Discovering new ML applications on Knowledge Graphs

Michael Galkin
Postdoctoral Fellow @ Mila & McGill

👋 About Mila

- Research Institute w/ UdeM, McGill, HEC Montreal, Polytechnique Montreal
  - Core Machine Learning, Deep Learning, Reinforcement Learning, NLP, Graph Learning
  - 100+ papers / year at top ML/AI conferences
- 900+ student researchers
- 40+ professors and CIFAR chairs
- 70+ industry partners
- 40+ Mila-affiliated startups

AMGEN intel AstraZeneca Roche
Meta FPT CN Hydro Québec
Montréal DRW QUANTUMBLACK A MCKINSEY COMPANY

SAMSUNG IBM Microsoft Google
HITACHI CDPQ Recursion NUANCE

Plan

→ **Vanilla KG Representation Learning: Re-cap**

→ The New Big Picture

→ NodePiece: Beyond Shallow Embeddings

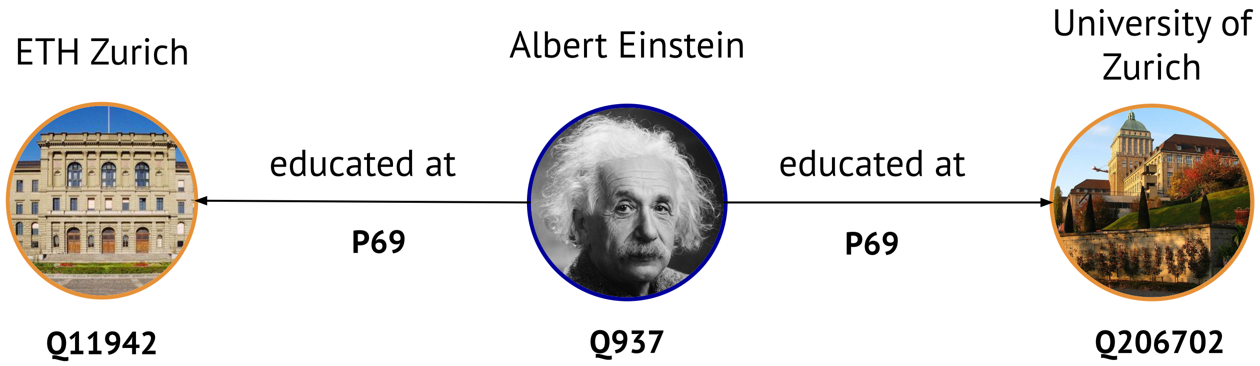
→ Hyper-Relational KGs

→ Inductive Link Prediction with HR KGs

→ Complex Query Answering with HR KGs

→ Past, Today, Future

Triple-based Knowledge Graphs

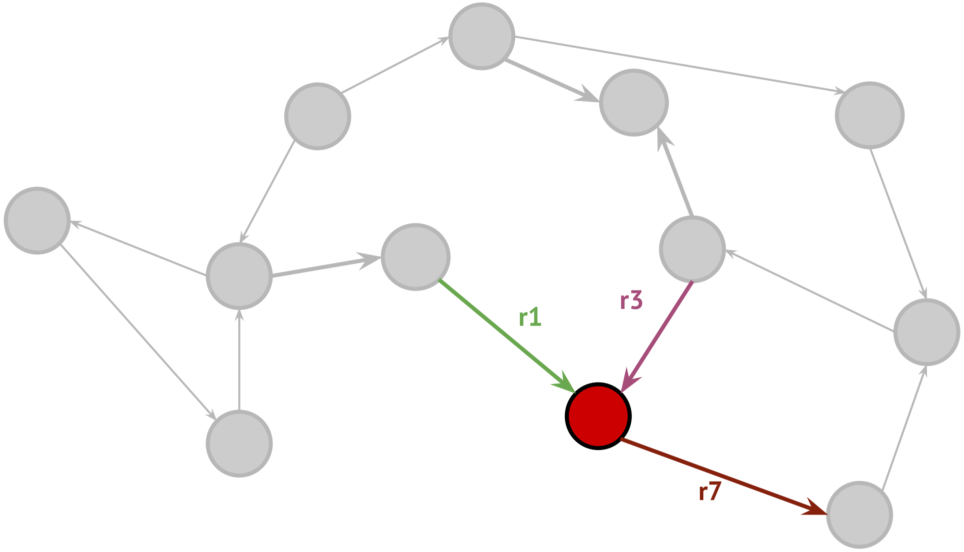


RDF

```
Albert Einstein    educatedAt    University of Zurich .
Albert Einstein    educatedAt    ETH Zurich .
```

Source: Albert Einstein on Wikidata

Knowledge Graphs: Setup



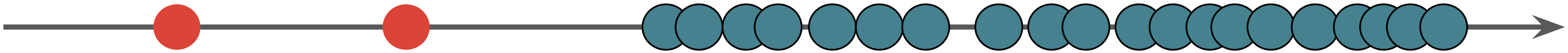
- Directed graphs
- Explicit relation types (learnable edge features)
- Input node features are **not** given

Brief History: 2011 -

RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]

100+ KG embedding models since then 😱



Brief History: 2011 -

Transductive

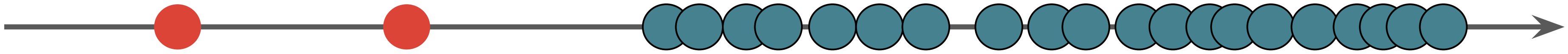
Triples

Supervised

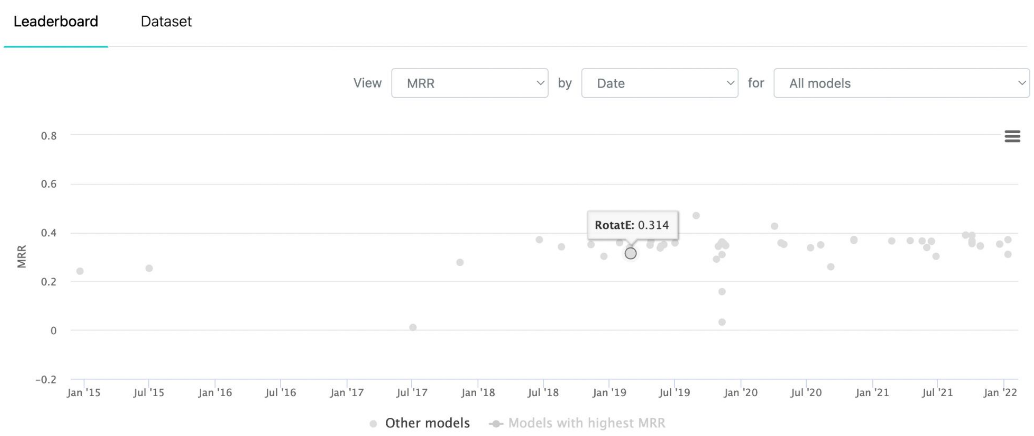
RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]

100+ KG embedding models since then 😱



Link Prediction on FB15k-237



No substantial progress since 2018

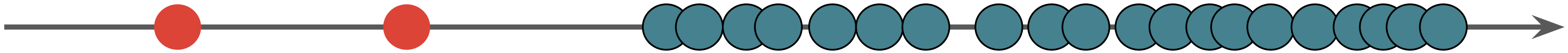
Brief History: 2011 -

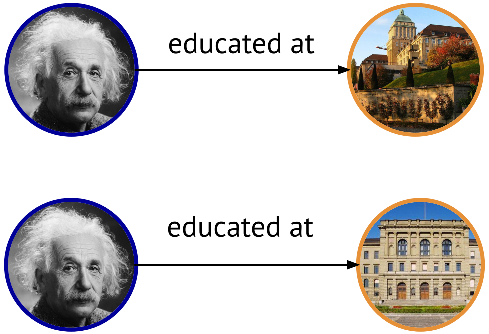
Transductive | Triples | Supervised

RESCAL
[Nickel et al, ICML 2011]

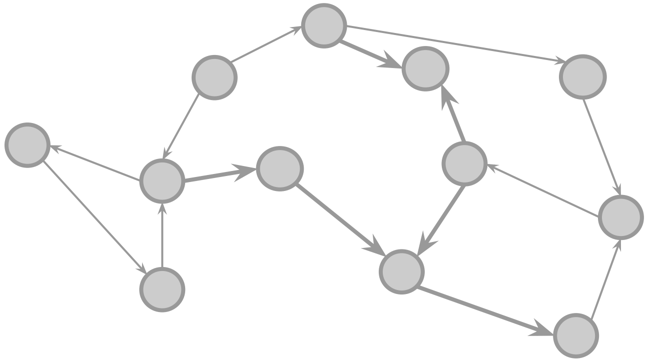
TransE
[Bordes et al, NeurIPS 2013]

100+ KG embedding models since then 😱





Can we go from *single triples* to encoding the whole *graph structure*?

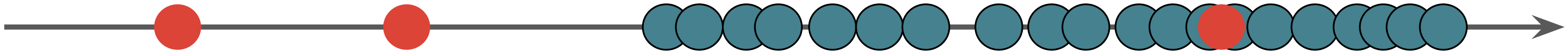


Brief History: 2011 -

Transductive | Triples | Supervised

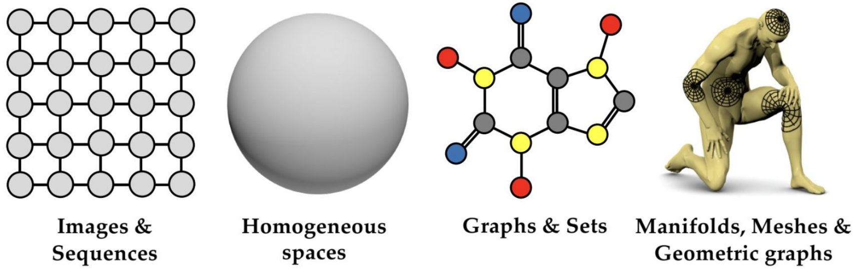
RESCAL
[Nickel et al, ICML 2011]

TransE
[Bordes et al, NeurIPS 2013]



Geometric DL
2018

The "5G" of Geometric Deep Learning



https://geometricdeeplearning.com/

Geometric Deep Learning

Study of symmetries and invariances that unifies many deep learning architectures

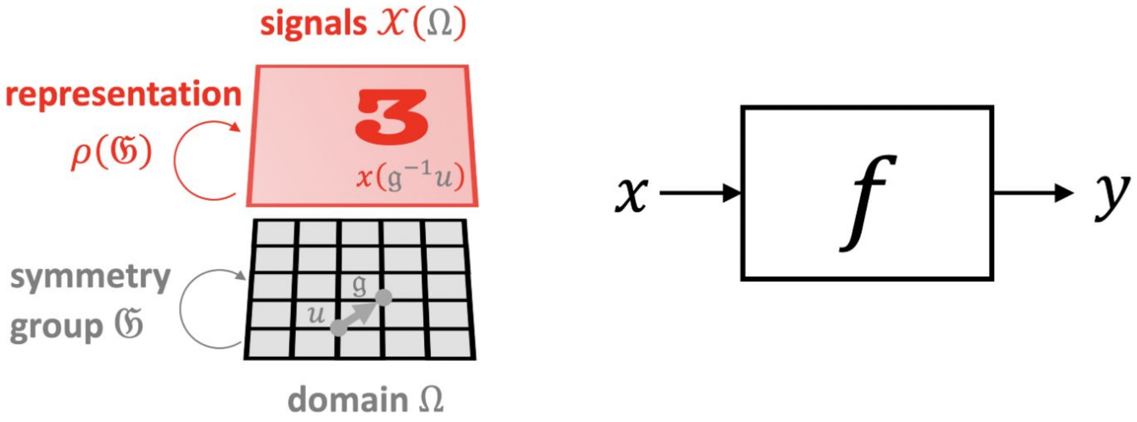
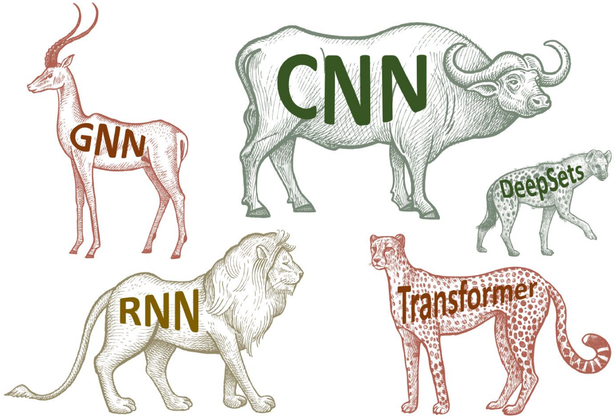


Illustration of geometric priors: the input signal (image $x \in \mathcal{X}(\Omega)$) is defined on the domain (grid Ω), whose symmetry (translation group $\mathfrak{G}$) acts in the signal space through the group representation $\rho(\mathfrak{g})$ (shift operator). Making an assumption on how the functions f (e.g. image classifier) interacts with the group restricts the hypothesis class.



