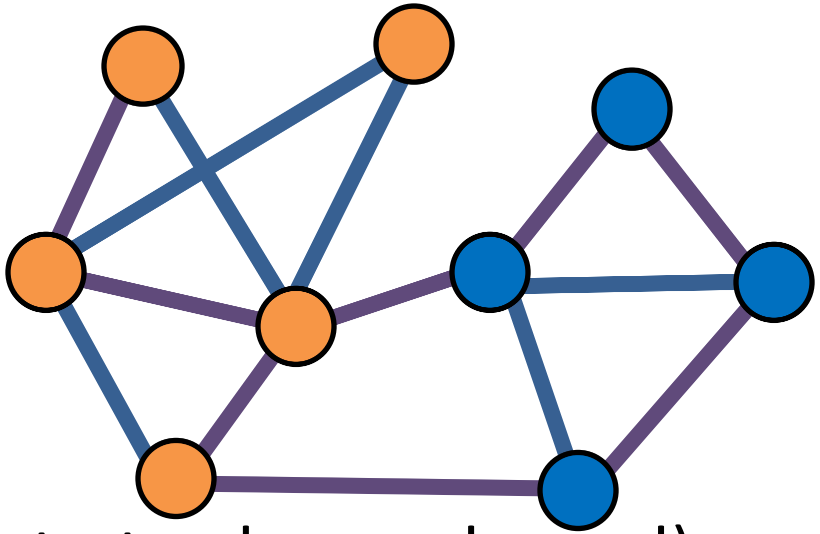
8 possible relation types!

(Paper, Cite, Paper)

(Paper, Like, Paper)

(Paper, Cite, Author)

(Paper, Like, Author)



(Author, Cite, Author)

(Author, Like, Author)

(Author, Cite, Paper)

(Author, Like, Paper)

Relation types: (node_start, edge, node_end)

- We use **relation type to describe an edge** (as opposed to edge type)
- Relation type better captures the interaction between nodes and edges

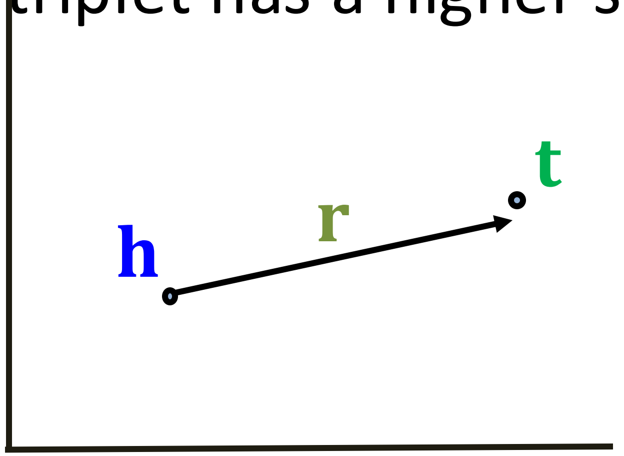
TransE

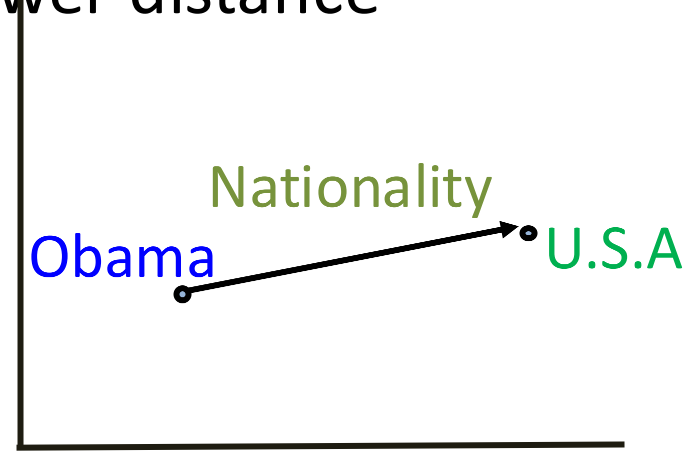
- **Intuition: Translation**
  For a triplet (h, r, t), let $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$
  be embedding vectors.
- **TransE:** $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ if the given link exists else $\mathbf{h} + \mathbf{r} \neq \mathbf{t}$

embedding vectors will appear in boldface

Entity scoring function: $f_r(h, t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||$

