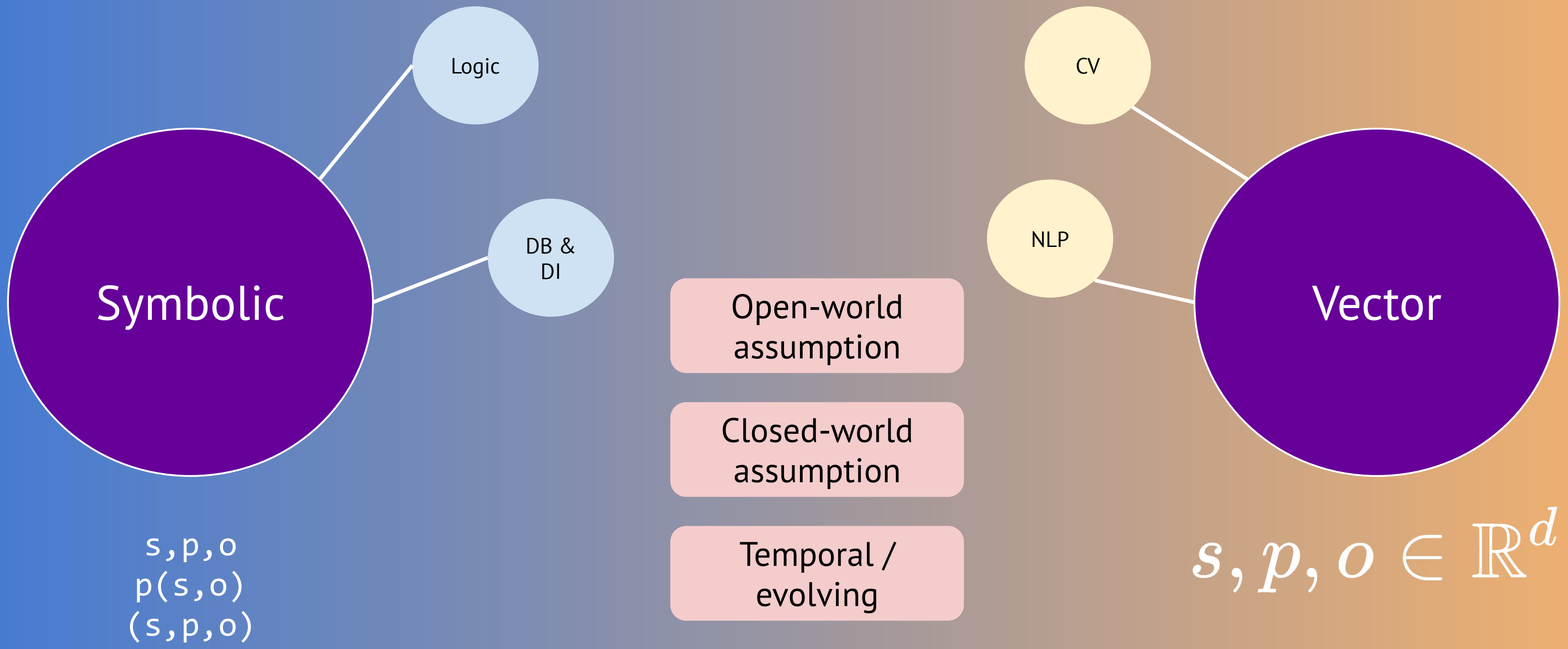
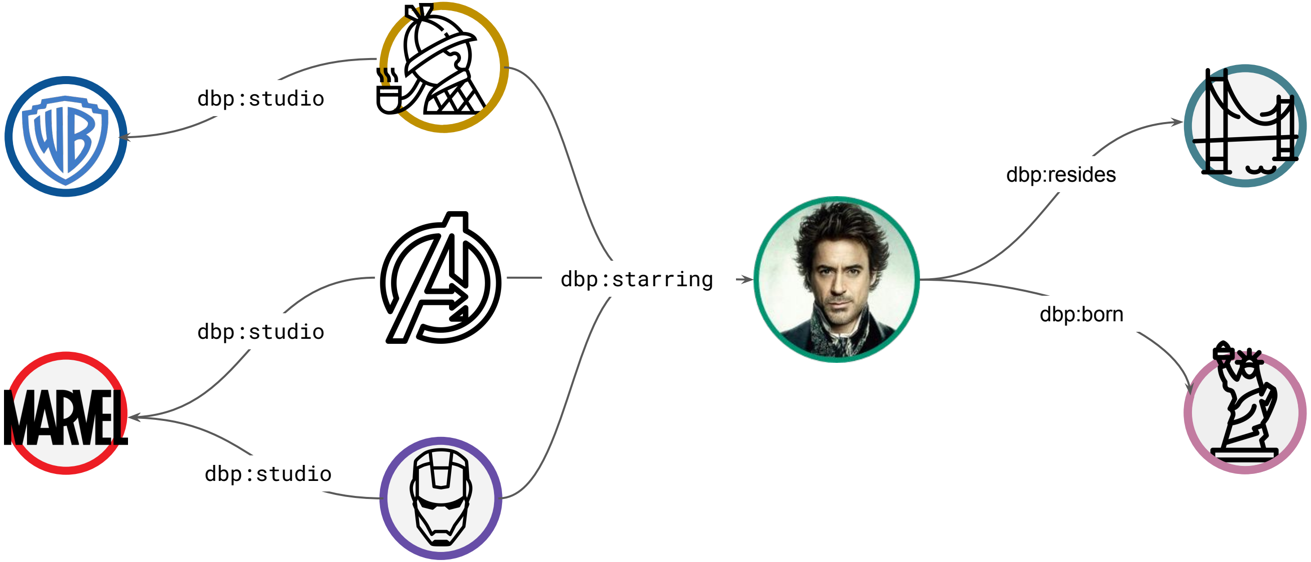
Placing Knowledge Graphs In Graph ML

On representation of Knowledge Graphs



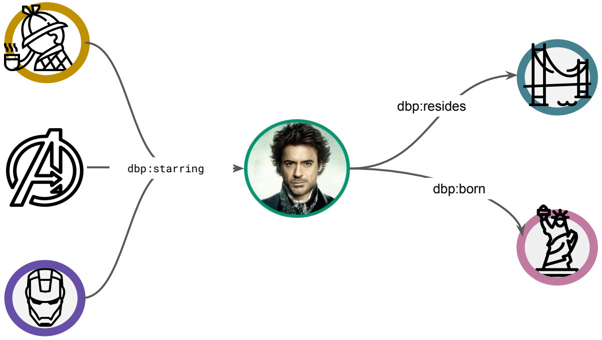
Vanilla Triples



RDJ	dbp:resides	SF .	Avengers	dbp:studio	Marvel .
RDJ	dbp:born	NY .	Avengers	dbp:starring	RDJ .
Sherlock_Holmes	dbp:studio	WB .	Iron_Man	dbp:studio	Marvel .
Sherlock_Holmes	dbp:starring	RDJ.	Iron_Man	dbp:starring	RDJ .