Source: Michael Bronstein's blog post https://towardsdatascience.com/geometric-foundations-of-deep-learning-94cdd45b451d

Plan

→ Vanilla Triple KGs: Re-cap

→ **The New Big Picture**

→ NodePiece: Beyond Shallow Embeddings

→ Hyper-Relational KGs

→ Inductive Link Prediction with HR KGs

→ Complex Query Answering with HR KGs

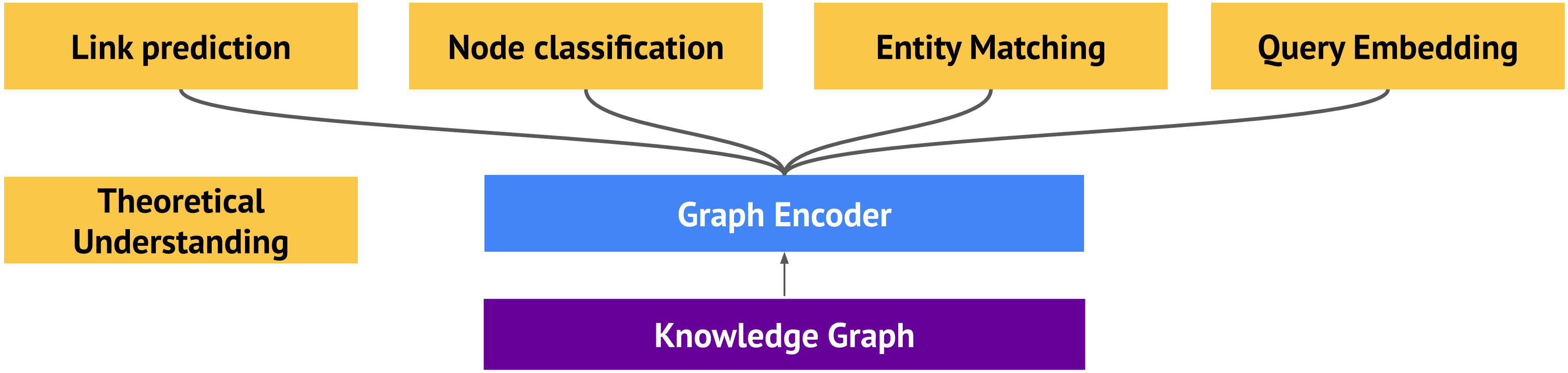
→ Past, Today, Future

Big Picture in $\mathbb{R}^5$

Transductive	Triples	Supervised	Unimodal	Small
Inductive	Hyper-relational	Unsupervised	Multimodal	Large (sampling)

SETTING

TASK



Big Picture in $\mathbb{R}^5$

Transductive	Triples	Supervised	Unimodal	Small
Inductive	Hyper-relational	Unsupervised	Multimodal	Large (sampling)

SETTING

TASK

Link prediction

Node classification

Entity Matching

Query Embedding

Theoretical Understanding

Graph Encoder

Knowledge Graph

Plan

→ Vanilla Triple KGs: Re-cap
→ The New Big Picture
→ **NodePiece: Beyond Shallow Embeddings**
→ Hyper-Relational KGs
→ Inductive Link Prediction with HR KGs
→ Complex Query Answering with HR KGs
→ Past, Today, Future

The ImageNet Moment for KGs

Self-supervised pre-training on a LARGE graph → **Fine-tuning on a downstream task**

- Graph ML
- NLP
- Vision

The ImageNet Moment for KGs

Self-supervised pre-training on a LARGE graph → **Fine-tuning on a downstream task**

Wikidata: 100M nodes
Embs: [100M, dim] ?

PyTorch BigGraph

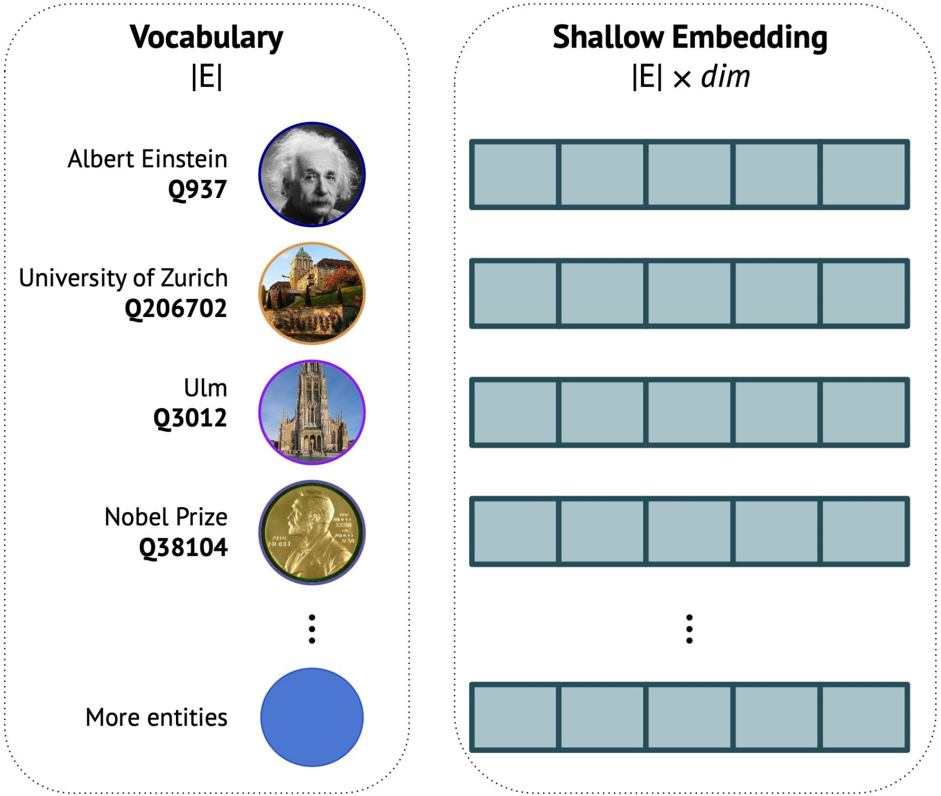
~200 GB

- Graph ML
- NLP
- Vision

Shallow Embedding

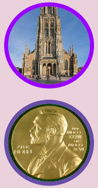
Looks like a Representation Learning challenge 🤔

Can we do better?



Transductive vs Inductive

Shallow embeddings

	Transductive	Inductive
Training	**Vocab** 	
Inference		 New, unseen nodes (entities) ● Added to the seen graph ● Completely new inference graph

OGB WikiKG: Just 2.5M nodes (June'21)

BERT-Large is ~340M params

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	Test MRR	Validation MRR	Contact	References	#Params	Hardware	Date
1	**PairRE (200dim)**	0.5208 ± 0.0027	0.5423 ± 0.0020	Linlin Chao	Paper, Code	500,334,800	Tesla P100 (16GB GPU)	Jan 28, 2021
2	RotatE (250dim)	0.4332 ± 0.0025	0.4353 ± 0.0028	Hongyu Ren – OGB team	Paper, Code	1,250,435,750	Quadro RTX 8000 (45GB GPU)	Jan 23, 2021
3	TransE (500dim)	0.4256 ± 0.0030	0.4272 ± 0.0030	Hongyu Ren – OGB team	Paper, Code	1,250,569,500	Quadro RTX 8000 (45GB GPU)	Jan 23, 2021
4	ComplEx (250dim)	0.4027 ± 0.0027	0.3759 ± 0.0016	Hongyu Ren – OGB team	Paper, Code	1,250,569,500	Quadro RTX 8000 (45GB GPU)	Jan 23, 2021

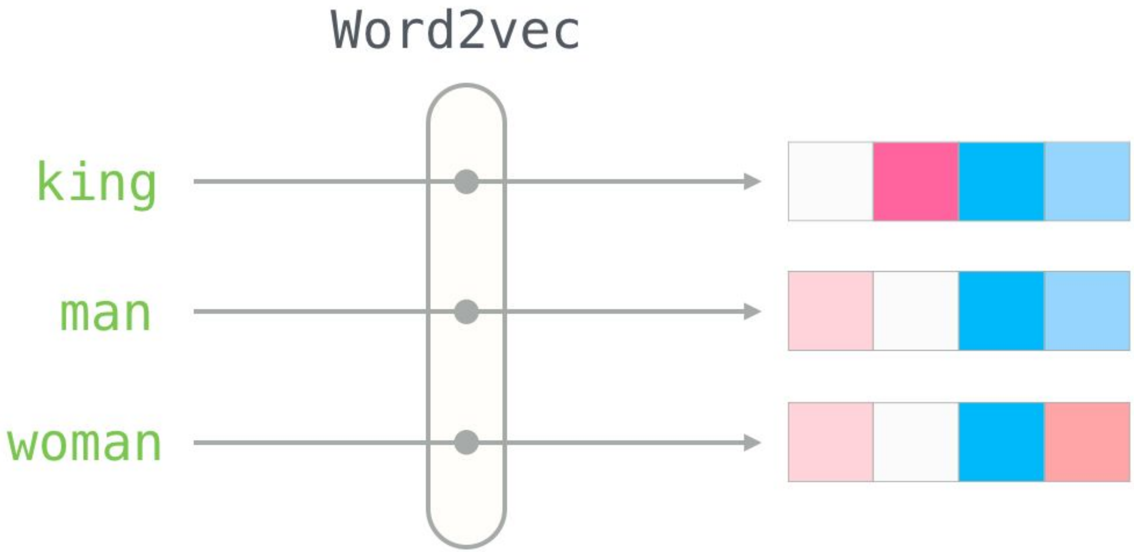
BERT (340M params) - disruption in NLP ✅
KG embs (>1B params) - 🥴

Life beyond shallow embedding?

Do we really need to learn & store the whole shallow embedding matrix $|E|$ x *dim* ?

Trying to fit a 100M x 200 tensor on a Tesla V100 ->

Back to 2014



Unseen words = [OOV] (out-of-vocabulary)

Source: The Illustrated Word2vec by Jay Alammar

Byte-Pair Encoding / WordPiece

Dictionary

5 l o w
2 l o w e r
6 n e w **est**
3 w i d **est**

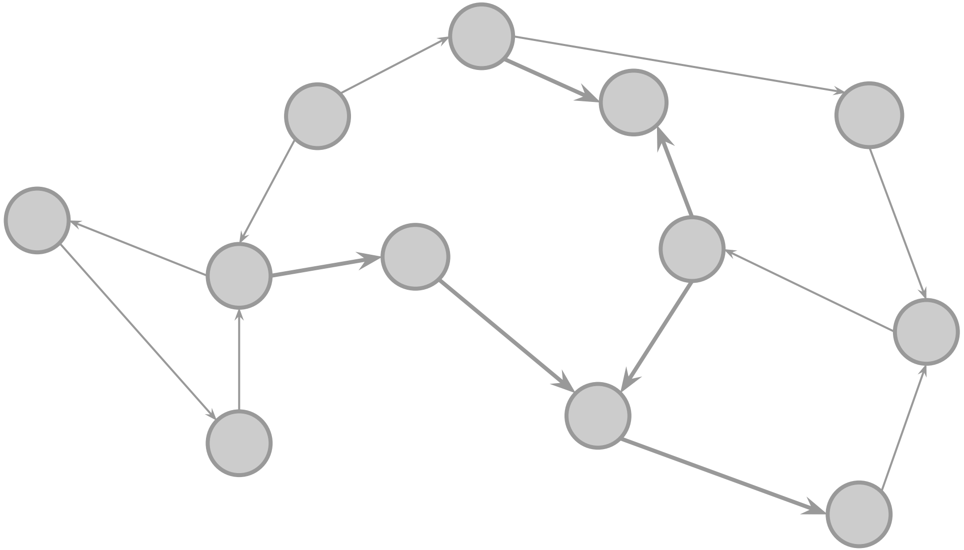
Vocabulary

l, o, w, e, r, n, w, s, t, i, d, es, **est**

Add a pair (es, t) with freq 9

Byte-Pair Encoding vocabulary. Source: Sennrich et al and CS224N

Tokenization + Graphs?