- A valid triplet has a higher score / lower distance



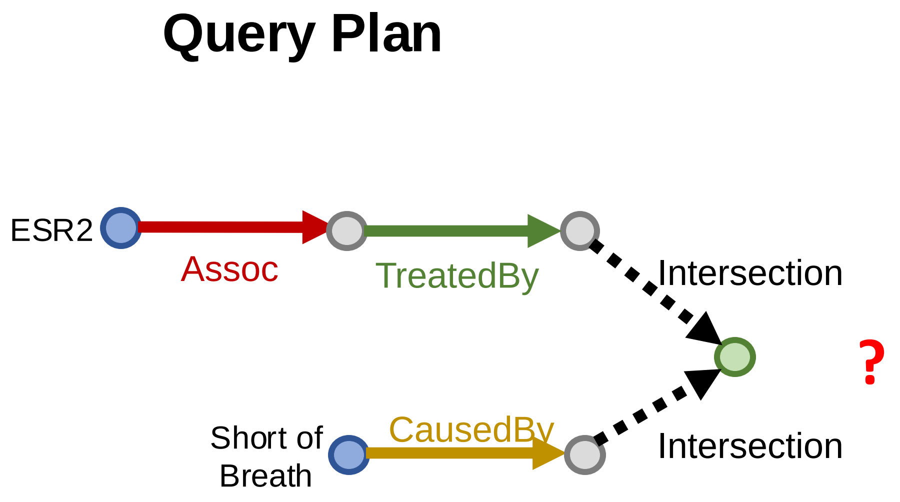


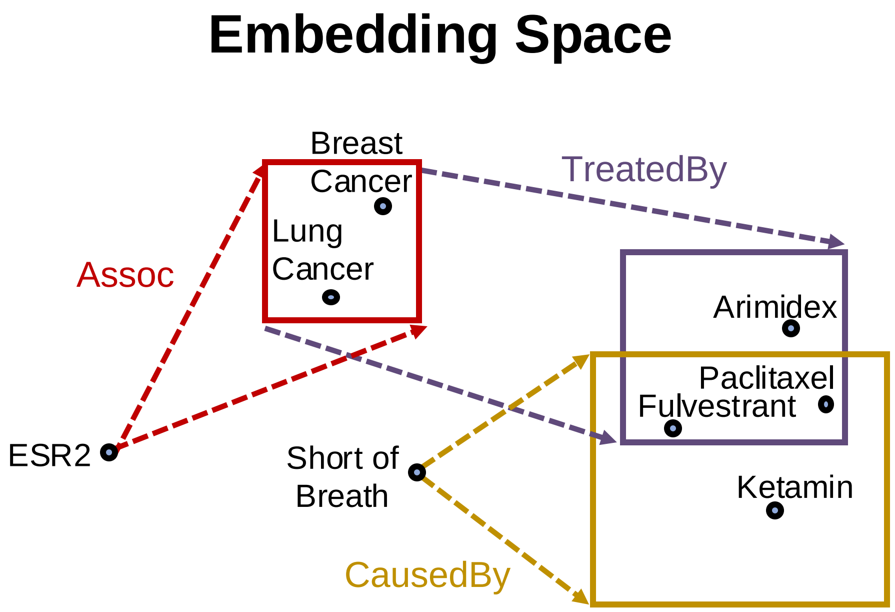
Embed with Box Embedding

"What is the drug that causes Short of Breath and treats disease associated with protein ESR2?"

((e:ESR2, (r:Assoc, r:TreatedBy)), (e:Short of Breath, (r:CausedBy))

