Graph Foundation Models for Knowledge Graph Reasoning and Beyond

Michael Galkin
Intel AI Lab

Foundation Models

A **single** model pre-trained (often) in the self-supervised fashion on **large amounts of data** that is applicable to **many downstream tasks**

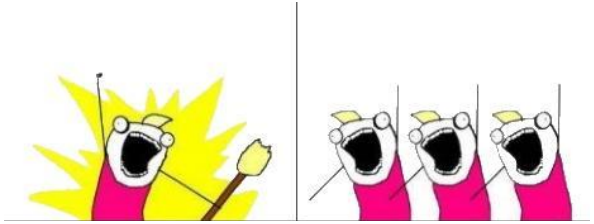
- By in-context learning
- By fine-tuning

We Want Graph Foundation Models!

- ... Large!
  - Non strong signal that GNNs or Graph Transformers benefit from depth / increasing # params
  - Scaling laws for GNNs / GTs are non-existent

- ... Self-supervised pre-training!
  - No unified task
  - Limited signal that pre-training helps

- ... Uniform featurizing and Multi-modal!
  - But different 2D / 3D graphs, periodic structures, geometry



Foundation Models at Intel AI

Knowledge Graph Reasoning

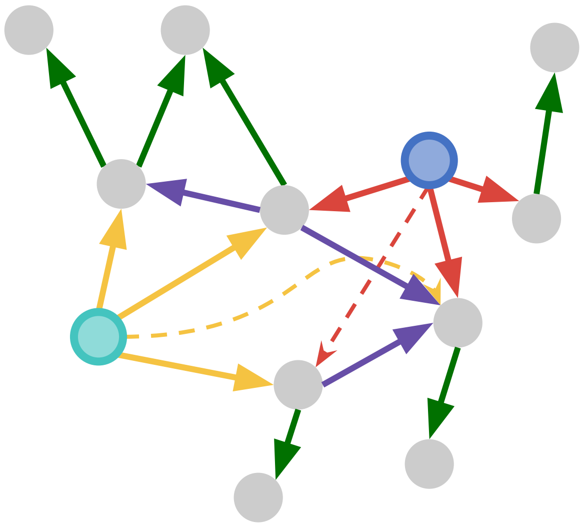
- At large-scale
- Inference on any domain
- All graph-level tasks (start from link prediction)

AI 4 Science

- Molecules, proteins, materials (crystals)
- Materials generation, eg, new catalysts

Foundation models: Graph Reasoning

- Simple link prediction
- Complex logical query answering
- ... and beyond



Knowledge Graphs

Multi-relational graphs with **(subject, predicate, object)** triples.

Multi-domain graphs:

- **Encyclopedias** (Wikidata, Freebase)

  In search and retrieval-augmented LLMs

London (Google)

About

London, the capital of England and the United Kingdom, is a 21st-century city with history stretching back to Roman times. At its centre stand the imposing Houses of Parliament, the iconic 'Big Ben' clock tower and Westminster Abbey, site of British monarch coronations. Across the Thames River, the London Eye observation wheel provides panoramic views of the South Bank cultural complex, and the entire city. — Google

Weather: 57°F (14°C), Wind W at 7 mph (11 km/h), 78% Humidity More on weather.com

Local time: Thursday 7:29AM

Neighborhoods: Elephant and Castle, Chiswick, Brent Cross, MORE

Elevation: 36 ft (11 m)

Local government districts: 32 London boroughs; and the City of London

Region: London (Greater London)

Settled by Romans: AD 47; 1976 years ago; as Londinium

Feedback

London (Bing)

London is the capital and largest city of England and the United Kingdom, with a population of around 8.8 million. It stands on the River Thames in south-east England at the head of a 50-mile es… +
Wikipedia

gov.uk

Country England
Region London (Greater London)
Elevation 36 ft (11 m)
Sovereign state United Kingdom
See more

Knowledge Graphs

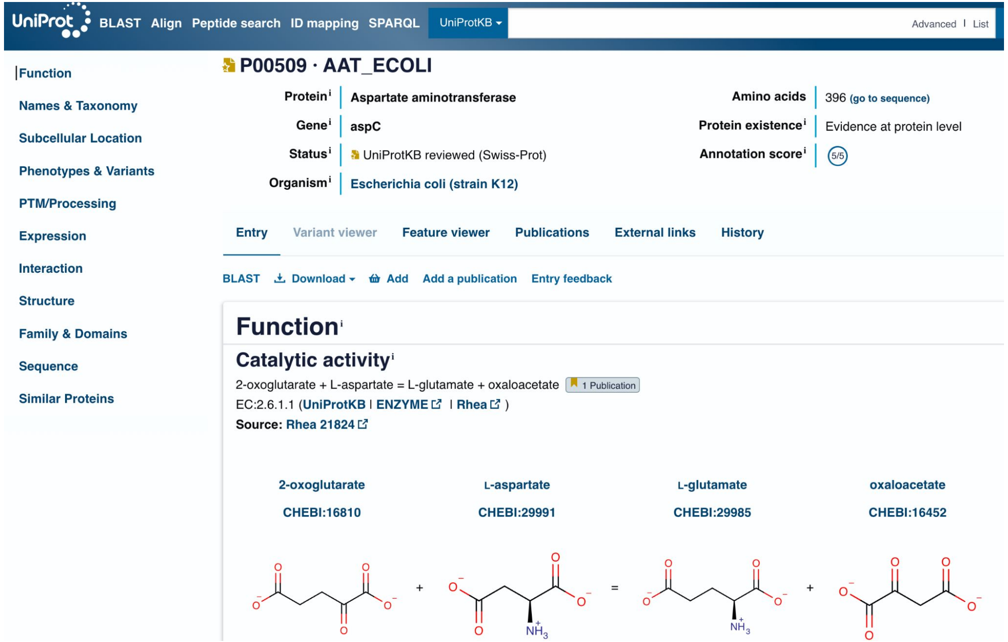
Multi-relational graphs with **(subject, predicate, object)** triples.

Multi-domain graphs:

- Encyclopedias (Wikidata, Freebase)
- **Sciences** (UniProt, DrugBank, Hetionet)

  eg, protein LMs are trained on UniProt

UniProt



Knowledge Graphs

Multi-relational graphs with **(subject, predicate, object)** triples.

Multi-domain graphs:

- Encyclopedias (Wikidata, Freebase)
- Sciences (UniProt, DrugBank, Hetionet)
- Thousands of **domain-specific KGs**

Spatiotemporal Urban KG

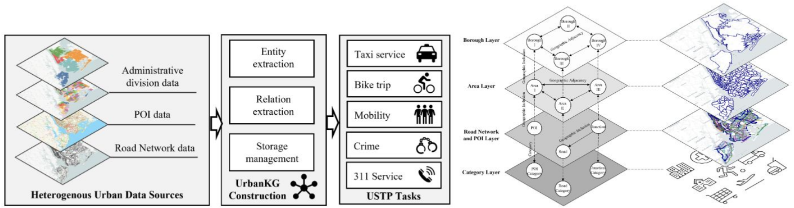
UUKG

The Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction. PDF

Overview • Installation • Dataset • How to Run • Directory Structure • Citation

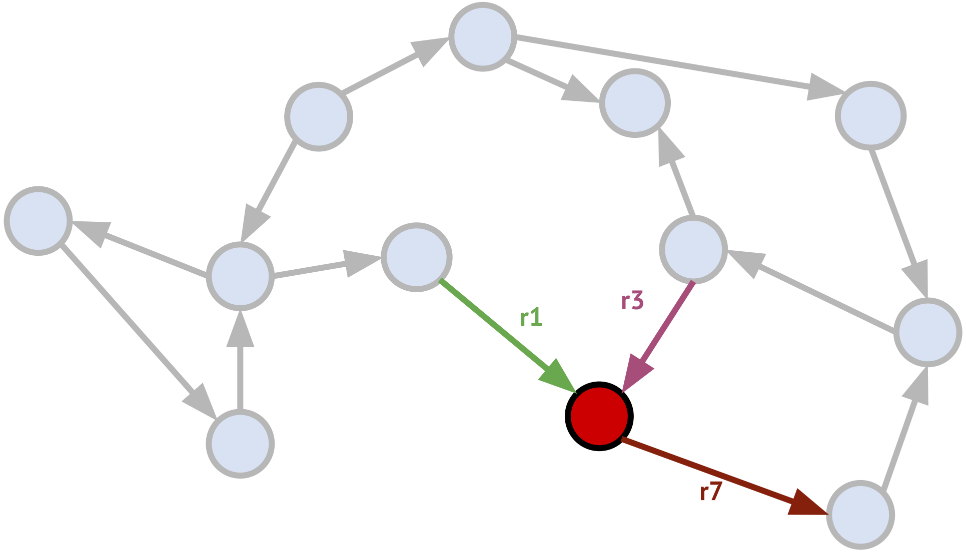
Official repository of NeurIPS 2023 Dataset and Benchmark Track paper "UUKG: The Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction". Please star, watch and fork our repo for the active updates!

1. Overview



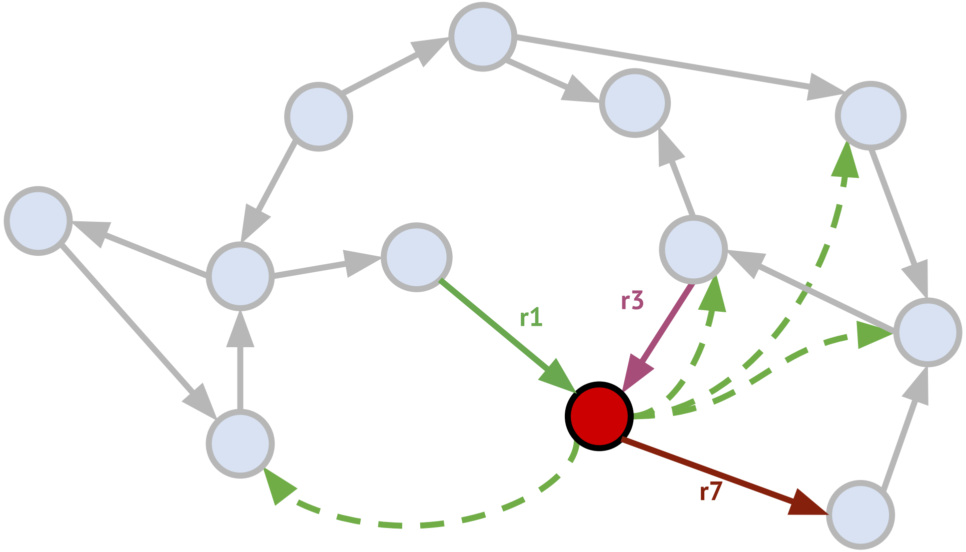
Knowledge Graphs: Setup

intel labs



- Directed graphs (V, E)
- Explicit relation types (R)
- Input node features are **not** given
- **Transductive**: the same graph at inference
- **Inductive**: different graph at inference

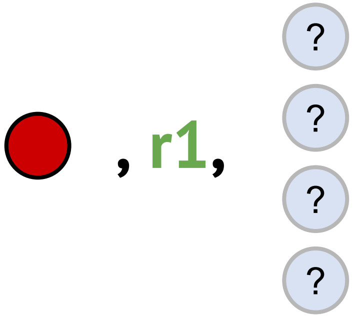
Basic Knowledge Graph Reasoning



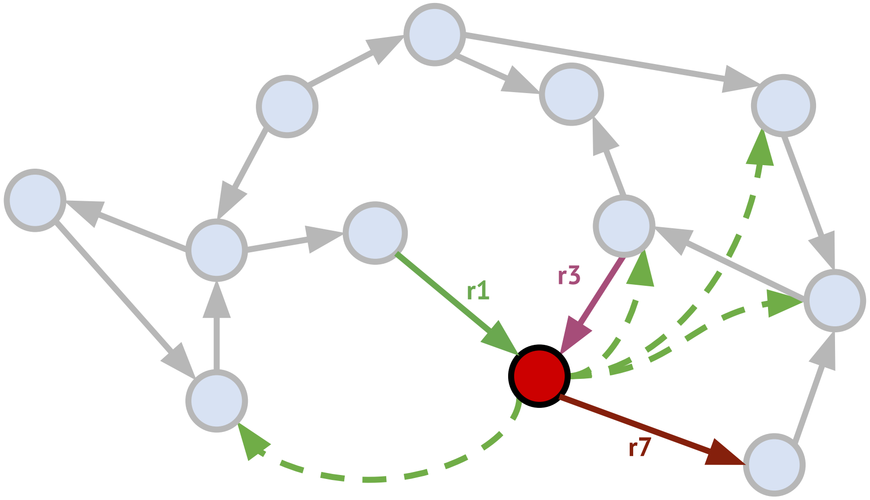
- Query: (head, relation, ?)

  , r1, ?

- Rank **all** entities as possible tails



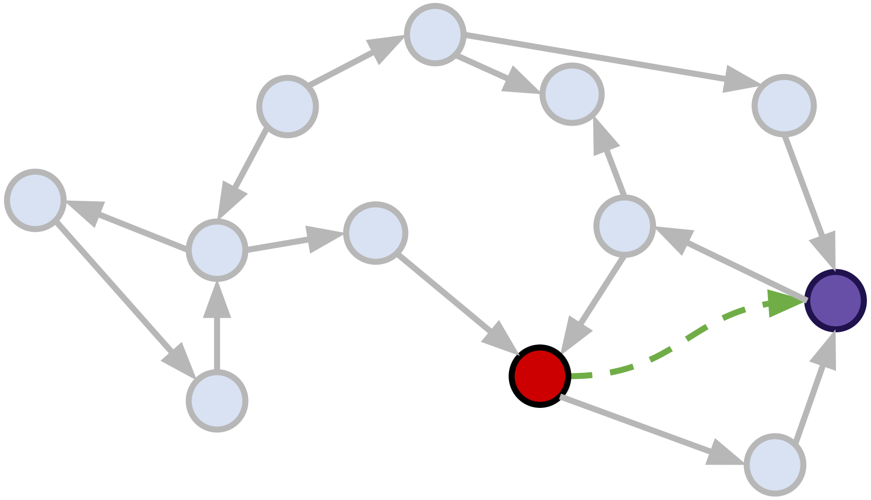
KG Completion vs Link Prediction



- Query: (head, relation, ?)