KGs in Graph ML

Original Graph	Embeddings	Downstream Task



Graph Encoder

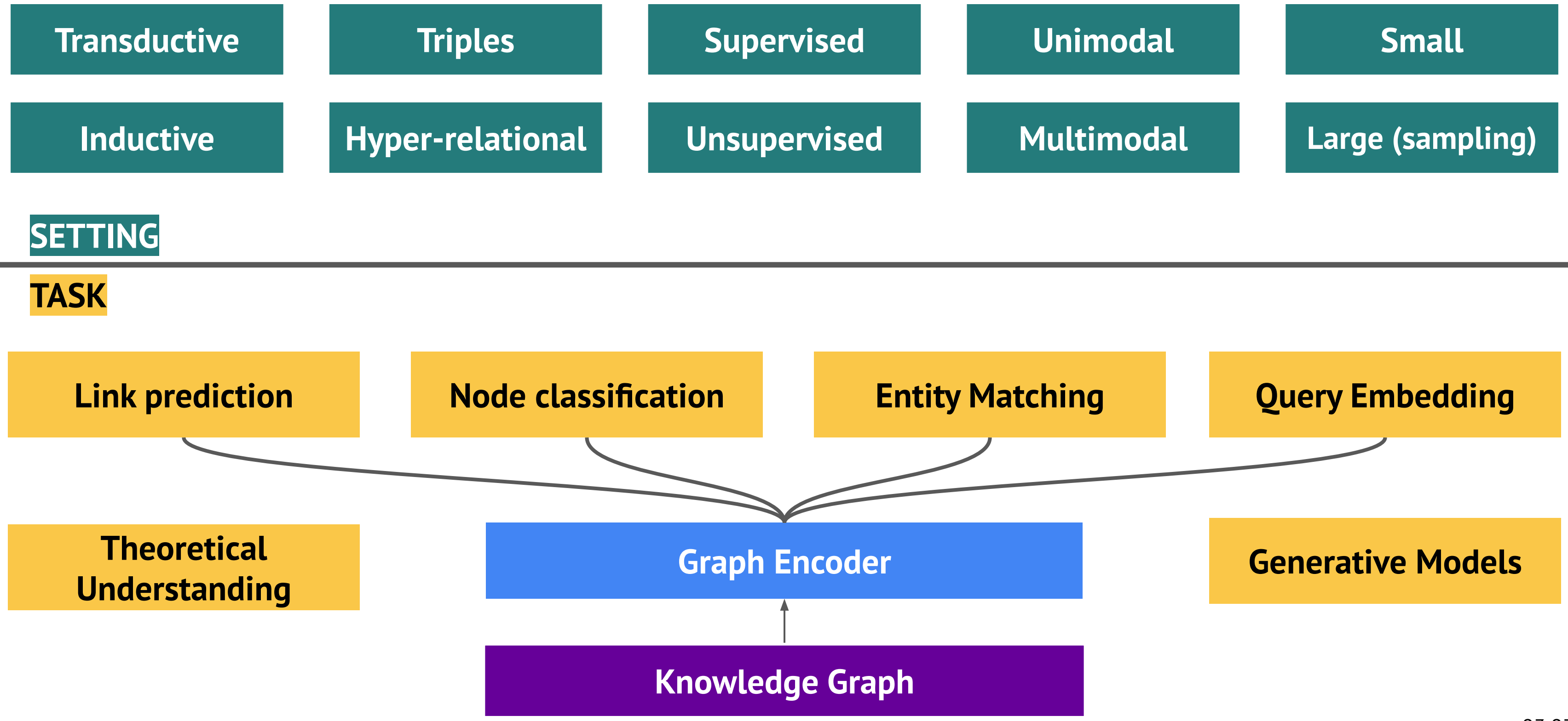
- Link prediction
- Node classification
- Graph classification

Often

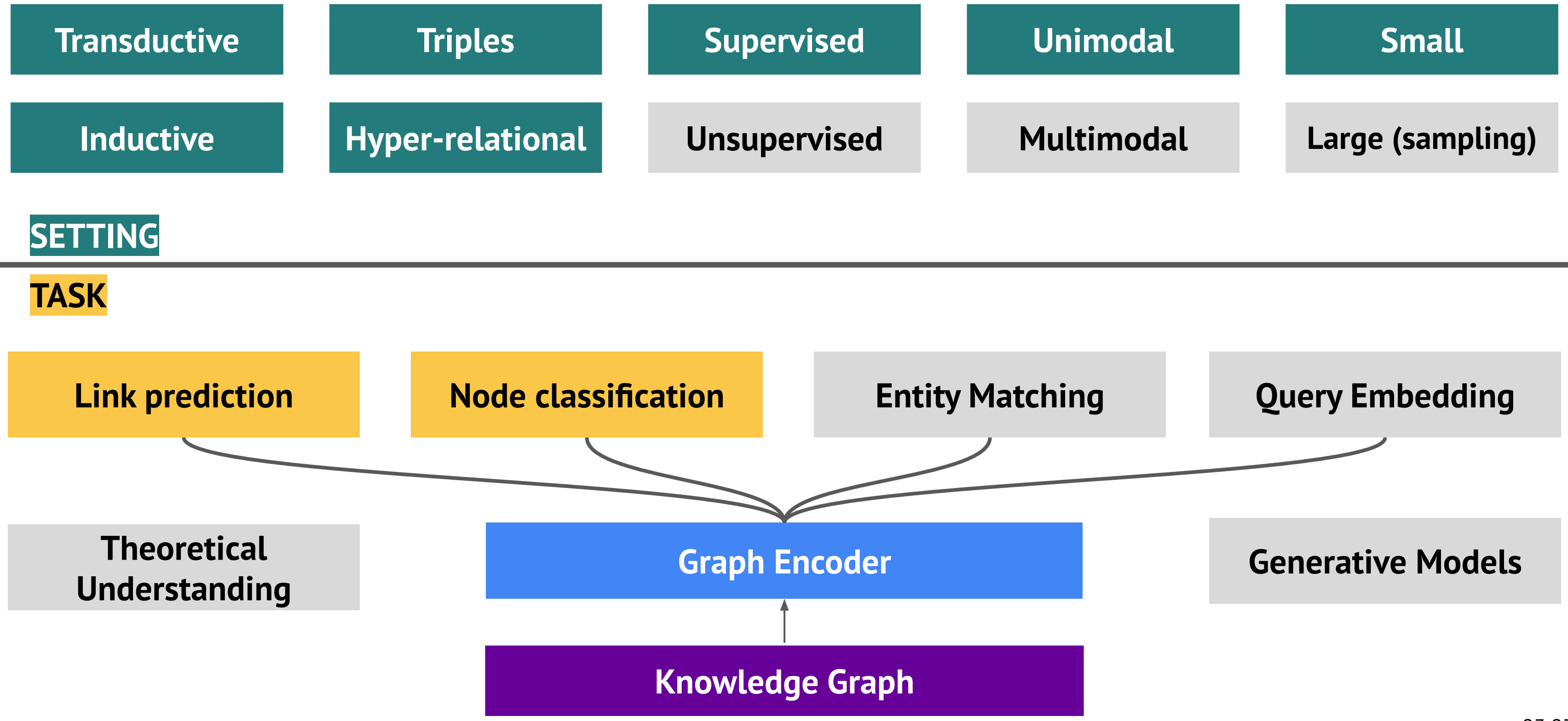
Rare

Almost never 👀

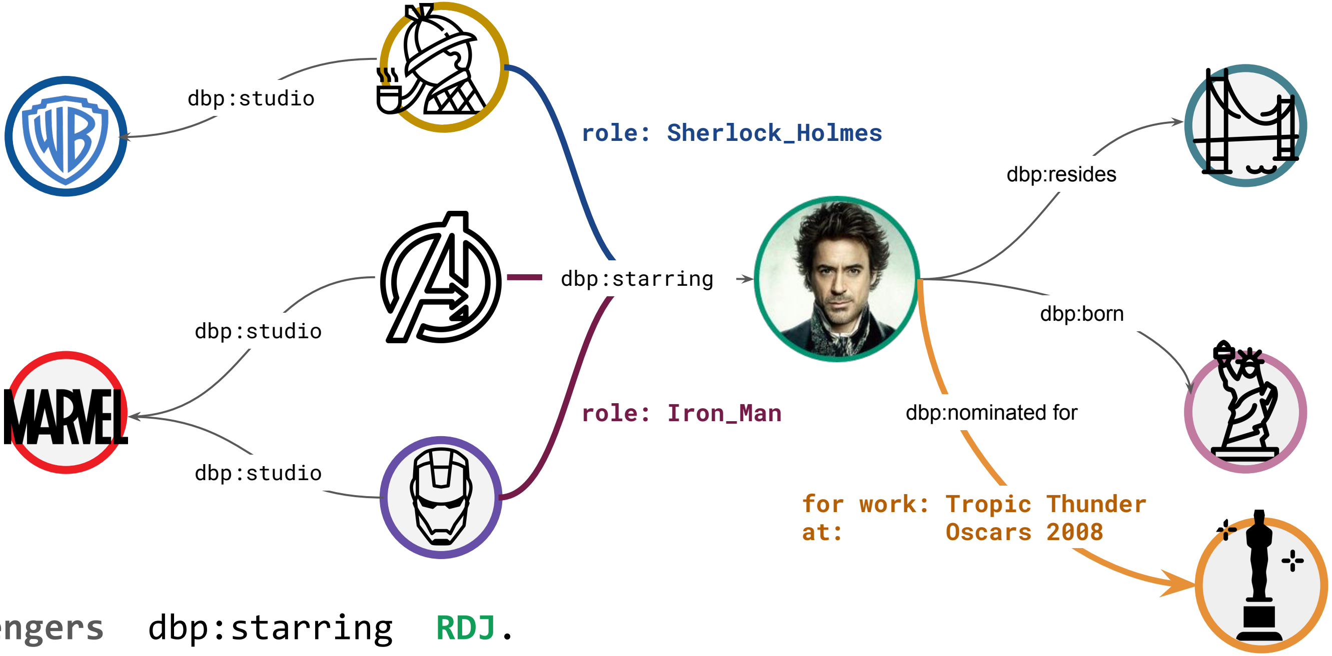
KGs in Graph ML: Big Picture in $\mathbb{R}^5$



