NEURAL INFORMATION
PROCESSING SYSTEMS

LOG
CONFERENCE

Mila

LOG
Meetup
at Mila

Plan for today

LOG 2022 accepted papers:

- Weisfeiler and Leman Go Relational
- Taxonomy of Benchmarks in Graph Representation Learning

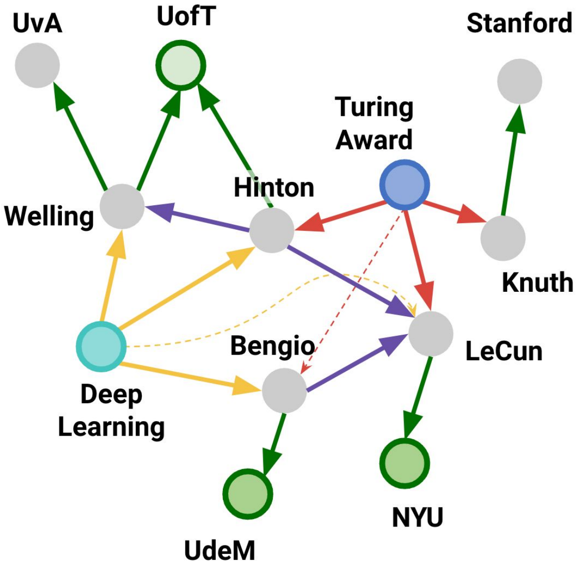
NeurIPS 2022 papers:

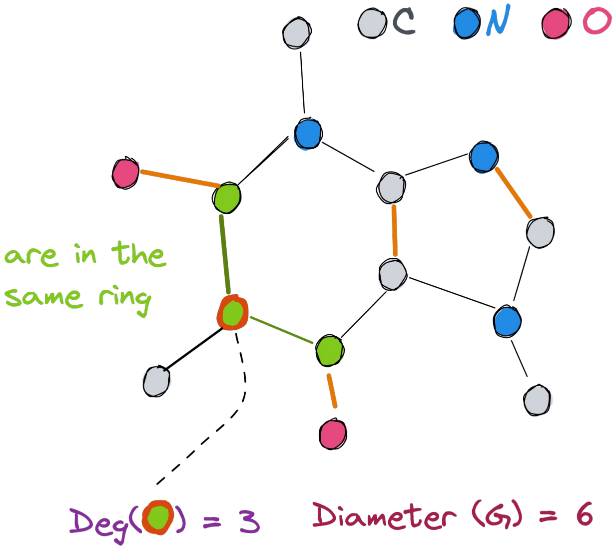
- Inductive Logical Query Answering in Knowledge Graphs
- A Recipe for a General, Powerful, and Scalable Graph Transformers
- Long-Range Graph Benchmark

Weisfeiler and Leman Go Relational

Pablo Barcelo, Mikhail Galkin, Christopher Morris, Miguel Romero Oorth

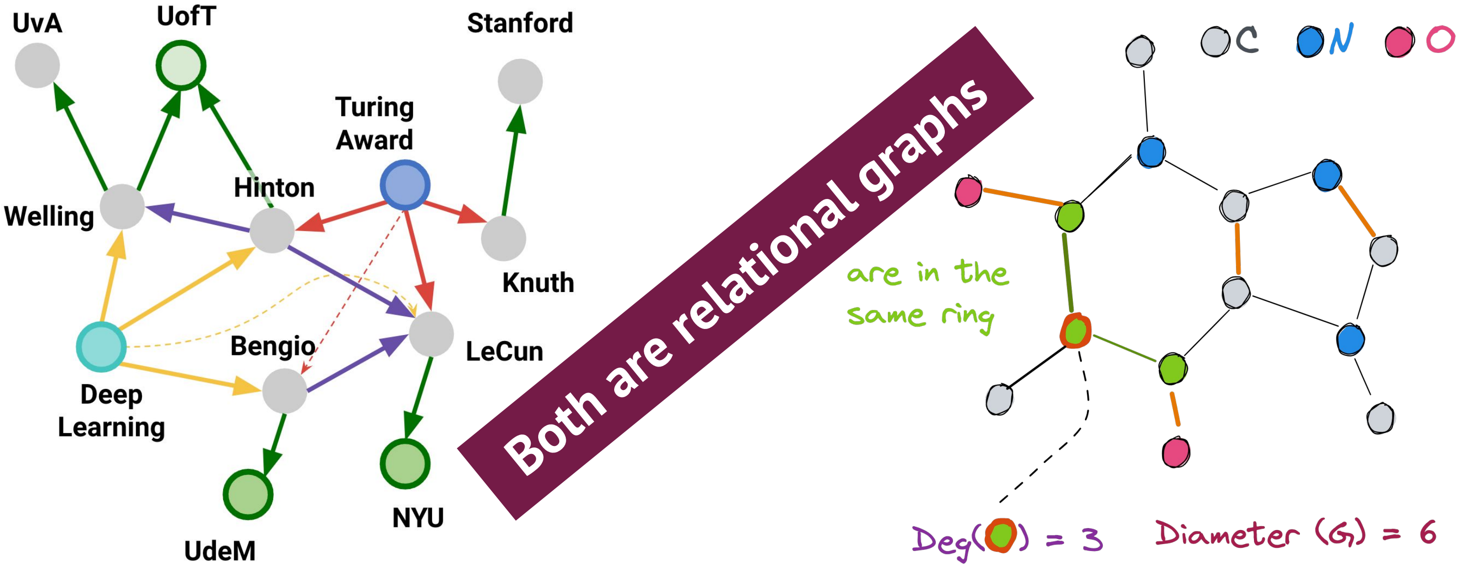
WL Go Relational





What's common between KG and molecular graph?

WL Go Relational



What's common between KG and molecular graph?

So how expressive are relational GNNs?

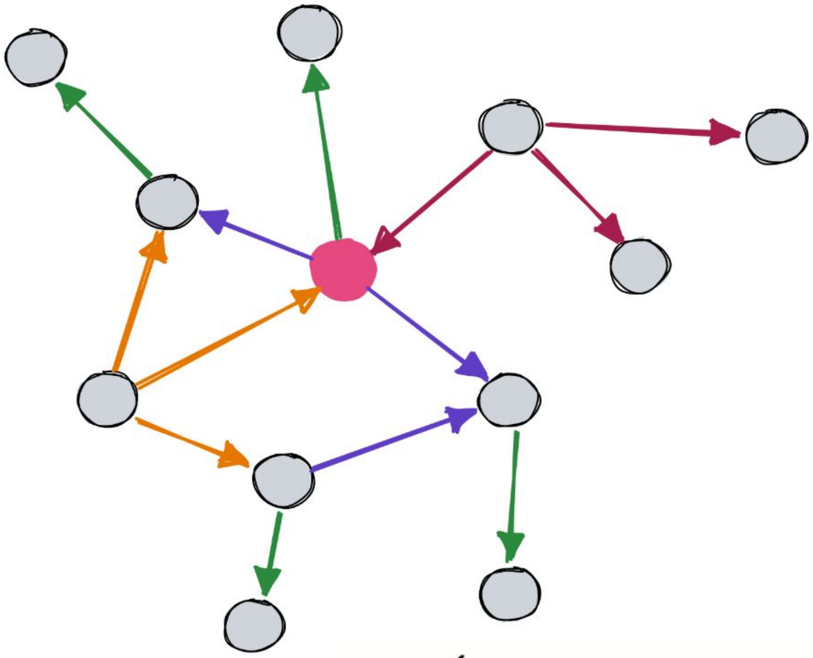
Some places our guys Weisfeiler and Leman have been to recently:

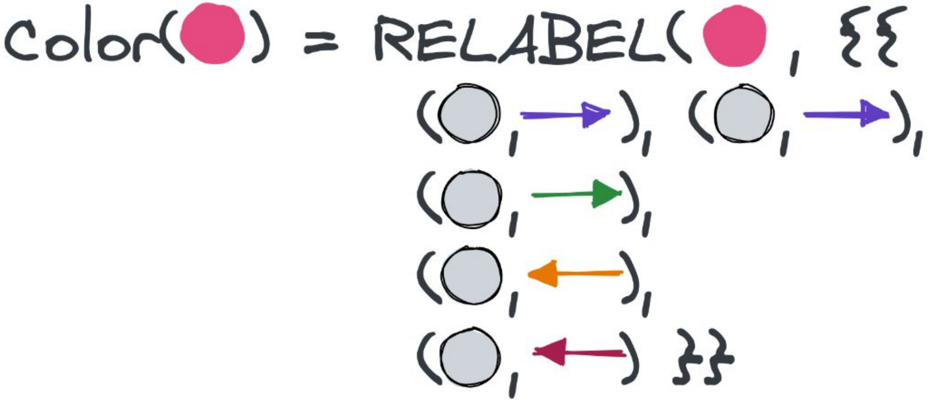
✅ Neural
✅ Sparse
✅ Topological
✅ Cellular
✅ Hyperbolic
✅ Infinite
❌ Relational :(- time to fix that!

A. Leman

B. Weisfeiler

Relational WL Iteration





$$\mathrm{RL}\left(\left(C_{\mathrm{R}}^{(t-1)}(v), \{\!\{(C_{\mathrm{R}}^{(t-1)}(u), i) \mid i \in [r], u \in N_i(v)\}\!\}\right)\right)$$

Relational WL Findings

✅ 1-RWL > 1-WL, provably

✅ Relational GCN (R-GCN) ≡ CompGCN and bounded by 1-RWL

✅ Multiplicative message functions is **the best** (generally, those that capture vector scaling)

Relational WL Findings

```python
# Current GINE Conv
def message(self, x_j: Tensor, edge_attr: Tensor) -> Tensor:

        if self.lin is not None:
            edge_attr = self.lin(edge_attr)

        return (x_j + edge_attr).relu()
```

❌ BAD

```python
# Best GINE Conv
def message(self, x_j: Tensor, edge_attr: Tensor) -> Tensor:

        if self.lin is not None:
            edge_attr = self.lin(edge_attr)

        return (x_j * edge_attr).relu()
```

👍 GOOD

Taxonomy of Benchmarks in Graph Representation Learning

Learning on Graphs (LoG) 2022

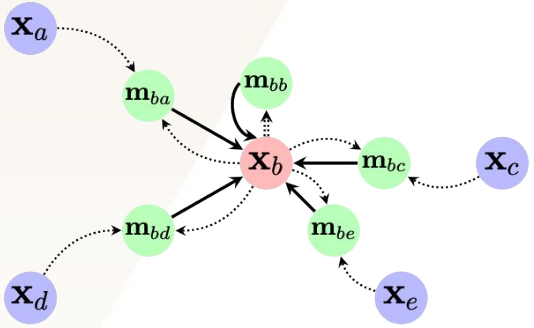
Renming Liu[3], **Semih Cantürk**[1,2],
Frederik Wenkel[1,2], Sarah McGuire[3], Xinyi Wang[3],
Anna Little[4], Leslie O'Bray[5], Michael Perlmutter[6], Bastian Rieck[7],
Matthew Hirn[3], Guy Wolf[1,2], and Ladislav Rampášek[1,2]

[1]Mila - Quebec AI Institute, [2]Université de Montréal, [3]Michigan State University, [4]University of Utah,
[5]ETH Zürich, [6]University of California, Los Angeles, [7]Helmholtz Zentrum München

Illustration credit: Bronstein, M.M., Bruna, J., Cohen, T. and Veličković, P., 2021. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv:2104.13478.