, **r1**, ?

- Rank **all** entities as possible tails

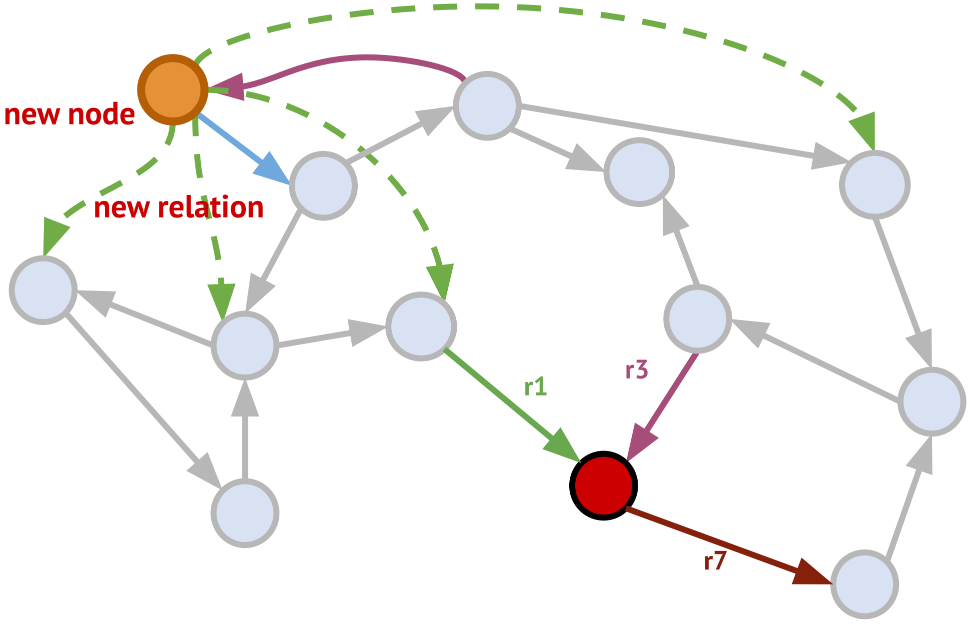


- Query: (head, tail)

?

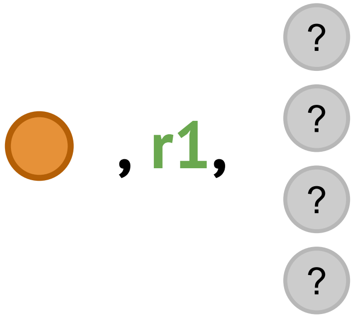
- Binary classification / Relation prediction

Inductive Graph Reasoning



- New nodes and relation types at inference time

- We still want to reason over new entities and relations



The Holy Grail

- One (pre)trained model
- 0-shot inference on any possible multi-relational graph
- Any simple or complex query reasoning
  - 1-hop KG completion
  - Multi-hop logical query answering

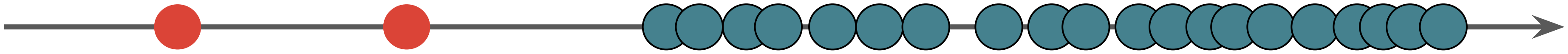
KG completion (simple queries)

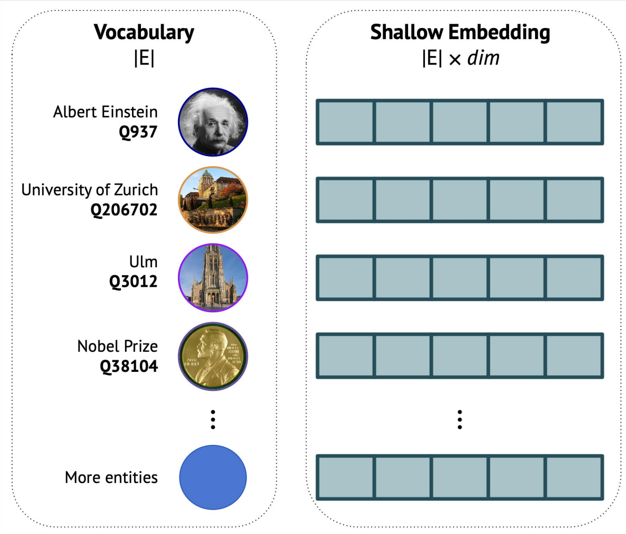
Brief History: 2011 -

RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]

100+ KG embedding models since then 😱





Transductive models only: they learn graph-specific

- Entity embeddings (|V| x d)
- Relation embeddings (|R| x d)

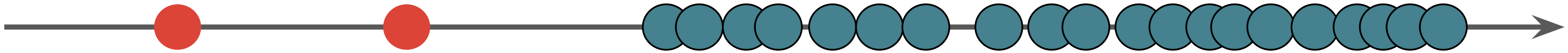
Brief History: 2011 -

Transductive | Triples | Supervised

RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]

100+ KG embedding models since then 😱



Link Prediction on FB15k-237

No substantial progress since 2018

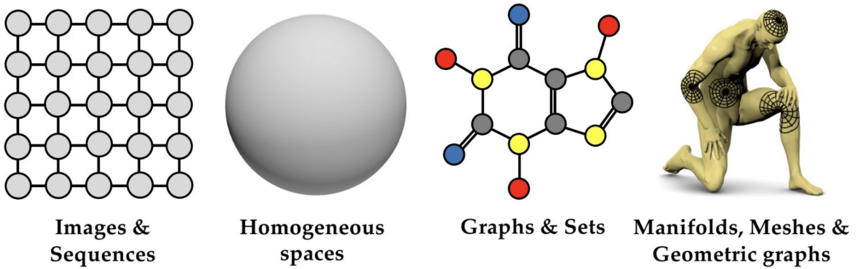
Brief History: 2011 -

Transductive | **Triples** | **Supervised**

RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]

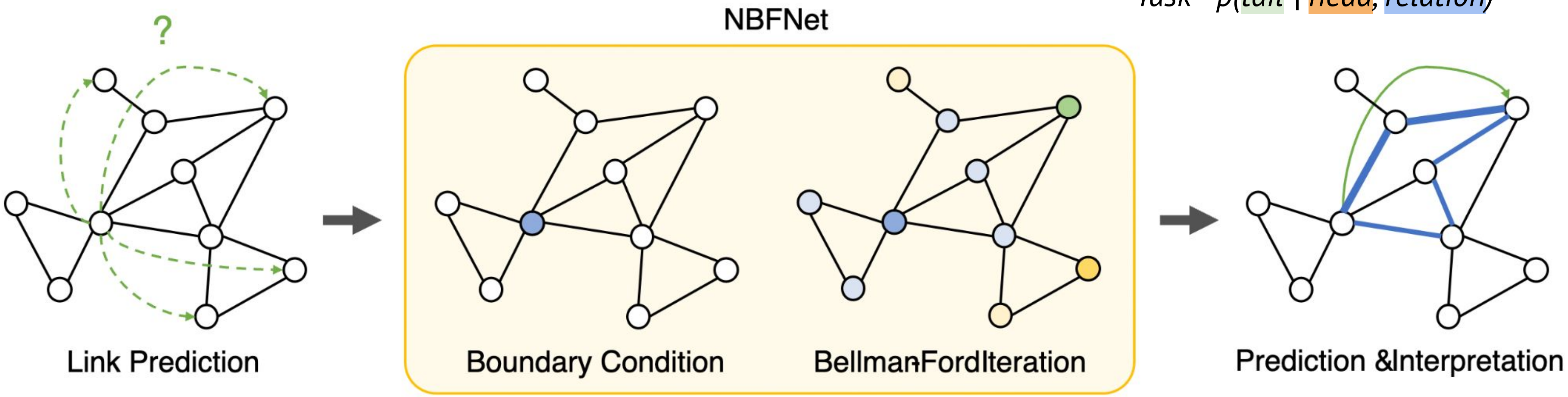
Geometric DL 🎇
2018



https://geometricdeeplearning.com/

Breakthrough: Neural Bellman-Ford (2021)

Task - p(tail | head, relation)



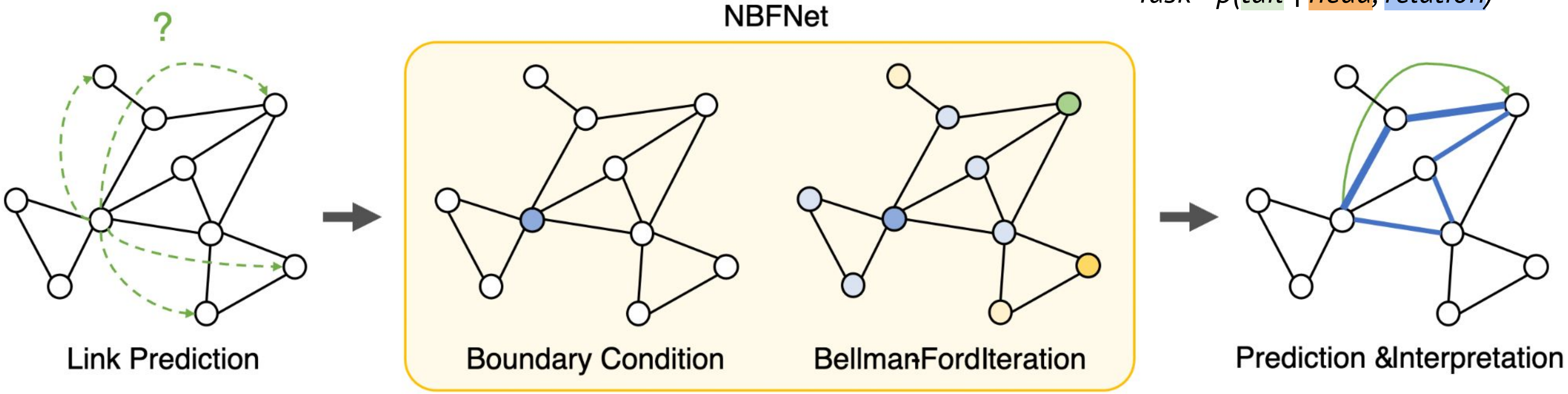
Idea:

1. Relations do not change at inference -> we can learn relation (edge type) embeddings
2. Initialize head node feature with the learnable relation vector (query)
3. Propagate for L layers, take final representations as final node features

Zhu et al. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. NeurIPS 2021

Breakthrough: Neural Bellman-Ford (2021)

Task - p(tail | head, relation)



$$\boldsymbol{h}^0_{v|u} = \textsc{Indicator}_e(u, v, q) = \mathbb{1}_{u=v} * \boldsymbol{R}_q[q]$$

$$\boldsymbol{h}^{t+1}_{v|u} = \textsc{Update}\Big(\boldsymbol{h}^t_{v|u}, \textsc{Aggregate}\big(\textsc{Message}(\boldsymbol{h}^t_{w|u}, g^{t+1}(\boldsymbol{r}))|w \in \mathcal{N}_r(v), r \in \mathcal{R}\big)\Big)$$

Zhu et al. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. NeurIPS 2021

Other Labeling Tricks

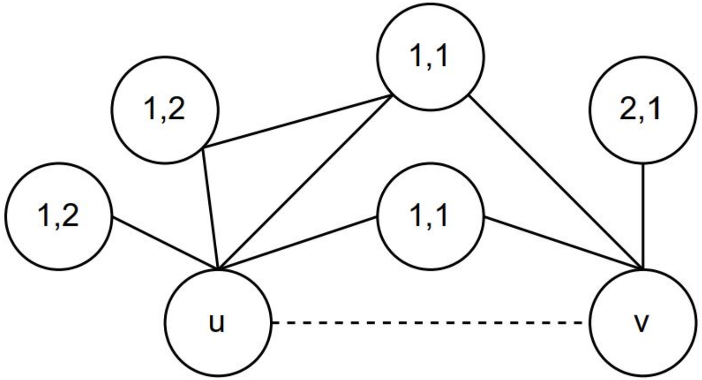
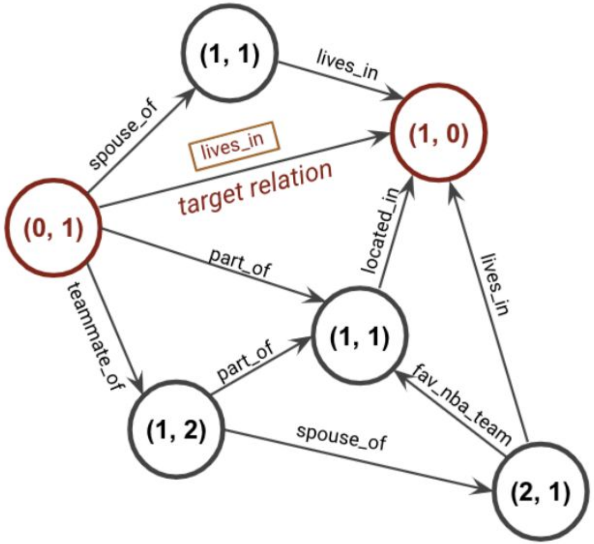


Figure 5: The DE node labeling scheme for link (u, v)

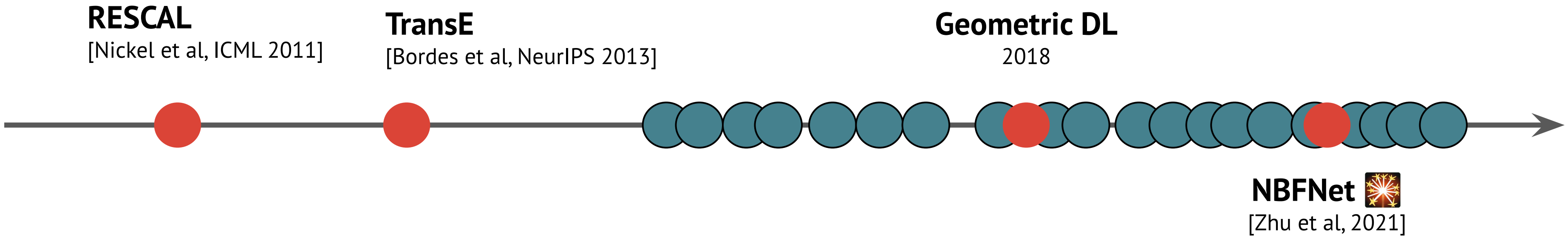


2. Label the nodes w.r.t the target nodes to identify their structural role. Uniquely labels target nodes to mark them for the model.