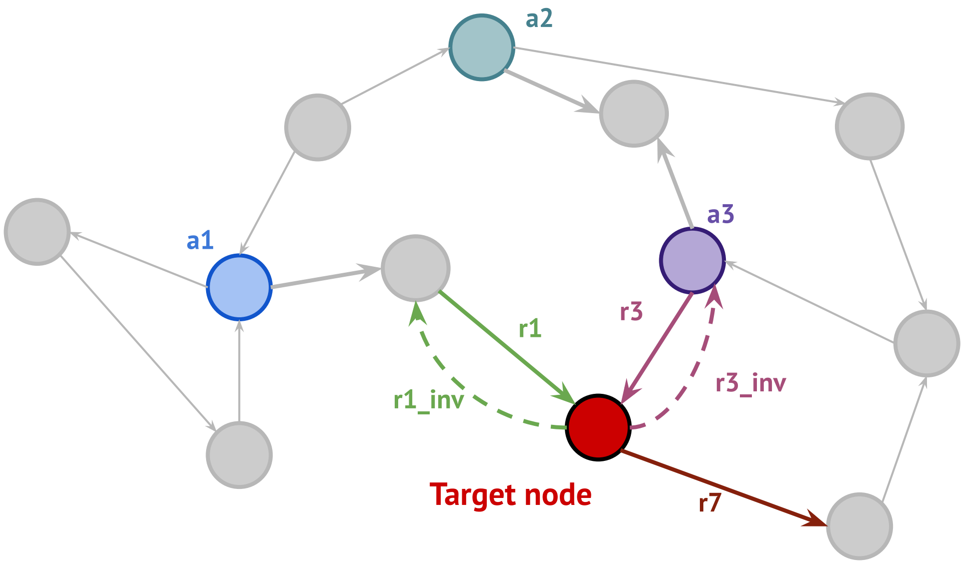


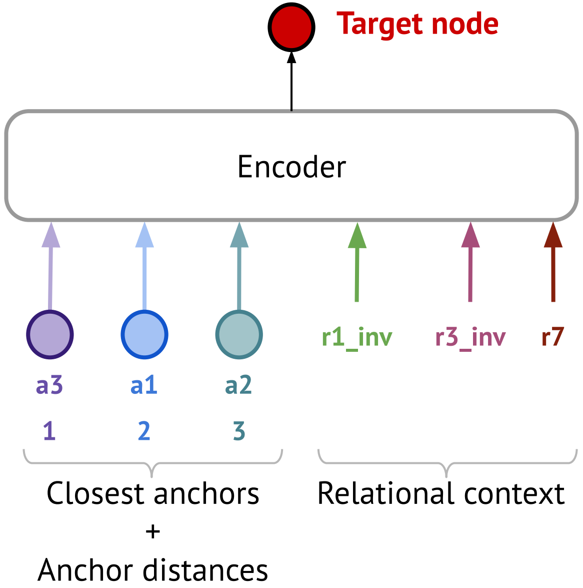
If nodes in a graph are "**words**", can we design a fixed-size vocab of "**sub-word**" units?

Tokenizing KGs

	Shallow embedding, only known words, otherwise OOV	Compositional representations, subword units
Language	Word2vec, GloVe	Byte-Pair Encoding, WordPiece
Graphs	All KG embedding algorithms (TransE, etc)	**NodePiece**

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

Tokenizing Einstein

Albert Einstein

= f (

Ulm ,

Nobel Prize ,

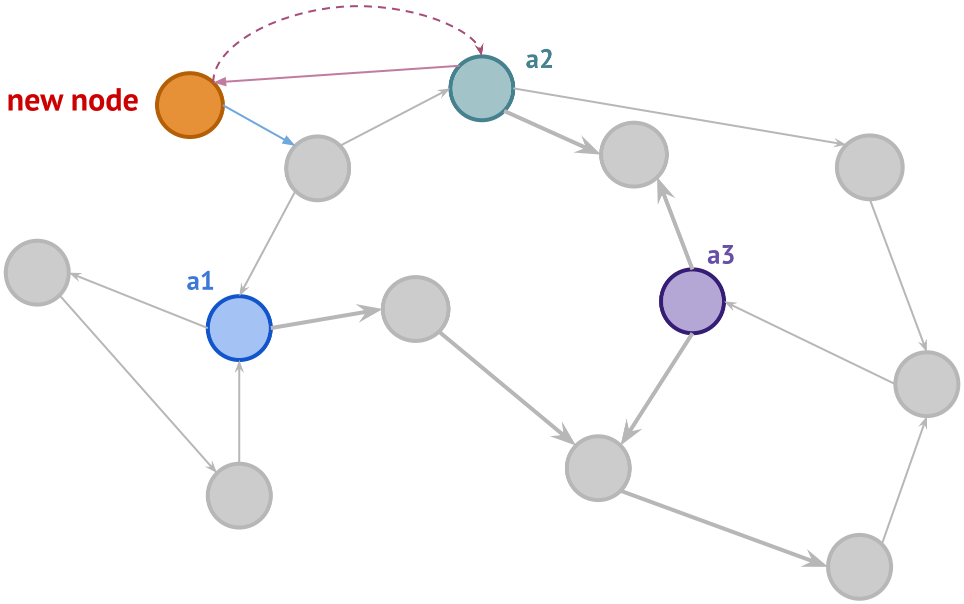
Theoretical Physics ,

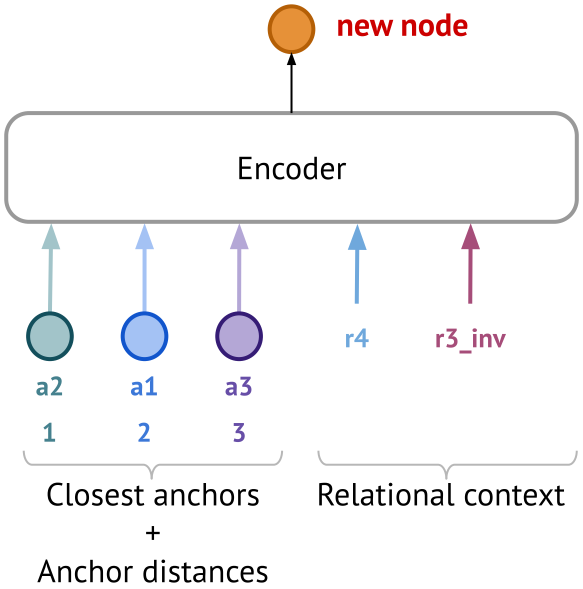
place of birth **P19** , award received **P166** , occupation **P106** , academic degree **P512**)

3 nearest anchors

4 unique outgoing relations in the context

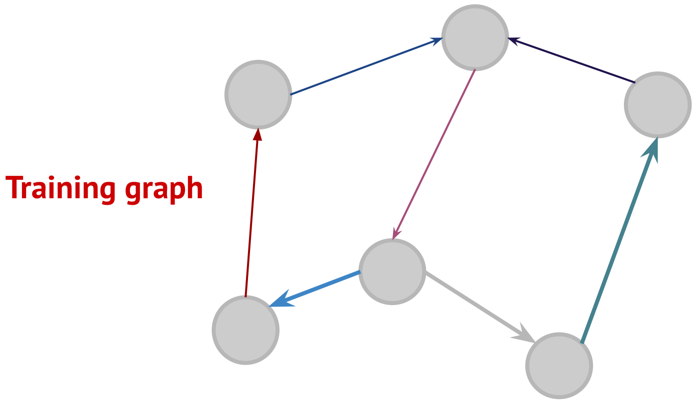
Unseen Node Tokenization

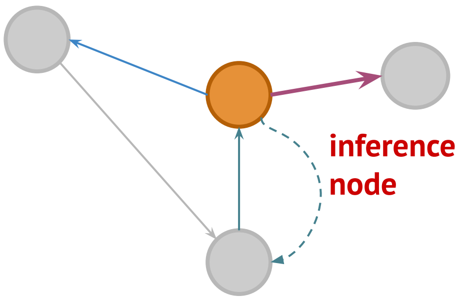


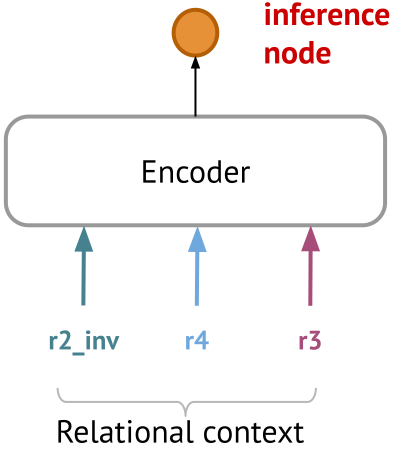


Galkin et al. NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs. ICLR 2022

Inductive Node Tokenization

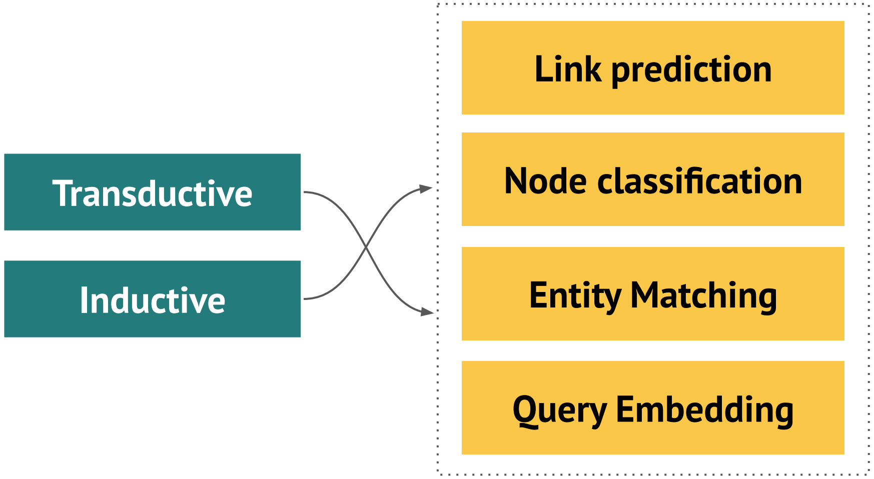






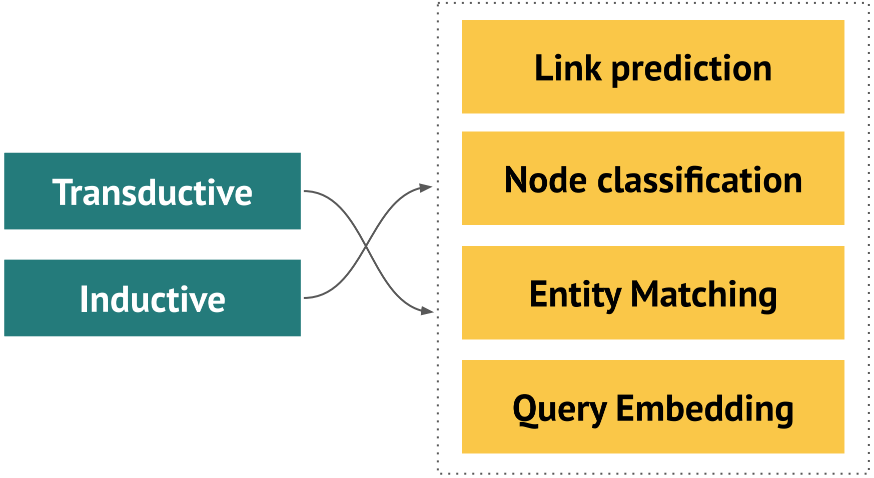
Galkin et al. NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs. ICLR 2022

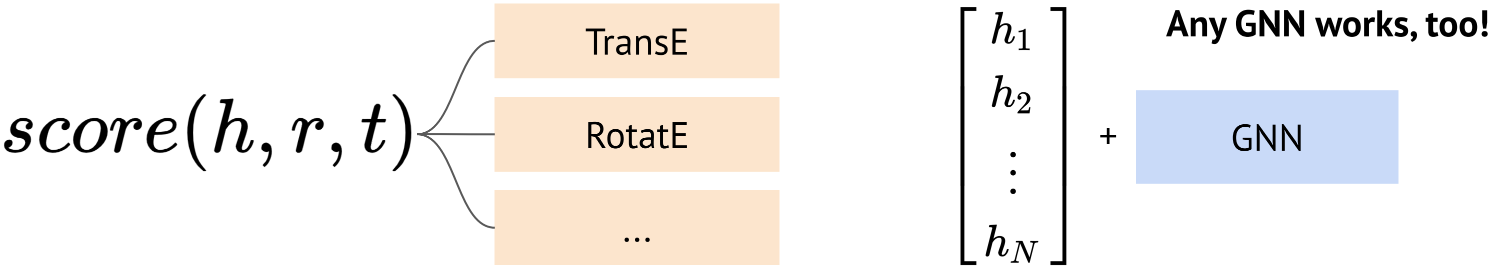
New Downstream Tasks



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

New Downstream Tasks





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

OGB WikiKG 2 : NodePiece is New SOTA

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.