Motivation

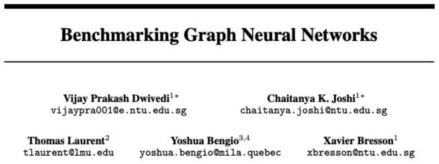
- Graph Neural Network (GNN) development is a hot topic!
  - GCN, GAT, GraphSAGE, GIN…
  - Recently: Graph Transformers, k-GNNs…



- With emerging collections of benchmarks:

TUDataset

OGB
OPEN GRAPH BENCHMARK


Benchmarking Graph Neural Networks

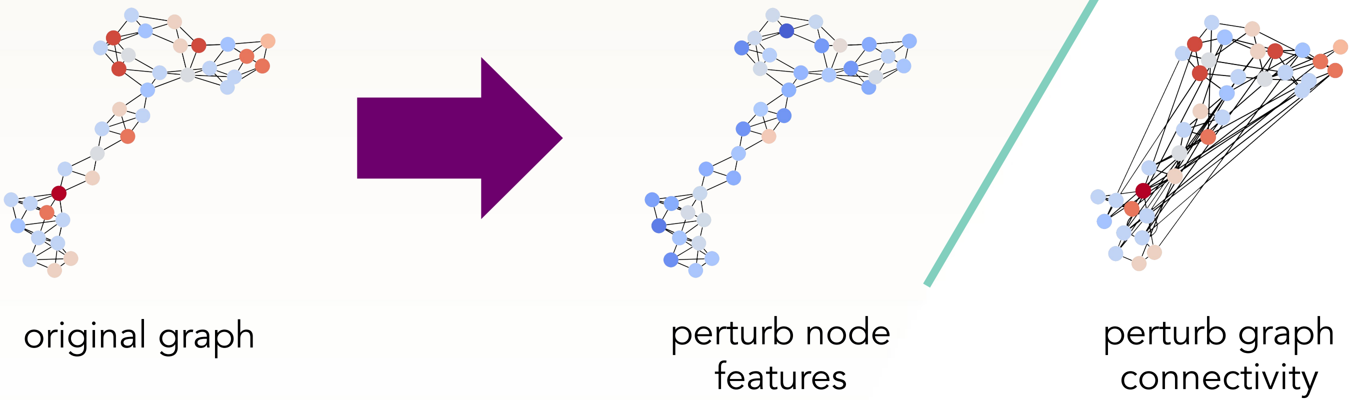
Vijay Prakash Dwivedi[1*] vijaypra001@e.ntu.edu.sg
Chaitanya K. Joshi[1*] chaitanya.joshi@ntu.edu.sg
Thomas Laurent[2] tlaurent@lmu.edu
Yoshua Bengio[3,4] yoshua.bengio@mila.quebec
Xavier Bresson[1] xbresson@ntu.edu.sg

- But what aspects of GNNs are actually tested by these?

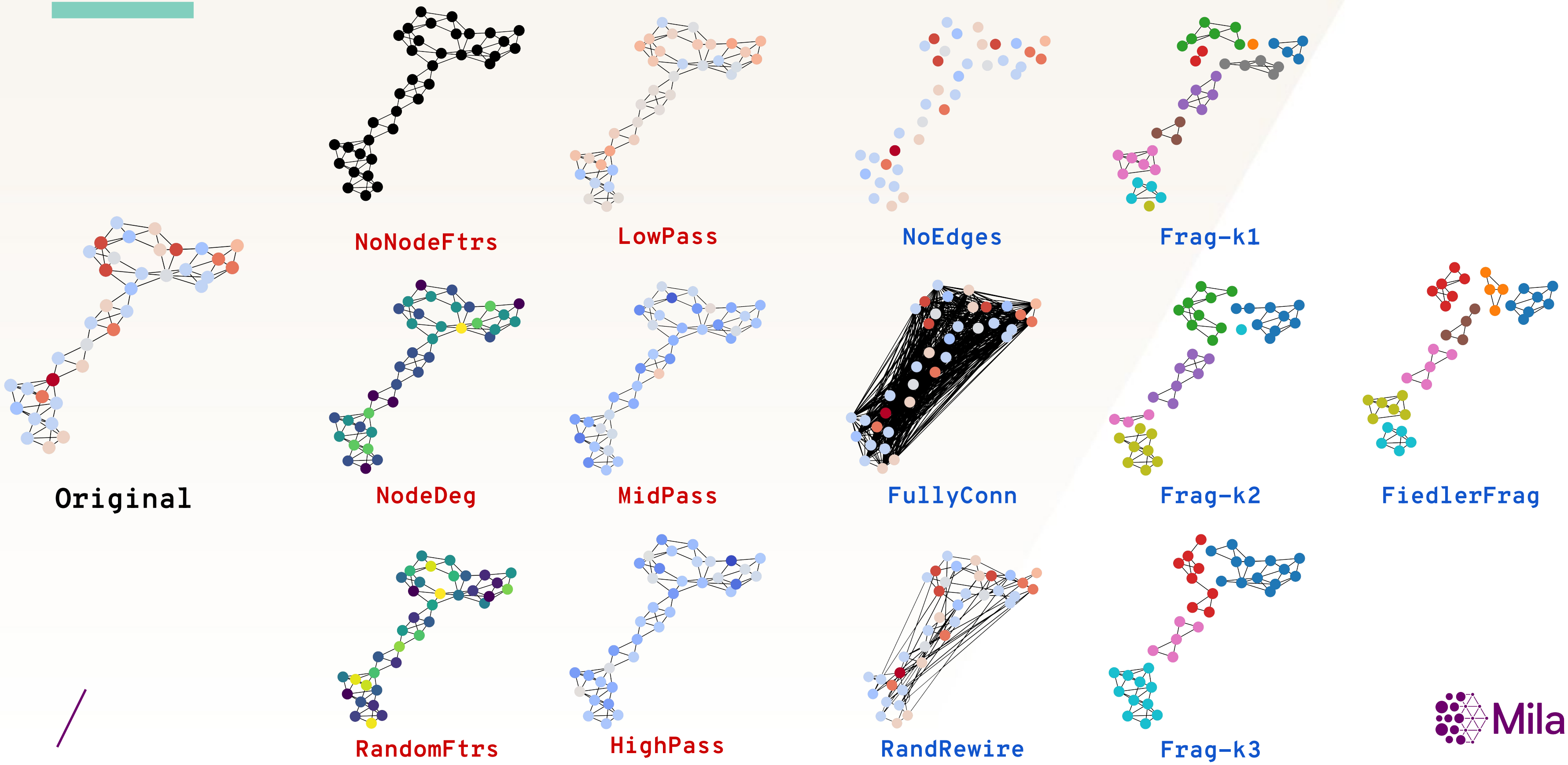
Mila

Approach

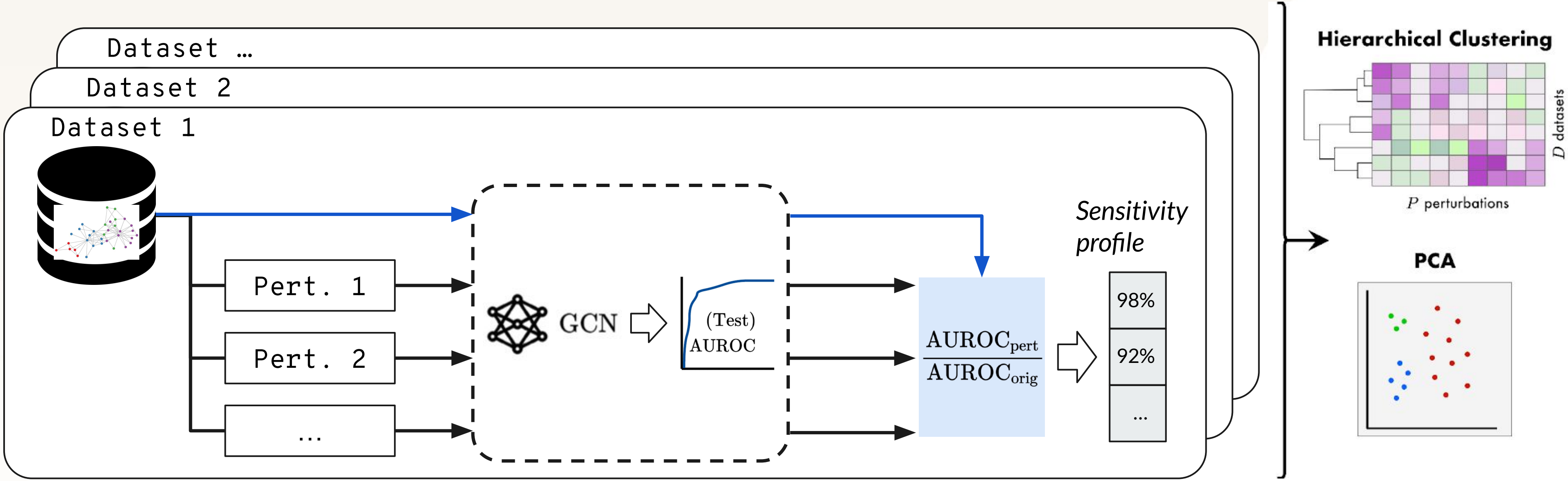
- Empirically study specific transformation sensitivity to gauge *how* task-related information is encoded in graph datasets:
    1. Perturb graph dataset to alter node-features or graph connectivity in a specific way



Perturbations - feature (6) vs. structure (7)



Taxonomy Framework



Key idea: Gauge *how* task-related information is encoded in graph datasets by empirically studying perturbation sensitivity and generate "fingerprints"

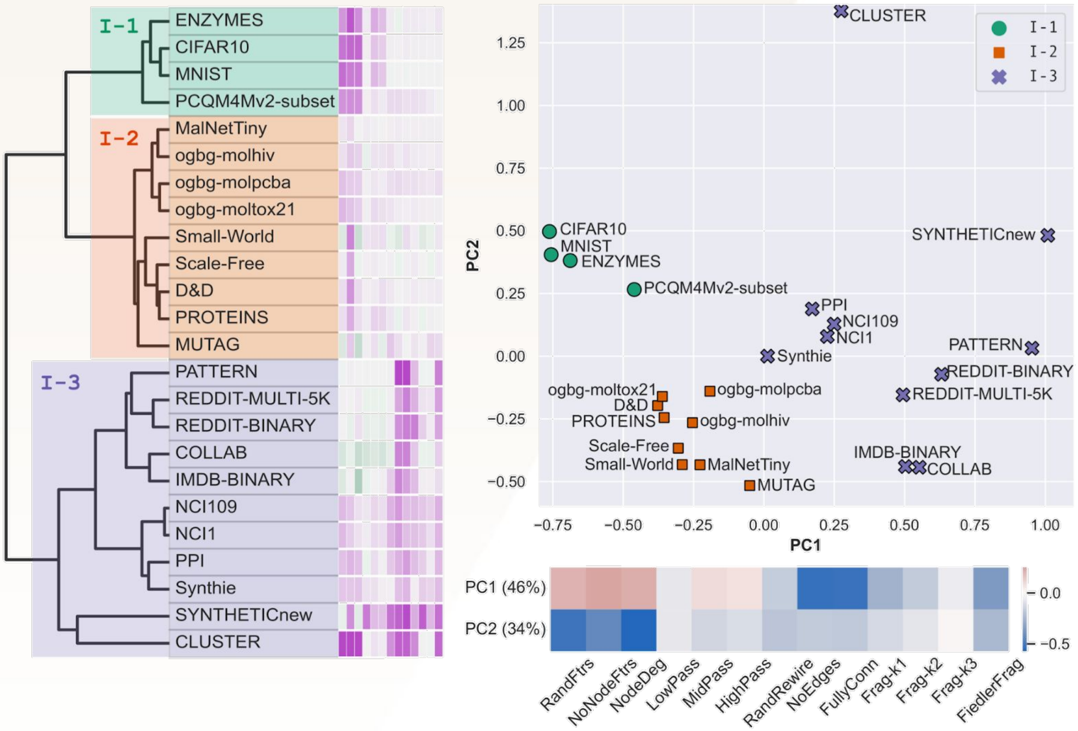
Datasets

- 49 datasets (24 inductive, 25 transductive)
  - Node- vs. Graph-level tasks
  - Inductive vs. Transductive
  - Real-world vs. Synthetic
  - Homophilic vs. Heterophilic
- Multiple domains: Biochemistry, image data, social graphs, collaboration graphs, citation & web graphs
- Dataset & graph sizes both ranging from $\sim 10^1$ to $\sim 10^5$