Chamberlain, Shirobokov et al. Graph Neural Networks for Link Prediction with Subgraph Sketching. ICLR 2023
Teru et al. Inductive Relation Prediction by Subgraph Reasoning. ICML 2020

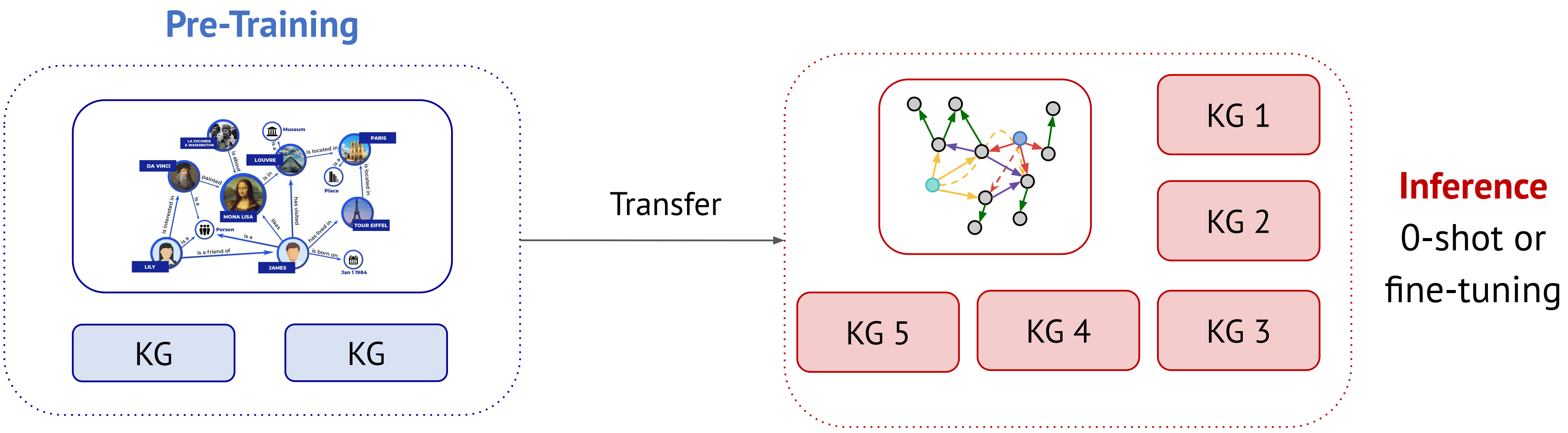
Brief History: 2011 -

Inductive (ent) | Triples | Supervised



- **NBFNet** and Labeling Trick GNNs generalize to new nodes given **fixed relation types**:
- Is is possible to generalize to **both new nodes and new relation types?**

Foundation Models for Graph Reasoning

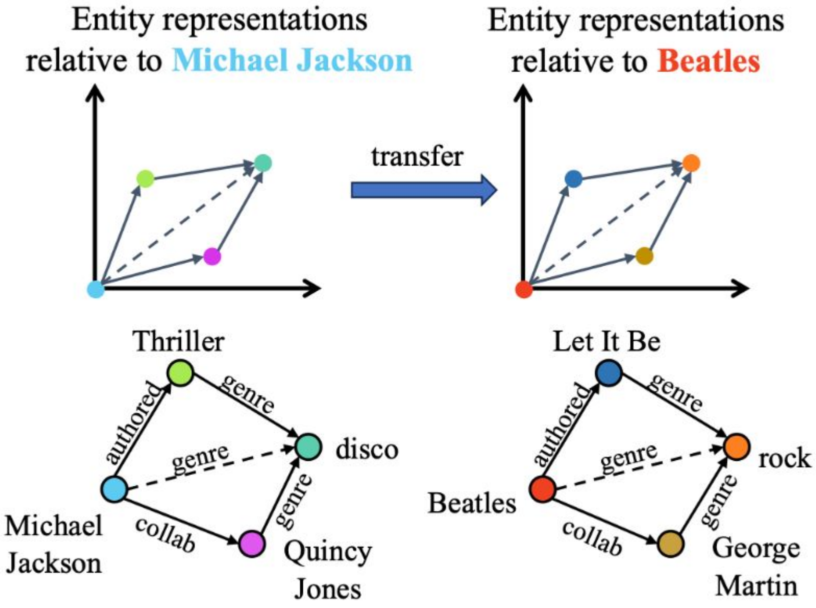


- ➔ We want to train a **single** model on one (or many) graph and run inference on **any other** possible KG
- ➔ Main problem: different entity and relation vocabularies
- ➔ For that, what is the transferable <u>invariance</u>?

Existing Inductive (entity) Models

Most of existing models after NBFNet:

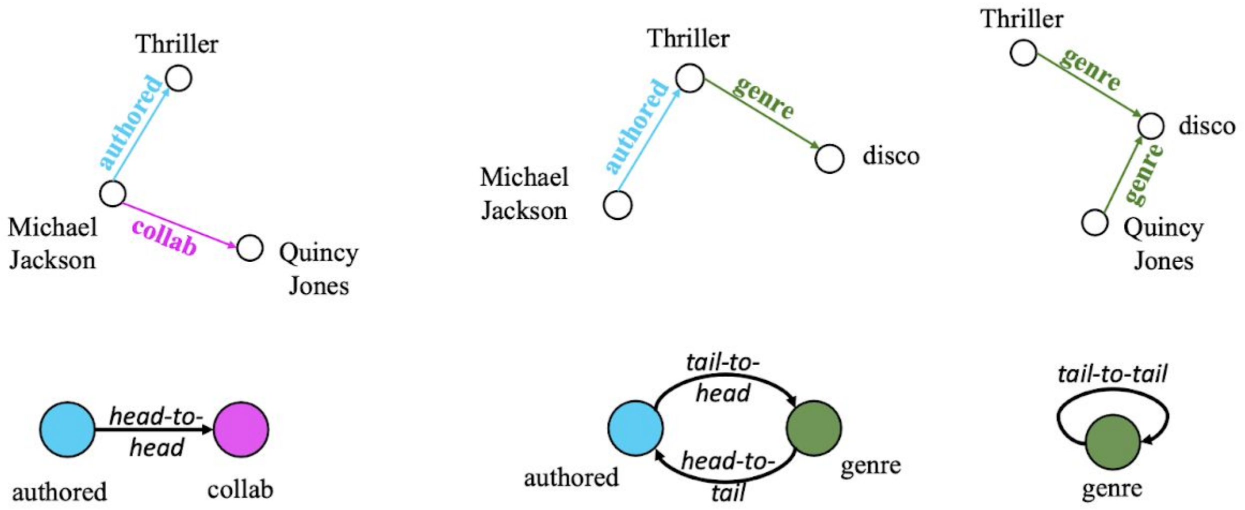
- learn relation embeddings
- build **relative** entity representations (using a labeling trick)
  - Initialize the head node with a learnable query vector $\boldsymbol{q}$
  - Other nodes <- 0
  - Message passing GNN
- Transfer to graphs with the **same relation types**



(a) Relative **entity** representations transfer to new entities (NBFNet, RED-GNN)

ULTRA: Unified, Learnable, Transferable

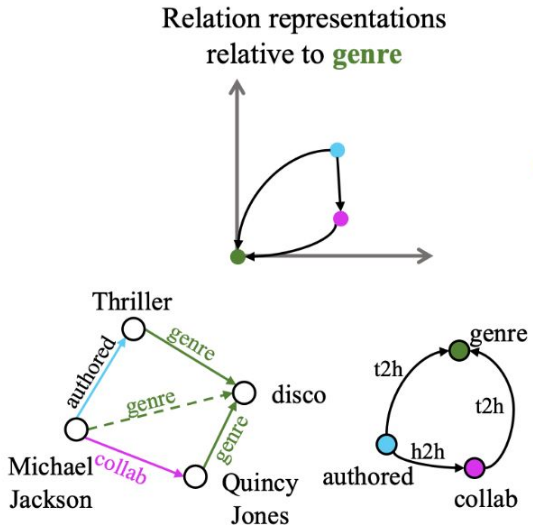
- Let's try building a graph of relations



ULTRA: Unified, Learnable, Transferable

- Let's try building a graph of relations

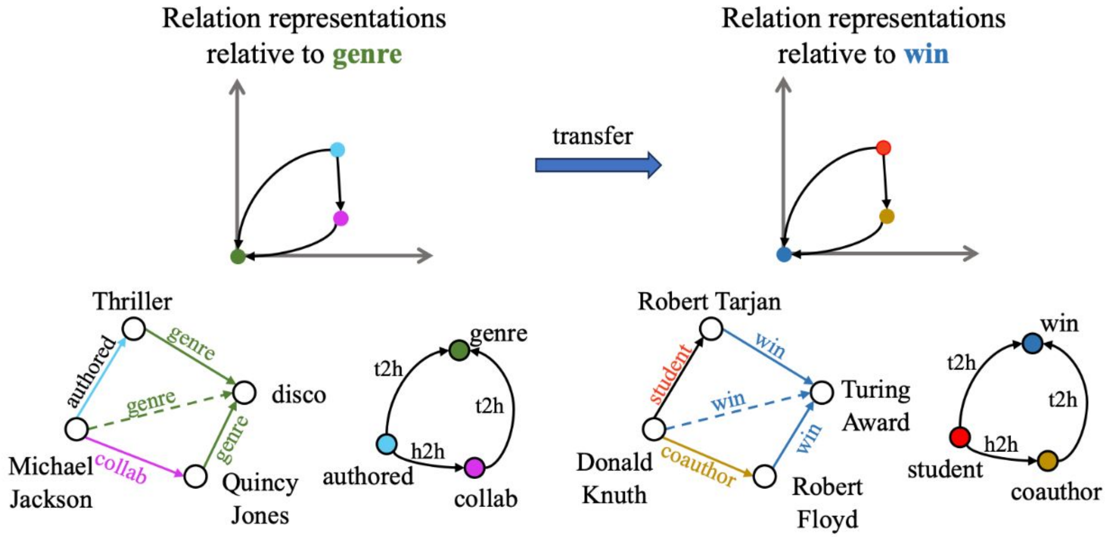
- 4 fundamental interactions:
    - Head-to-head (*h2h*)
    - Tail-to-head (*t2h*)
    - Tail-to-tail (*t2t*)
    - Head-to-tail (*h2t*)



Observation: fundamental relations between relations remain the same!

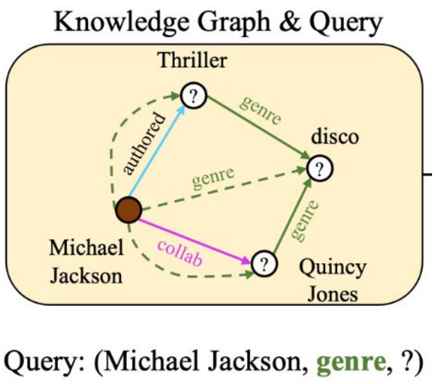
ULTRA: Unified, Learnable, Transferable

- Let's try building a graph of relations
- 4 fundamental interactions:
  - Head-to-head (*h2h*)
  - Tail-to-head (*t2h*)
  - Tail-to-tail (*t2t*)
  - Head-to-tail (*h2t*)
- Can be used to infer **relative relation representations** of **new** relations



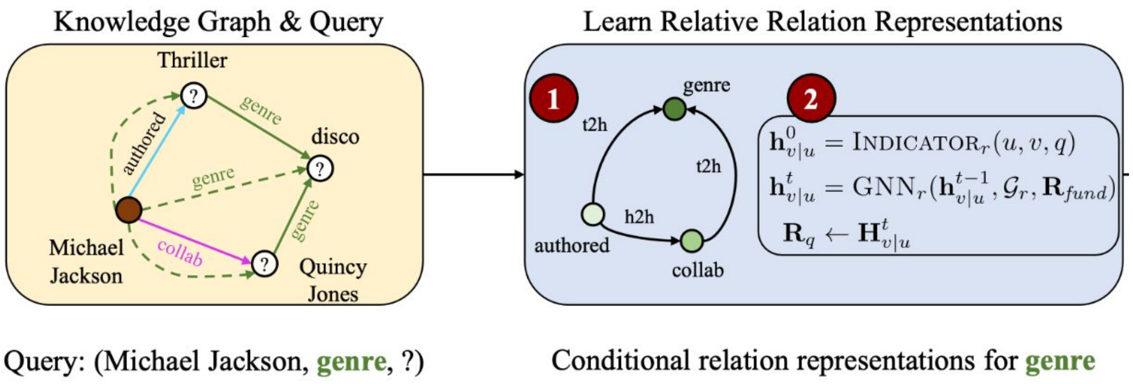
(b) Relative **relation** representations transfer to new relations (ULTRA)

Step 0: Input graph and query



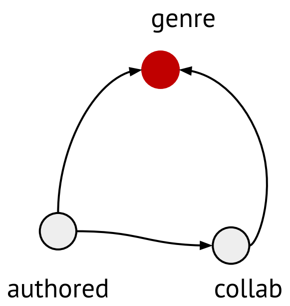
- ➔ Literally any multi-relational graph
- ➔ No input node/edge features are needed

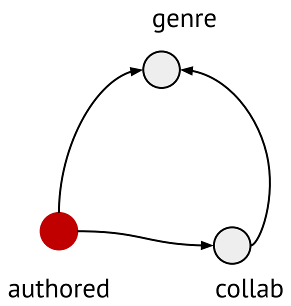
Steps 1+2 : graph of relations + labeling trick

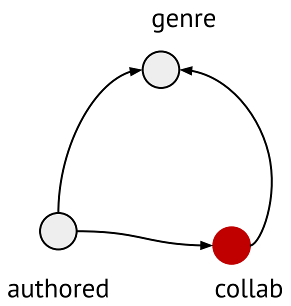


- ➔ Nodes = unique relations, edge types = 4 fundamental interactions
- ➔ Initialize the query relation node with $\mathbf{1}^d$
- ➔ Initialize the rest nodes with $\mathbf{0}^d$
- ➔ Message passing yields relative relation representations
- ➔ **Each relation = Unique relation representations |R| x d**