Hyper-Relational KGs

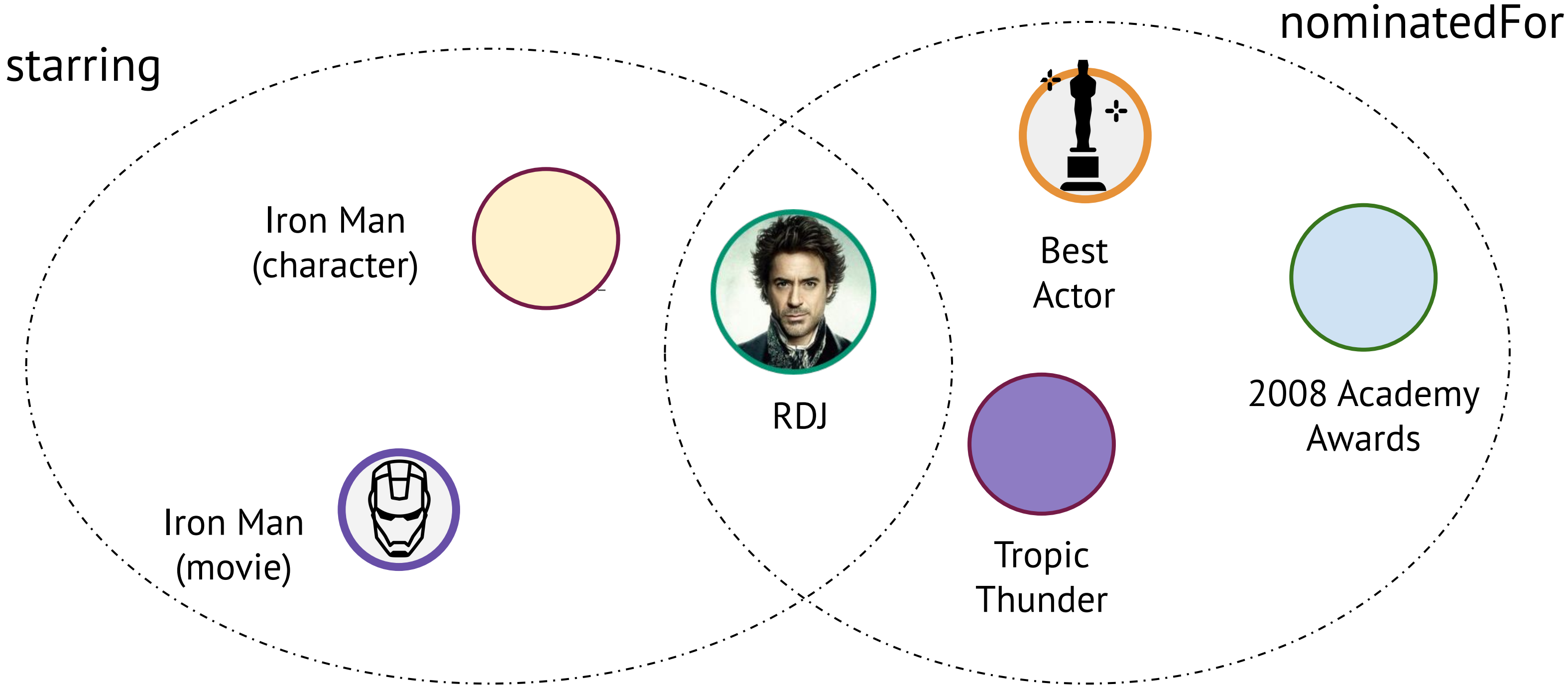


Hyper-Relational KGs: RDF* and SPARQL*




```
The_Avengers  dbp:starring  RDJ.
<< The_Avengers dbp:starring RDJ >> role Iron_Man .
<< RDJ dbp:nominated_for Oscar >> for_work Tropic_Thunder;
                                  at       Oscars_2008 .
```

Hyper-Relational != Hypergraphs



Immediate loss of the fine-grained predicates & e-r attribution

Multirelational GNN Encoders for KGs

$$\mathbf{h}_v^{(k)} = f\left(\sum_{u \in \mathcal{N}(v)} \mathbf{W}^{(k)} \mathbf{h}_u^{(k-1)}\right)$$

Vanilla GCN: no relations

$$\mathbf{h}_v^{(k)} = f\left(\sum_{(u,r) \in \mathcal{N}(v)} \mathbf{W}_r^{(k)} \mathbf{h}_u^{(k-1)}\right)$$

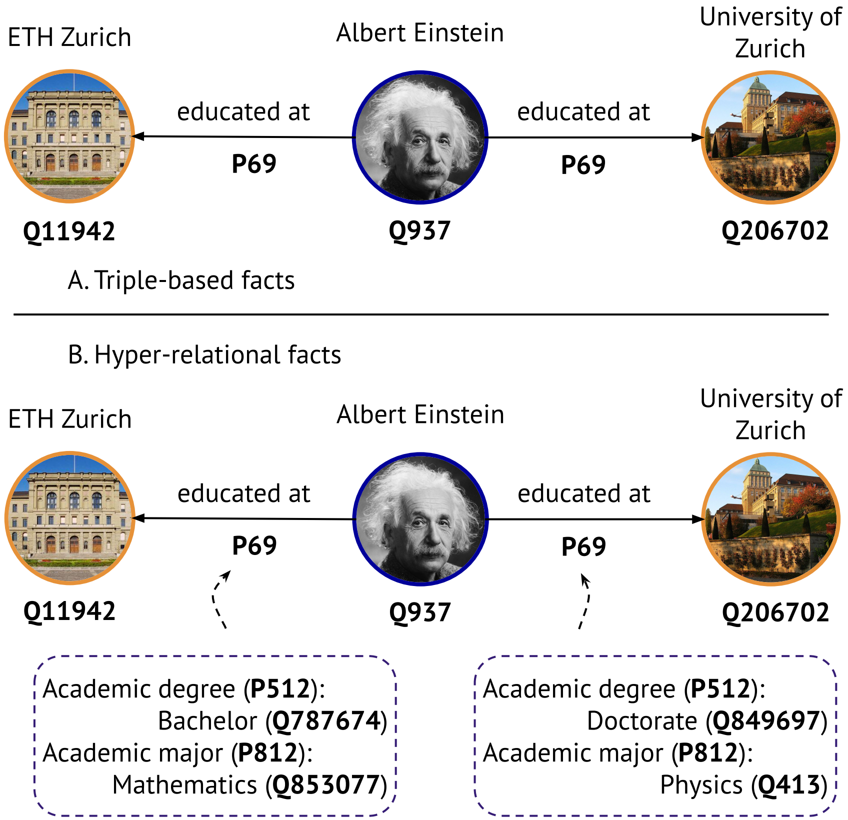
R-GCN [1]: a whole matrix W per relation

$$\boldsymbol{h}_v = f\left(\sum_{(u,r) \in \mathcal{N}(v)} \boldsymbol{W}_{\lambda(r)} \phi(\boldsymbol{x}_u, \boldsymbol{z}_r)\right)$$