Results: Inductive Tasks

Three main clusters:

- I-1 is sensitive to node feature perturbations
- I-2 is robust to either type of perturbations
- I-3 is very sensitive to graph structure perturbations



Mila

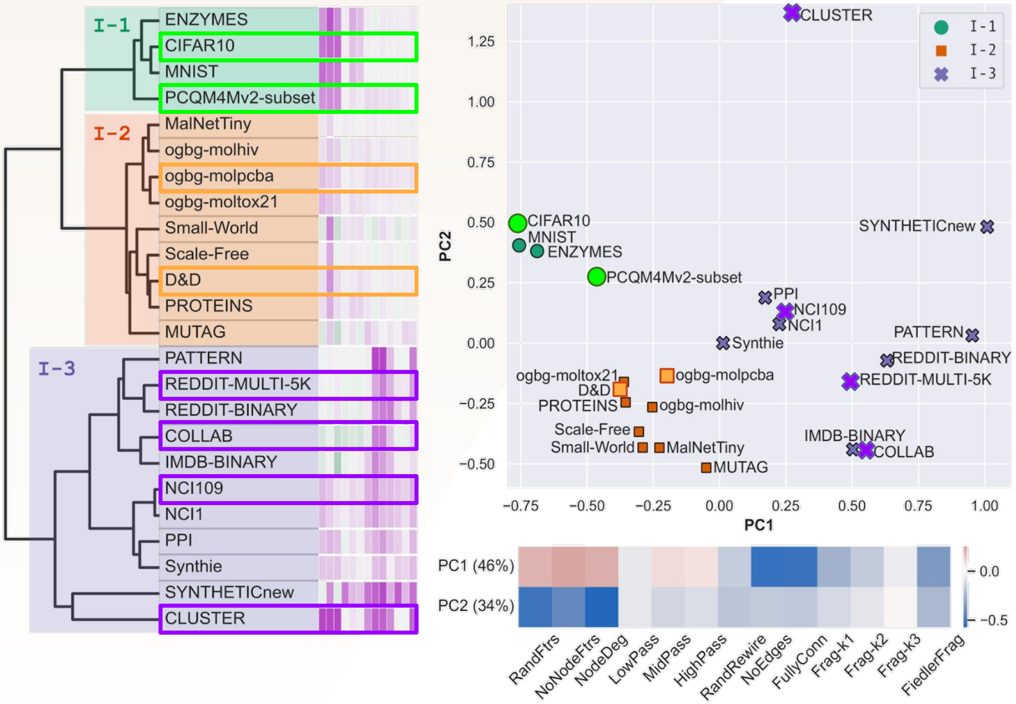
Results: Inductive Tasks

REPRESENTATIVE SET

- CIFAR10
- PCQM4Mv2-subset

- ogbg-molpcba
- D&D

- REDDIT-MULTI-5K
- COLLAB
- NCI1
- CLUSTER

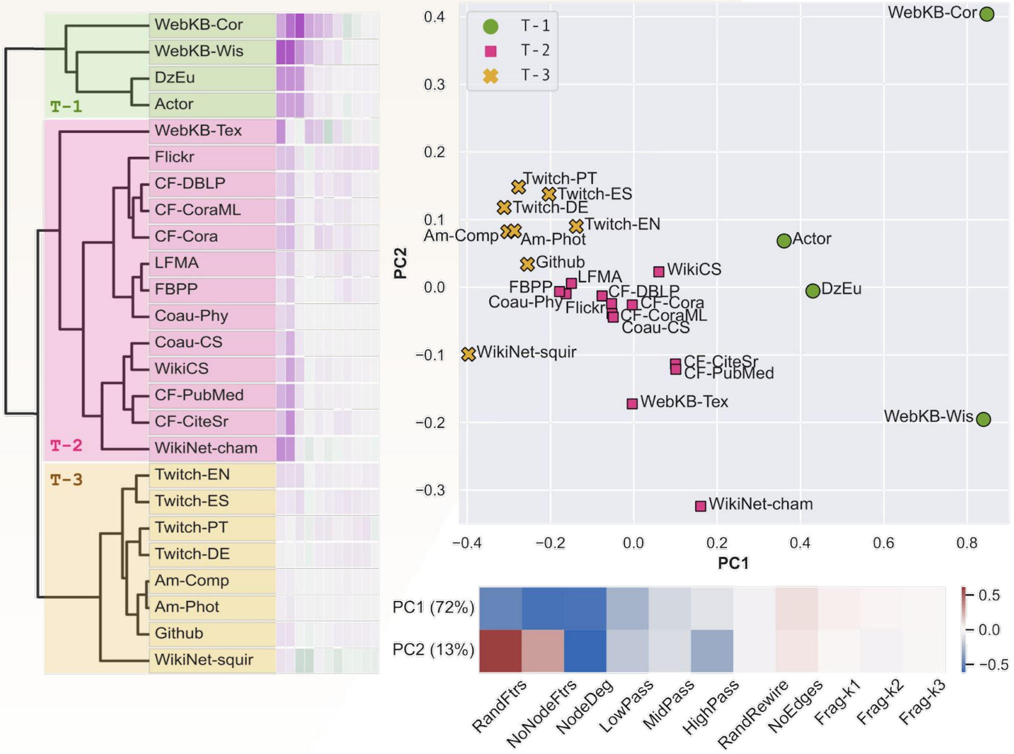


Mila

Results: Transductive Tasks

Three main clusters:

- **T−1** contains *heterophilic* datasets
- **T−2** relies strongly on node features
- **T−3** is robust to either type of perturbations

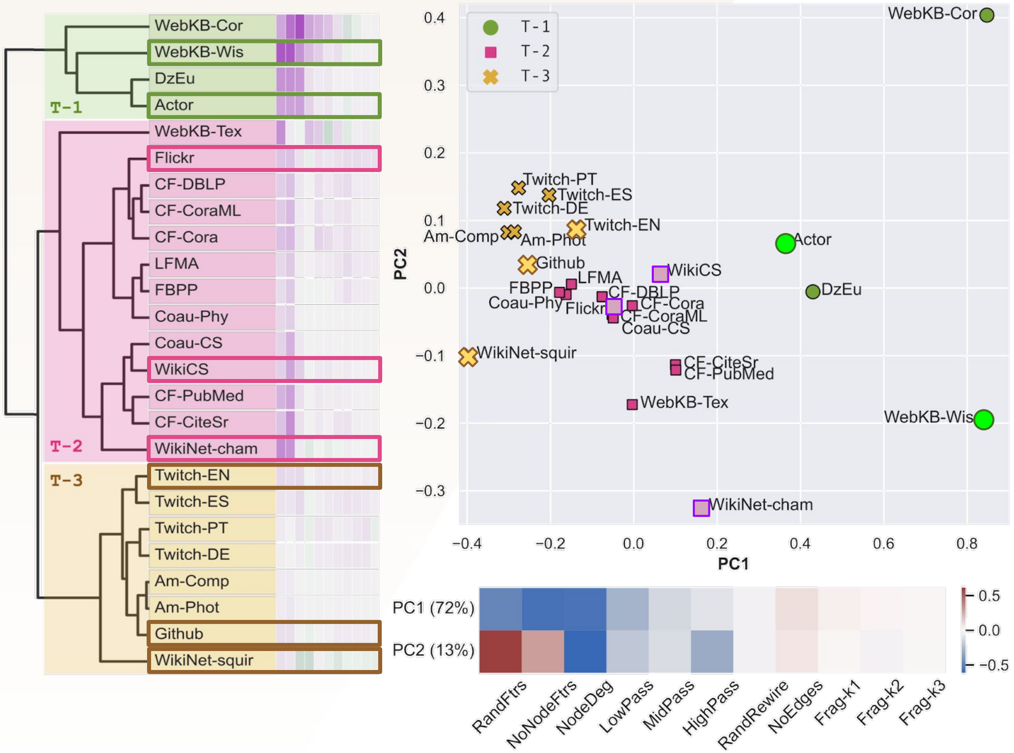


Mila

Results: Transductive Tasks

REPRESENTATIVE SET

- WebKB-Wisconsin
- Actor
- Flickr
- WikiCS
- WikiNet-chameleon
- Twitch-EN
- GitHub
- WikiNet-squirrel



Mila

NEURAL INFORMATION PROCESSING SYSTEMS

35th Conference on Neural Information Processing Systems 2022

Inductive Logical Query Answering in Knowledge Graphs



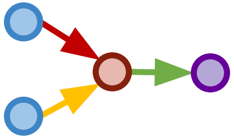

Mikhail Galkin[1,2] Zhaocheng Zhu[1,3] Hongyu Ren[4] Jian Tang[1,5]

[1]Mila, [2]McGill University, [3]Unversité de Montreal, [4]Stanford University, [5]HEC Montreal

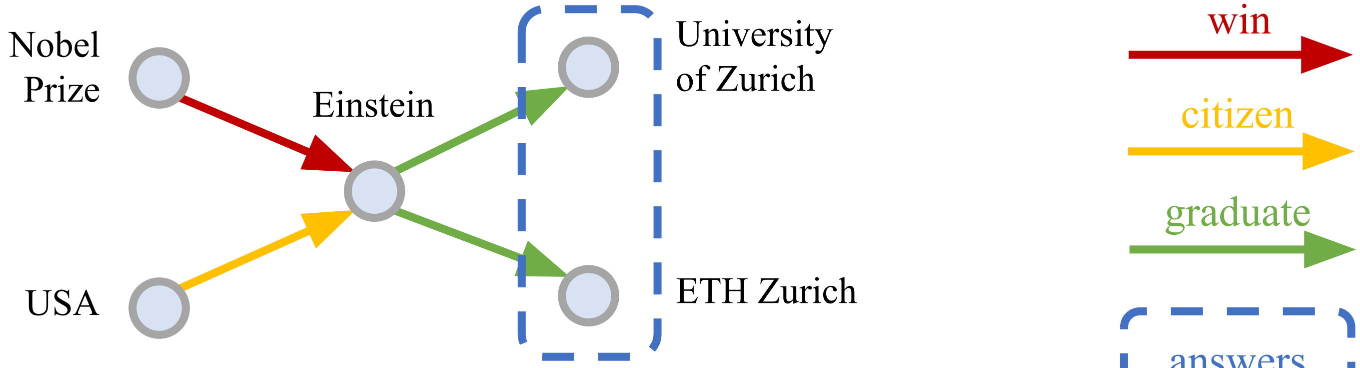
Query Answering in KGs

Where did US citizens with Nobel Prize graduate from?

$$q = ?U.\exists V : \text{Win}(\text{NobelPrize}, V) \wedge \text{Citizen}(\text{USA}, V) \wedge \text{Graduate}(V, U)$$



Variables **Constants** **Projections R(a,b)** **Logical Operators $\wedge, \vee, \neg$**