Each query relation = Unique representations



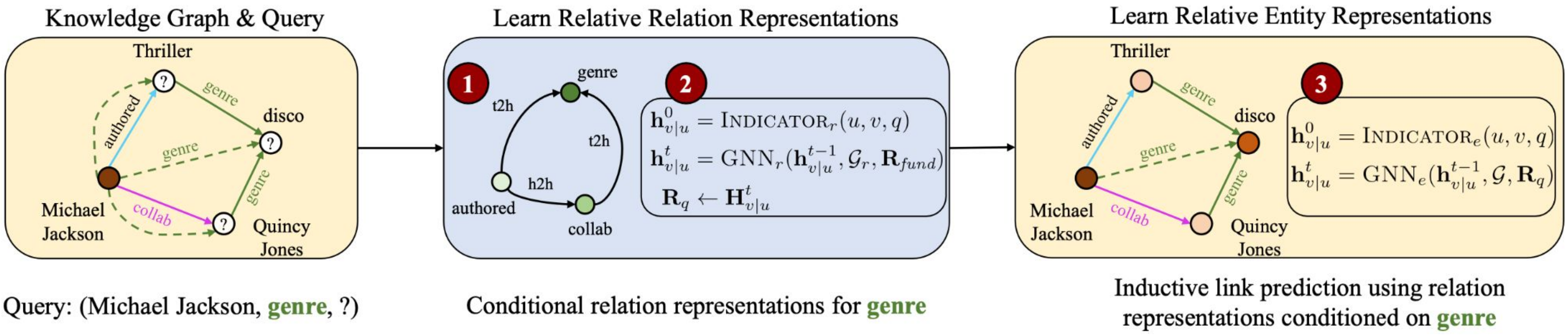




	Conditional MPNN	Conditional MPNN	Conditional MPNN
genre	**[[0.5, 1.2, 1.3]**	**[[0.7, 2.2, 0.2]**	**[[1.3, 0.5, 2.7]**
authored	**[0.8, 0.4, 1.0]**	**[1.2, 0.9, 3.0]**	**[0.3, 0.8, 1.0]**
collab	**[1.1, 2.0, 0.4]]**	**[0.1, 1.4, 2.6]]**	**[0.6, 2.4, 3.1]]**

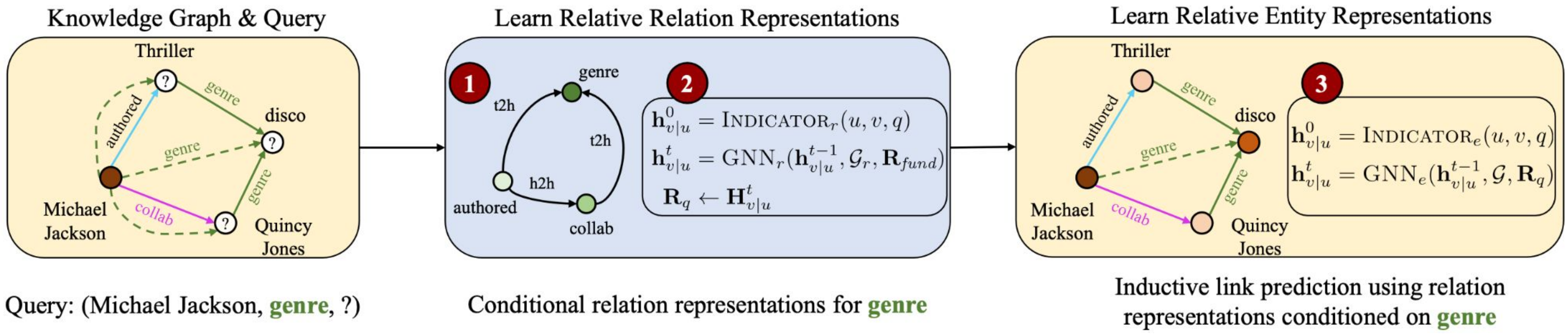
Huang et al. A theory of link prediction via relational Weisfeiler-Leman on knowledge graphs. NeurIPS 2023

Step 3: run any inductive GNN



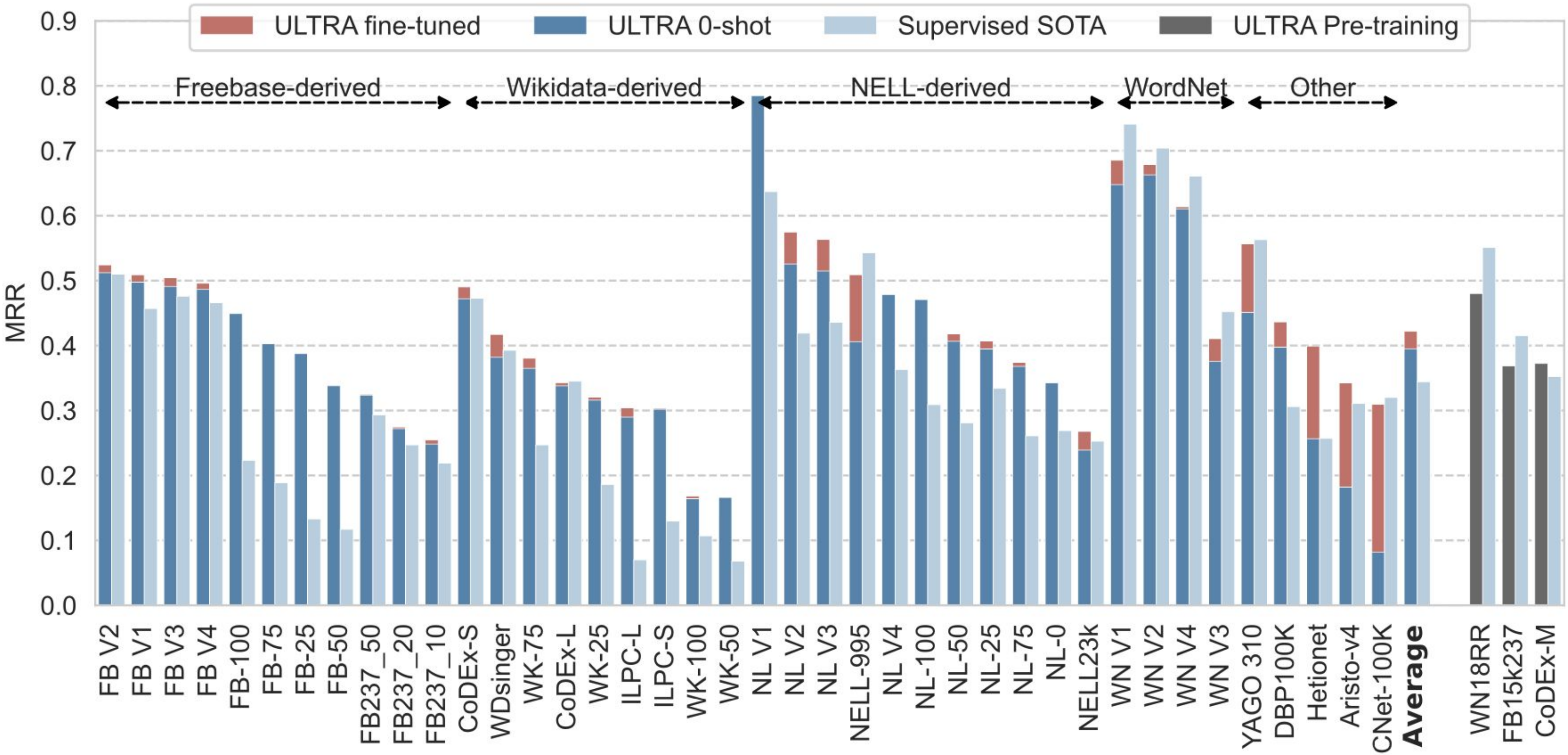
- ➔ **Each relation = Unique relation representations |R| x d**
- ➔ Use those relational representations for any inductive GNN (like NBFNet)

ULTRA: Foundation Model for KG Reasoning



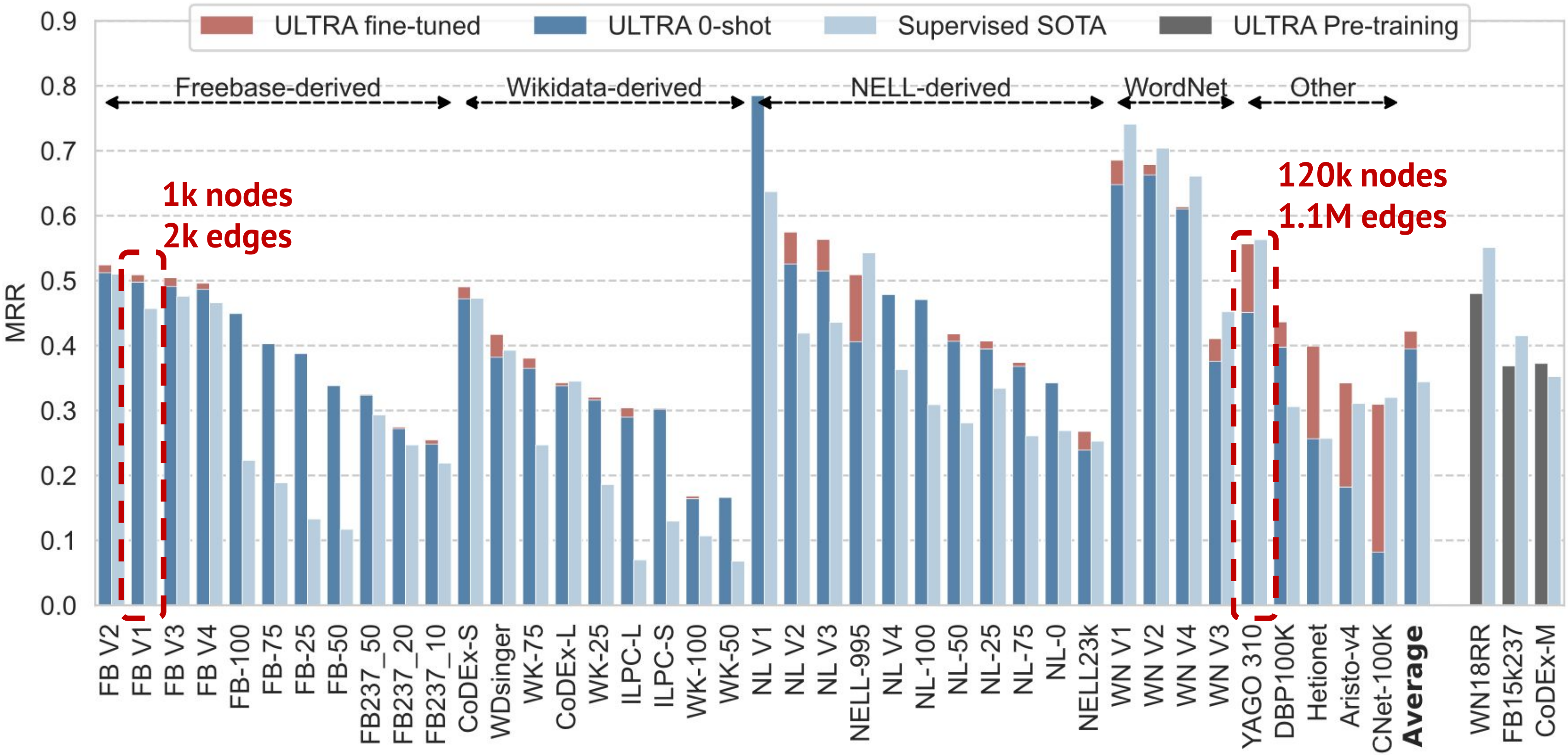
- ✓ Doesn’t need any input entity/relation features
- ✓ Learnable parameters: 4 fundamental relations (*h2t, t2t, t2h, h2h*) + GNN weights
- ✓ Generalizes to any graph of any size with any relation vocabulary
- ✓ Allows 0-shot inference and fine-tuning on any graph

Pre-trained ULTRA beats supervised SOTA in 0-shot inference on 50+ KGs



May 15th 2024

Generalization to different graph sizes



May 15th 2024

Generalization to New Unseen Domains

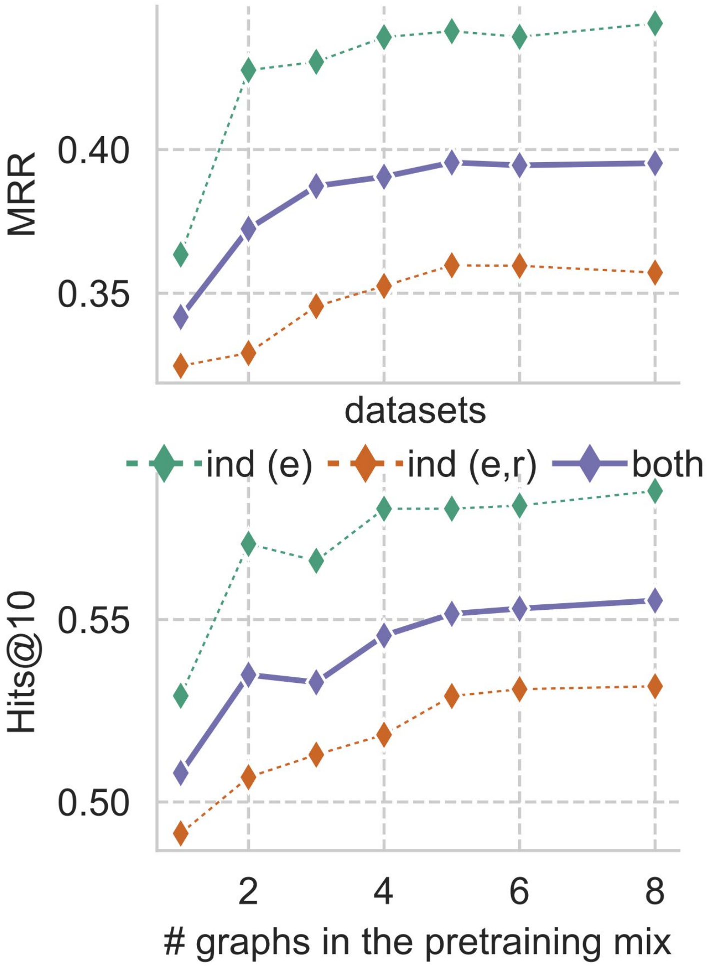
- ➢ Pre-trained on mostly general encyclopedia data (Freebase, Wikidata)

Graph	Domain	Supervised SOTA (MRR)	ULTRA (0-shot / ft) (MRR)
Hetionet	Biology, drugs	0.257	0.257 / **0.399**
ConceptNet	Commonsense reasoning	**0.320**	0.082 / <u>0.310</u>
Urban KG	Geography, location	0.552	0.556 / **0.618**

- ➢ Let us know more domain-specific KGs!