CompGCN [2]: a vector z_r per relation +
composition of (h,r) +
only 3 different W: input/output/self-loop

[1] Schlichtkrull et al. Modeling Relational Data with Graph Convolutional Networks. ESWC 2018
[2] Vashishth et al. Composition-Based Multi-Relational Graph Convolutional Networks. ICLR 2020

Embedding Hyper-Relational KGs

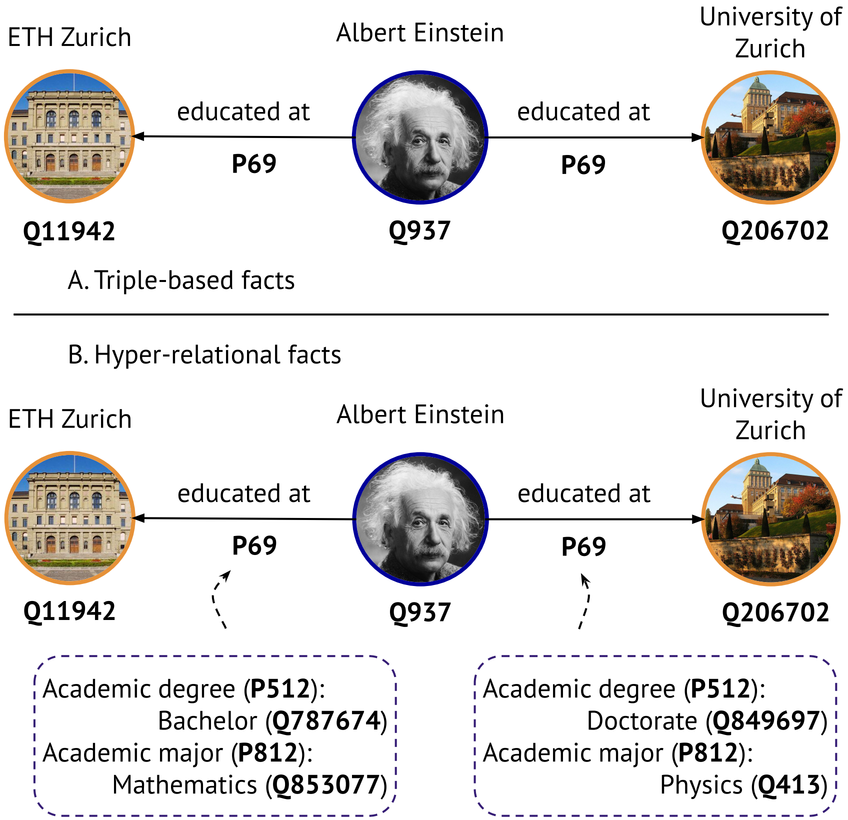


$$\boldsymbol{h}_v = f\left(\sum_{(u,r)\in\mathcal{N}(v)} \boldsymbol{W}_{\lambda(r)} \phi(\boldsymbol{x}_u, \boldsymbol{z}_r) \right)$$

?

- Qualifying relations and entities can be used as main terms in other facts
- Not all facts might have qualifiers

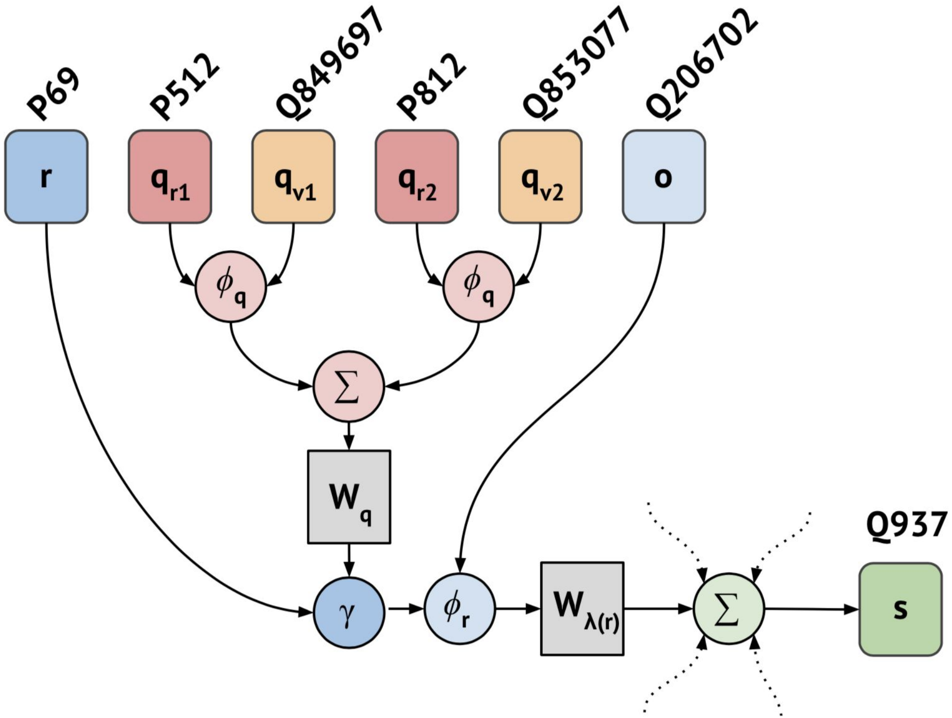
Embedding Hyper-Relational KGs

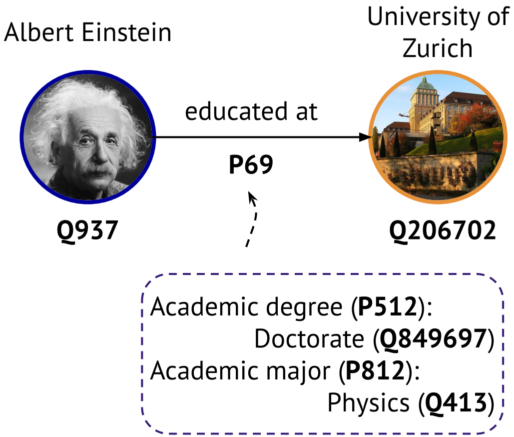


$$\boldsymbol{h}_v = f\left(\sum_{(u,r)\in\mathcal{N}(v)} \boldsymbol{W}_{\lambda(r)}\phi(\boldsymbol{x}_u, \boldsymbol{z}_r)\right)$$

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

Embedding Hyper-Relational KGs





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

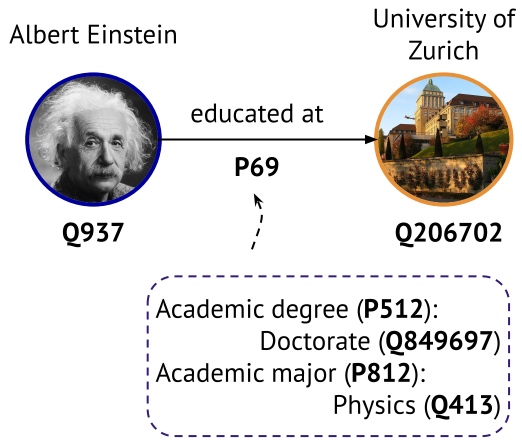
Encoding Hyper-Relational KGs

Sparse Triple Representation

s	Q937	...	...
o	Q206702	...	...
r	P69	...	...
index	*k*	*k+1*	*k+2*

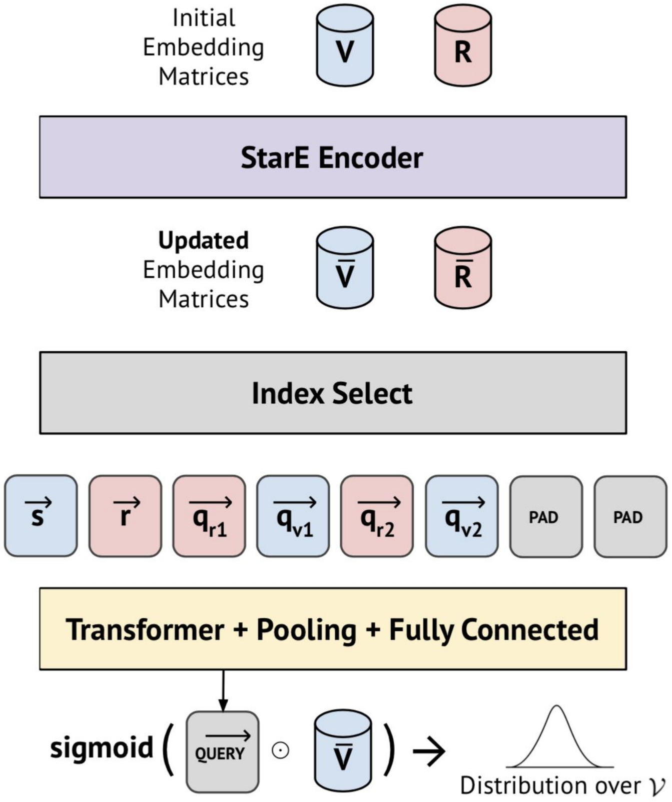
Sparse Qualifier Representation

index	*k*	*k*	...
qr	P812	P512	...
qv	Q413	Q849697	...