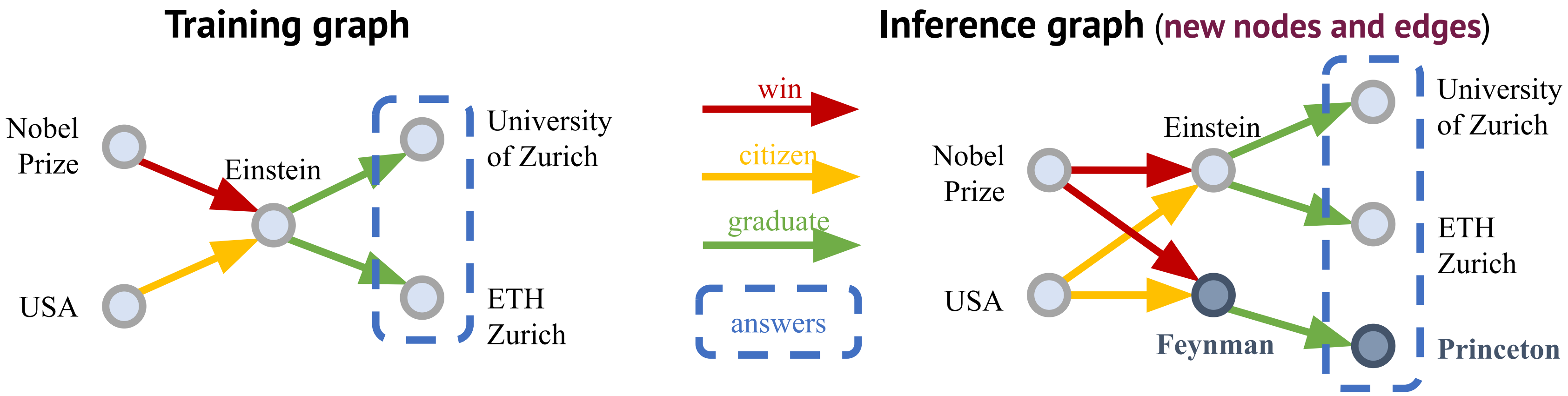


Training Graph

Inductive Query Answering

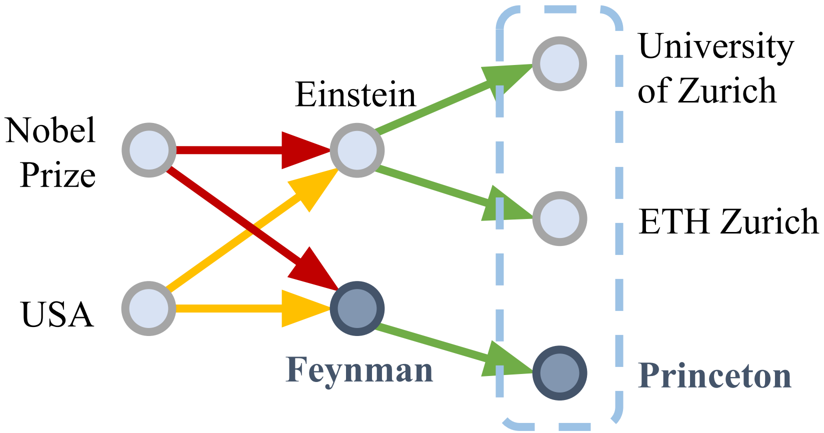
The same query executed against a new graph with new nodes and edges

$$q = ?U.\exists V : \text{Win}(\text{NobelPrize}, V) \boxed{\land} \text{Citizen}(\text{USA}, V) \boxed{\land} \text{Graduate}(V, U)$$



New correct answers at inference time

Setup



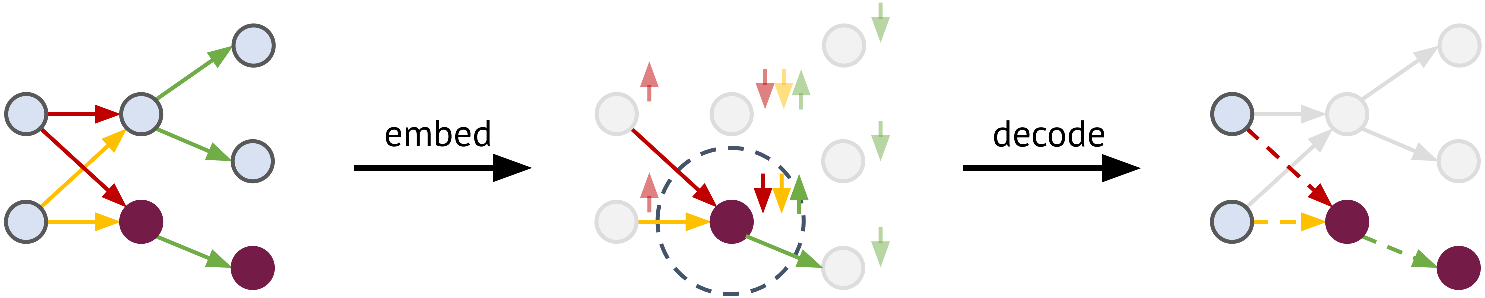
Any model pipeline typically needs input features:

$$X' = \mathrm{GNN}(X, A, W)$$

- Input node features are not given
- Learning shallow node embeddings is useless for unseen inference nodes
- How to get inductive features?

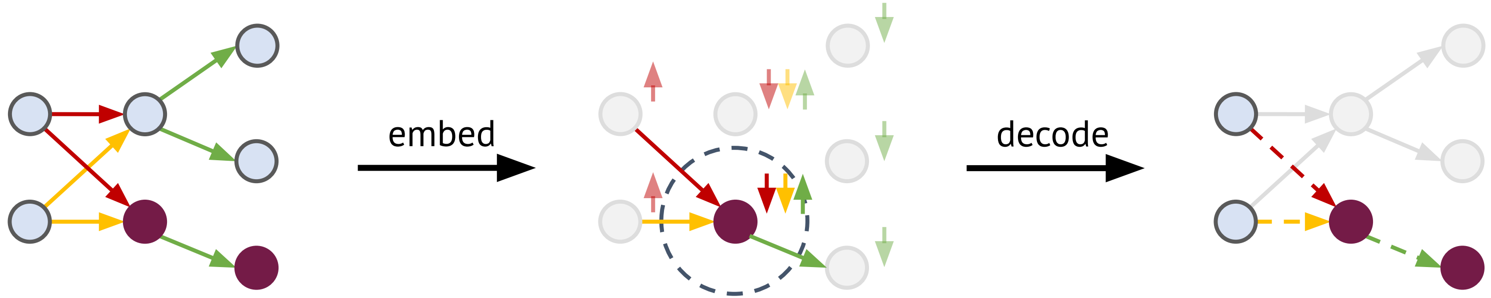
The Essence of Inductiveness

1. Inductive representations of each **entity**

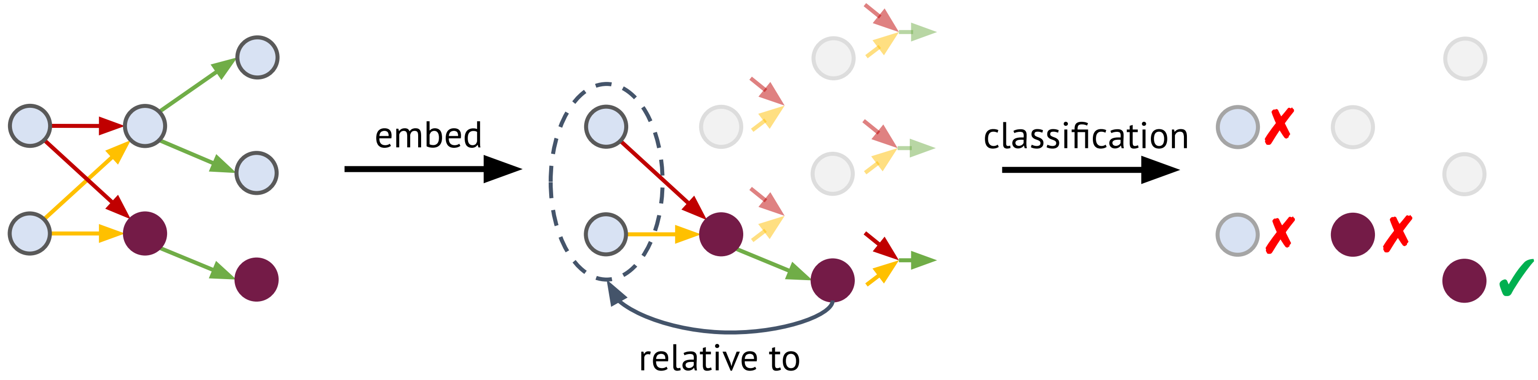


The Essence of Inductiveness

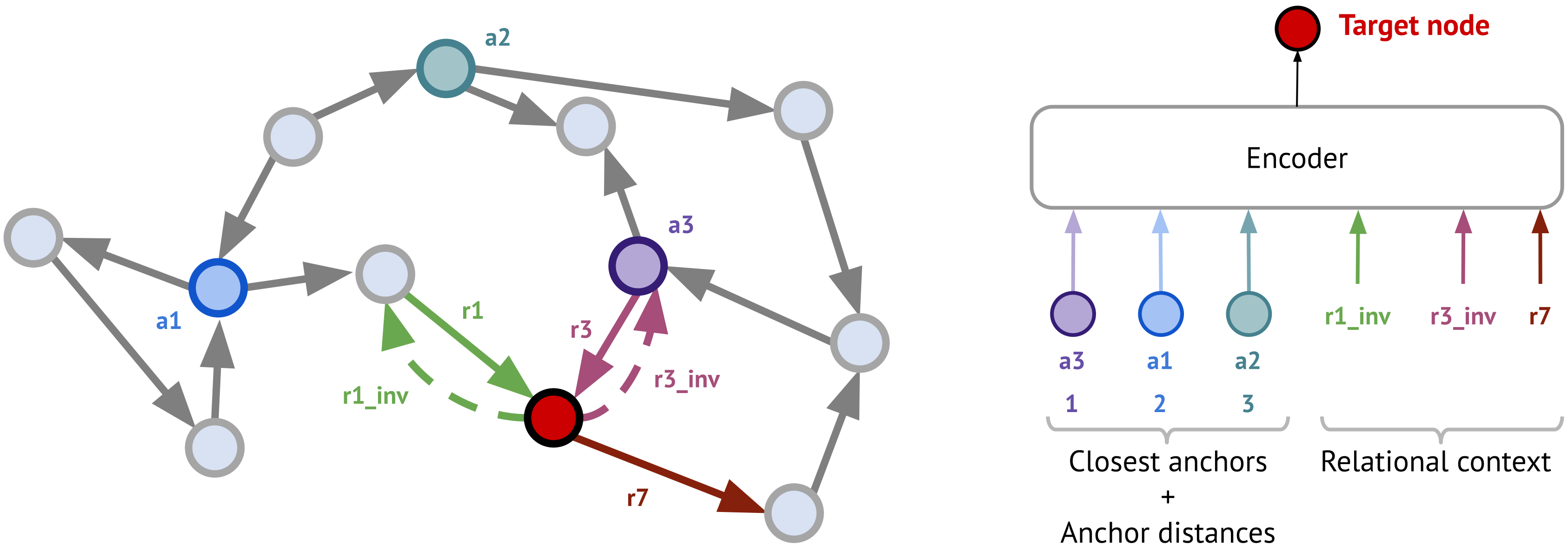
1. Inductive representations of each **entity**



2. Inductive representations of **the relative relational structure**



NodePiece - "*subword units*" for KGs



Vocabulary = Anchors + Relation types

Inductive out-of-the-box: unseen nodes are "tokenized" with the same Vocab

Galkin et al. NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs. ICLR 2022

Leaderboard for ogbl-wikikg2

The MRR score on the test and validation sets. The higher, the better.

Package: >=1.2.4

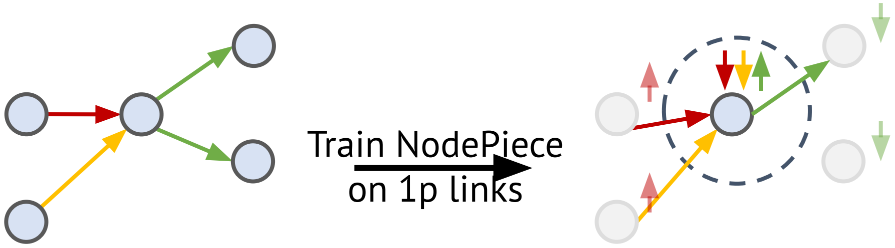
Deprecated `ogbl-wikikg` leaderboard can be found here.

