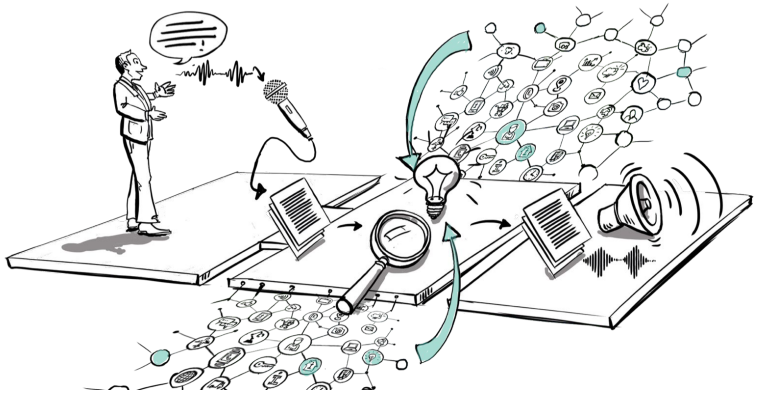


Knowledge Graphs

and Conversational AI

Dr. Michael Galkin

Fraunhofer IAIS & Smart Data Analytics
Dresden

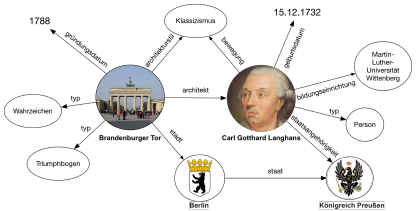
Outline

- Knowledge Graphs
- Knowledge Graph Embeddings (KGE) : ML + Logic
- Question Answering over KGs (KGQA)
- Conversational AI