$O(|\mathcal{E}| + |\mathcal{Q}|)$ Space complexity

Decoders for Downstream Tasks

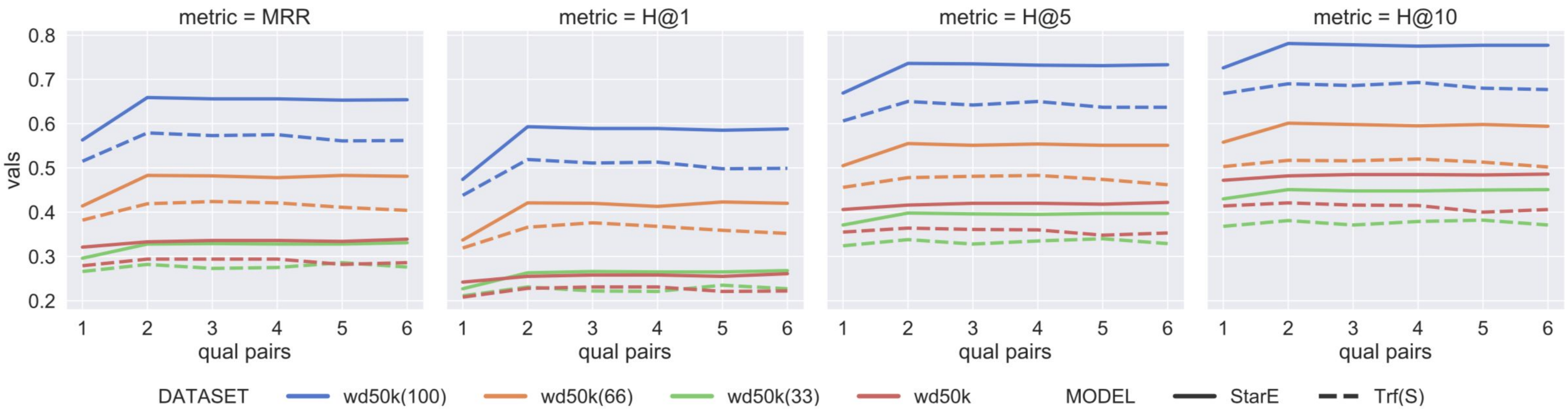


← A decoder for link prediction

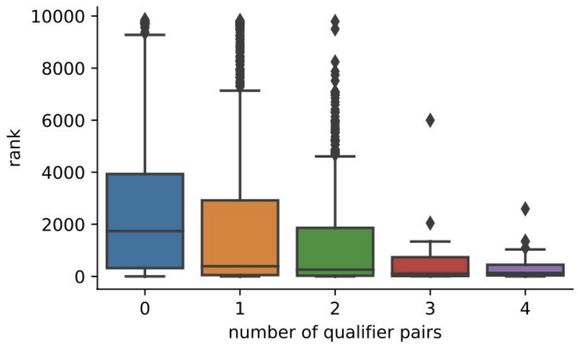
- Datasets issue:
  - No such graphs in OGB
  - Very few LP datasets

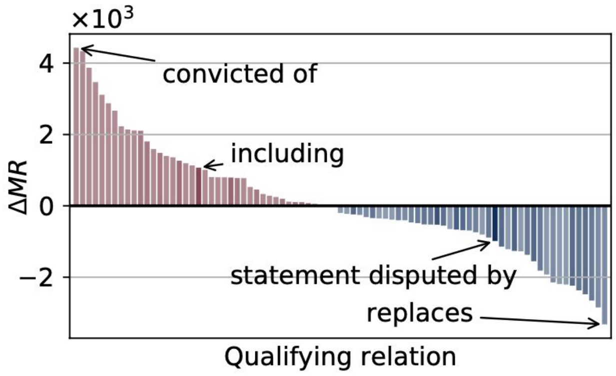
Hyper-Relational KGs: Link Prediction

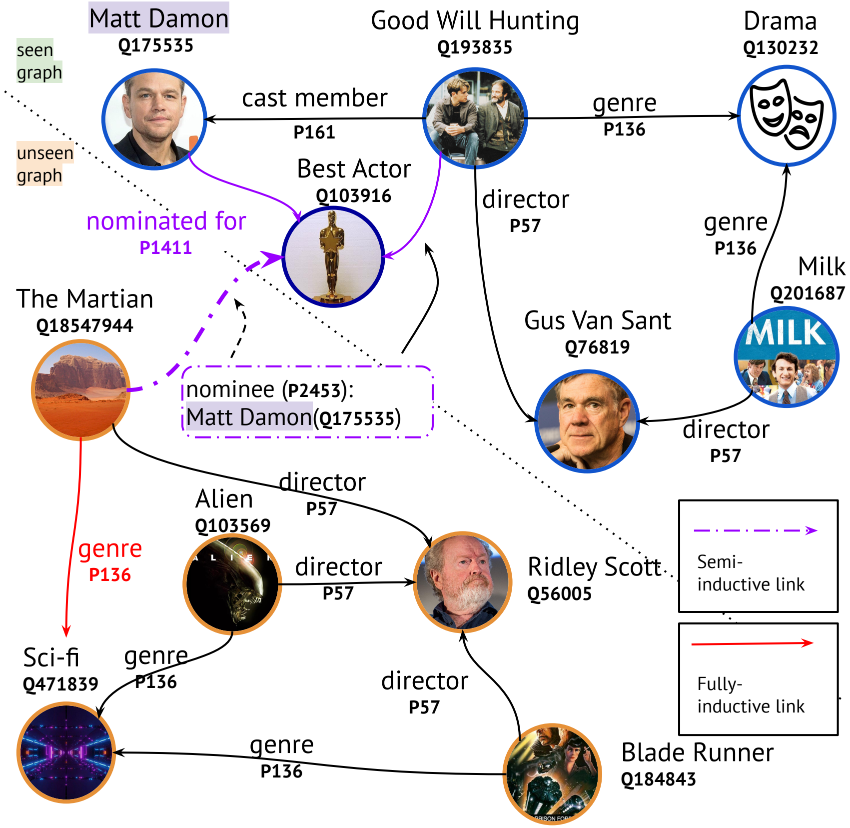
Exp #	Dataset → / Method ↓	WD50K			WD50K (33)			WD50K (66)			WD50K (100)		
		MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
4	Baseline (Transformer (T))	0.275	0.207	0.404	0.218	0.158	0.334	0.270	0.197	0.417	0.351	0.261	0.530
4	StarE (T) + Transformer(T)	0.308	0.228	0.465	0.246	0.173	0.388	0.297	0.212	0.470	0.380	0.276	0.584
1,2,4	Baseline (Transformer (H))	0.286	0.222	0.406	0.276	0.227	0.371	0.404	0.352	0.502	0.562	0.499	0.677
1,2,4	StarE (H) + Transformer(H)	**0.349**	**0.271**	**0.496**	**0.331**	**0.268**	**0.451**	**0.481**	**0.420**	**0.594**	**0.654**	**0.588**	**0.777**