February 2022

NodePiece-enabled models

Rank	Method	Ext. data	Test MRR	Validation MRR	Contact	References	#Params	Hardware	Date
1	**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
2	**InterHT**	No	0.6779 ± 0.0018	0.6893 ± 0.0015	Baoxin Wang (HFL)	Paper, Code	19,215,402	Tesla V100 (32GB)	Feb 10, 2022
3	**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
4	**ComplEx-RP (50dim)**	No	0.6392 ± 0.0045	0.6561 ± 0.0070	Yihong Chen (UCL NLP & FAIR London)	Paper, Code	250,167,400	Tesla V100 (32GB)	Nov 23, 2021
5	**tripleRE**	No	0.5794 ± 0.0020	0.6045 ± 0.0024	Long Yu (360AI)	Paper, Code	500,763,337	Tesla P40(24GB)	Dec 17, 2021
6	**NodePiece + AutoSF**	No	0.5703 ± 0.0035	0.5806 ± 0.0047	Mikhail Galkin (Mila)	Paper, Code	6,860,602	Tesla V100 (32 GB)	Jul 17, 2021

Inductive Link Prediction Challenge 2022

https://github.com/pykeen/ilpc2022

ILPC'22 Small

Split	Entities	Relations	Triples
Train	10,230	96	78,616
Inference	6,653	96 (subset)	20,960
Inference validation	6,653	96 (subset)	2,908
Inference test	6,653	96 (subset)	2,902
Hold-out test set	6,653	96 (subset)	2,894

ILPC'22 Large

Split	Entities	Relations	Triples
Train	46,626	130	202,446
Inference	29,246	130 (subset)	77,044
Inference validation	29,246	130 (subset)	10,179
Inference test	29,246	130 (subset)	10,184
Hold-out test set	29,246	130 (subset)	10,172

Model	MRR	H@100	H@10	H@5	H@3	H@1	AMRI
InductiveNodePieceGNN	**0.1326**	**0.4705**	**0.2509**	**0.1899**	**0.1396**	**0.0763**	**0.730**
InductiveNodePiece	0.0381	0.4678	0.0917	0.0500	0.0219	0.007	0.666

Plan

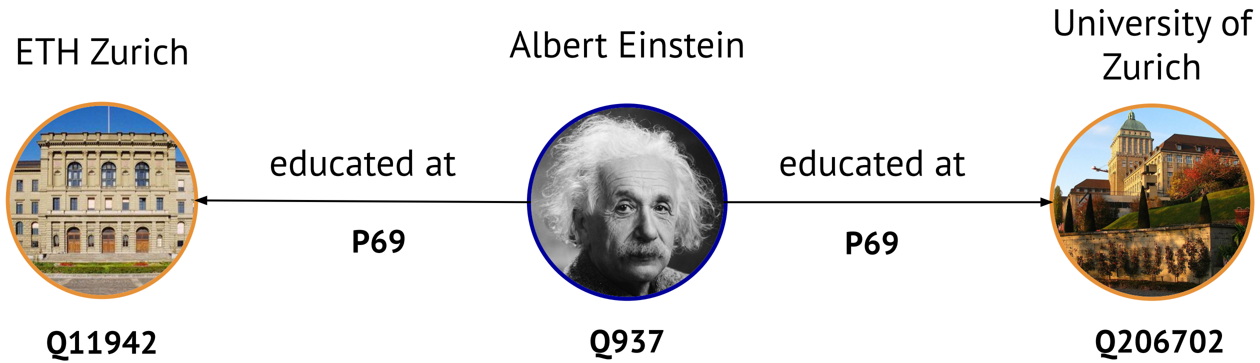
→ Vanilla Triple KGs: Re-cap
→ The New Big Picture
→ NodePiece: Beyond Shallow Embeddings
→ **Hyper-Relational KGs**
→ **Inductive Link Prediction with HR KGs**
→ **Complex Query Answering with HR KGs**
→ Past, Today, Future

Triple-based Knowledge Graphs

Wait a sec

Studied the same subject?

Got the same degree?



RDF

```
Albert Einstein    educatedAt    University of Zurich .
Albert Einstein    educatedAt    ETH Zurich .
```

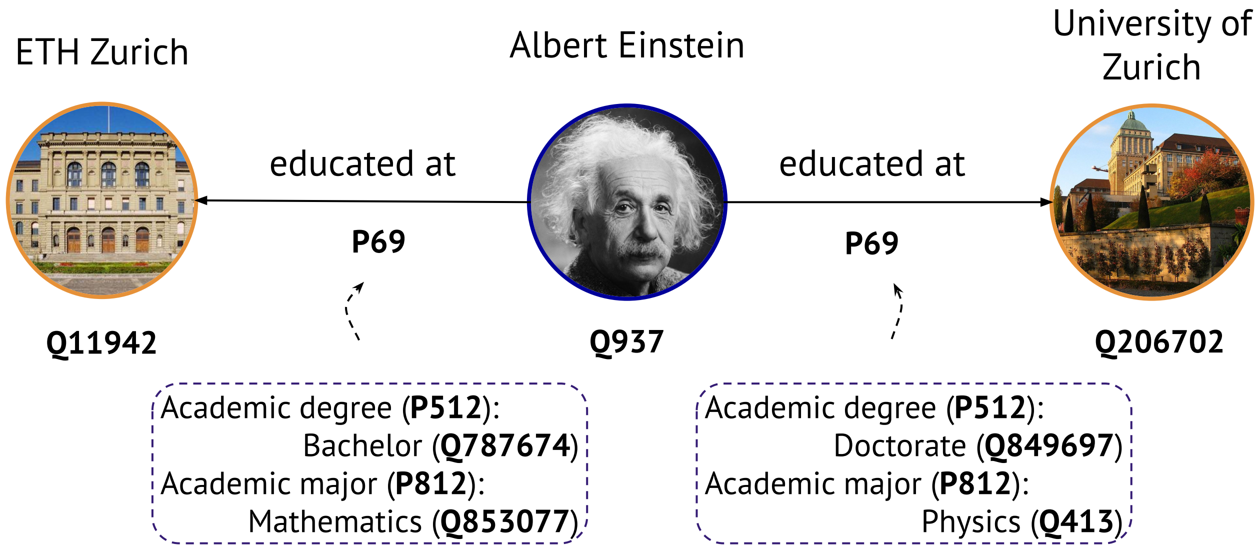
Source: [Albert Einstein](#) on Wikidata

Hyper-relational KGs

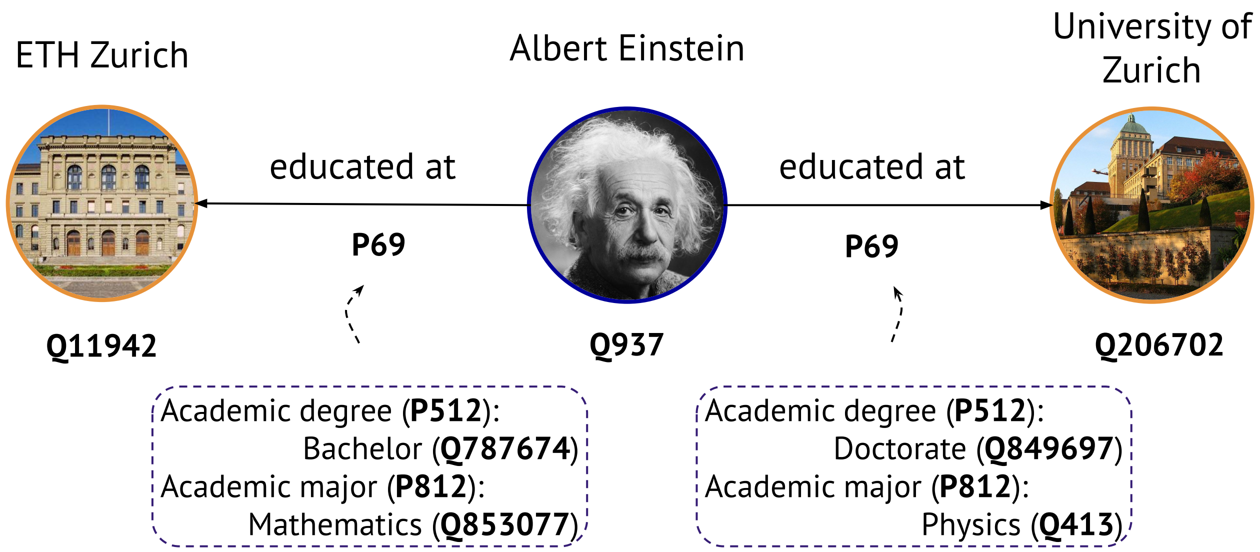
Transductive | Triples

Inductive | Hyper-relational

Link Prediction



Hyper-relational KGs & RDF*



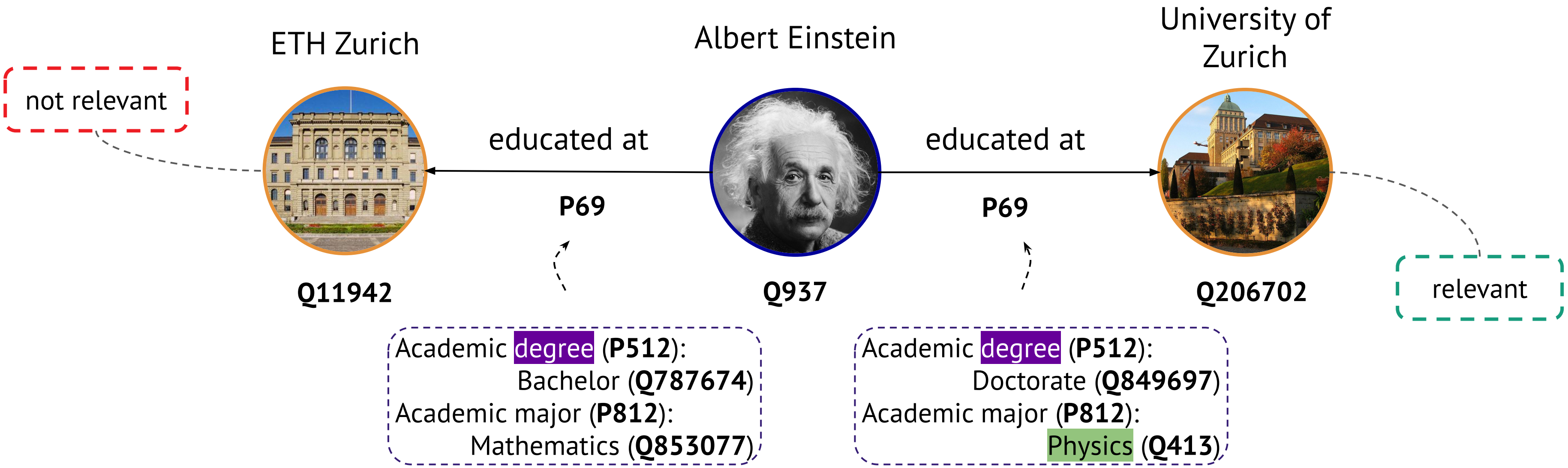
RDF*

```
<< Albert Einstein educatedAt University of Zurich >>
    academic degree Doctorate;
    academic major  Physics .
<< Albert Einstein educatedAt ETH Zurich >>
    Academic degree Bachelor ;
    Academic major  Mathematics .
```

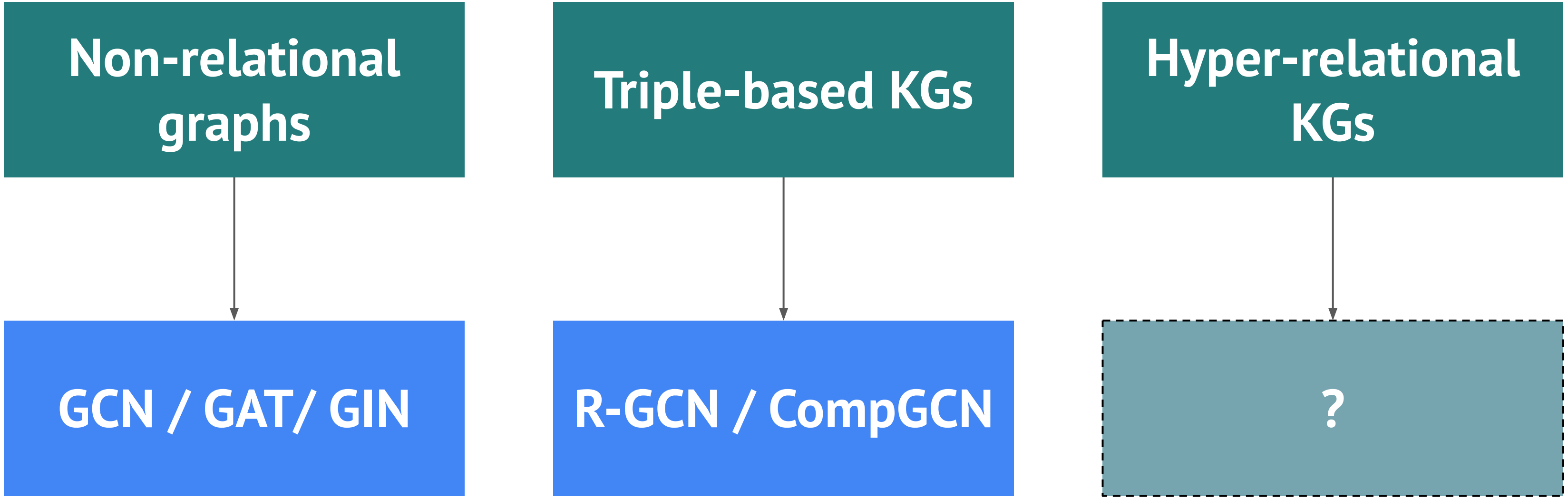
Statements with Qualifiers in Wikidata
Entity-relation Edge Attributes
Edge Instances