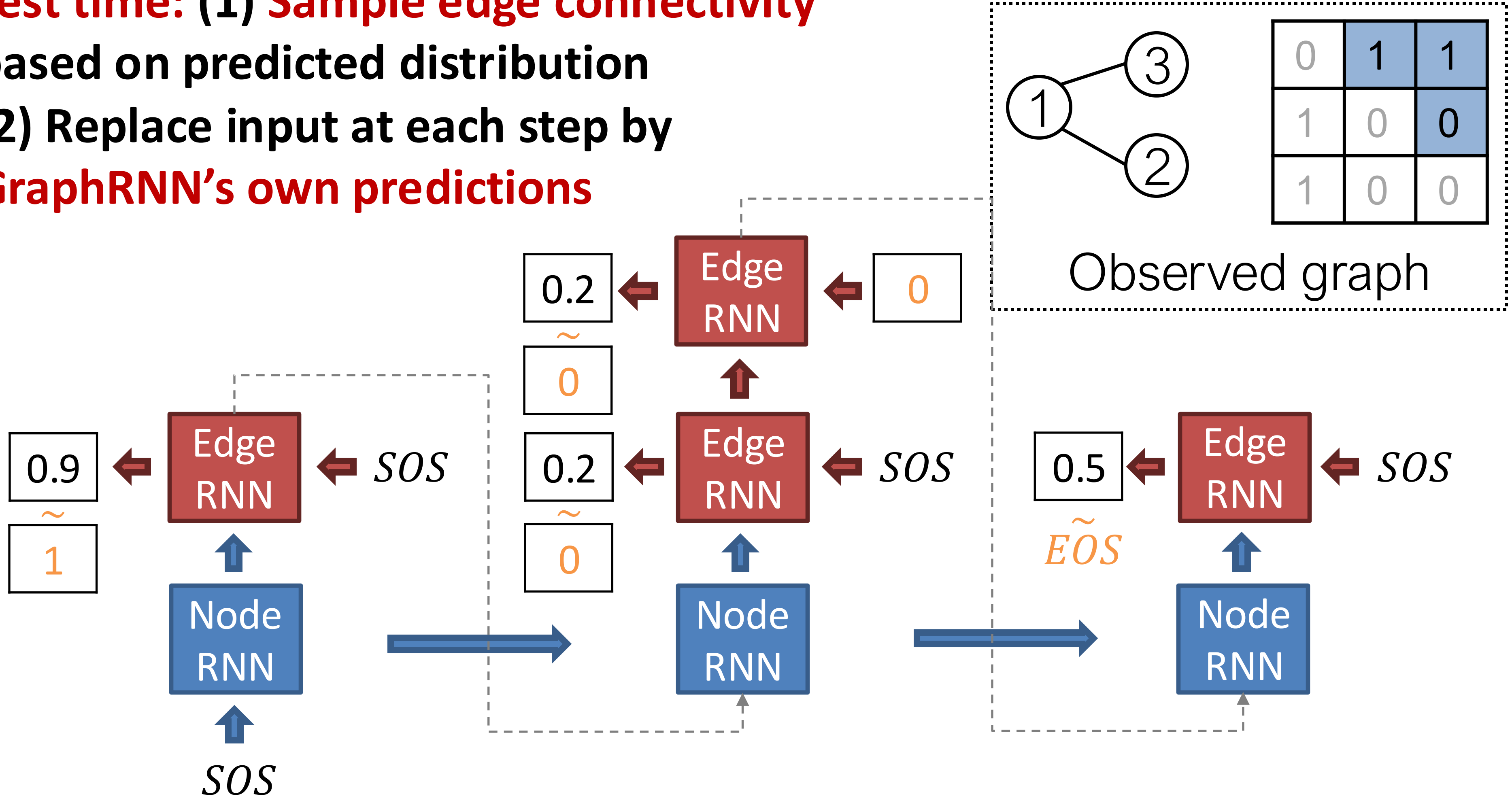
- **How do we take intersection of boxes?**