Inductive Scenarios





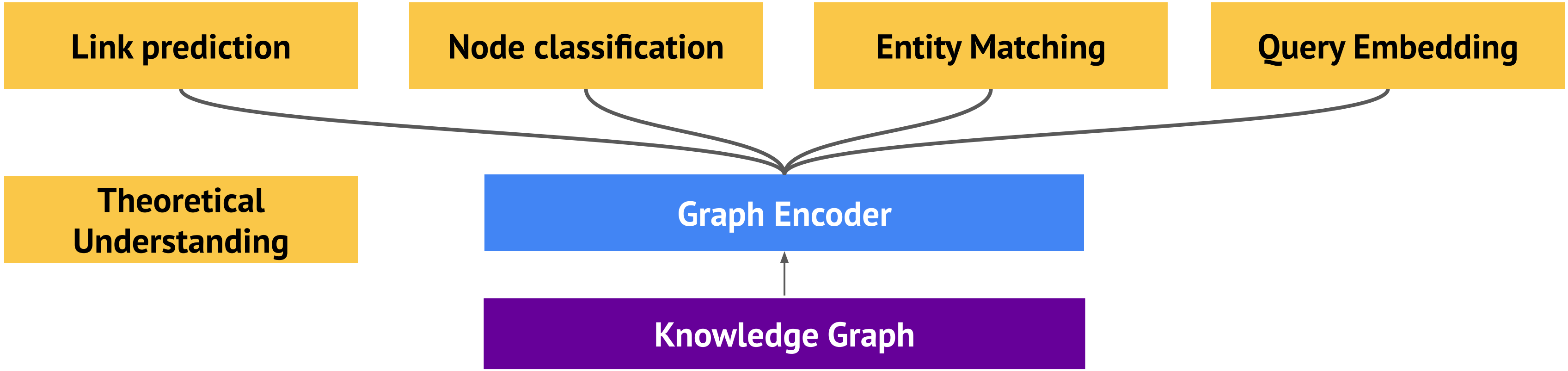


Big Picture in $\mathbb{R}^5$

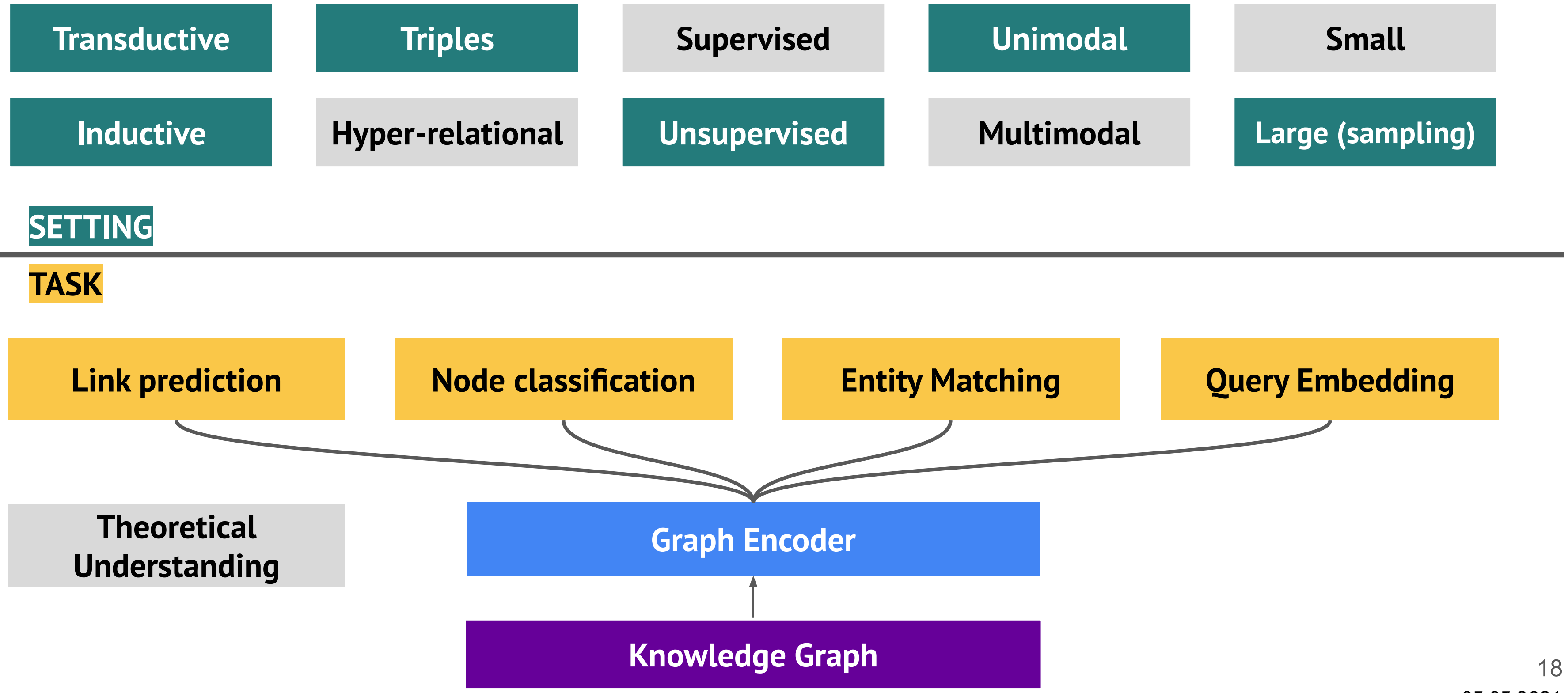
Transductive	Triples	Supervised	Unimodal	Small
Inductive	Hyper-relational	Unsupervised	Multimodal	Large (sampling)

SETTING

TASK



Big Picture in $\mathbb{R}^5$ - Scaling Up



The ImageNet Moment for KGs

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

- Graph ML
- NLP
- Vision

The ImageNet Moment for KGs

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

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

PyTorch BigGraph

~200 GB

- Graph ML
- NLP
- Vision

OGB WikiKG: Just 2.5M nodes

Leaderboard for ogbl-wikikg2

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

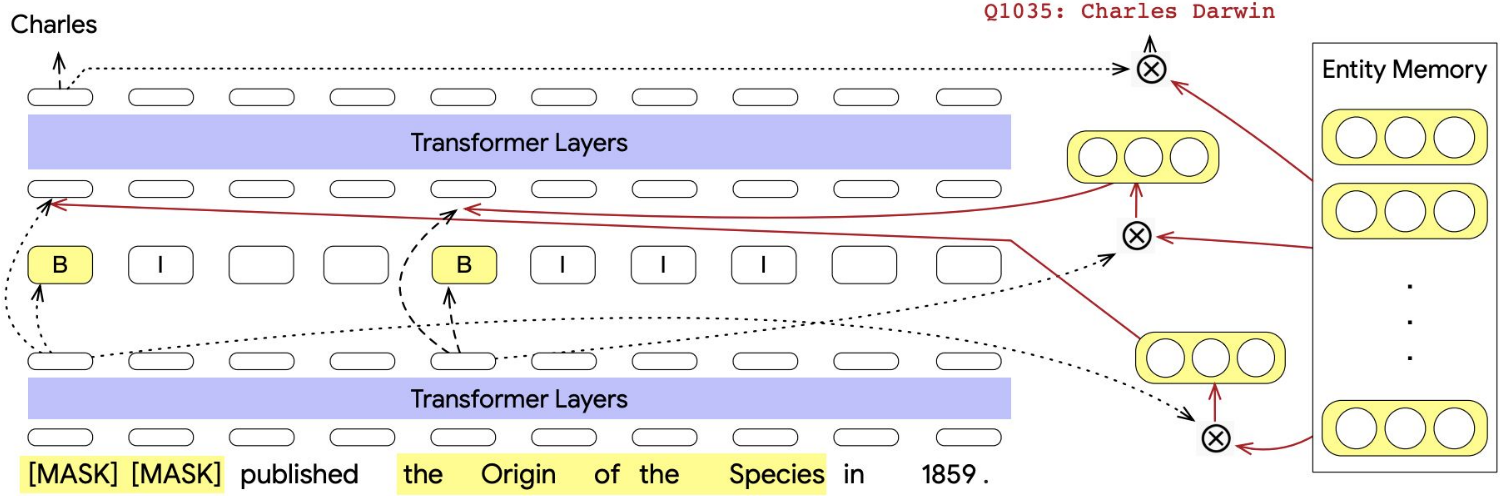
Package: >=1.2.4

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

BERT-Large is ~340M params

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

Explicit embeddings in LMs



Wikidata-scale (100M nodes, 2B edges):