Hyper-relational KGs & RDF*

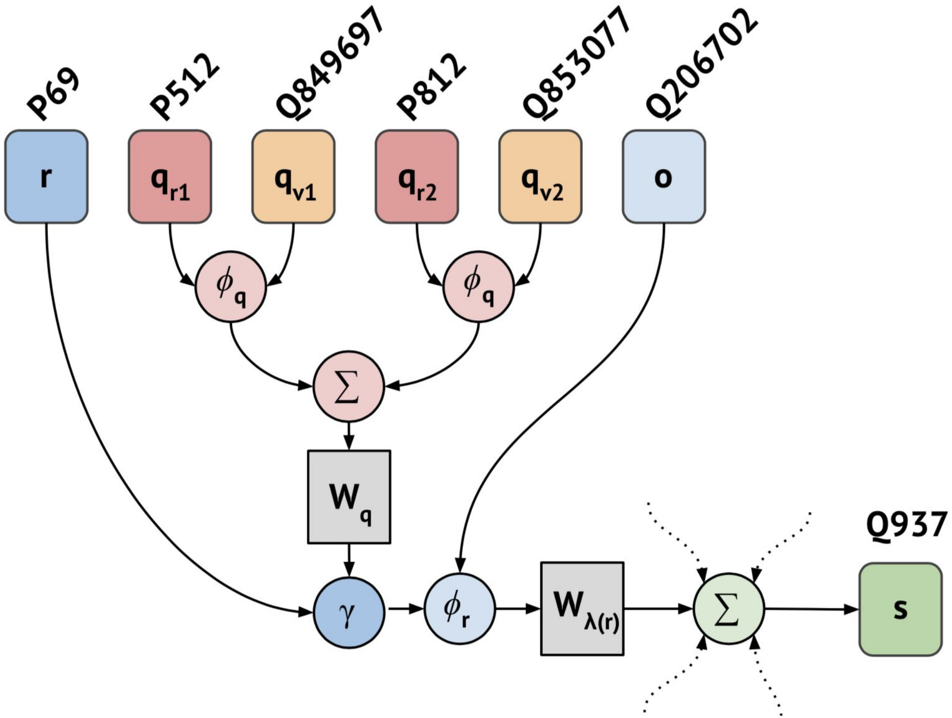
Where did Albert Einstein receive his degree in physics?

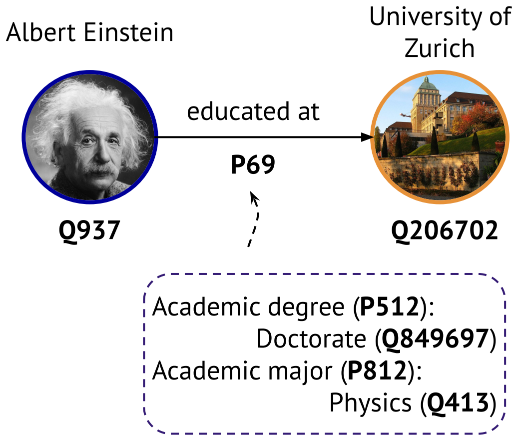


GNN Encoders



★ StarE: GNN architecture for HR KGs

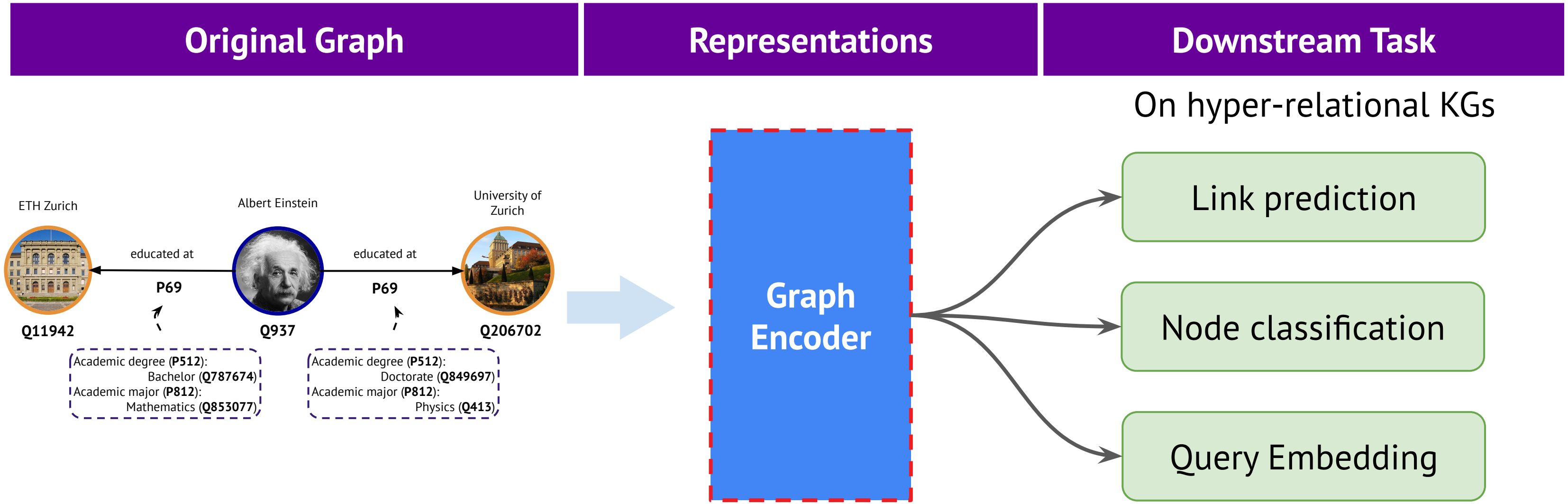




$$\mathbf{h}_v = f\left(\sum_{(u,r)\in\mathcal{N}(v)} \mathbf{W}_{\lambda(r)} \phi_r(\mathbf{h}_u, \gamma(\mathbf{h}_r, \mathbf{h}_q)_{vu}) \right)$$

Galkin et al. Message Passing for Hyper-Relational Knowledge Graphs. EMNLP 2020

Enabling New Tasks



Galkin et al. Message Passing for Hyper-Relational Knowledge Graphs. EMNLP 2020

★ StarE Summary

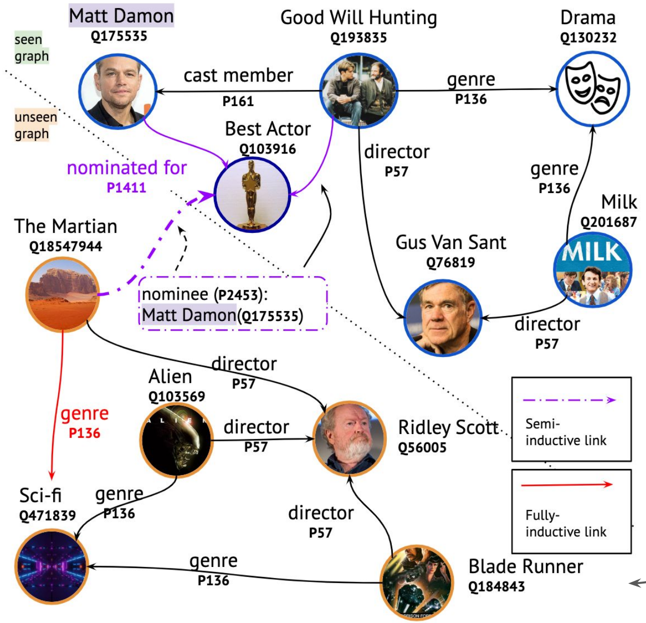
- A GNN encoder for embedding HR KGs and many downstream tasks
- Sparse Qualifier Representation
- As small as 1 qualifier per triple gives boosts
- The more qualifiers - the better

Hyper-relational KGs + Inductive Link Prediction

Transductive | Triples

Inductive | Hyper-relational

Link Prediction



Training Graph

disjoint

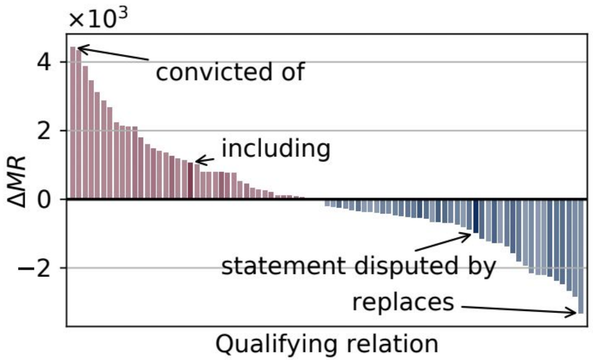
Inference Graph (new entities)

Use RoBERTa encoding of entity descriptions as node features

Ali et al. Improving Inductive Link Prediction Using Hyper-Relational Facts. ISWC 2021, **Best Research Paper Award**

★ Inductive StarE Summary

- Qualifiers help in Inductive Link Prediction
- Features as encoded RoBERTa entity descriptions are good enough
- Some qualifiers give a lot of boost, some do not



Better predictions

Ali et al. Improving Inductive Link Prediction Using Hyper-Relational Facts. ISWC 2021, **Best Research Paper Award**

Complex Logical Query Answering: Why?

Transductive | Triples

Inductive | Hyper-relational

Query Embedding

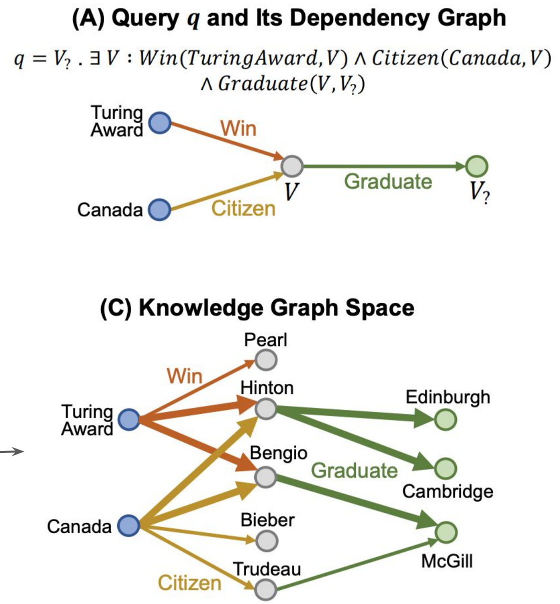
Where did Canadian citizens with Turing Award graduate?

Logical Form

Query (SPARQL)

Database

Typical KGQA pipeline



Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020

Complex Logical Query Answering: Why?

Transductive | Triples

Inductive | Hyper-relational

Query Embedding

Where did Canadian citizens with Turing Award graduate?

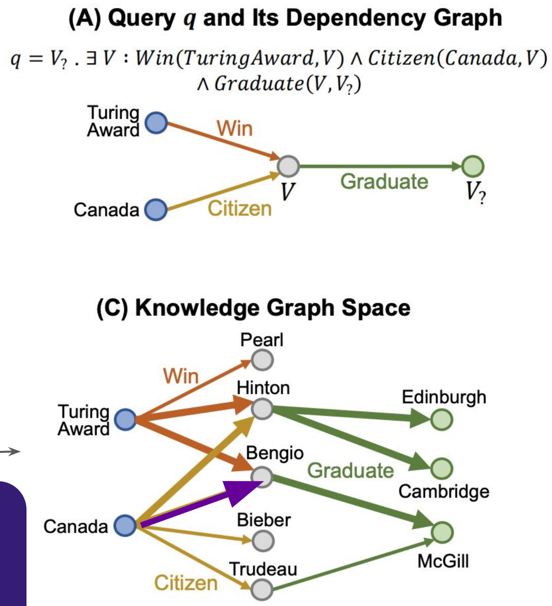
Logical Form

Query (SPARQL)

Database

Typical KGQA pipeline

What if this edge is missing?