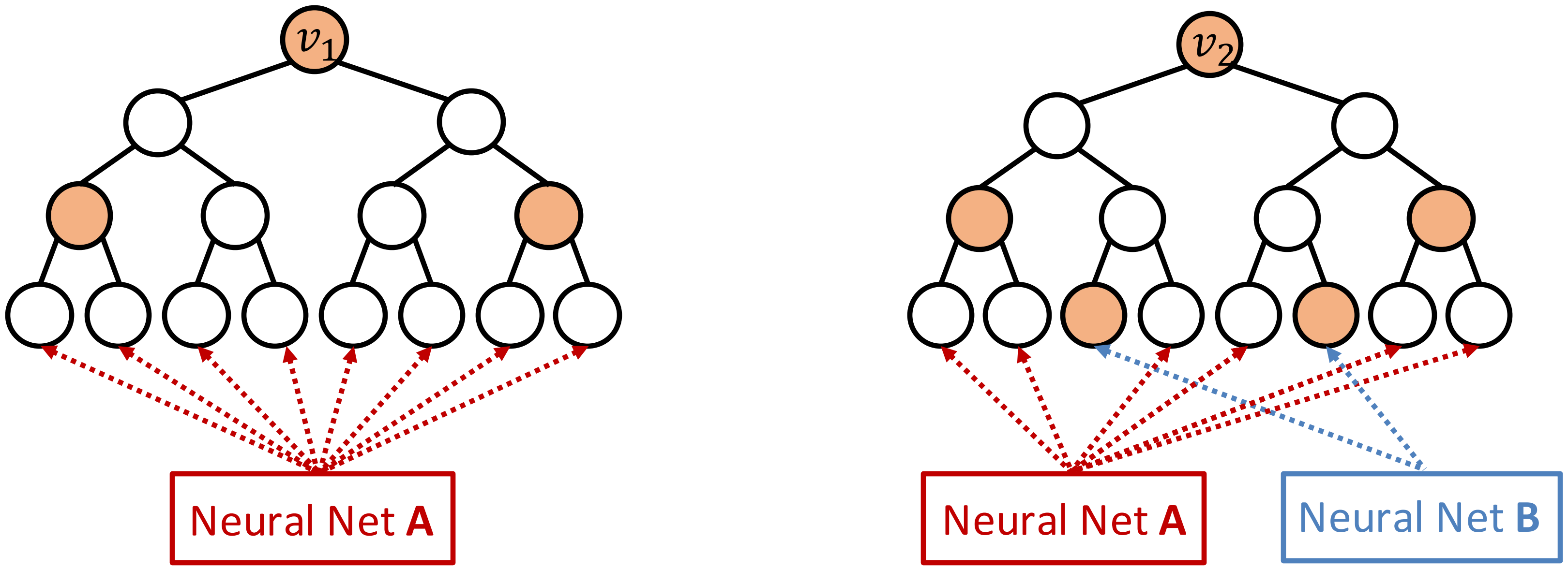
Put Things Together: Test

Test time: (1) Sample edge connectivity based on predicted distribution
(2) Replace input at each step by GraphRNN's own predictions



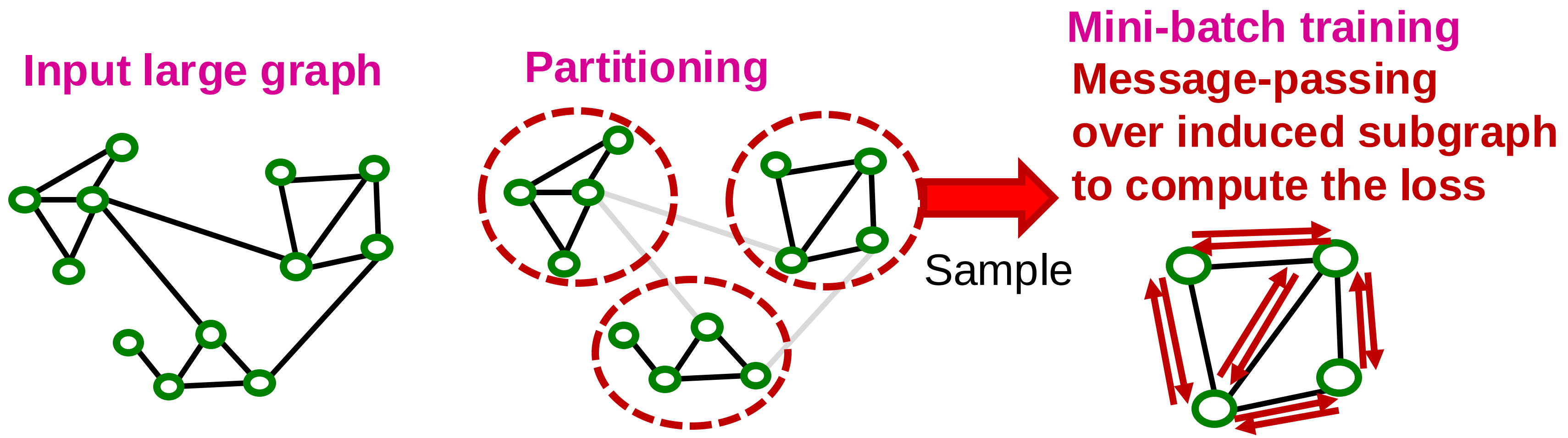
Identity-aware GNN

- **Why does heterogenous message passing work:**
  - Suppose two nodes v_1, v_2 have the same computational graph structure, but have different node colorings
  - Since we will apply different neural network for embedding computation, their embeddings will be different



Cluster-GCN: Overview

- **We first introduce "vanilla" Cluster-GCN.**
- **Cluster-GCN** consists of two steps:
  - **Pre-processing**: Given a large graph, partition it into groups of nodes (i.e., subgraphs).
  - **Mini-batch training**: Sample one node group at a time. Apply GNN's message passing over the **induced subgraph**.