More data helps
0-shot inference

- 👀 Aggregated results over 40 KGs
- 👀 More diverse KGs in the pre-training data mix help
  - ○ More relational graphs and their interactions
- 🤔 Saturation after training on 3-4 graphs
- 🤔 Scaling behavior to be investigated



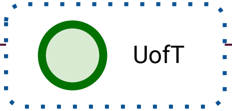
Complex logical queries

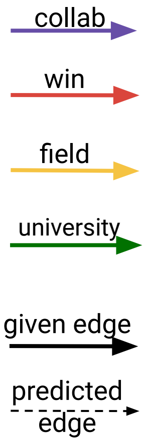
At what universities do the Turing Award winners in the field of Deep Learning work?

$$q = U_?.\exists V : win(TuringAward, V) \wedge field(DeepLearning, V) \wedge university(V, U_?)$$

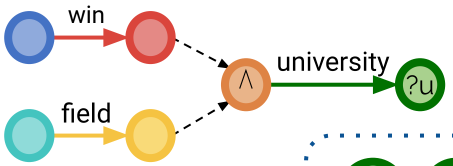
SPARQL query (edge traversal)

```
SELECT ?uni WHERE
{
  TuringAward     win        ?person .
  DeepLearning    field      ?person .
  ?person   university       ?uni    .
}
```





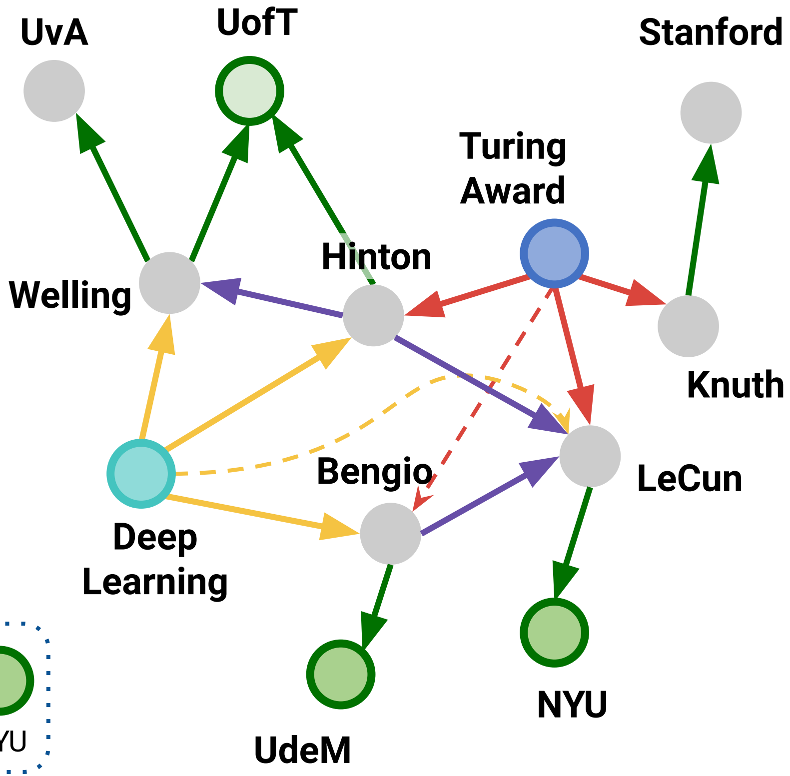
Neural query execution (+ link prediction)



Answer set: UofT UdeM NYU

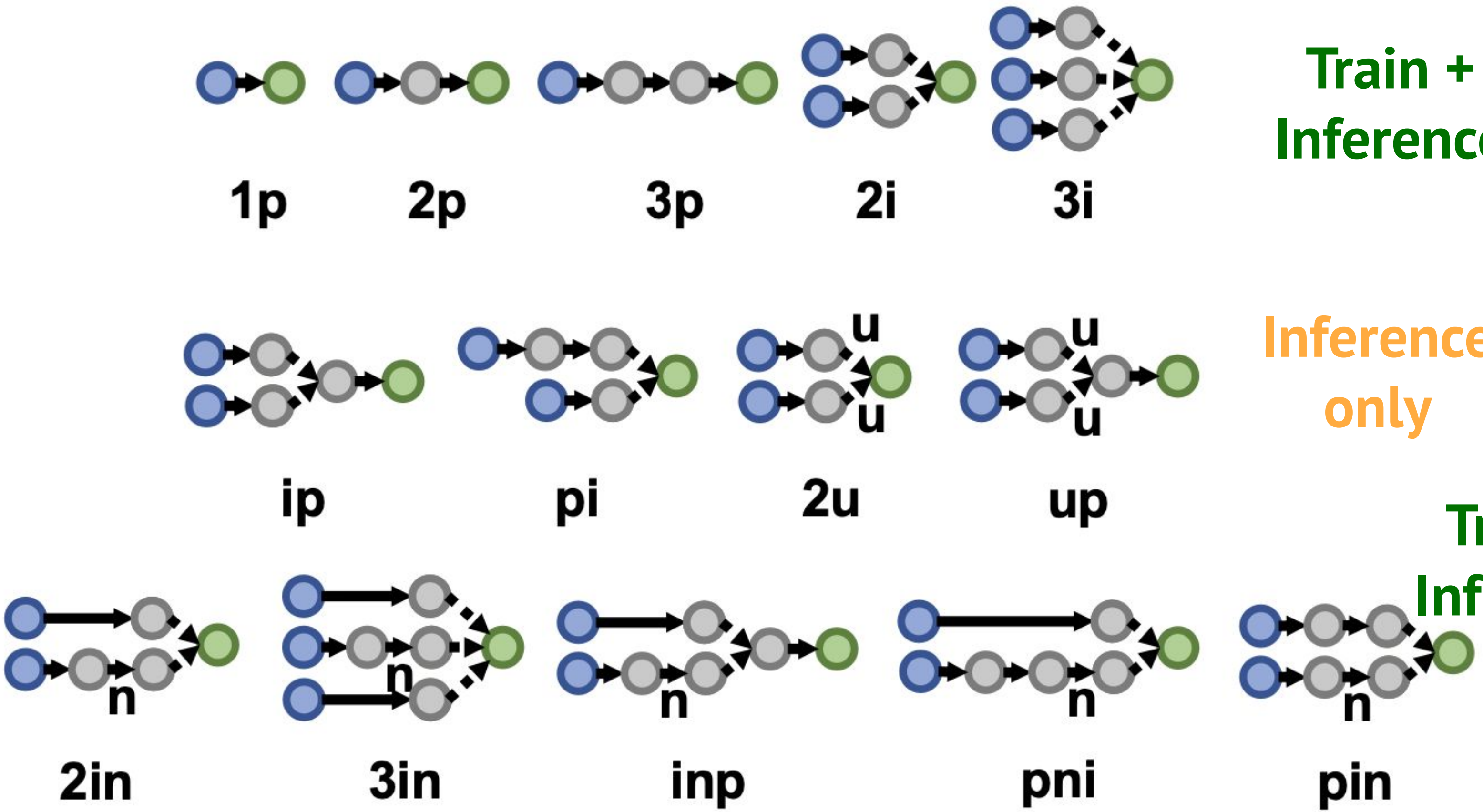
Easy: UofT

Hard: UdeM NYU



Ren et al. Neural Graph Reasoning: Complex Logical Query Answering meets Graph Databases. arxiv 2023

Query patterns

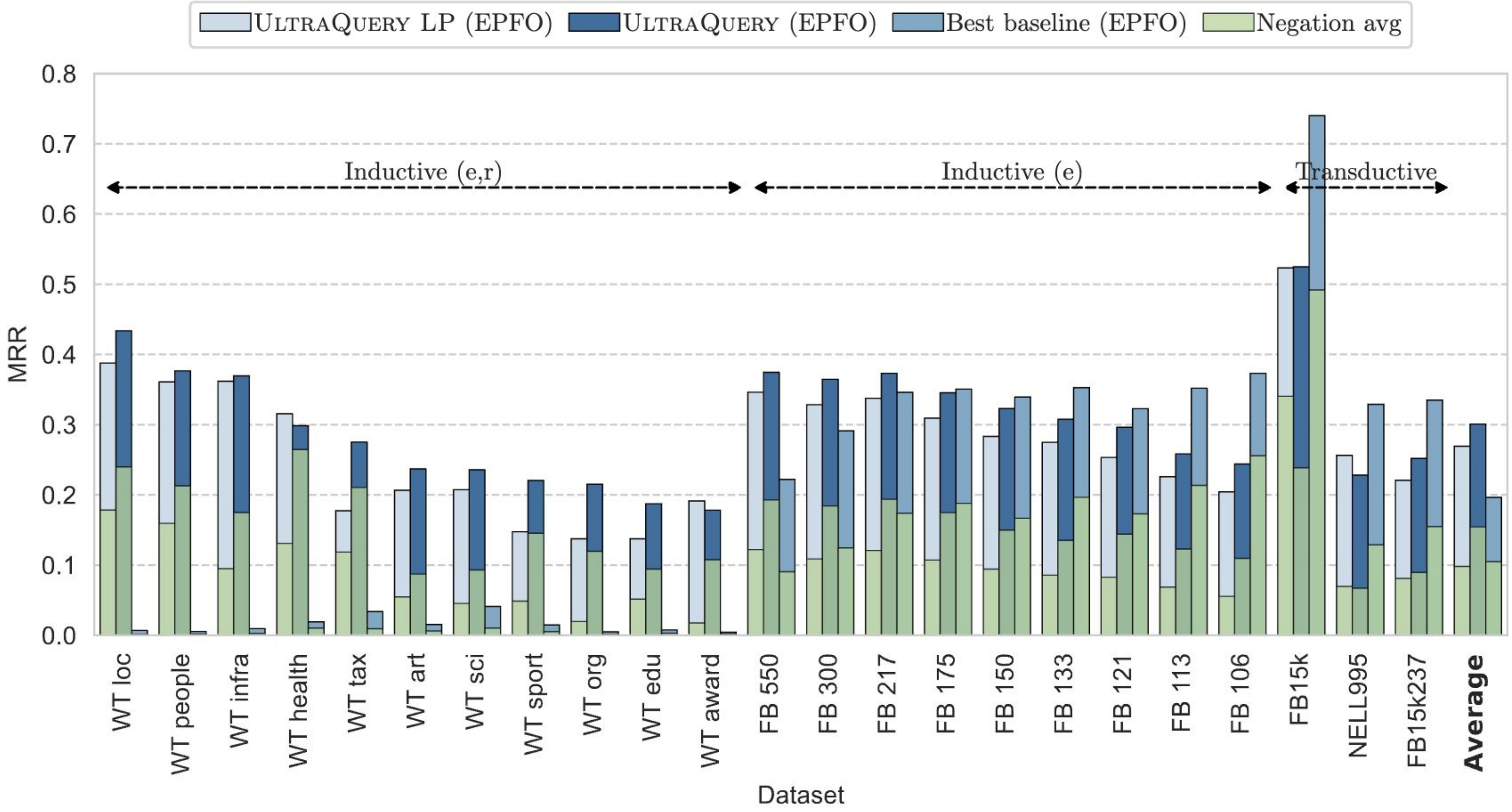


Train + Inference

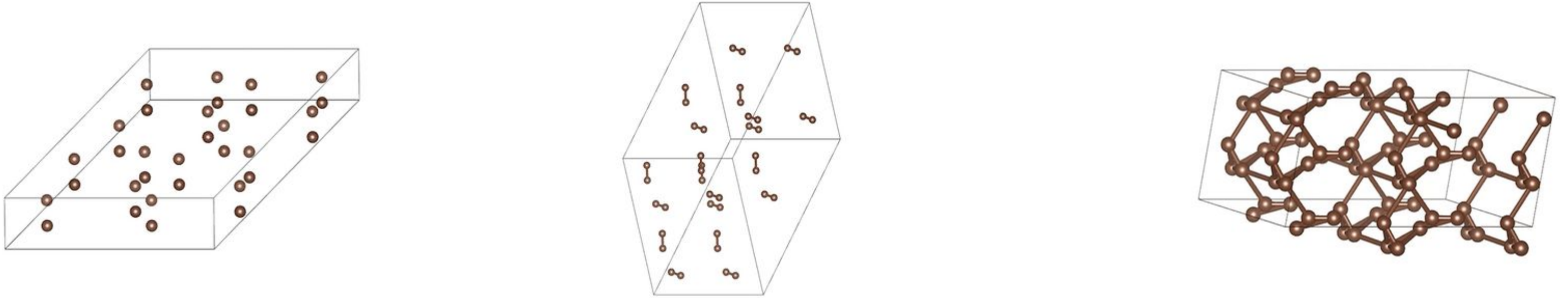
Inference only

Train + Inference

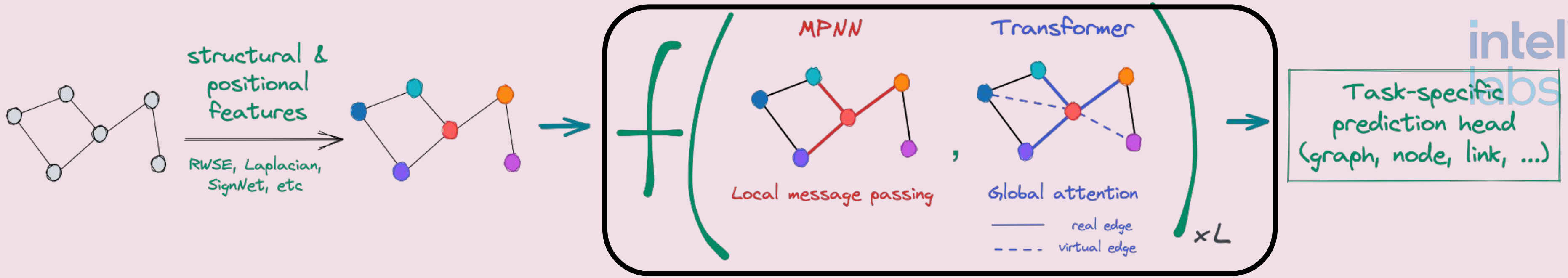
The same pre-trained ULTRA for complex, multi-hop queries