Rank	Method	Ext. data	Test MRR	Validation MRR	Contact	References	#Params	Hardware	Date
1	**StarGraph + TripleRE**	No	0.7201 ± 0.0011	0.7288 ± 0.0008	Hongzhu Li (360AI)	Paper, Code	86,762,146	Tesla A100(40GB)	May 30, 2022
2	**TranS**	No	0.6939 ± 0.0011	0.7058 ± 0.0018	Xuanyu Zhang (DXM AI)	Paper, Code	38,430,804	Tesla V100 (16GB)	Apr 19, 2022
	TranS	No	0.6882 ± 0.0019	0.6988 ± 0.0006	Xuanyu Zhang (DXM AI)	Paper, Code	19,215,402	Tesla V100 (16GB)	Apr 28, 2022
4	**TripleRE + NodePiece**	No	0.6866 ± 0.0014	0.6955 ± 0.0008	Long Yu (360AI)	Paper, Code	36,421,802	Tesla A100(40GB)	Feb 24, 2022
5	**InterHT**	No	0.6779 ± 0.0018	0.6893 ± 0.0015	Baoxin Wang (HFL)	Paper, Code	19,215,402	Tesla V100 (32GB)	Feb 10, 2022
6	**TripleRE + NodePiece**	No	0.6582 ± 0.0020	0.6616 ± 0.0018	Long Yu (360AI)	Paper, Code	7,289,002	Tesla A100(40GB)	Dec 25, 2021

All use NodePiece

Galkin et al. NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs. ICLR 2022

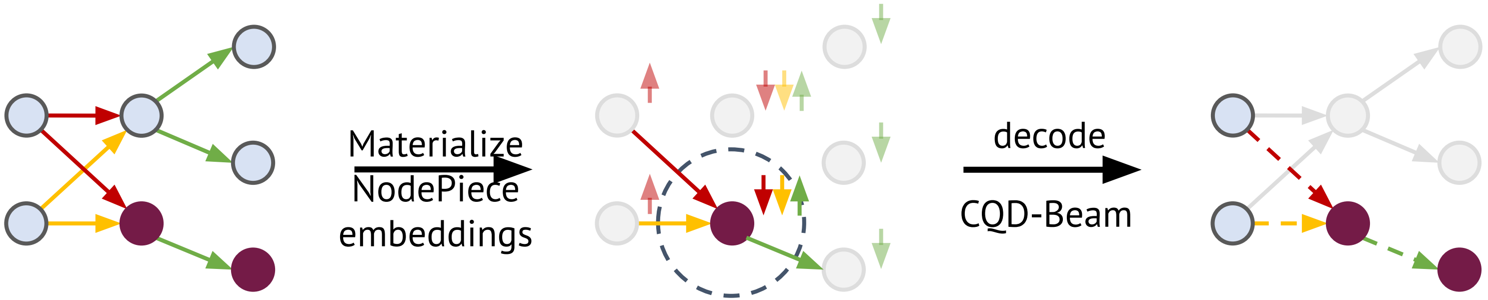
Inductive NodePiece-QE



1. Projection operator: scoring function (ComplEx)

2. Logical operators: t-norms

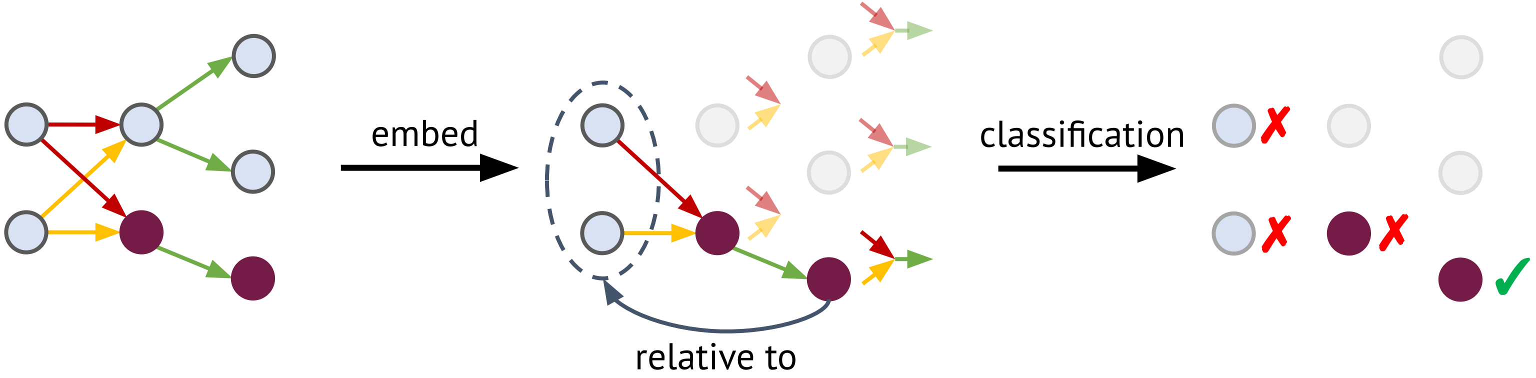
3. Inference-only decoder: non-parametric CQD-Beam



[1] Arakelyan et al. Complex Query Answering with Neural Link Predictors. ICLR 2021

The Essence of Inductiveness

2. Inductive representations of **the relative relational structure**

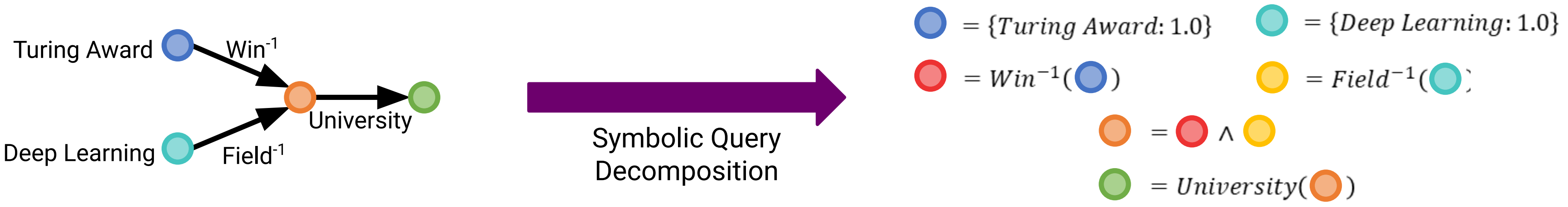


In simple link prediction, such inductive representations are studied by NBFNet.

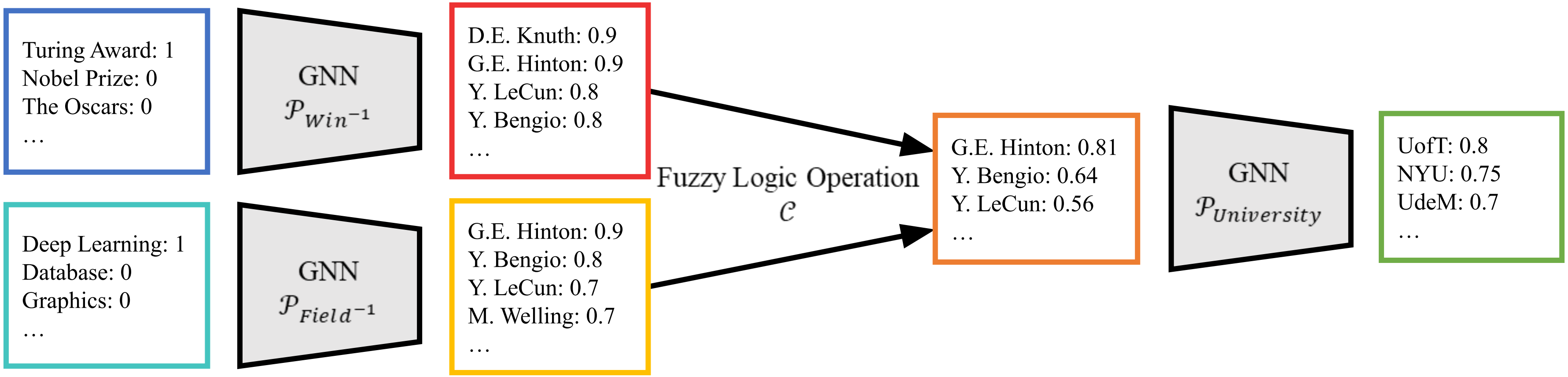
How to extend NBFNet to inductive complex queries?

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

GNN-QE: NBFNet + T-norms



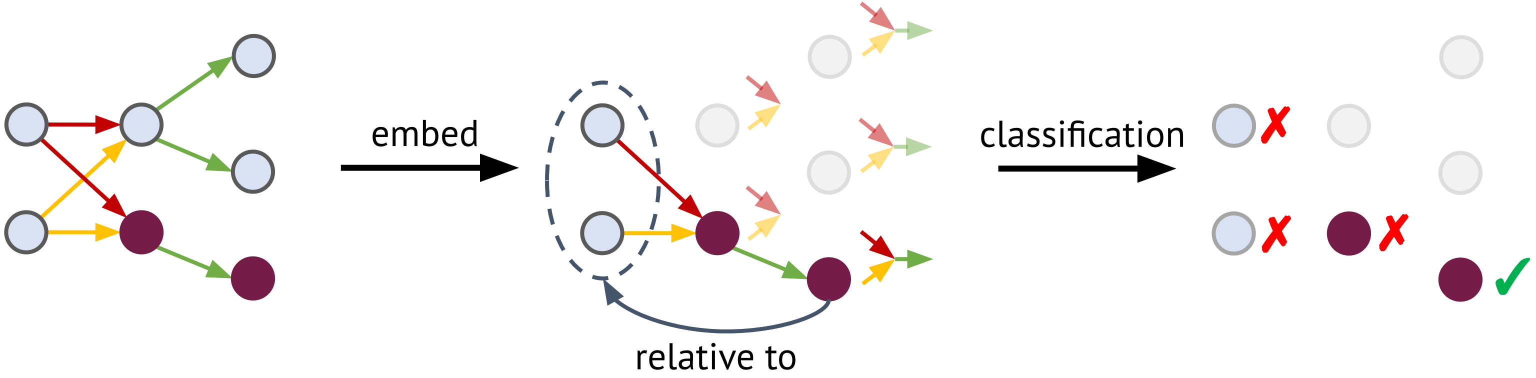
Each variable is a fuzzy set of entities, where each element in the set has a probability.