How to inject PE/SE

- Inject local PE/SE with node inputs and treat relative PE/SE as additional attention bias
  - Example: Graphormer

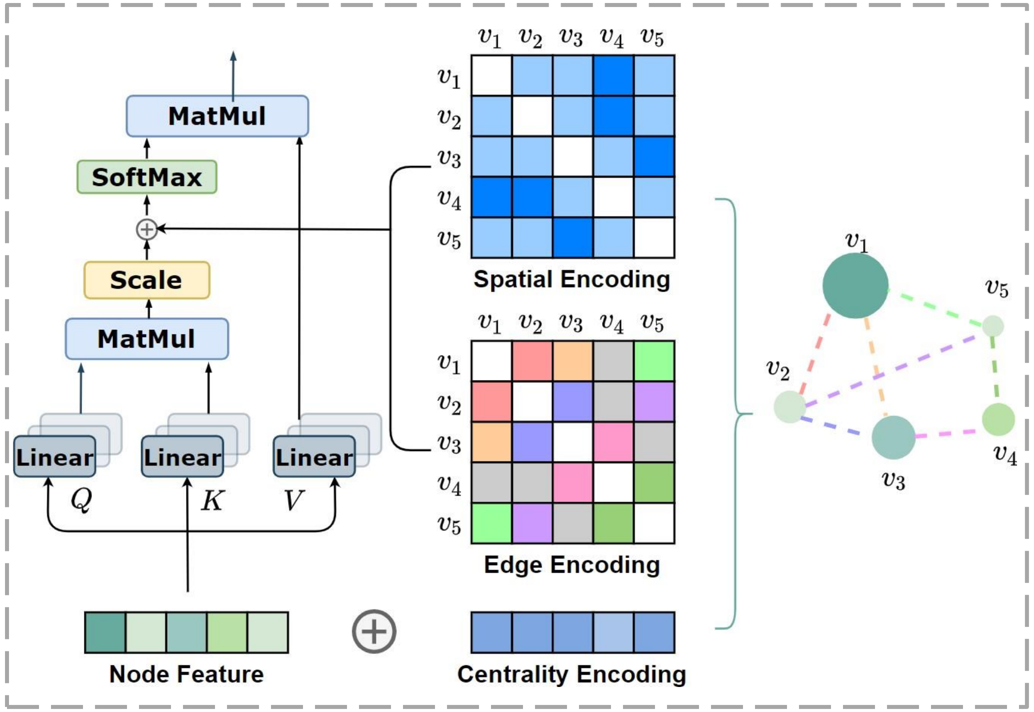
Attention: $e_{ij} = \frac{(\boldsymbol{h}_i \boldsymbol{W}_q)(\boldsymbol{h}_j \boldsymbol{W}_k)^T}{\sqrt{d}} + b_{\phi(v_i, v_j)} + c_{ij}$

Spatial Encoding:

Shortest path between v_i, v_j

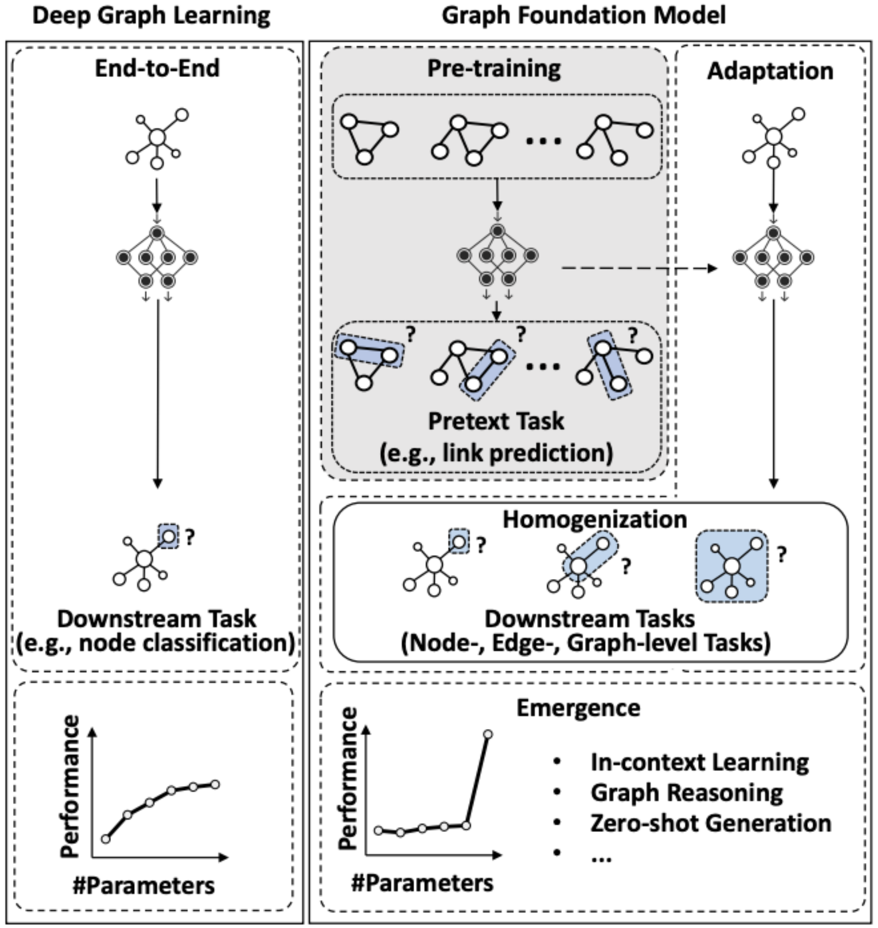
Edge Encoding:

Average all edge features along the shortest path between v_i, v_j, $c_{ij} = \frac{1}{N} \sum_{e \in SP(i,j)} x_e w_e$



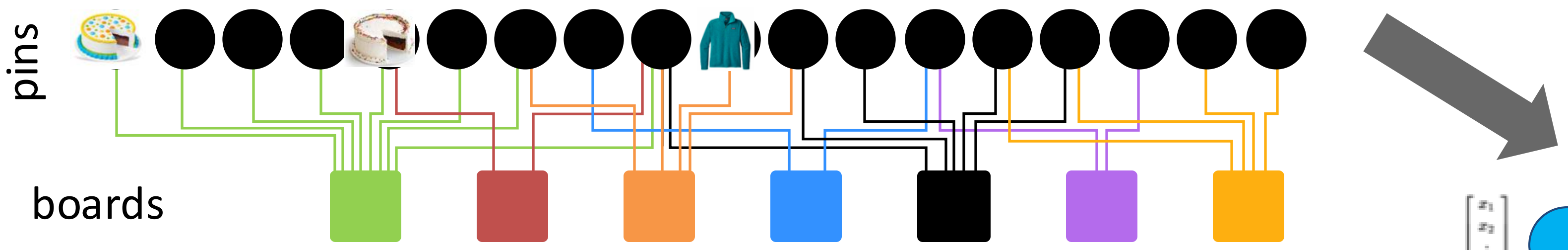
Graph Foundation Models

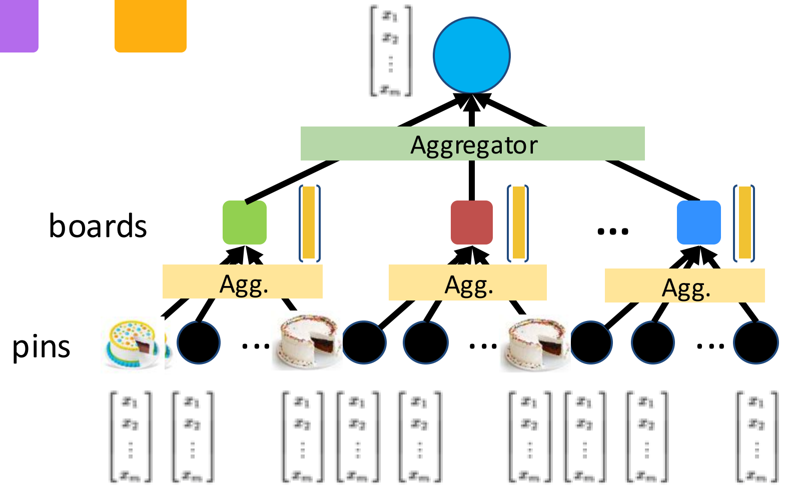
- A graph foundation model (GFM) is a model pre-trained on **extensive graph data**, adapted for **diverse downstream graph tasks**.