Foundation Models: AI 4 Science



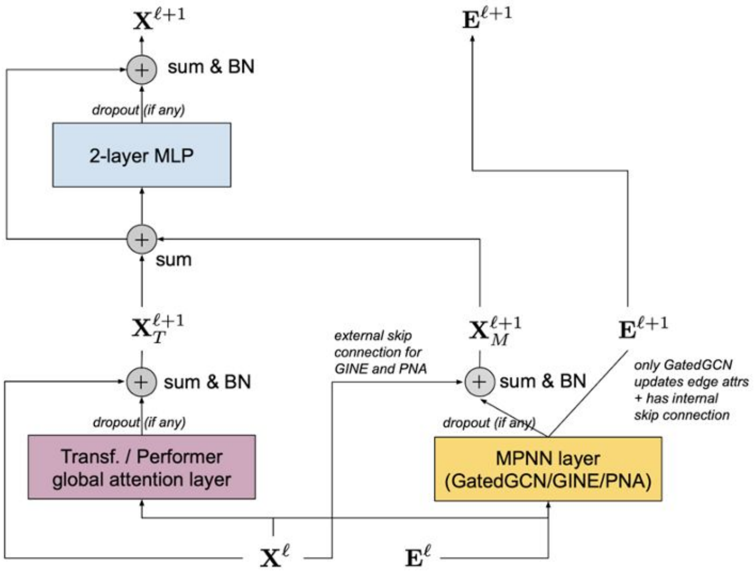
Bandgap-guided carbon structure generation
Source: https://distributionalgraphormer.github.io/



GraphGPS [Rampasek et al, 2022]

stack of *L* GPS layers

Entrance to the molecular ML



Combines Local MPNN and Transformer:

- Sum aggregation of the two representations
- Followed by a 2-layer MLP and skip-connections

Shameless plug: Best Graph Transformer of 2022

Recipe for a General, Powerful, Scalable Graph Transformer

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, A. Luu, Guy Wolf, D. Beaini · Computer Science ·
Neural Information Processing Systems · 25 May 2022

TLDR This paper proposes the first architecture with a complexity linear in the number of nodes and edges $O(N+E)$ by decoupled the local real-edge aggregation from the fully-connected Transformer, and argues that this decoupling does not negatively affect the expressivity, with the architecture being a universal function approximator on graphs. Expand

GraphGPS Public Watch 9 Fork 77 Starred 455

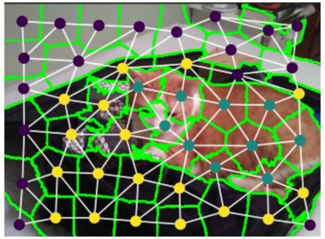
Model	PCQM4Mv2		
	Validation MAE ↓	Training MAE	# Param.
GCN-virtual	0.1153	n/a	4.9M
GIN-virtual	0.1083	n/a	6.7M
GRPE [48]	0.0890	n/a	46.2M
EGT [29]	0.0869	n/a	89.3M
Graphormer [51]	0.0864	0.0348	48.3M
GPS-small	0.0938	0.0653	6.2M
GPS-medium	0.0858	0.0726	19.4M

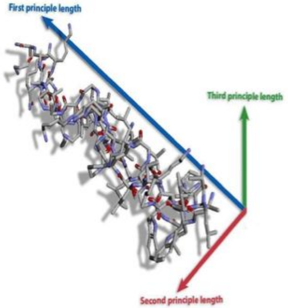
Model	ZINC MAE ↓
GCN [33]	0.367 ± 0.011
GIN [60]	0.526 ± 0.051
GatedGCN [7, 15]	0.282 ± 0.015
PNA [13]	0.188 ± 0.004
DGN [3]	0.168 ± 0.003
CIN [5]	0.079 ± 0.006
CRaWl [53]	0.085 ± 0.004
GIN-AK+ [67]	0.080 ± 0.001
SAN [36]	0.139 ± 0.006
Graphormer [62]	0.122 ± 0.006
K-Subgraph SAT [9]	0.094 ± 0.008
EGT [29]	0.108 ± 0.009
GPS (ours)	0.070 ± 0.004

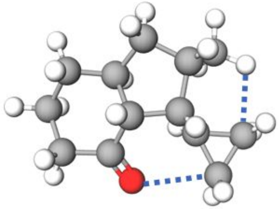
Long Range Graph Benchmark (LRGB) Results

- A new collection of datasets that require long range modeling for a network to perform well.

`reg-bound` graph







Model	PascalVOC-SP	COCO-SP	Peptides-func	Peptides-struct	PCQM-Contact
	F1 score ↑	F1 score ↑	AP ↑	MAE ↓	MRR ↑
GCN	0.1268 ± 0.0060	0.0841 ± 0.0010	0.5930 ± 0.0023	0.3496 ± 0.0013	0.3234 ± 0.0006
GINE	0.1265 ± 0.0076	0.1339 ± 0.0044	0.5498 ± 0.0079	0.3547 ± 0.0045	0.3180 ± 0.0027
GatedGCN	0.2873 ± 0.0219	**0.2641 ± 0.0045**	0.5864 ± 0.0077	0.3420 ± 0.0013	0.3218 ± 0.0011
GatedGCN+RWSE	0.2860 ± 0.0085	0.2574 ± 0.0034	0.6069 ± 0.0035	0.3357 ± 0.0006	0.3242 ± 0.0008
Transformer+LapPE	0.2694 ± 0.0098	**0.2618 ± 0.0031**	0.6326 ± 0.0126	**0.2529 ± 0.0016**	0.3174 ± 0.0020
SAN+LapPE	**0.3230 ± 0.0039**	0.2592 ± 0.0158*	**0.6384 ± 0.0121**	0.2683 ± 0.0043	**0.3350 ± 0.0003**
SAN+RWSE	**0.3216 ± 0.0027**	0.2434 ± 0.0156*	**0.6439 ± 0.0075**	**0.2545 ± 0.0012**	**0.3341 ± 0.0006**
GPS (ours)	**0.3748 ± 0.0109**	**0.3412 ± 0.0044**	**0.6535 ± 0.0041**	**0.2500 ± 0.0005**	**0.3337 ± 0.0006**

Dwivedi V.P., Rampášek L., Galkin M., Parviz A., Wolf G., Luu A.T. and Beaini D., *Long Range Graph Benchmark*. NeurIPS Datasets and Benchmarks 2022.

GraphGPS++: ensembling 112 models

- **GraphGPS** hybrid architecture with Laplacian PEs and Random Walk SEs
- **Transformer-M** biased global attention with 2D/3D grouped input masking
- Denoising autoencoding auxiliary task (**Noisy Nodes**)

Table 4: Ensembled model performance on PCQM4Mv2 dataset. Models in the proxy set are trained on the `train+half_valid` data split whereas those in the full set are trained on all available data.

	Proxy Set			Main Set	
		Valid MAE			Ensembling
Case	**# Models**	**Avg.**	**Ensembled**	**# Models**	**Weight**
1: Baseline	10	0.0755	0.0725	35	1
2: No Atomic Number	4	0.0761	0.0734	16	0.5
3: FNN Dropout = 0.412	8	0.0759	0.0729	14	1
4: FNN Dropout = 0.412; No Atomic Number	5	0.0761	0.0736	7	0.5
5: Feature Set 2†	4	0.0755	0.0731	15	1
6: Feature Set 3†	4	0.0754	0.0731	14	1
7: Masking Weights = [1,2,2]	4	0.0754	0.0730	15	1
All	**39**	**0.0756**	**0.0722**	**112**	

† As defined in Table 2.

GPS++ is OGB LSC 2022 Winner in PCQM4M v2

intel labs

Private Test Challenge

Leaderboard for PCQM4Mv2

Mean Absolute Error (MAE). The lower, the better.

Rank	Team	Test-challenge MAE
1	WeLoveGraphs	0.0719
2	ViSNet	0.0723
2	NVIDIA-PCQM4Mv2	0.0723

Public Test

Leaderboard for PCQM4Mv2

MAE on the test-dev and validation sets. The lower, the better.

Package: >=1.3.2

Rank	Method	Ensemble	Test-dev MAE	Validation MAE	Team	Contact	References	#Params	Hardware	Date
1	**GPS++**	Yes	0.0720	0.0778	GraphcoreValenceMILA	Dominic Masters (Graphcore/Valence/MILA)	Paper, Code	44,291,413	Graphcore BOW-POD16	Nov 18, 2022
2	**MolNet_Ensemble**	Yes	0.0753	0.0797	polixir.ai	zouxiaochuan (polixir.ai)	Paper, Code	32,047,874	8 RTX3090	Nov 1, 2022
3	**Global-ViSNet**	No	0.0766	0.0784	ViSNet	Tong Wang (Microsoft Research AI4Science)	Paper, Code	78,450,692	4 NVIDIA A100 GPUs	Oct 26, 2022

How much molecular and scientific
data is there?

Enormous LLM datasets vs scientific data

Beaini et al, *Towards Foundational Models for Molecular Learning on Large-Scale Multi-Task Datasets,* 2023.

How much data is there?

Fresh release: 100M molecules, 3000 tasks, 13B labels

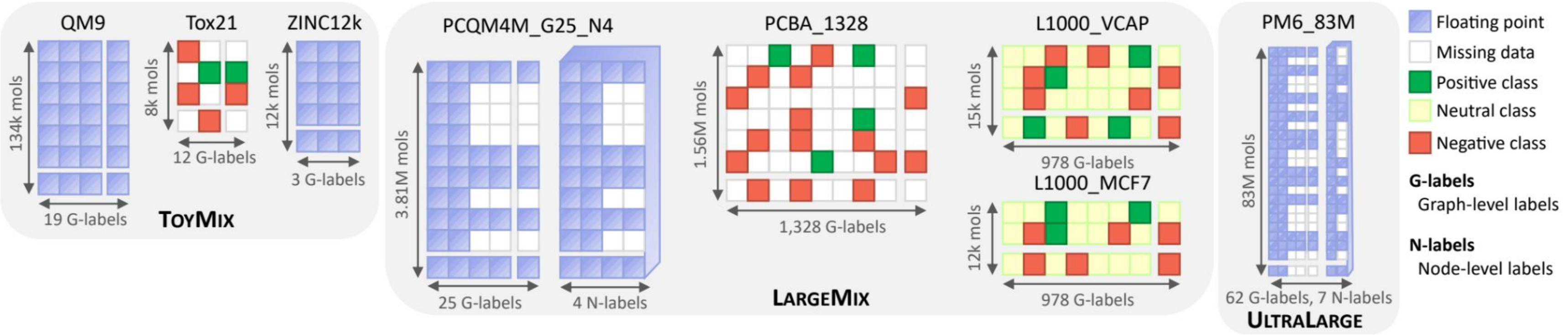
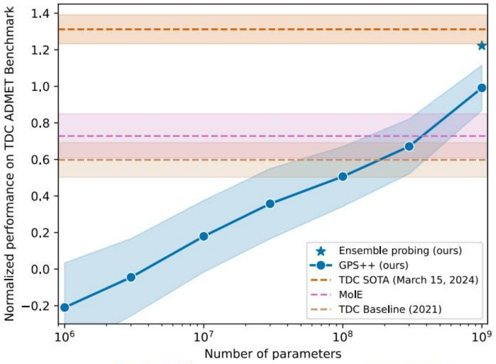
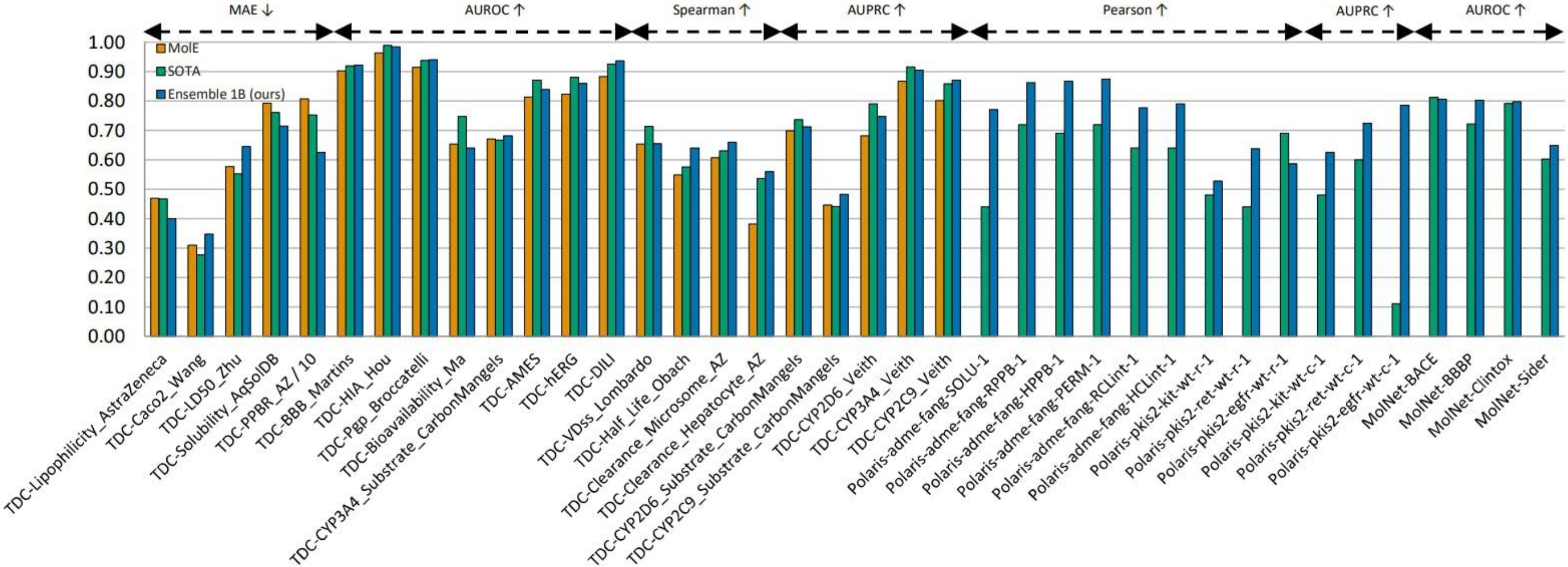


Figure 1: Visual summary of the proposed collections of molecular datasets. The "mixes" are meant to be predicted simultaneously in a multi-task fashion. They include quantum, chemical, and biological properties, categorical and continuous data points, graph-level and node-level tasks.