Zhu et al. Neural-Symbolic Models for Logical Queries on Knowledge Graphs. ICML 2022

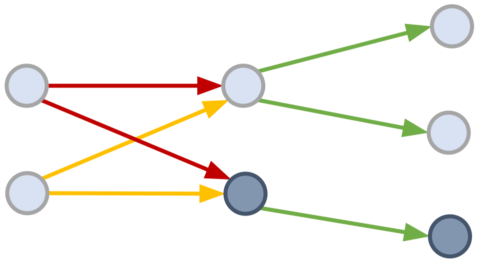
Inductive GNN-QE

2. Inductive representations of **the relative relational structure**



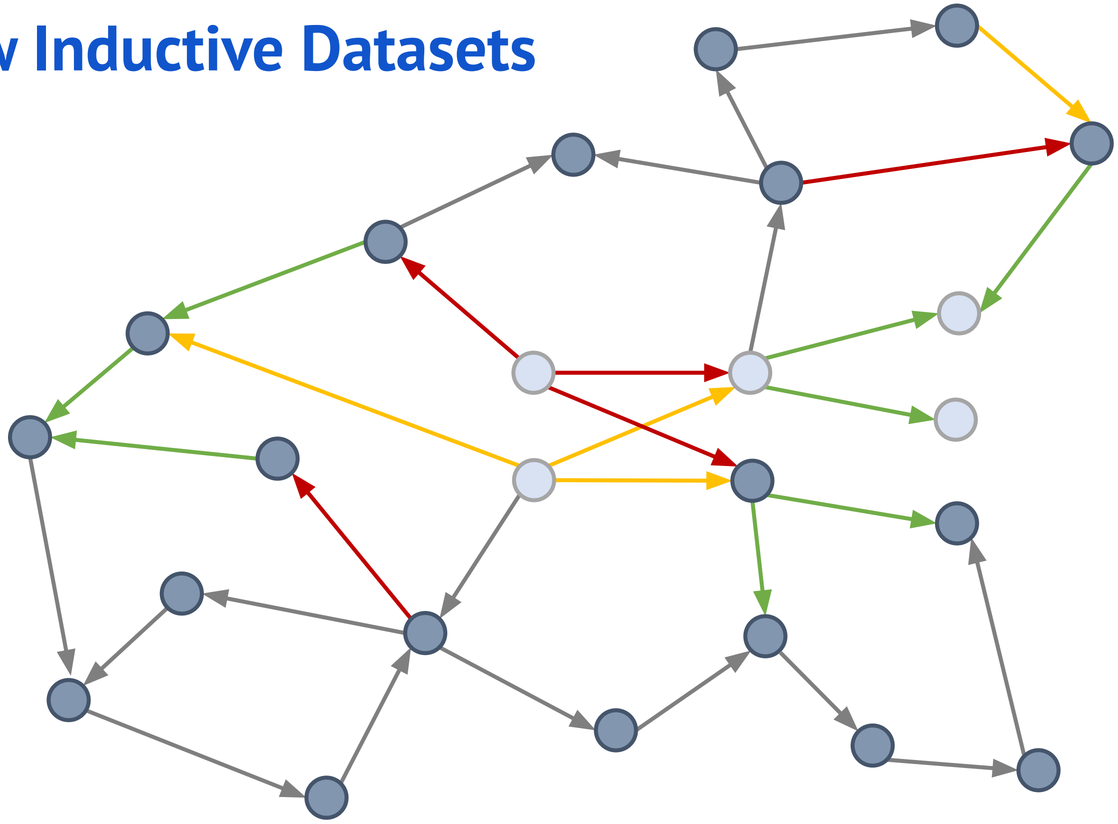
- NBFNet as learnable projection
- Non-parametric t-norms as logical operators
- Learning relation (query) embeddings only, no entity embeddings

New Inductive Datasets



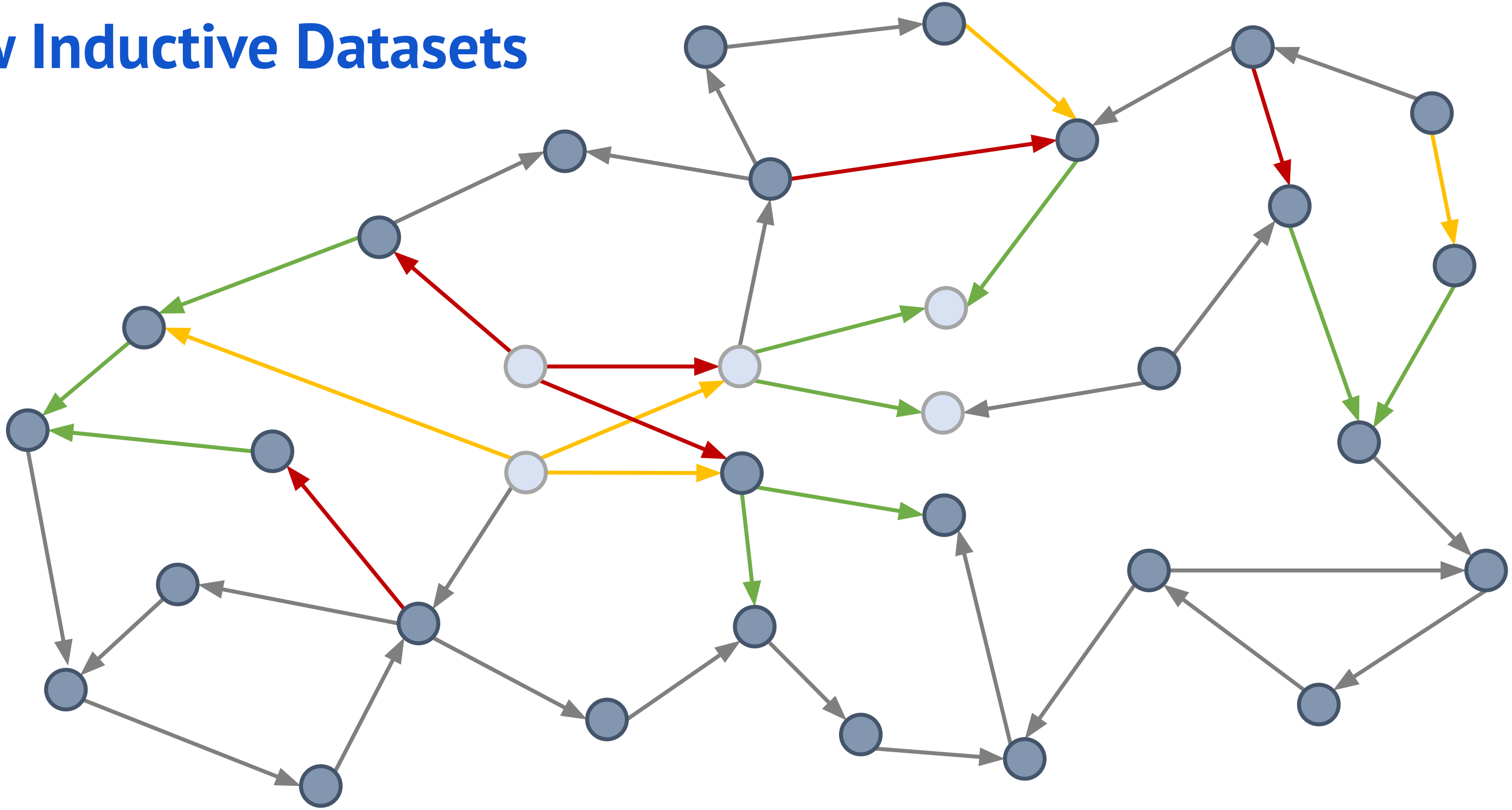
More new nodes and edges at inference time (**105%**)

New Inductive Datasets



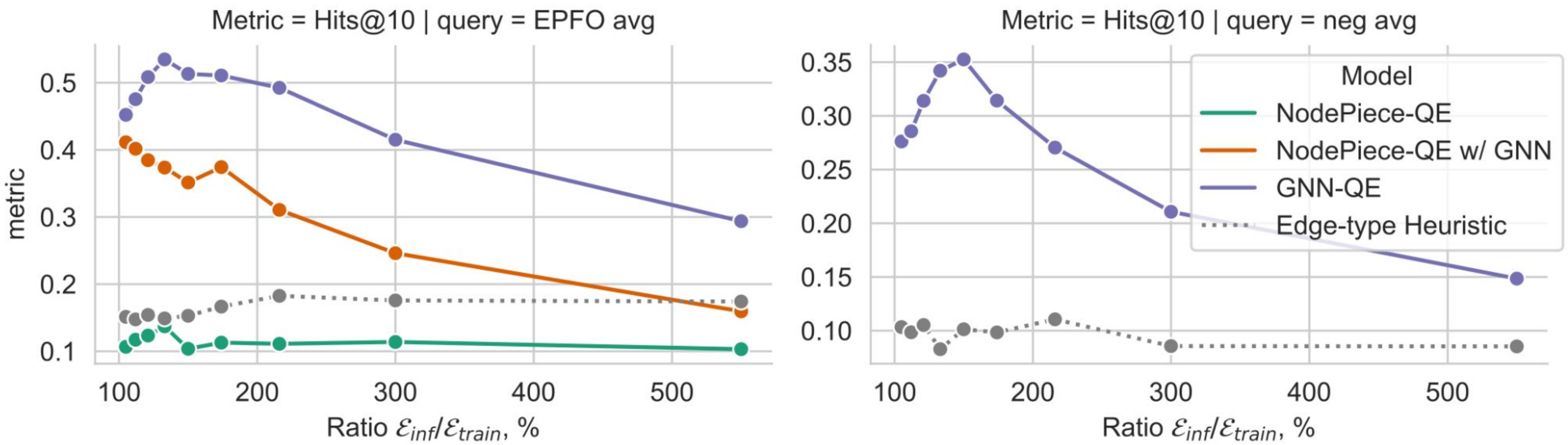
More new nodes and edges at inference time (**300%**)

New Inductive Datasets



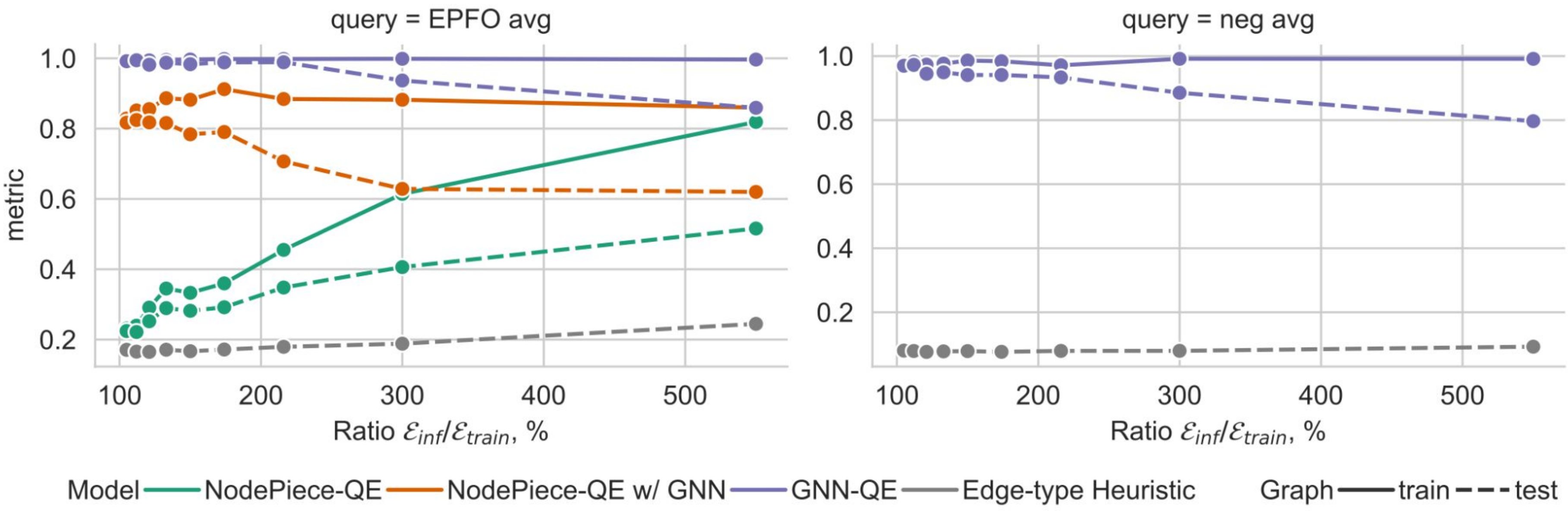
More new nodes and edges at inference time (**550%**)

Inductive Generalization to Larger Test Graphs is Still a Problem

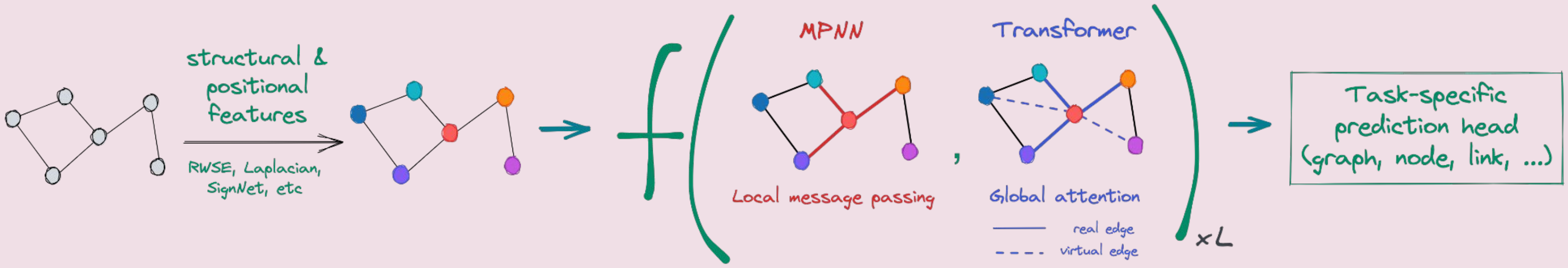


All GNN-based models are affected

Finding New Answers to Train Queries in Larger Graphs



Inductive models can infer new correct answers but still struggle on larger graphs



Recipe for a General, Powerful, Scalable (GPS) Graph Transformer

Ladislav Rampášek

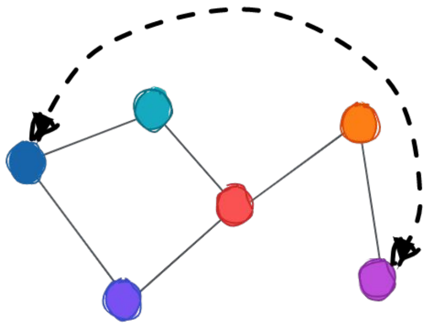
Mila
Université de Montréal

Rampášek L., Galkin M., Dwivedi V. P., Luu A. T., Wolf G., & Beaini D. *Recipe for a General, Powerful, Scalable Graph Transformer*. NeurIPS 2022.