PinSAGE: Graph Neural Network

- Graph has tens of billions of nodes and edges
- Further resolves embeddings across the Pinterest graph





Order Constraint (2)

- **We specify the order constraint to ensure that the subgraph properties are preserved in the order embedding space**

Embedding dimension

$$\forall_{i=1}^{D} z_q[i] \le z_t[i] \quad \text{iff} \quad G_Q \subseteq G_T$$

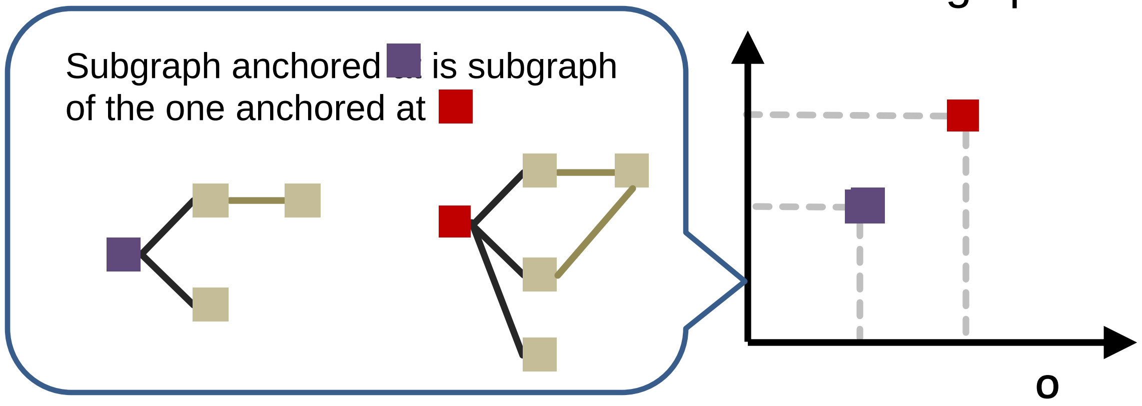
Query embedding

Target embedding

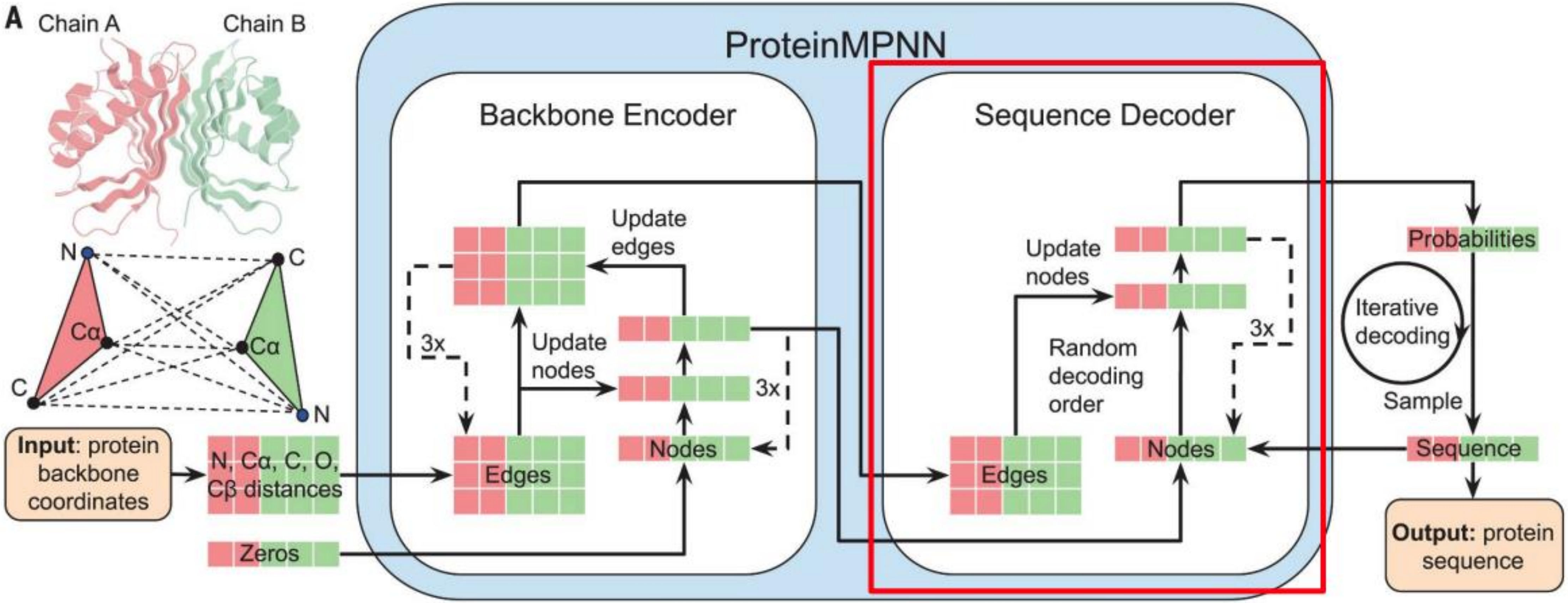
Subgraph Relation

trained with max-margin loss

Embedding space



Autoregressive Decoder in ProteinMPNN

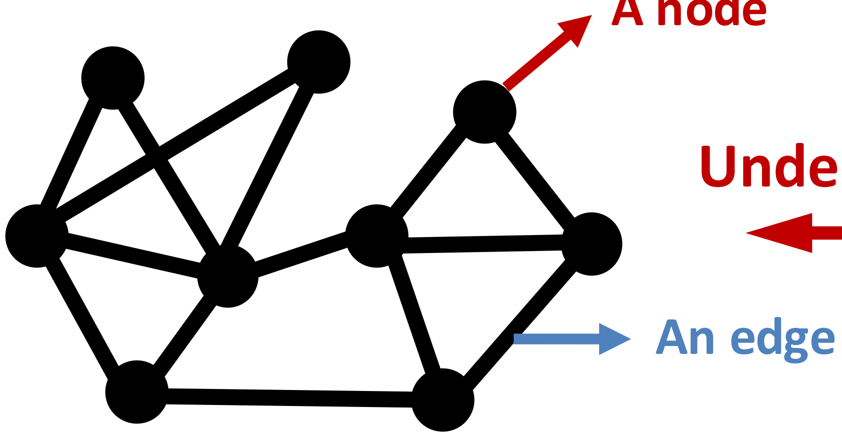


The Bottom Line

- **There is exciting relational structure in many many real-world problems**
- **Identifying and harnessing this relational structure leads to better AI**

Interconnected world

Represent



Graph-structured data

Understand

AI agent

Conclusion

Training for AI research

The 5 Questions for A Good Paper

(Originally for writing paper introduction)

- What is the problem?
- Why is it interesting and important?
- Why is it hard? (E.g., why do naive approaches fail?)
- Why hasn't it been solved before? (Or, what's wrong with previous proposed solutions? How does mine differ?)
- What are the key components of my approach and results? Also include any specific limitations.

What's Next?

- **Review and Response** (Due on **Dec 8**)
  - Everyone will be **assigned 2 papers to review** by this weekend.
  - Each group is required to **participate in the public discussion** of the papers. We will discuss general principles and practical strategies today.
  - Review and response will count towards **10%** of your final grade. We will evaluate both your reviews and your responses.

What's Next?

- **Presentation** (Due on **Dec 4 & Dec 6**)
  - Each group is required to prepare slides for the presentation.
  - We will have **5 minutes** for presentation and **3 minutes** for comments.
  - TA will assign each group into the Wednesday or Friday session by Dec 1.
  - Presentation will count towards **10%** of your final grade.

Conclusion

Review & Response