Beaini et al, *Towards Foundational Models for Molecular Learning on Large-Scale Multi-Task Datasets,* 2023.

MolGPS: scales to 1B params!





Sypetkowski, Wenkel, et al. On the Scalability of GNNs for Molecular Graphs, arxiv 2024

What is the best pre-training objective?

Noisy Nodes [Godwin et al., 2022]
Input: 2D / 3D molecules
Output: Energy

- Aims to tackle the oversmoothing and overfitting problem in MPNNs
- Auxiliary denoising autoencoding
- Can be applied just to node and edge features, which is what we do
- 3D-based distance denoising didn't improve GPS++ performance :(

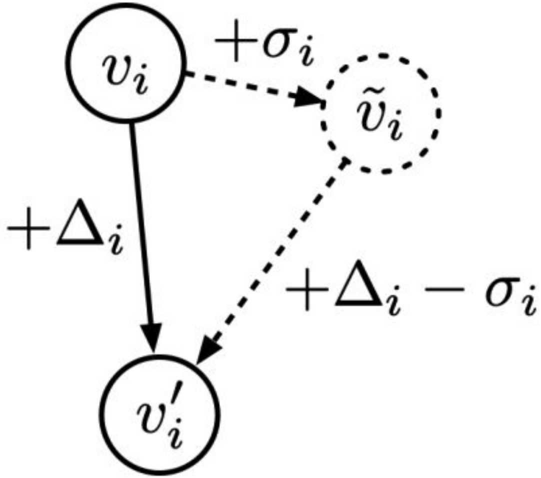


Figure 1: Noisy Node mechanics during training. Input positions are corrupted with noise σ, and the training objective is the node-level difference between target positions and the noisy inputs.

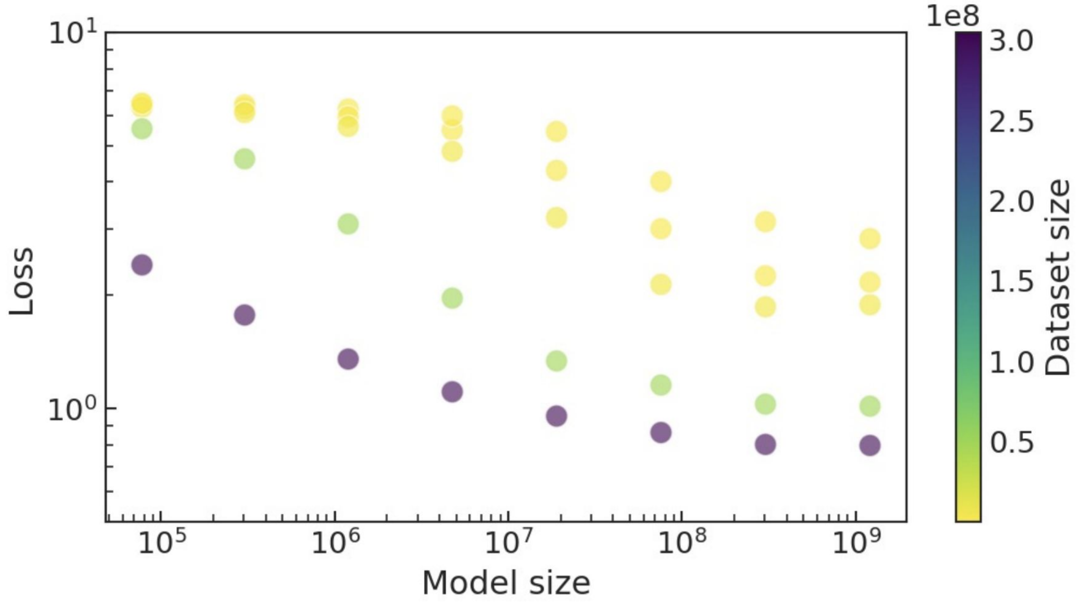
Godwin et al, *Simple gnn regularisation for 3d molecular property prediction & beyond,* ICLR 2022.

What is the best pre-training objective?

ChemGPT [Frey et al., 2022]
Input: SELFIES
Output: Next token

- Slap a transformer over string representations

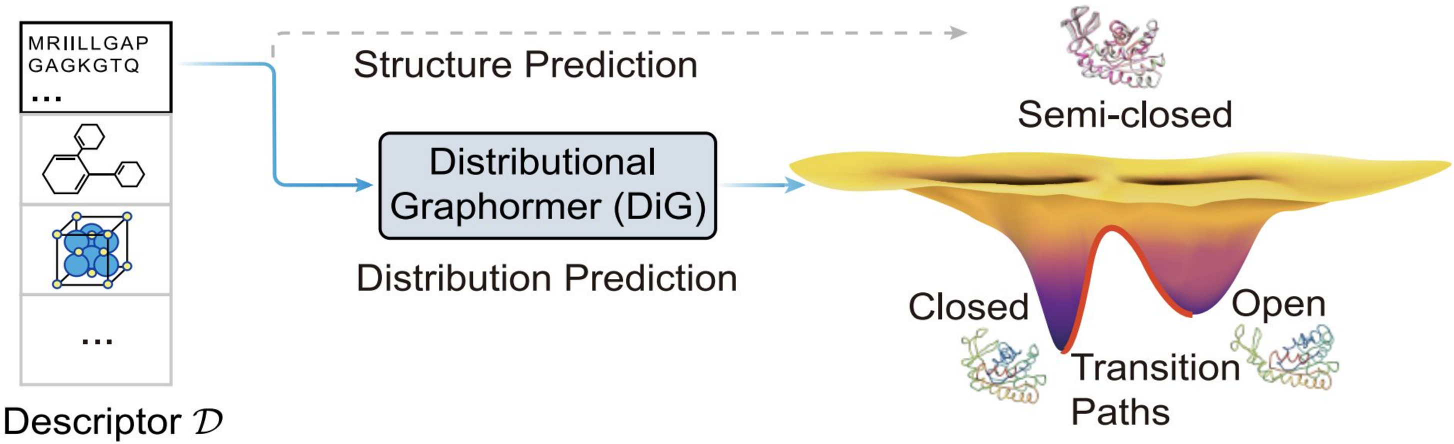
- Some scaling laws can be derived



N Frey, R Soklaski, S Axelrod, S Samsi, R Gomez-Bombarelli, C Coley, V Gadepally, *Neural Scaling of Deep Chemical Models,* 2022.

What is the best pre-training objective?

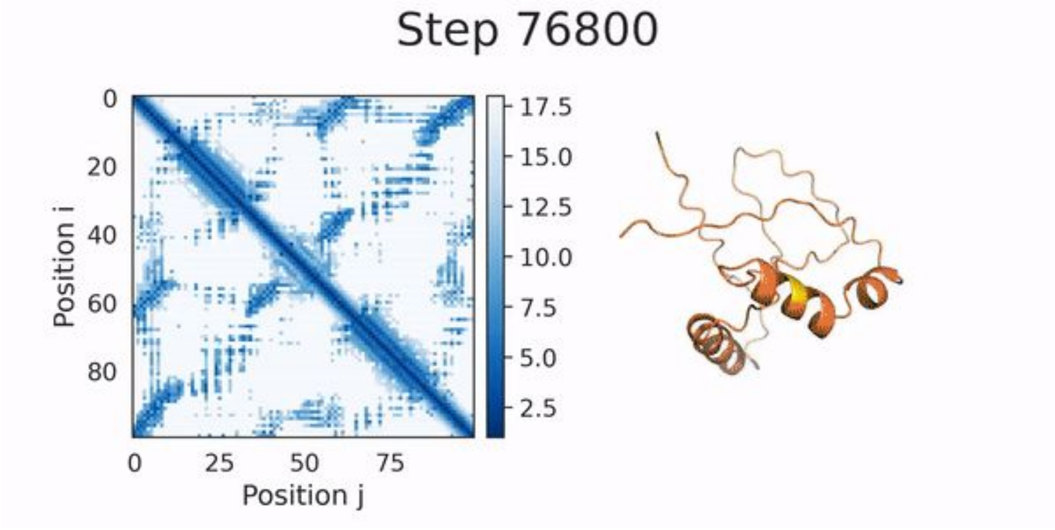
Distributional Graphormer [Frey et al., 2022]
Input: 3D structures (molecules, proteins, crystals)
Output: Equilibrium energy distribution + nice generative model



Zheng et al, *Distributional Graphormer: Towards Predicting Equilibrium Distributions for Molecular Systems with Deep Learning*, 2023.

Proteins: ESM-2 as a Foundation Model

ESM-2, ESMFold [Lin et al., 2022]
MLM on protein sequences
Bonus: 3D structure (folding) emerges from LM representations!



ESM Fold https://github.com/facebookresearch/esm

Lin, Akin, Rao, Hie et al, *Language models of protein sequences at the scale of evolution enable accurate structure prediction,* 2022.

Proteins: ESM-2 as a Foundation Model

ESM-2, ESMFold [Lin et al., 2022]
MLM on protein sequences
Bonus: 3D structure (folding) emerges from LM representations!

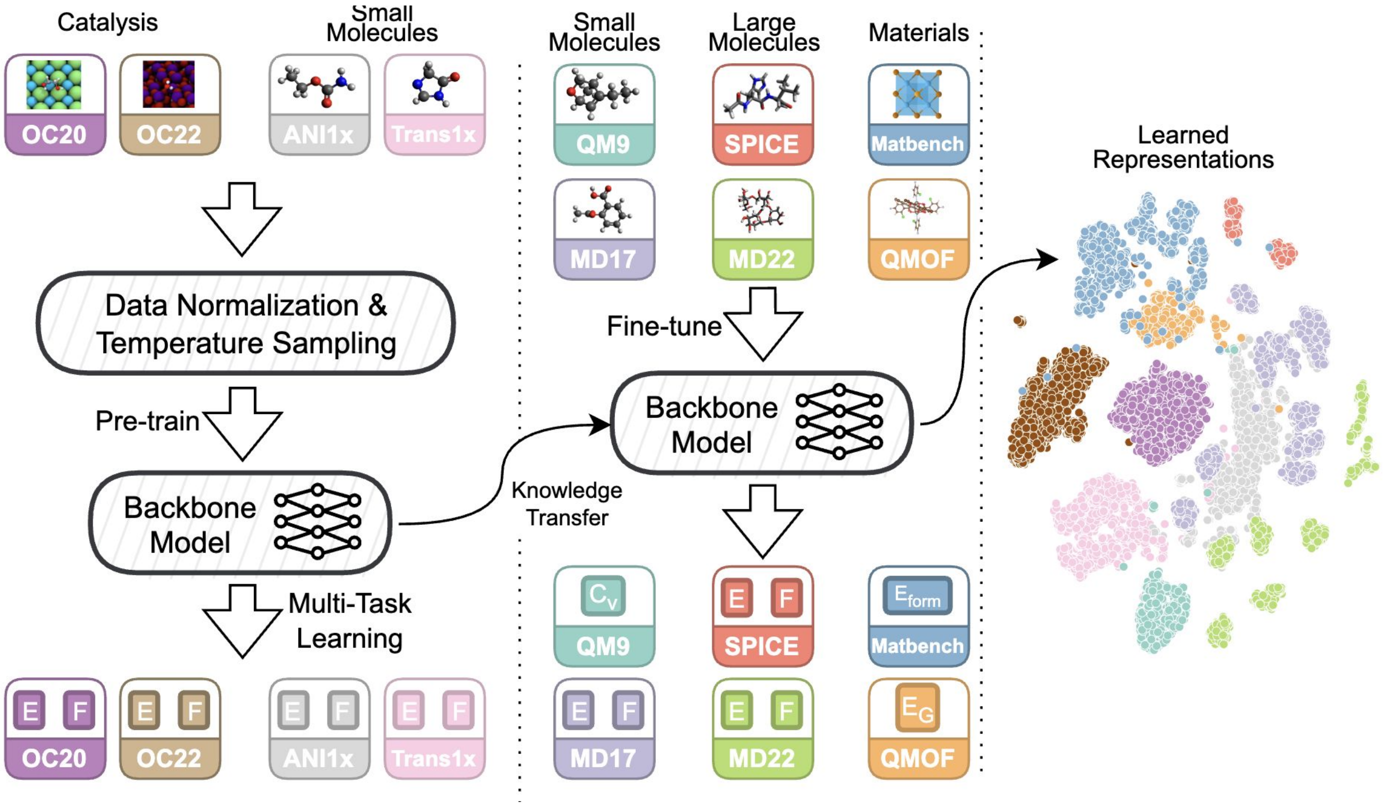
ESM-2 embeddings are used in a variety of protein models:

- **DiffDock** [Corso et al, ICLR 2023] - a diffusion model for protein-ligand docking

- **ProtST** [Xu, Yuan, et al, ICML 2023 Oral] - text-to-protein retrieval

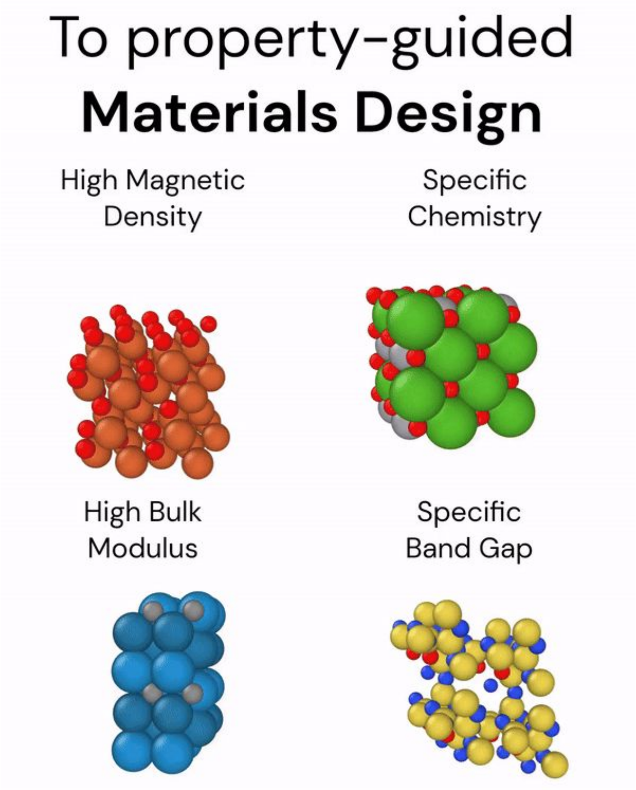
Corso et al, *DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking,* ICLR 2023
Xu, Yuan, et al, *ProtST: Multi-Modality Learning of Protein Sequences and Biomedical Texts,* ICML 2023.