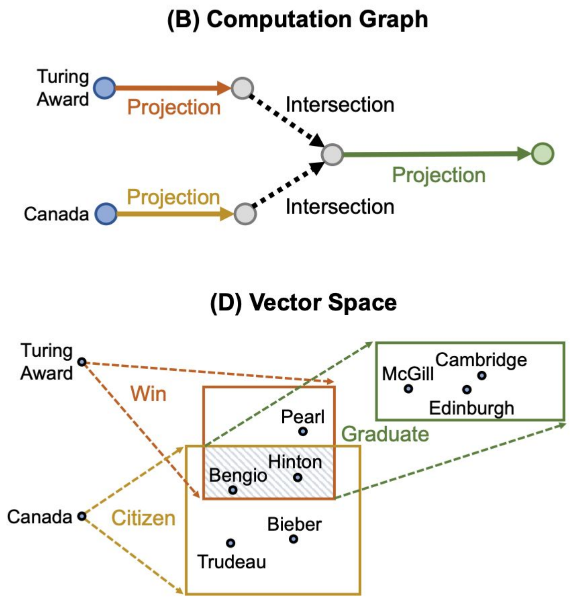
Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020

Complex Logical Query Answering

Transductive | Triples | **Query Embedding**

Inductive | Hyper-relational

- Databases assume KGs are **complete**
    - In reality - they are not
- We want to answer FOL queries over **incomplete** graphs with **neural** operators
- Embed a query in a latent space, MIPS decoder for kNN answers

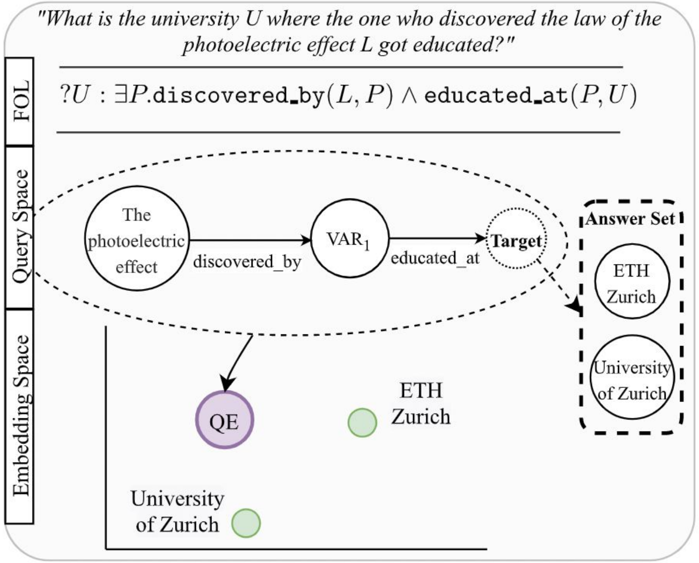


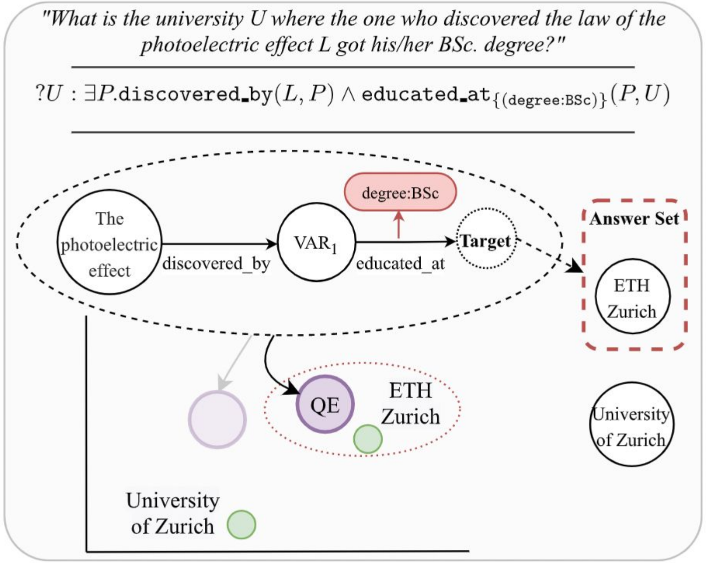
Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020

StarQE: Complex Logical Query Answering on HR KGs

Transductive | Triples | Inductive | Hyper-relational

Query Embedding



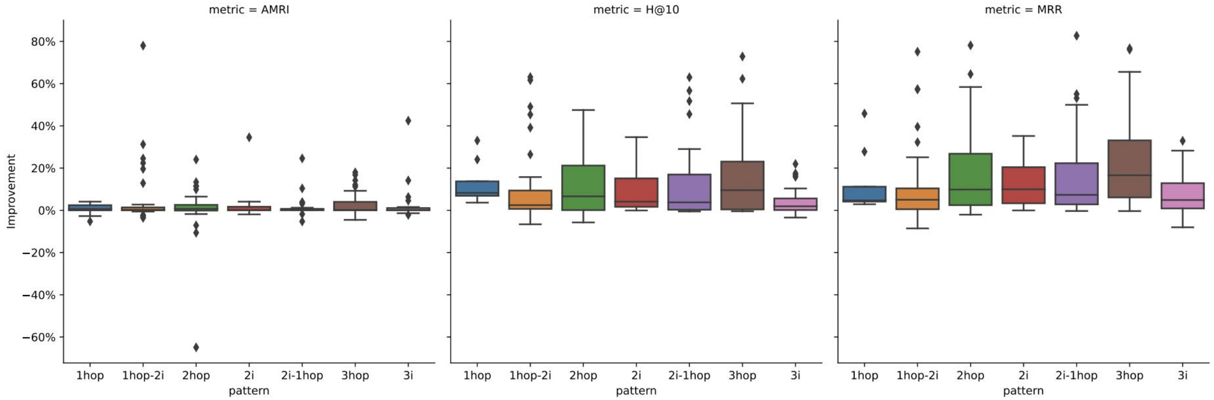


1. Extending FOL to HR KGs
2. Qualifiers help A LOT
3. New query types are enabled

Alivanistos et al. Query Embedding on Hyper-Relational Knowledge Graphs. ICLR 2022

★ StarQE for Logical Queries: Summary

- Extend FOL to hyper-relational graphs with qualifiers
- Enabling new query types (eg, joins over qualifier entities)
- Robust to inner representation: RDF* vs reified RDF
- Qualifiers help A LOT in answering complex queries



Alivanistos et al. Query Embedding on Hyper-Relational Knowledge Graphs. ICLR 2022

Plan

→ Vanilla Triple KGs: Re-cap
→ The New Big Picture
→ Hyper-Relational KGs
→ Inductive Link Prediction with HR KGs
→ Complex Query Answering with HR KGs
→ **Past, Today, Future**

Space of KG Tasks in 2019

Transductive

Triples

SETTING

TASK

Link prediction

Space of KG Tasks Today

Mila

Transductive	Triples	Supervised	Unimodal	Small
Inductive	Hyper-relational	Unsupervised	Multimodal	Large (sampling)

SETTING

TASK

Link prediction

Node classification

Entity Matching

Query Embedding

Theoretical Understanding

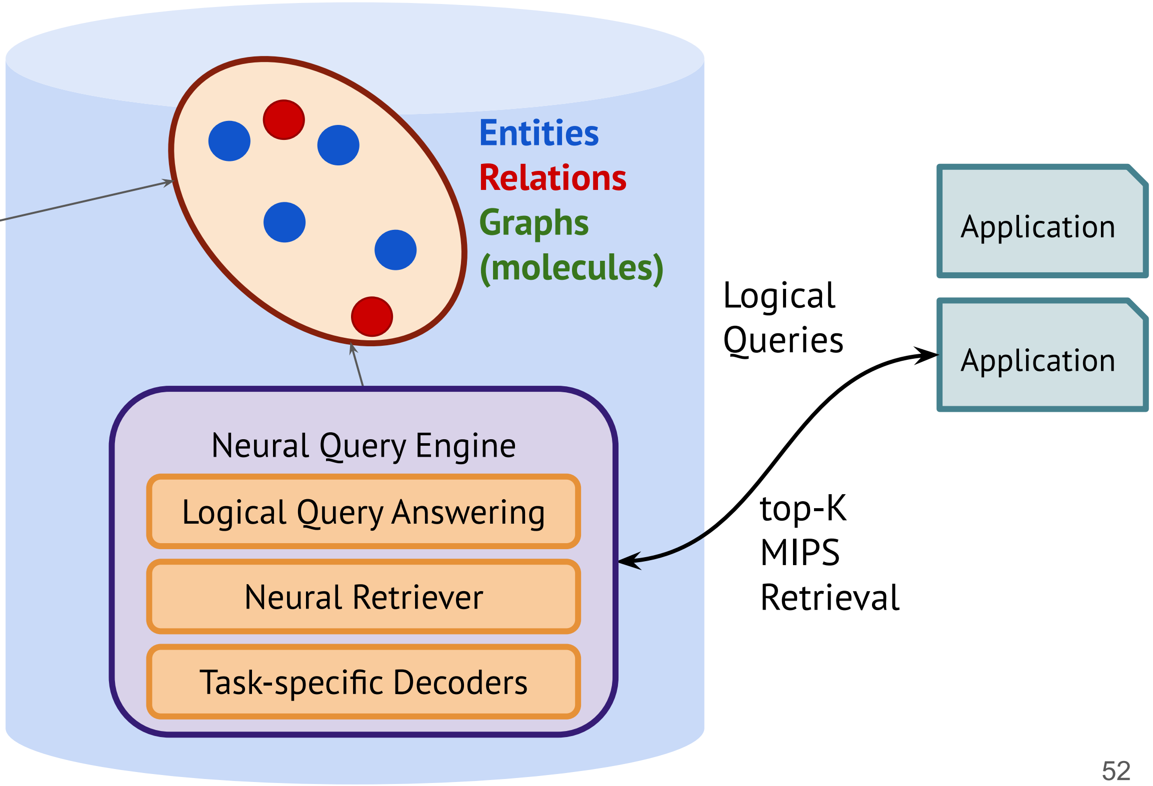
Graph Encoder

Knowledge Graph

Future: Neural Graph Databases

```
(Montreal, location, Quebec)
(Quebec, location, Canada)
(Canada, bordersWith, USA)
```

- No symbolic storage
- Embedding-based storage
- Inferring Missing Links
- Complex Query Answering
- Updatable



Acknowledgements and Papers

1. **Mikhail Galkin**, Priyansh Trivedi, Gaurav Maheshwari, Ricardo Usbeck, Jens Lehmann. **Message Passing for Hyper-Relational Knowledge Graphs.** EMNLP 2020

2. Mehdi Ali, Max Berrendorf, **Mikhail Galkin**, Veronika Thost, Tengfei Ma, Volker Tresp, Jens Lehmann. **Improving Inductive Link Prediction Using Hyper-relational Facts**. ISWC 2021. **Best Research Paper Award**

3. Dimitrios Alivanistos, Max Berrendorf, Michael Cochez, **Mikhail Galkin**. **Query Embedding on Hyper-relational Knowledge Graphs**. ICLR 2022

4. **Mikhail Galkin**, Etienne Denis, Jiapeng Wu, William L Hamilton. **NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs**. ICLR 2022

Q&A Time!