Reviewer Guide

- **An outline of the main reviewer tasks**
- **Step-by-step reviewing instructions**
- **Review Examples**

High Quality Review

- **1 Summary**: Provide a brief, accurate summary of the paper, emphasizing its key contributions, methodology, and findings.
- Example: *"This paper proposes a novel method for optimizing transformer architectures using a learned attention pruning mechanism. The authors demonstrate improvements in both computational efficiency and downstream performance across three benchmark datasets."*

High Quality Review

- **2 Strengths**: Highlight the paper's major contributions and strengths in detail.
- Example: *"The approach is innovative and addresses a significant problem in transformer scalability. The experimental results are compelling, with a 20% reduction in computational cost and consistent performance gains."*

High Quality Review

- **3 Weaknesses**: Clearly state limitations or potential issues, such as incomplete comparisons, insufficient analysis, or unclear writing.
- Example: *"The method lacks an ablation study to disentangle the effects of the pruning mechanism from other components of the model."*

High Quality Review

- **4 Constructive Feedback**: Provide **actionable** suggestions for improvement.
- Example: *"Adding a comparison with recent pruning methods like XYZ et al. (2023) would strengthen the claims. Also, explaining why Dataset A was chosen over other common benchmarks would improve clarity."*

High Quality Review

- **Detailed and Specific Comments**: Avoid vague statements. Back your claims with evidence or examples.
- Weak: *"The experiments are insufficient."*
- Strong: *"The experiments focus only on synthetic data. Real-world datasets like ABC would provide stronger evidence for the method's generalizability."*

High Quality Rebuttal Tip

- Think about yourself as an LLM agent
- **“RAG”**
- Retrieve evidence
  - From your paper
    - Line xxx to xxx, Figure xxx, Table xxx
  - From famous paper
    - Relevant works, famous works
- Action
  - Improve your writing
  - Add new results

Conclusion

Open-ended Discussion - AMA