[100M, 200d]

Fevry et al. Entities as Experts: Sparse Memory Access with Entity Supervision. EMNLP 2020

Sparsifying / Tokenizing KGs

BERT-Large (340M)

Transformer
~300M

Vocab: 30K x 1024d

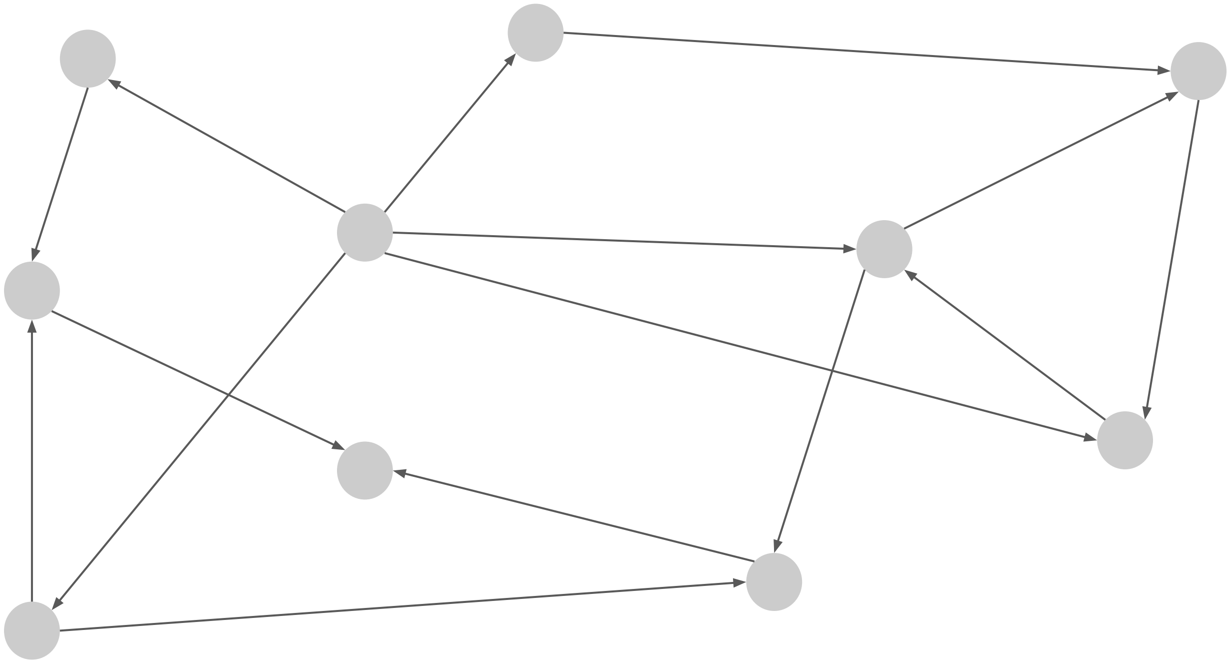
KGE (1250M)

Vocab:
2.5M x 500d

Sparsifying / Tokenizing KGs

BERT-Large (340M)	**Tokenized KG (100M)**	KGE (1250M)
Transformer ~300M	Encoder	Vocab: 2.5M x 500d
Vocab: 30K x 1024d	Vocab: 30K entities, All relations	