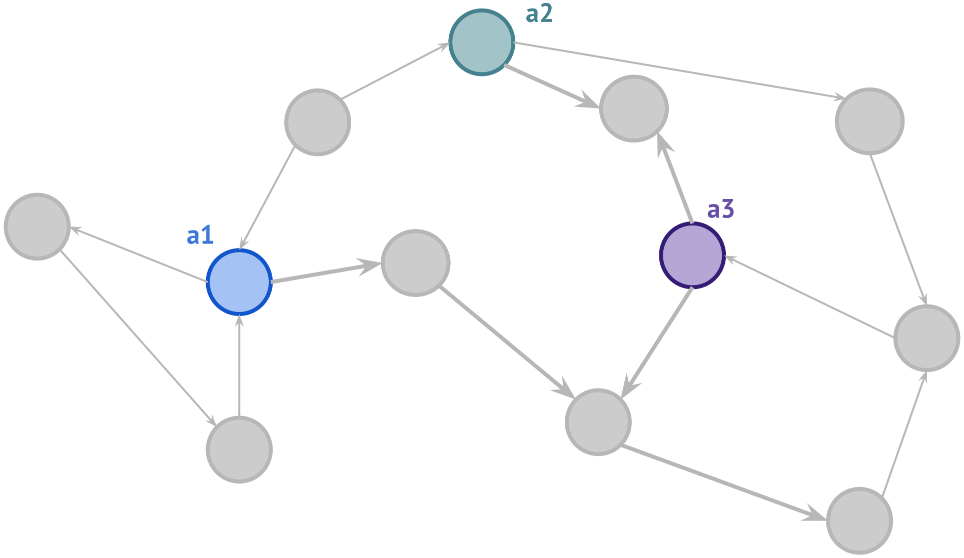
@michael_galkin

@mgalkin

mikhail.galkin@mila.quebec

migalkin.github.io

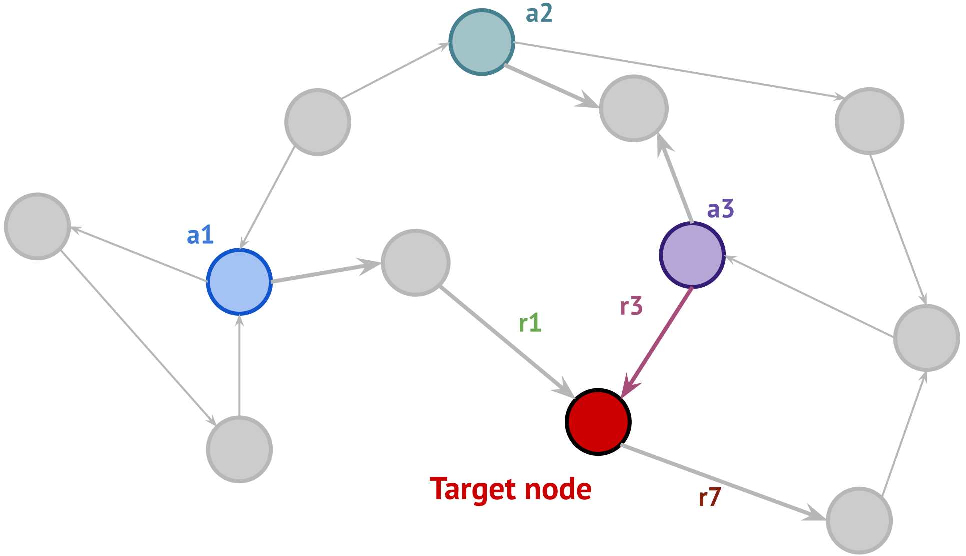
Anchor Node Selection

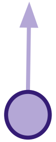


Current strategy:
40% top degrees
40% top PPR
20% random

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

Tokenization



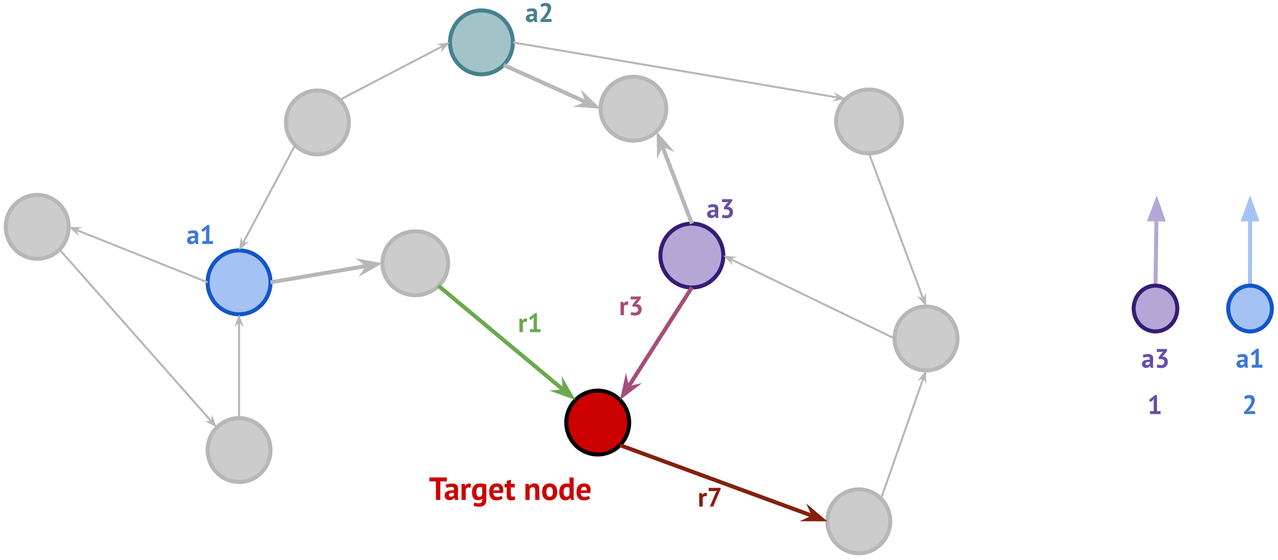


BFS from the target node until we reach |K| anchors

- Can be done in forward pass
- Can be pre-processed and saved

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

Tokenization

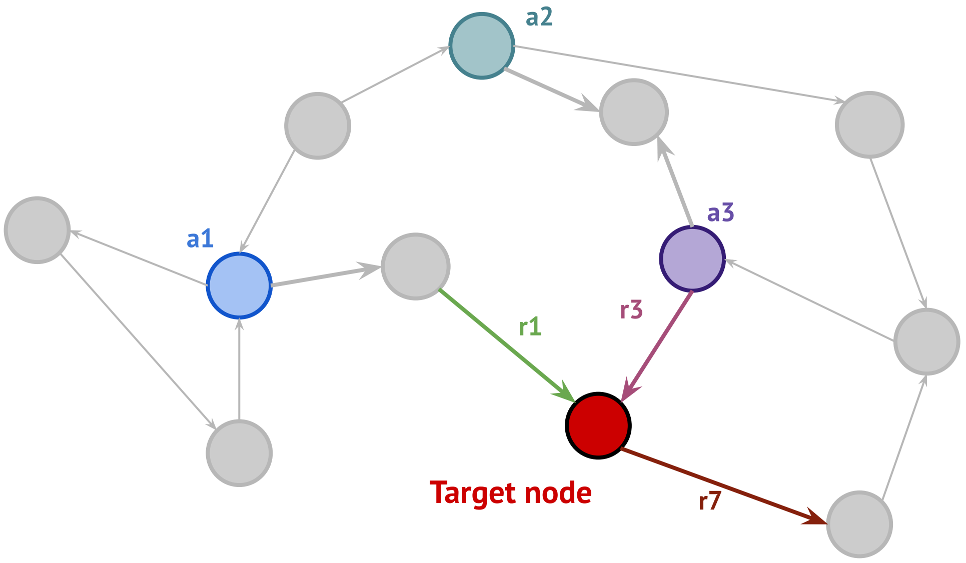


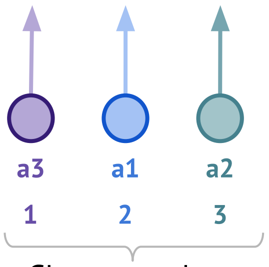
BFS from the target node until we reach |K| anchors

- Can be done in forward pass
- Can be pre-processed and saved

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

Tokenization



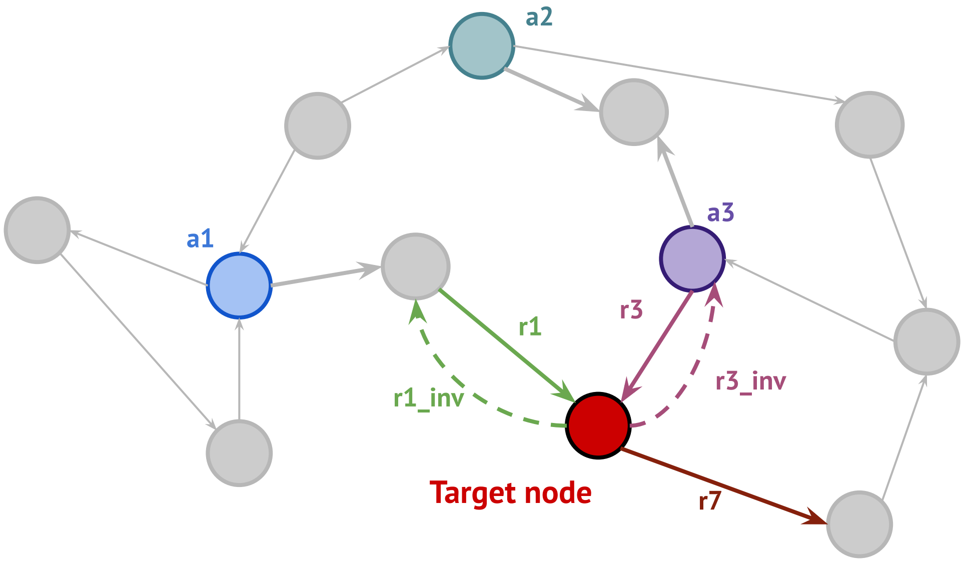


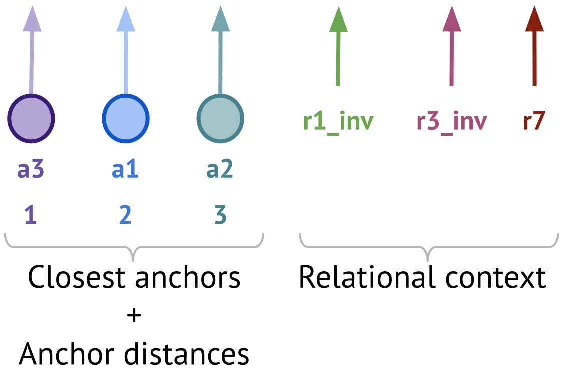
BFS from the target node until we reach |K| anchors

- Can be done in forward pass
- Can be pre-processed and saved

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

Tokenization



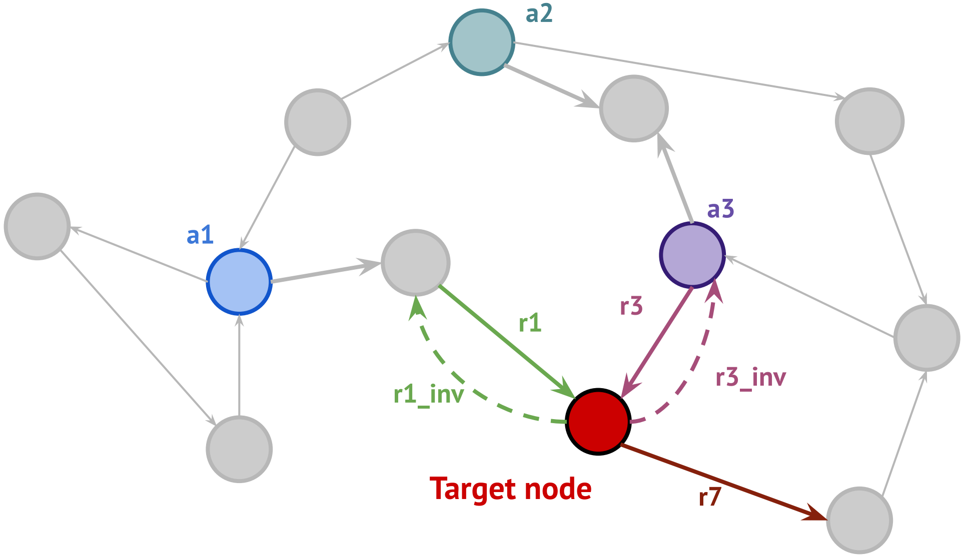


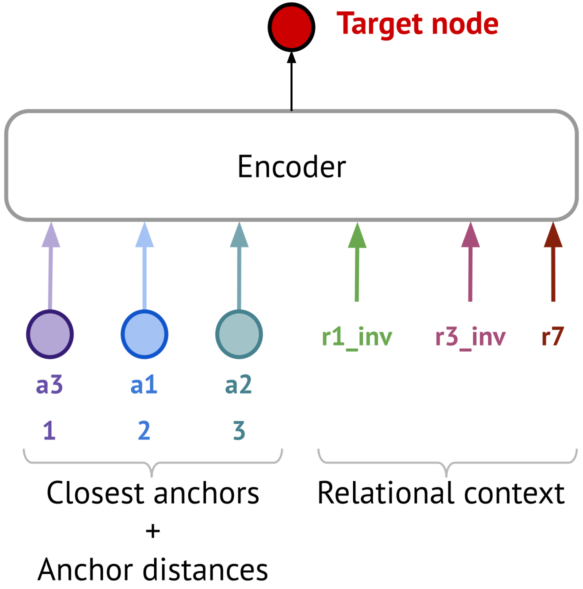
BFS from the target node until we reach |K| anchors