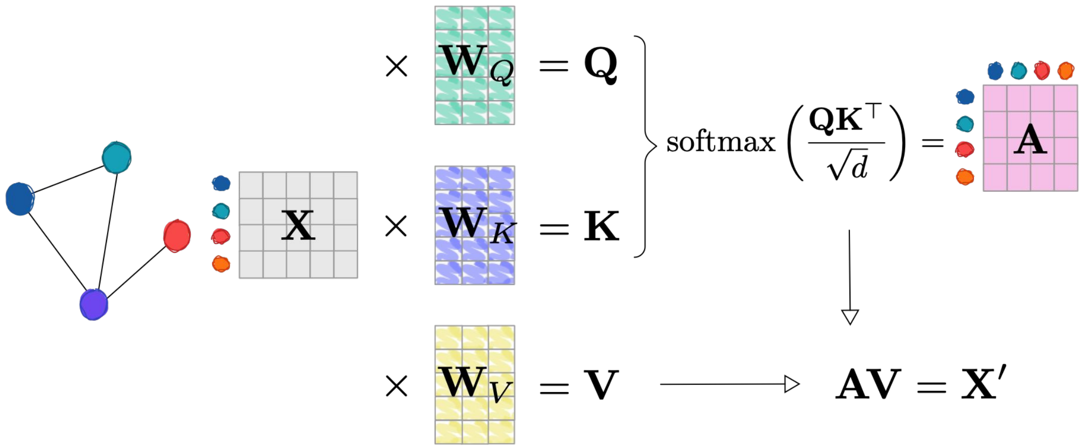
Message Passing Neural Networks

Drawbacks:

- 1-order MPNNs have **limited expressivity** (Weisfeiler-Leman test perspective)
- **Over-smoothing**:
  With increasing the number of GNN layers,
  the features tend to converge to the same value
- **Over-squashing:**
  Losing information when trying to aggregate
  messages from many neighbors into a single vector
- **Poor capturing of long-range dependencies**



Pros of Transformers on Graphs

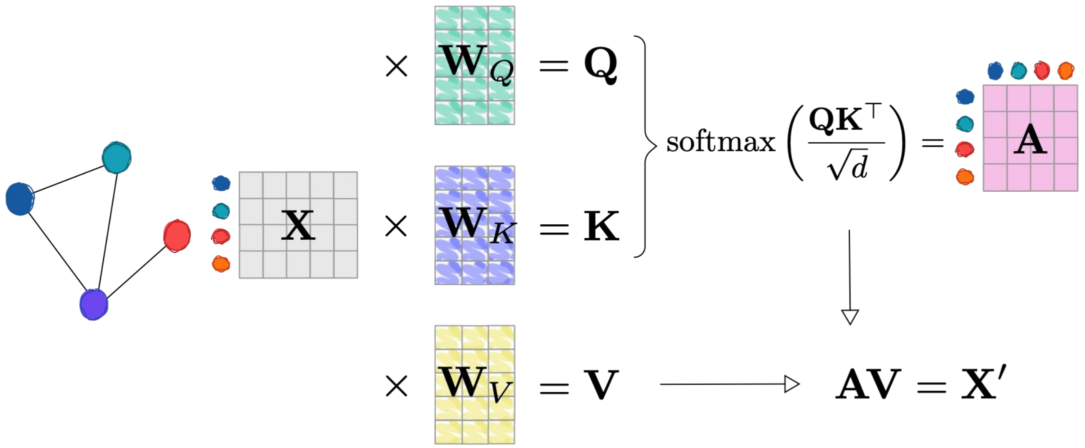


✅ **Decoupled** computation graph structure from the input graph structure

✅ No problem handling **long-range connections** as all nodes are now connected to each other.

✅ Under some assumptions graph Transformers are universal function approximators on graphs [Kreuzer et al., 2021].

Cons of Transformers on Graphs



- **Loss of graph structure**. We need better identifiability of nodes in a graph.
- **Loss of locality inductive bias**. MPNNs work well on graphs with pronounced locality.
- **$O(N^2)$ computational complexity** in the number of nodes whereas MPNNs are linear in the number of edges $O(E)$.

General, Powerful, Scalable Graph Transformer

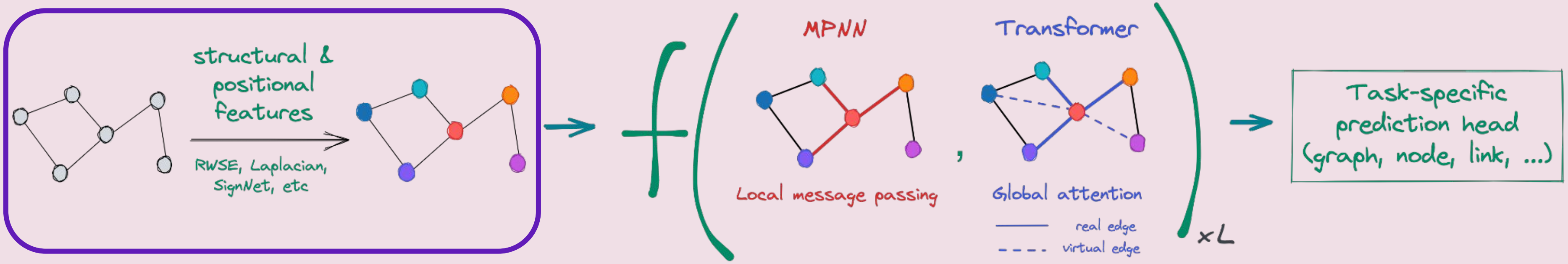
We provide a recipe for building Graph Transformers that are:

🧩 **General**: Our modular recipe consists of 3 main building blocks:
positional and structural encodings,
local message passing, and
global attention into a single pipeline

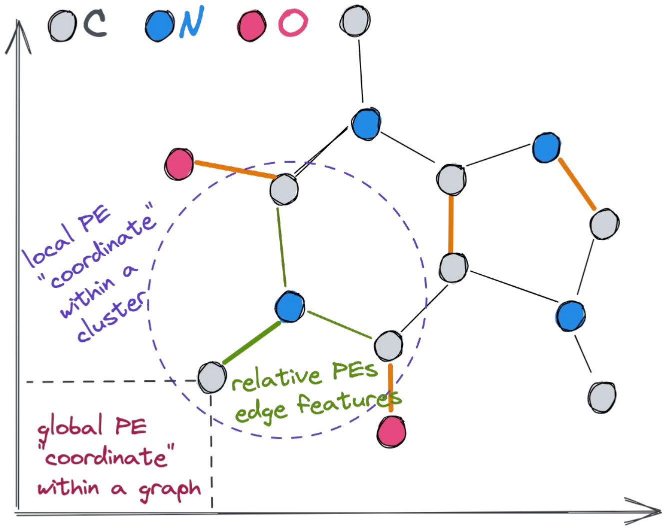
🏆 **Powerful**: More than 1-WL expressive when paired with appropriate positional and structural features.

📈 **Scalable**: The design allows linear global attention modules, hence scaling to graphs of many thousands of nodes each.

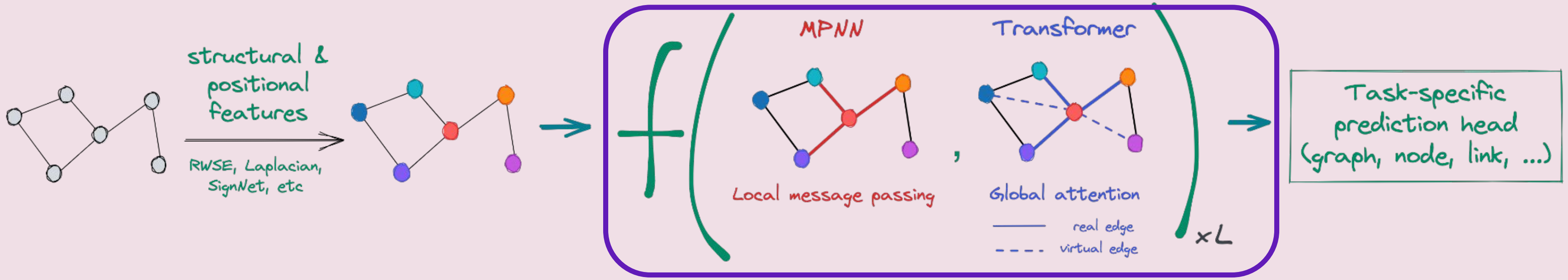
Rampášek L., Galkin M., Dwivedi V. P., Luu A. T., Wolf G., & Beaini D. *Recipe for a General, Powerful, Scalable Graph Transformer*. NeurIPS 2022.



Positional and Structural Encodings

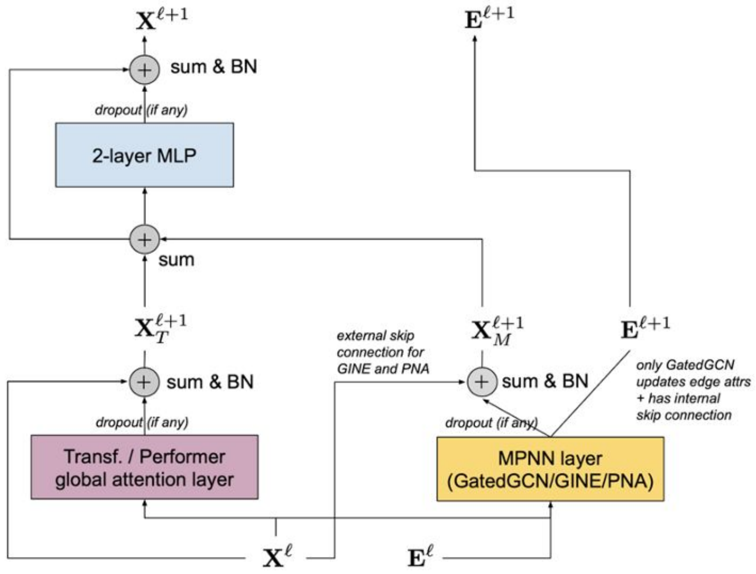


- **Positional**: *Where am I?*
  - Laplacian PE, SignNet, PEG, …
- **Structural**: *What does my neighborhood look like?*
  - Random-walk SE, subgraph patterns, …
- Can categorized as: ***local***, ***global***, or ***relative***