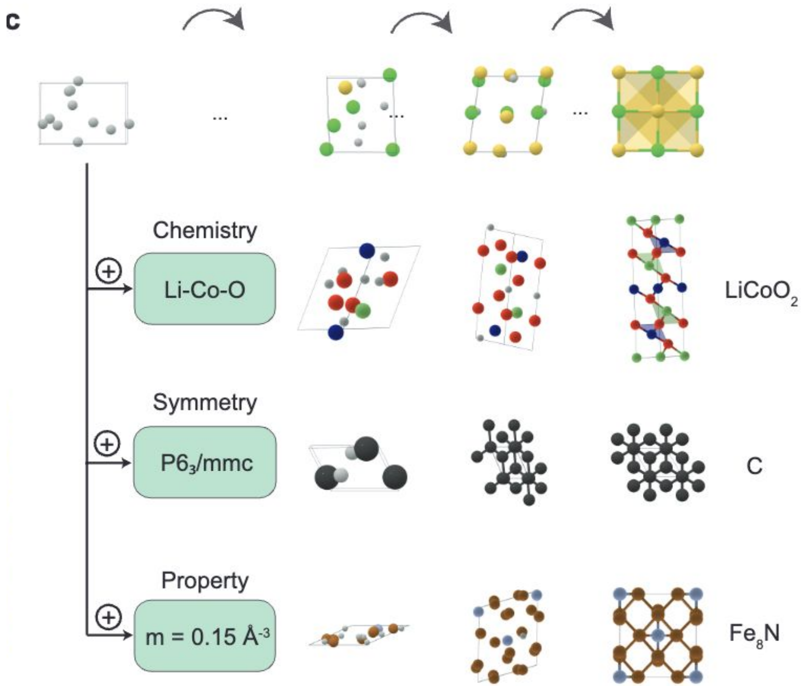
JMP-1, DPA-2: Geometric GNNs for Molecules and Crystals



Shoghi et al. From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction, ICLR 2024
Zhang et al. DPA-2: Towards a universal large atomic model for molecular and material simulation. Arxiv 2023

MatterGen: a conditional generative model for materials

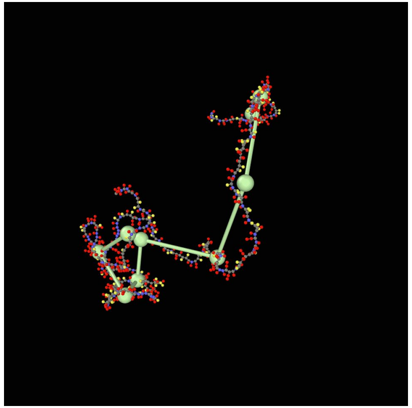


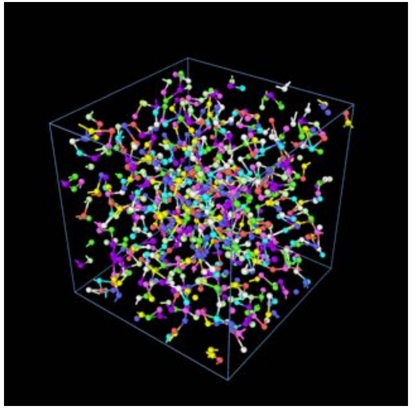


Zeni et al. MatterGen: a generative model for inorganic materials design, arxiv 2023

Molecular Dynamics Simulations (MD)

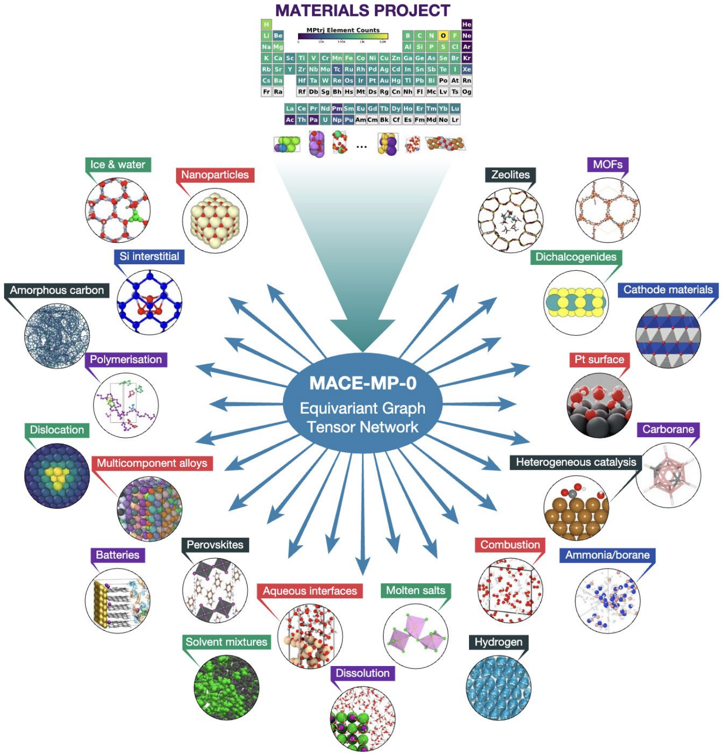
- *aka* ML potentials, ML force fields
- Predict how a structure changes over time
  - eg, atoms 3D coordinates
  - you'd need to obtain energy, forces, acceleration, and integrate over the desired time period
- Can be applied to molecules, proteins, crystals, and materials in general
- Classic models: slow
  ML models: fast but no silver bullet





Fu et al. *Simulate Time-integrated Coarse-grained Molecular Dynamics with Multi-scale Graph Networks*. TMLR 2023

MACE MP-0 and MatterSim: foundational MD models



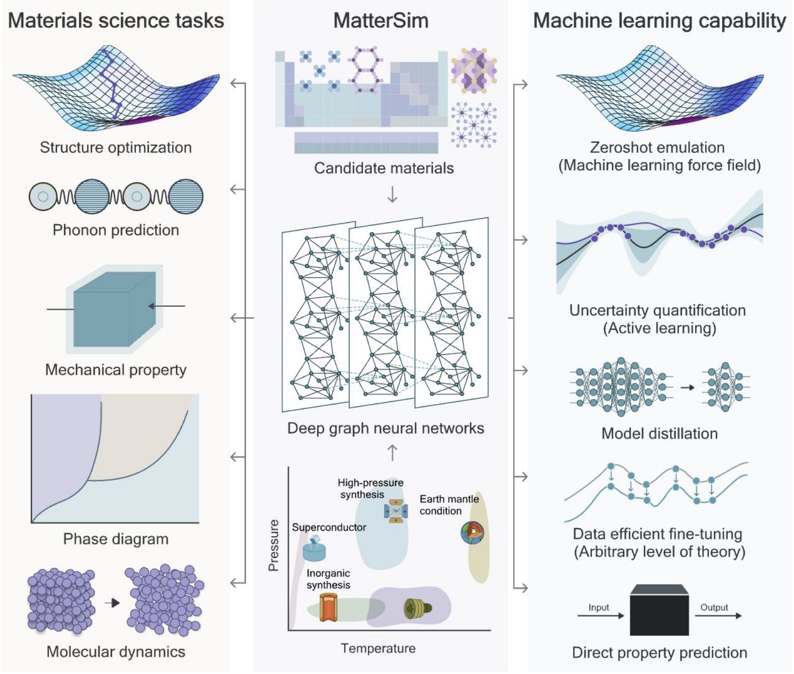


Fig. 1: **MatterSim is a deep learning atomistic model for predicting materials properties with high predictive accuracy across chemical elements, temperatures and pressures, enabling a wide range of applicability and functionality.**

Batatia et al. A foundation model for atomistic materials chemistry, arxiv 2024
Yang et al. MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures. arxiv 2024

Back to Materials and Crystals

Open MatSci ML Toolkit : A Broad, Multi-Task Benchmark for Solid-State Materials Modeling

TMLR Open MatSciML Toolkit | OpenReview AI4Mat 2022 HPO | Lightning v1.8.6+ | PyTorch v1.12+ | DGL v0.9+ | PyG 2.3.1 | License MIT

https://github.com/IntelLabs/matsciml

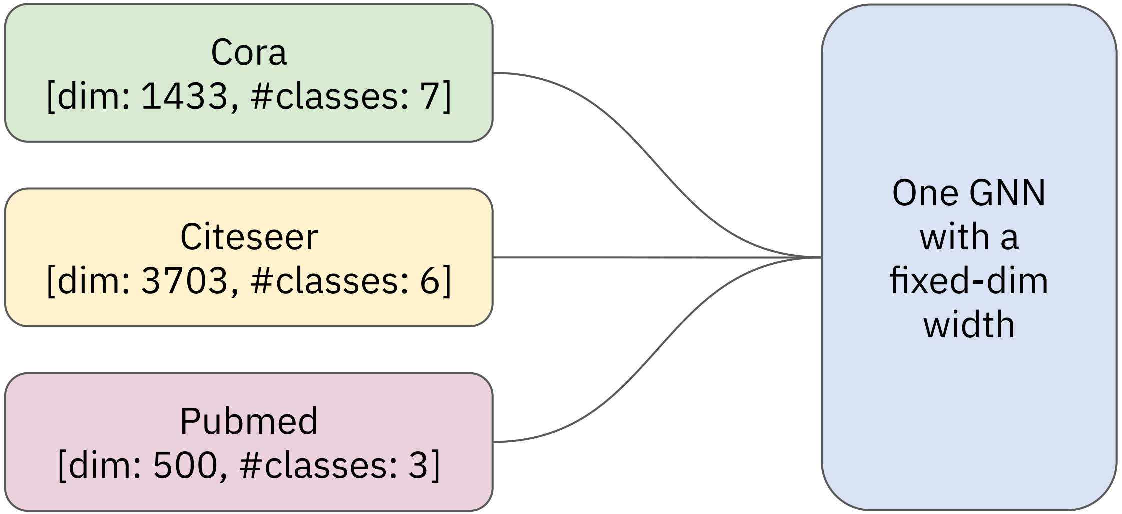
Announcement Blog Post (Oct 9th)

- 6 datasets (1.5M materials)
- 3 baseline models
- Many training tasks incl. generative pipeline

Miret, Lee, Gonzales, Nassar, Spellings. *The Open MatSci ML Toolkit: A Flexible Framework for Machine Learning in Materials Science*. TMLR, 2023.
Lee, Gonzales, Nassar, Spellings, Galkin, Miret. *MatSciML: A Broad, Multi-Task Benchmark for Solid-State Materials Modeling*. 2023

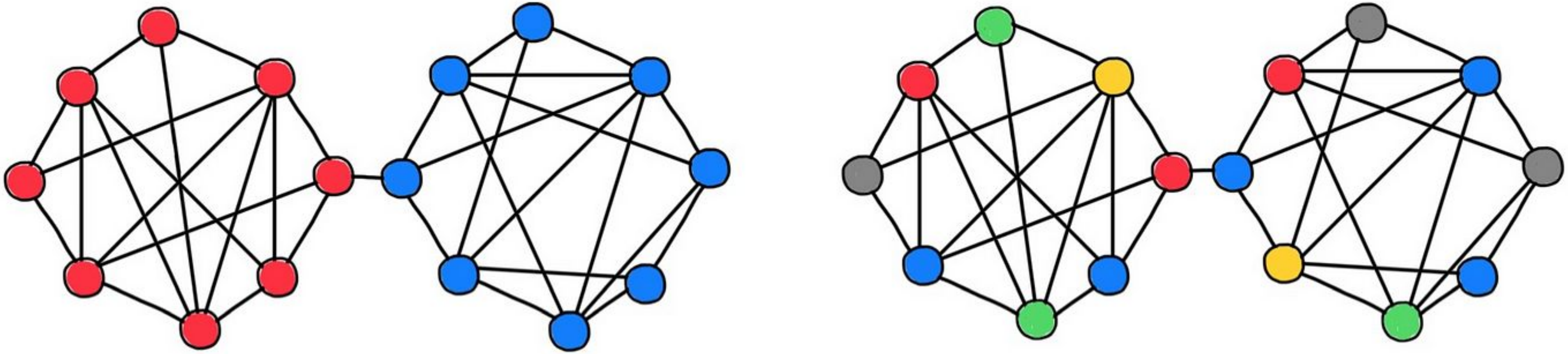
A single model for node classification?

- Different feature dimensions and # of labels



A single model for node classification?

- Different feature dimensions and # of labels
- Homophilic and heterophilic graphs exhibit different inductive biases
  - Homophilic like label propagation
  - Heterophilic depend more on node features



Michael Bronstein. Graph Neural Networks beyond Weisfeiler-Lehman and vanilla Message Passing, Medium 2022

A single model for node classification?

- Different feature dimensions and # of labels
- Homophilic and heterophilic graphs exhibit different inductive biases
  - Homophilic like label propagation
  - Heterophilic depend more on node features

Ideas?

🚀 > Run ULTRA on your own graph < 🚀

It’s only 177k params

Galkin et al. Towards Foundation Models for Knowledge Graph Reasoning, ICLR 2024

Mao, Chen, et al. Graph Foundation Models, ICML 2024 (new!)

Code & Data https://github.com/DeepGraphLearning/ULTRA

Contact mikhail.galkin@intel.com

Socials @michael_galkin