- Can be done in forward pass
- Can be pre-processed and saved

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

Tokenization



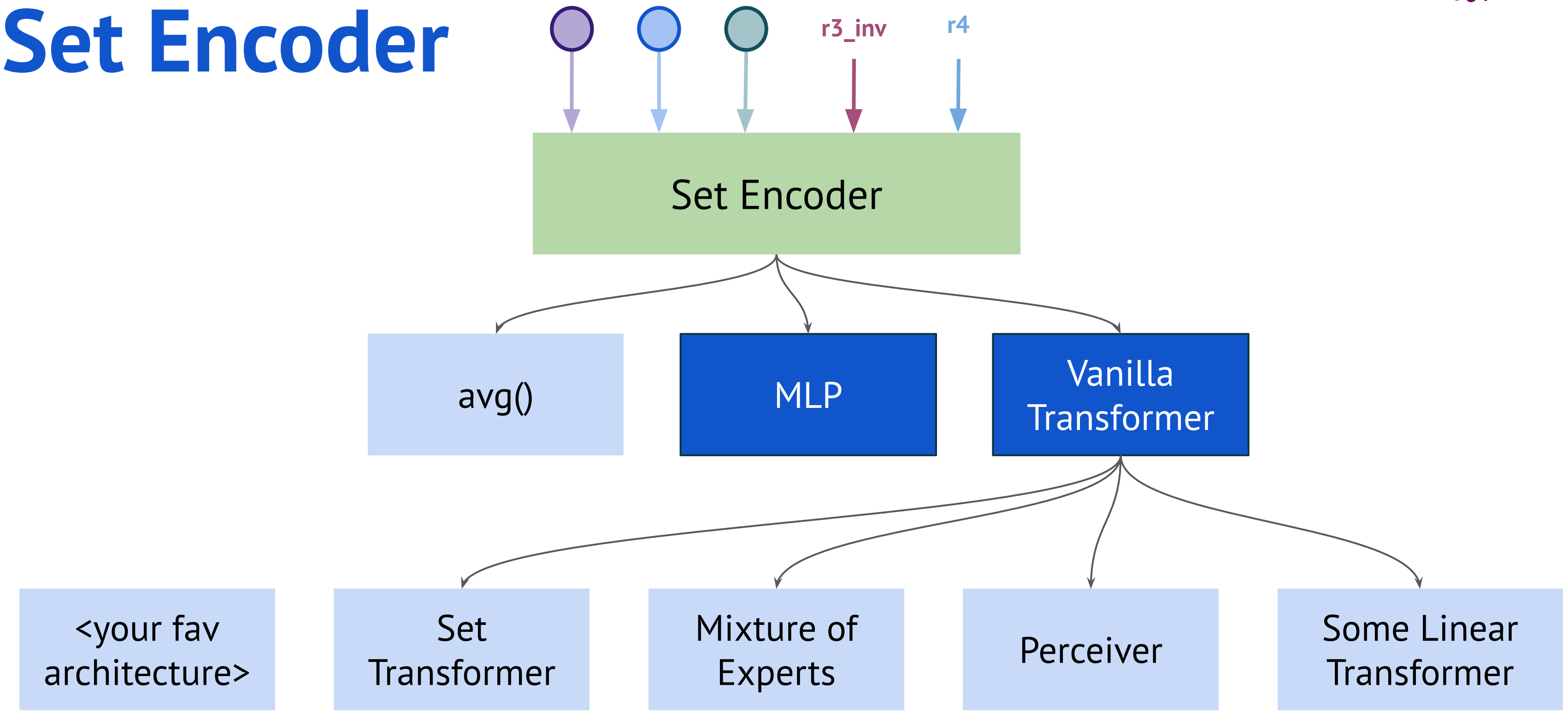


BFS from the target node until we reach |K| anchors

- Can be done in forward pass
- Can be pre-processed and saved

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

Set Encoder



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

Transductive Link Prediction

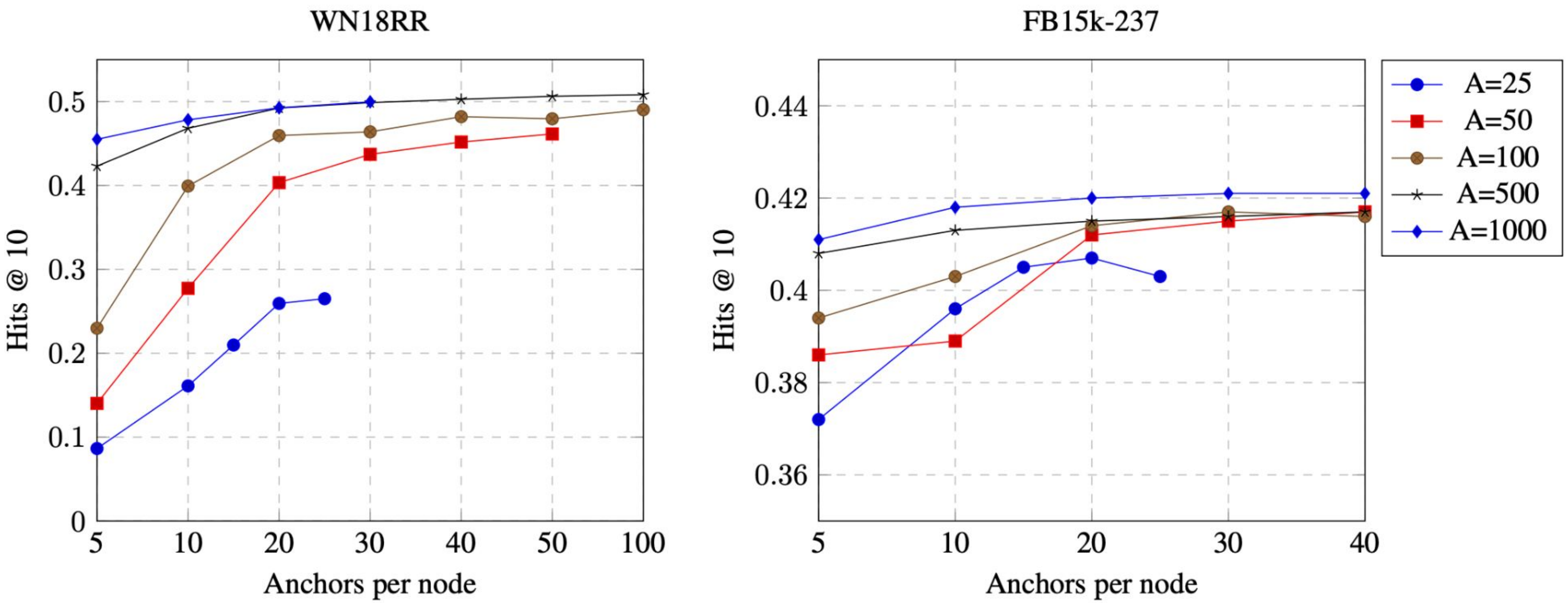


Figure 2: Combinations of total anchors A and anchors per node. Denser FB15k-237 saturates faster on smaller A while sparse WN18RR saturates at around 500 anchors.

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

NodePiece Experiments: Summary

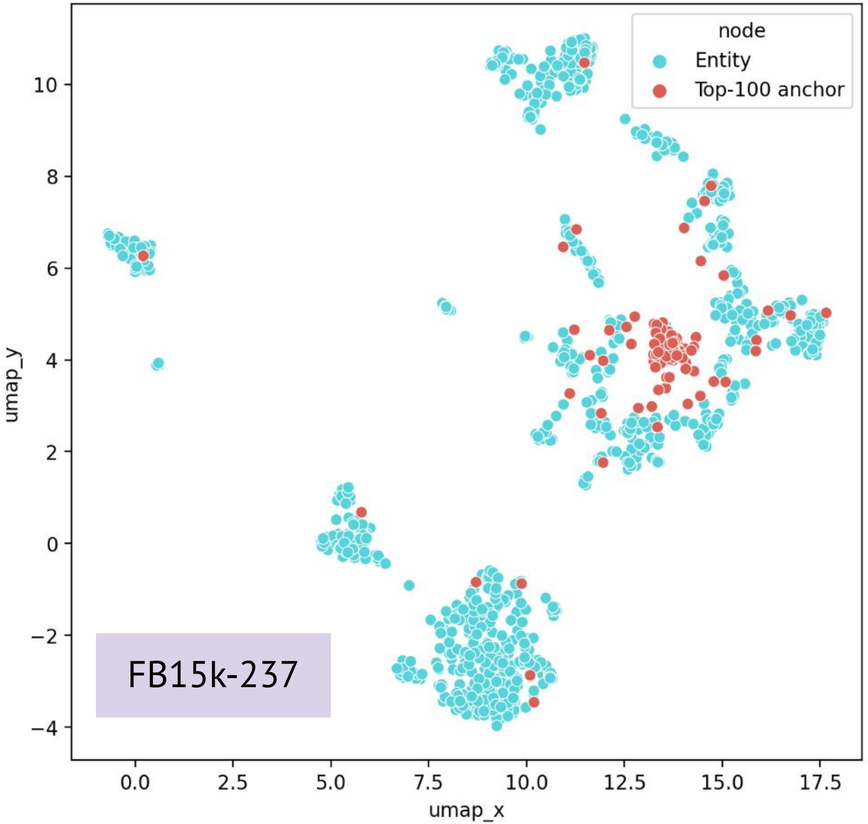
- 👌 **10x** fewer parameters while retaining **90%** of transductive LP
- 📈 **2x** better compared to shallow models of similar #params
- ⚓ Relation Prediction and Node Classification: no anchors is better!
- 🗃️ **Inductive out-of-the-box** and very competitive

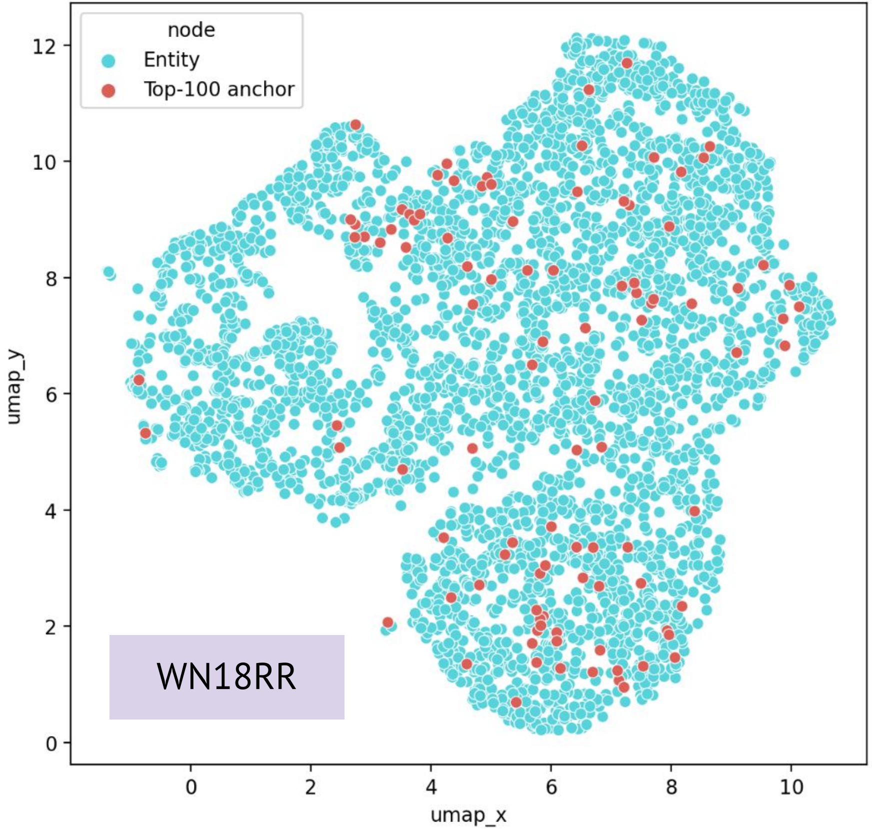
Table 6: Node classification results. $|V|$ denotes vocabulary size (anchors + relations), #P is a total parameter count (millions).

			WD50K (5% labeled)			WD50K (10% labeled)		
	$\|V\|$	#P (M)	ROC-AUC	PRC-AUC	Hard Acc	ROC-AUC	PRC-AUC	Hard Acc
MLP	46k + 1k	4.1	0.503	0.016	0.001	0.510	0.017	0.002
CompGCN	46k + 1k	4.4	0.836	0.280	0.176	0.834	0.265	0.161
NodePiece + GNN	50 + 1k	0.75	0.981	0.443	0.513	0.981	0.450	0.516
- no rel. context	50 + 1k	0.64	0.982	0.446	0.534	0.982	0.449	0.530
- no distances	50 + 1k	0.74	0.981	0.448	0.516	0.981	0.448	0.513
- no anchors, rels only	0 + 1k	0.54	0.984	0.453	0.532	0.984	0.456	0.533

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

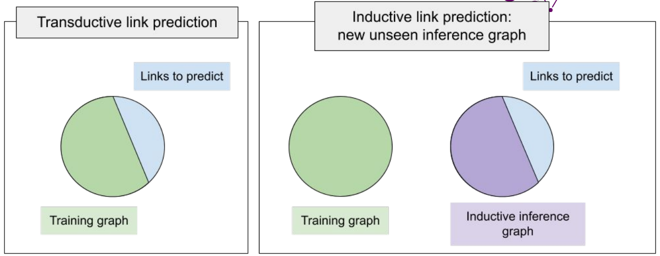
Visualizations: Anchors + Entities





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

Inductive Link Prediction



Inference graphs are disjoint with training (new nodes)
NodePiece + CompGCN encoder = SOTA on many tasks on relation-rich graphs

Table 5: Inductive link prediction results, Hits@10. Best results are in **bold**, second best are underlined. † results taken from Teru et al. (2020). NBFNet results taken from Zhu et al. (2021).

Class	Method	FB15k-237				WN18RR				NELL-995			
		V1	V2	V3	V4	V1	V2	V3	V4	V1	V2	V3	V4
Path	Neural LP †	0.529	0.589	0.529	0.559	0.744	0.689	0.462	0.671	0.408	0.787	0.827	0.806
	DRUM †	0.529	0.587	0.529	0.559	0.744	0.689	0.462	0.671	0.194	0.786	0.827	0.806
	RuleN †	0.498	0.778	0.877	0.856	0.809	0.782	0.534	0.716	0.535	0.818	0.773	0.614
GNN	GraIL †	0.642	0.818	0.828	0.893	0.825	0.787	0.584	0.734	0.595	**0.933**	0.914	0.732
	NBFNet	0.834	**0.949**	**0.951**	**0.960**	**0.948**	**0.905**	**0.893**	**0.890**	-	-	-	-
	NP + CompGCN	**0.873**	**0.939**	**0.944**	**0.949**	0.830	**0.886**	0.785	0.807	**0.890**	0.901	**0.936**	**0.893**

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