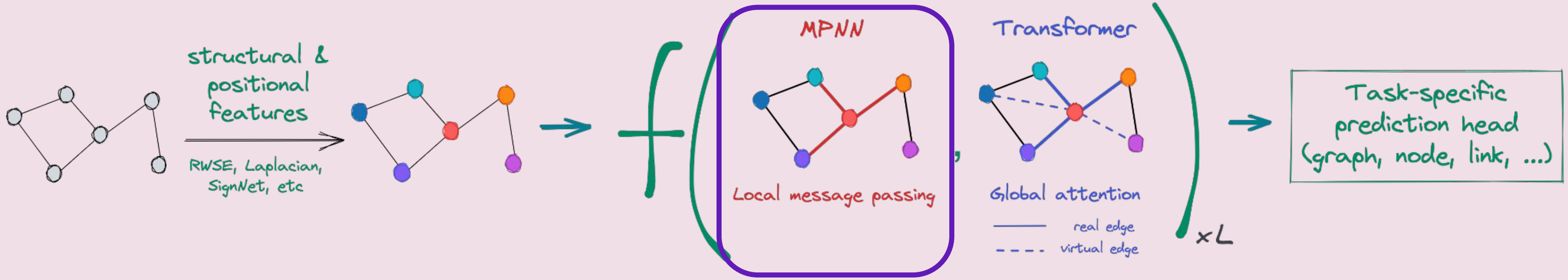
stack of L GPS layers

GPS layer



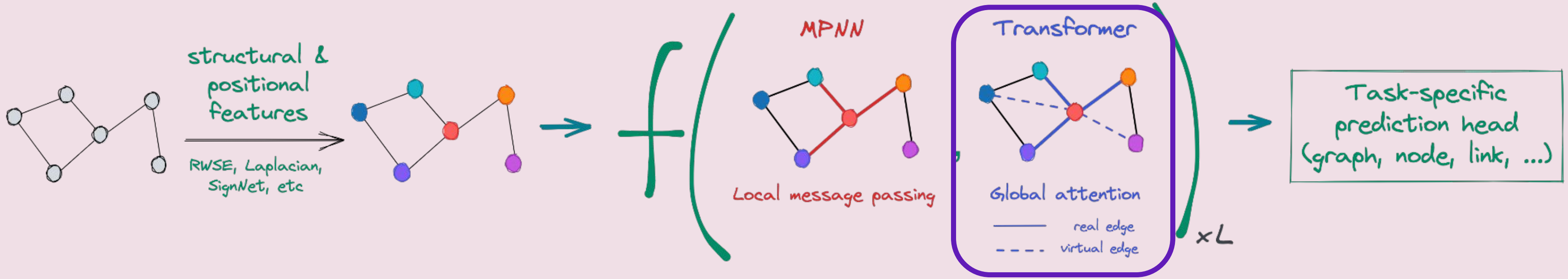
Combines Local MPNN and Transformer:

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



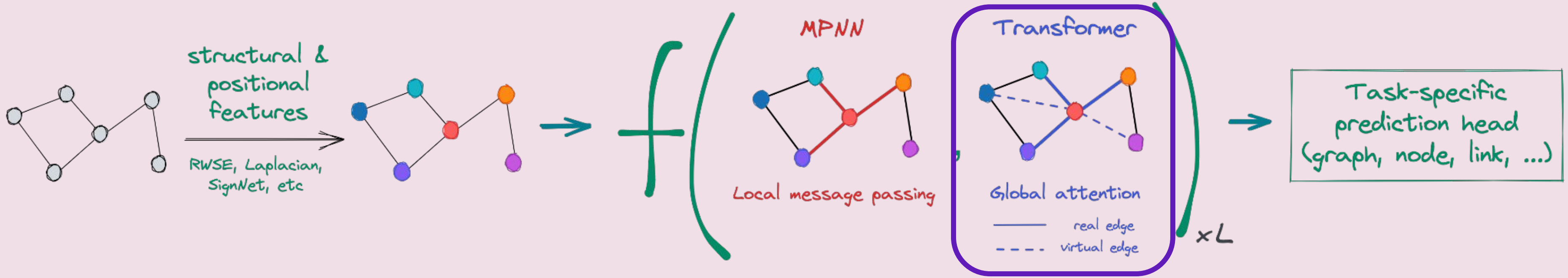
GPS layer: 1. Local Message Passing

- **Provides locality bias** that is difficult or expensive to achieve in Transformer
- **Processes features of real edges**:
  - Encodes edge features into the node features
  - Updates real edge features:
- Examples:
  - **GatedGCN** [Bresson & Laurent, 2017]
  - **GINE** [Hu et al., 2019]
  - **PNA** [Corso et al., 2020]



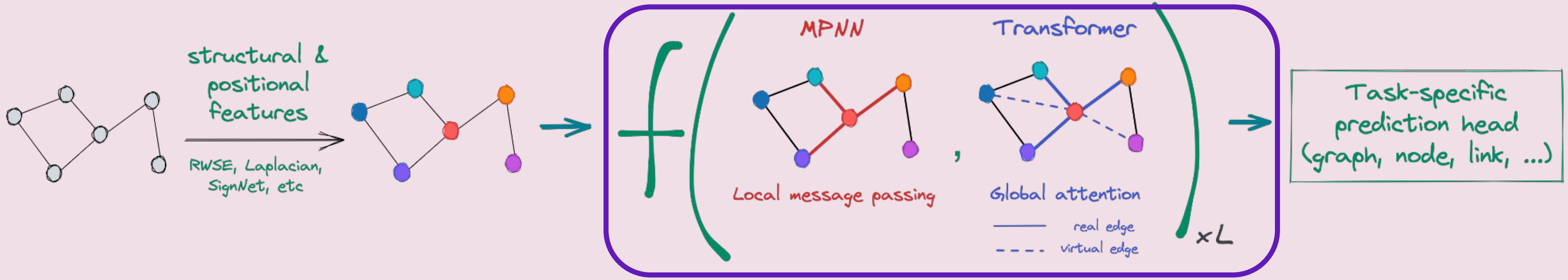
GPS layer: 2. Global Attention (Transformer)

- Fully connected computational graph
- Can utilize PE/SE and local MPNN encoding
- $O(N^2)$ computational complexity with vanilla Transformer
- As we don't need to consider edge features, we can use existing linear Transformer architectures:
  - **Performer** [Choromanski et al., 2021]
  - **BigBird** [Zaheer et al., 2020]



GPS layer: 2. Global Attention (Transformer)

- Fully connected computational graph
- Can utilize PE/SE and local MPNN encoding
- $O(N^2)$ computational complexity with vanilla Transformer
- As we don't need to consider edge features, we can use existing linear Transformer architectures:
  - **Performer** [Choromanski et al., 2021]
  - **BigBird** [Zaheer et al., 2020]



GPS layer

GPS layer keeps the benefits and remedies the cons of a Transformer:

- ✅ **Loss of graph structure** – solved by precomputed PE/SE, local MPNN module
- ✅ **Loss of locality inductive bias** – solved by local MPNN module
- ✅ **$O(N^2)$ computational complexity** – solved by Performer global attention module

Selected Results

- Particularly noteworthy is the performance on ZINC and OGB-LSC PCQM4Mv2.

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**

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

GPS doesn't use any molecular 3D information

GPS++ is OGB LSC 2022 Winner in PCQM4M v2

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.

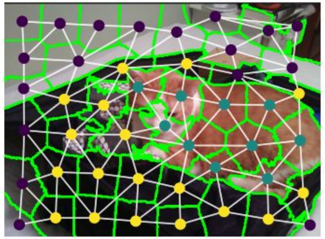
Package: >=1.3.2

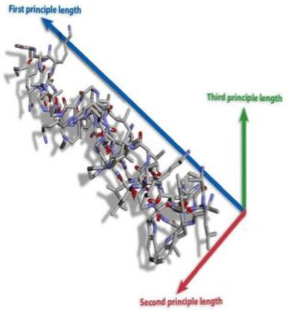
Rank	Method	Ensemble	Test-dev MAE	Validation MAE	Team	Contact	References	#Params	Hardware	Date
1	**GPS++**	Yes	0.0720	0.0778	GraphcoreValenceMILA	Dominic Masters (Graphcore/Valence/MILA)	Paper, Code	44,291,413	Graphcore BOW-POD16	Nov 18, 2022
2	**MolNet_Ensemble**	Yes	0.0753	0.0797	polixir.ai	zouxiaochuan (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

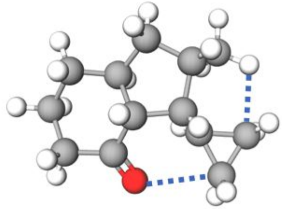
Long Range Graph Benchmark (LRGB) Results

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

`reg-bound` graph



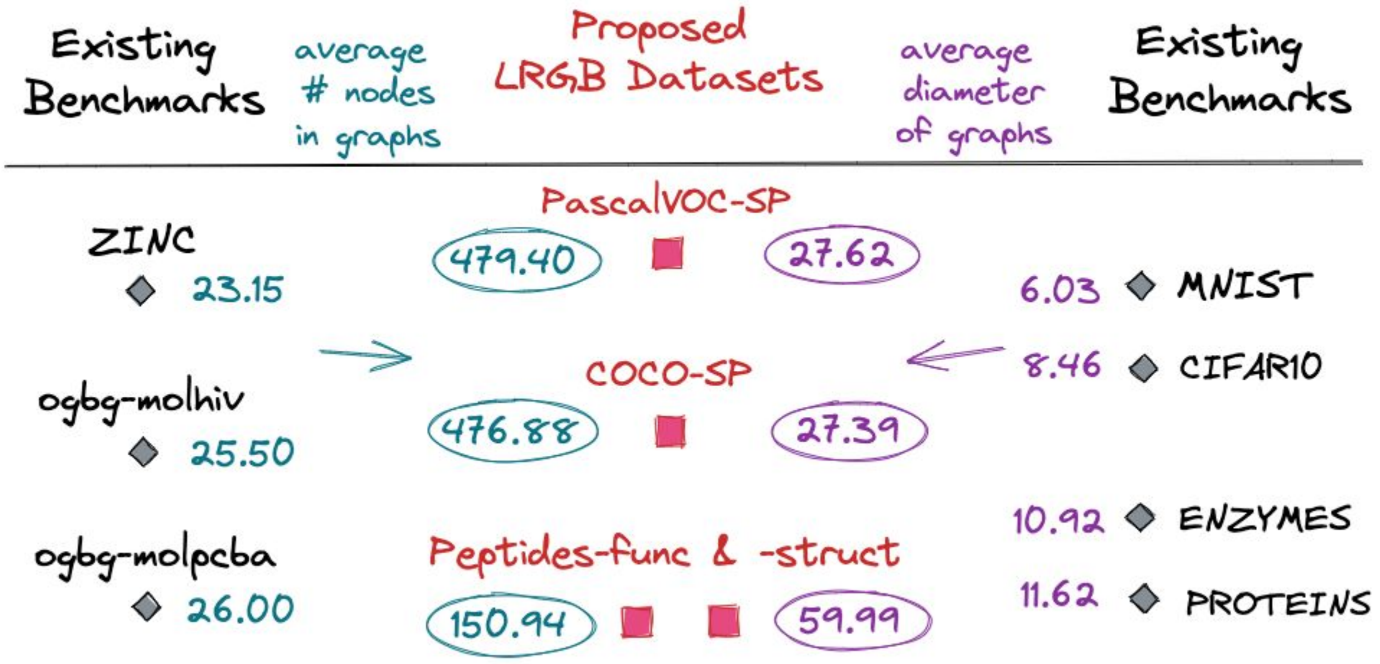




Model	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	Peptides-func AP ↑	Peptides-struct MAE ↓	PCQM-Contact MRR ↑
GCN	0.1268 ± 0.0060	0.0841 ± 0.0010	0.5930 ± 0.0023	0.3496 ± 0.0013	0.3234 ± 0.0006
GINE	0.1265 ± 0.0076	0.1339 ± 0.0044	0.5498 ± 0.0079	0.3547 ± 0.0045	0.3180 ± 0.0027
GatedGCN	0.2873 ± 0.0219	**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.

Long Range Graph Benchmark (LRGB) Results



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.