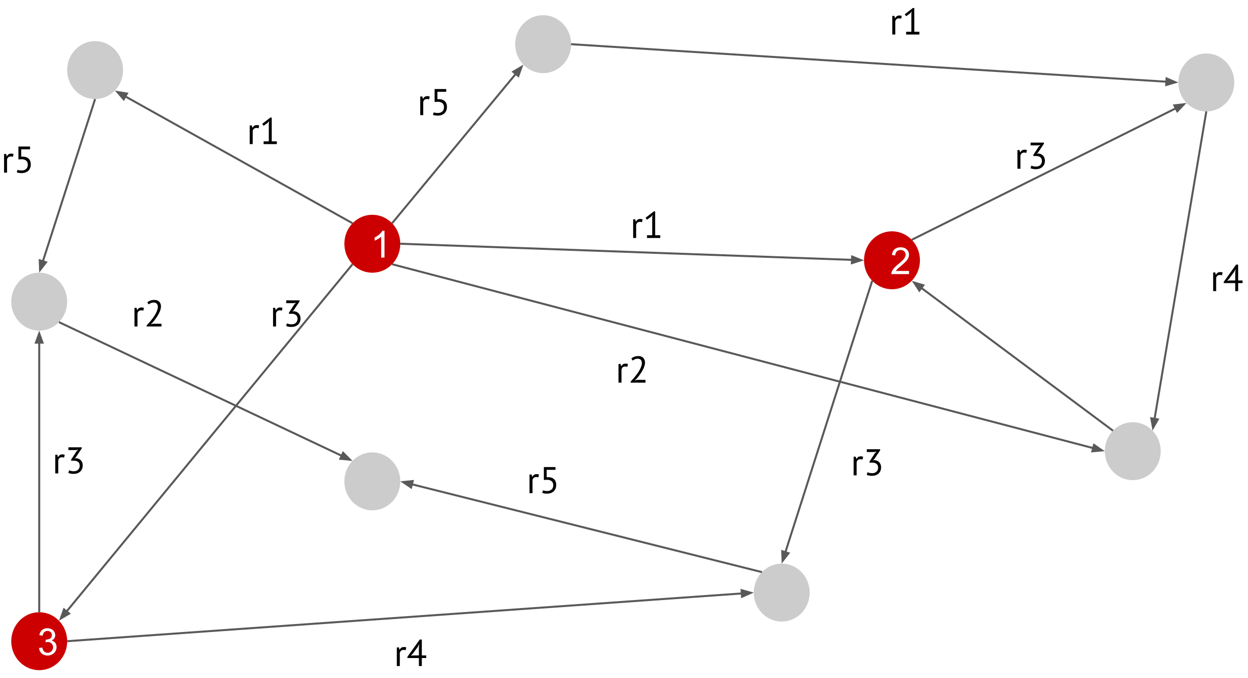
Sparsifying / Tokenizing KGs



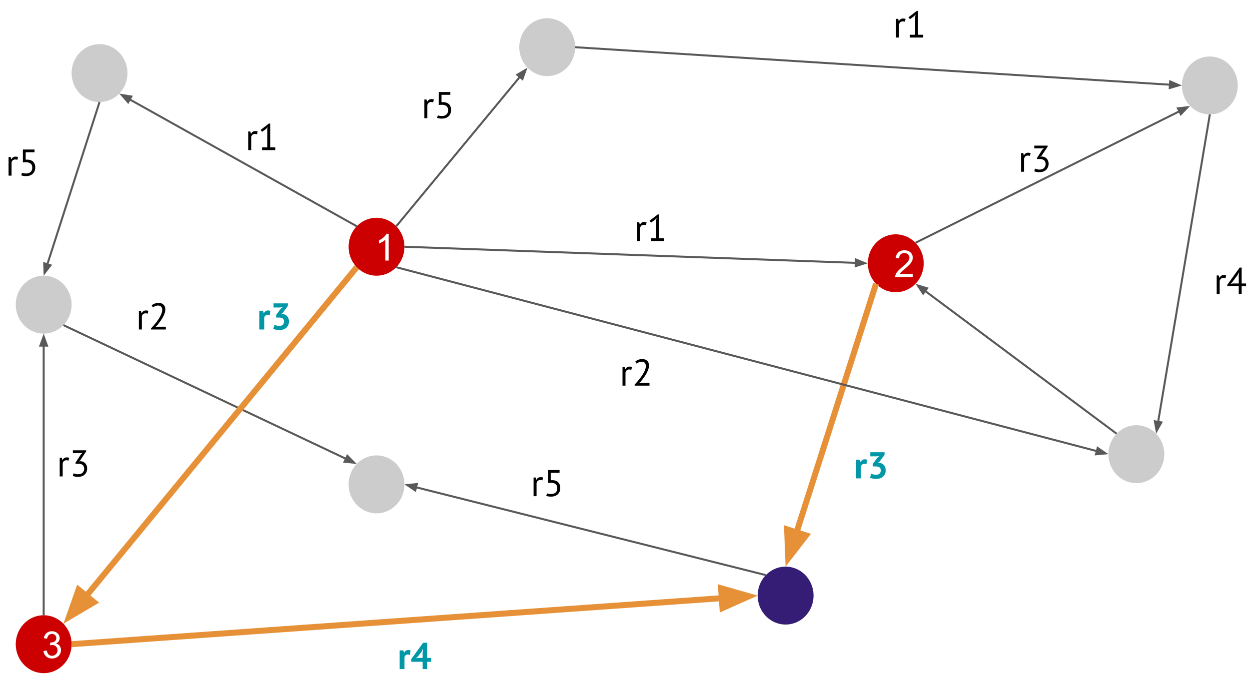
Sparsifying / Tokenizing KGs

k anchors

Conjunctive query of anchor paths

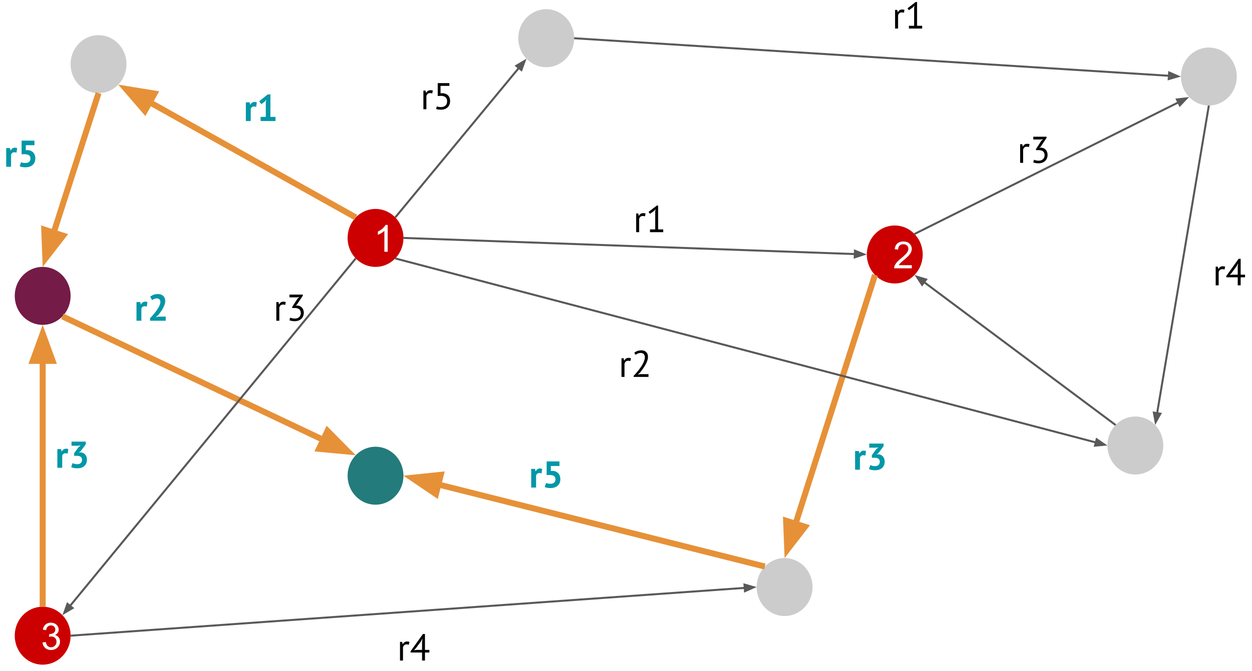


Sparsifying / Tokenizing KGs



a2 r3 ∧ **a3** r4 ∧ **a1** r3 r4

Sparsifying / Tokenizing KGs



a3 r3 ∧ **a1** r1 r5 ∧ **a2** r3 r5 r2_inv

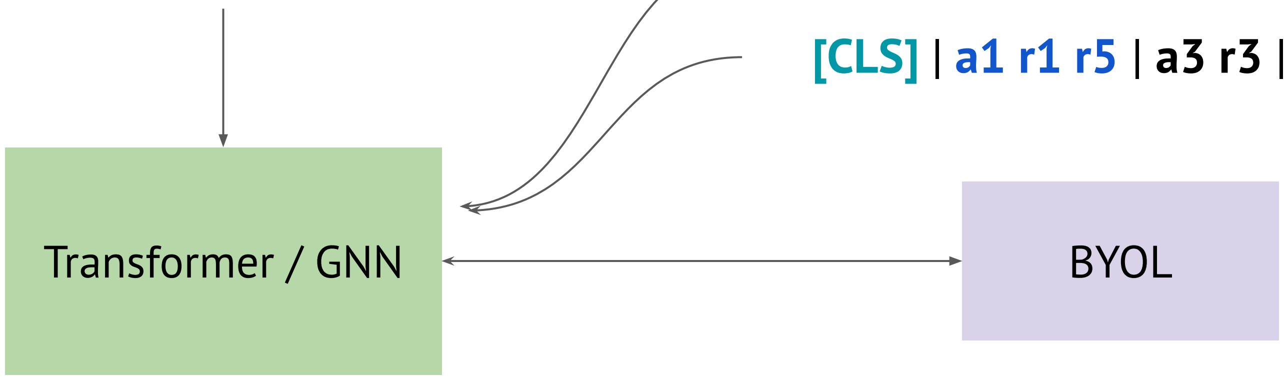
Sparsifying / Tokenizing KGs

a3 r3 ∧ **a1** r1 r5 ∧ **a2** r3 r5 r2i

[CLS] | a3 r3 | [MASK] r1 r5 | a2 r3 r5 r2i

[CLS] | **a3 r3** | **a1 r1 r5** | a2 r3 r5 r2i

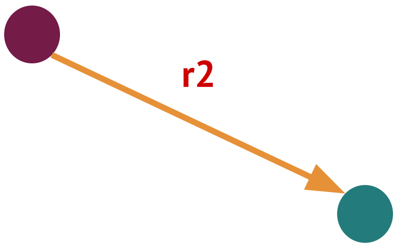
[CLS] | **a1 r1 r5** | **a3 r3** | a2 r3 r5 r2i



MLM loss

BYOL loss

Fine-tuning decoders



[CLS] | a3 r3 | a1 r1 r5 | a2 r3 r5 r2i

[CLS] | a3 r4 r1 | a2 r1

Old TransE

[h] + **r2** - [t]

hr

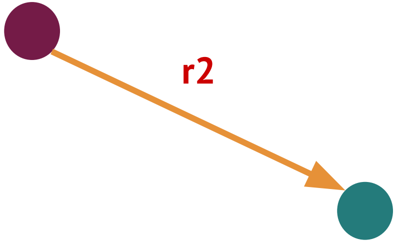
[CLS] | a3 r3 **r2** | a1 r1 r5 **r2** | a2 r3 r5 r2i **r2**

[CLS] | a3 r4 r1 | a2 r1

New Cosine

$$\frac{hr \cdot t}{||hr|| \cdot ||t||}$$

Positional encodings



hr

[CLS] | a3 r3 **r2** | a1 r1 r5 **r2** | a2 r3 r5 r2i **r2**

[CLS]	a3	r3	r2	a1	r1	r5	r2	a2	r3	r5	r2i	r2	
0	1	2	3	1	2	3	4	1	2	3	4	5	Op1
0	1	1	1	2	2	2	2	3	3	3	3	3	Op2
0	2	p	p	3	p	p	p	4	p	p	p	p	Op3
0	1	2	3	4	5	6	7	8	9	10	11	12	Op4

Even more!

- The ImageNet moment for KGs
- Compositional generalization with approximate query embedding
- KGs + links to the spectral theory
- Logical expressiveness of triple-based and hyper-relational KGs