Background

We build conversational AI platforms

Fraunhofer Dresden IAIS



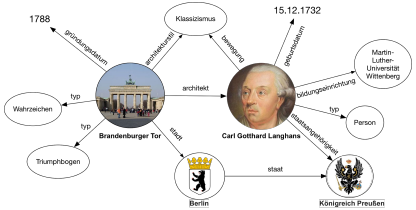
Background

We build conversational AI platforms

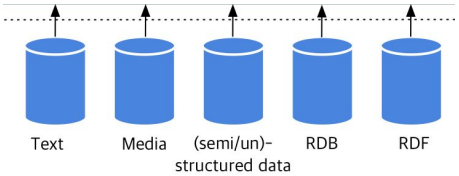
Powered by knowledge graphs

Background

We build conversational AI platforms

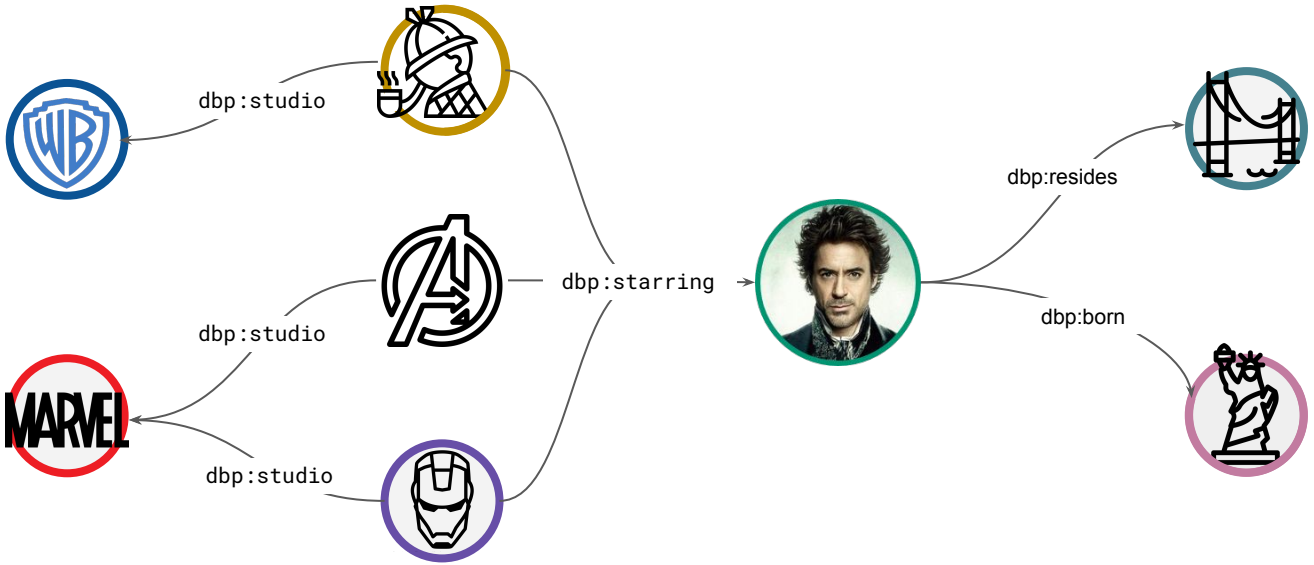


Powered by knowledge graphs

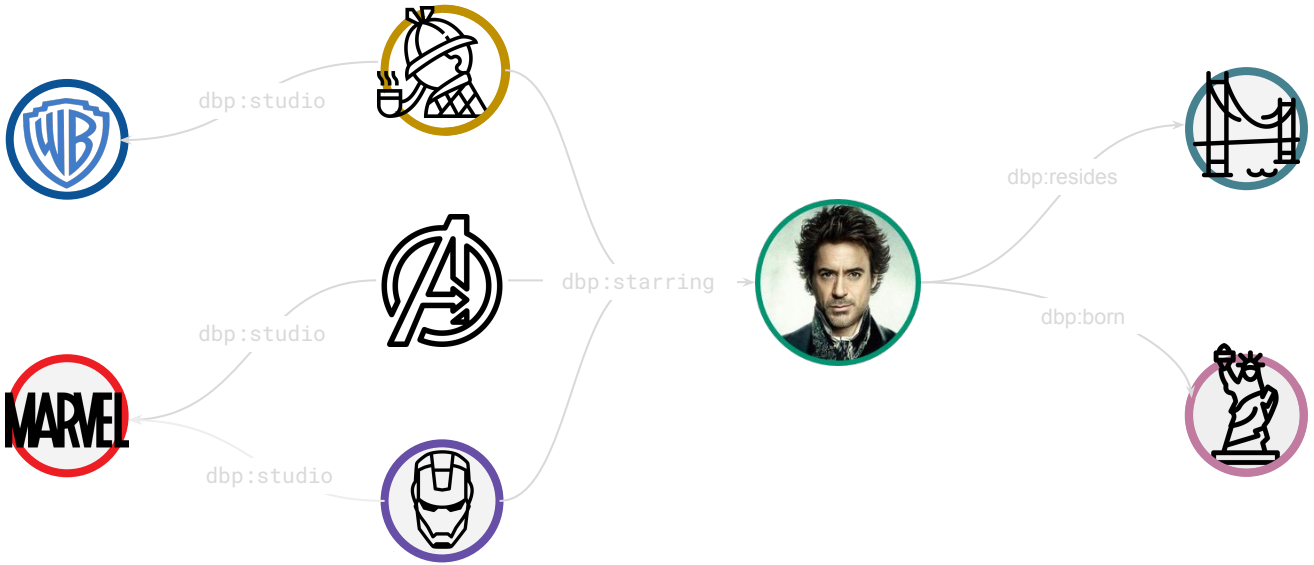


Obtained by integrating heterogeneous data

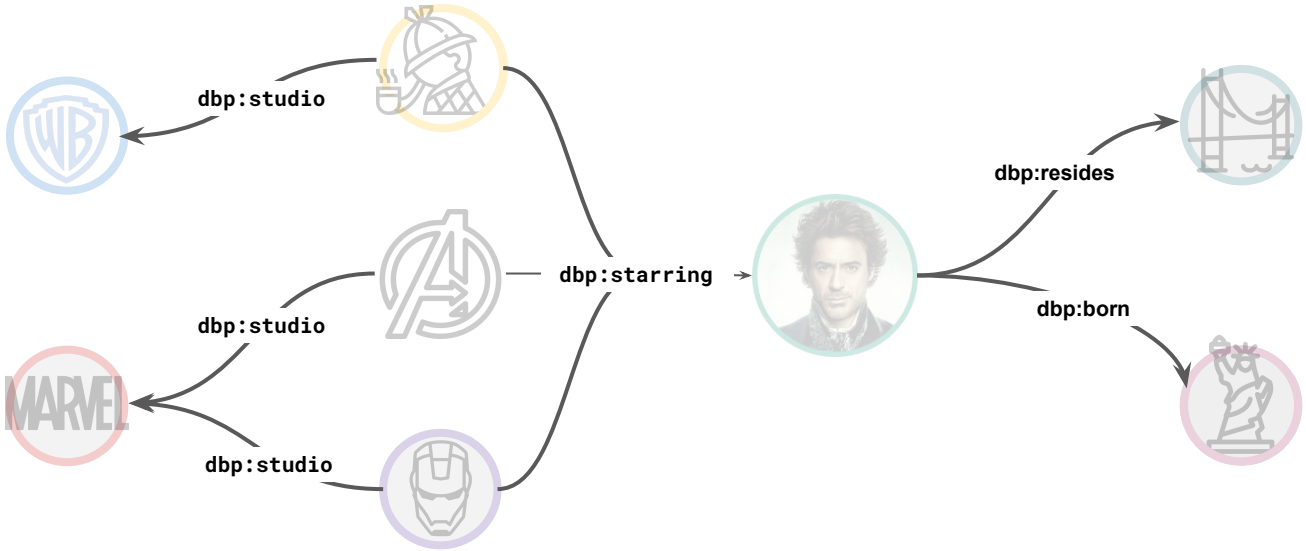
Knowledge Graph



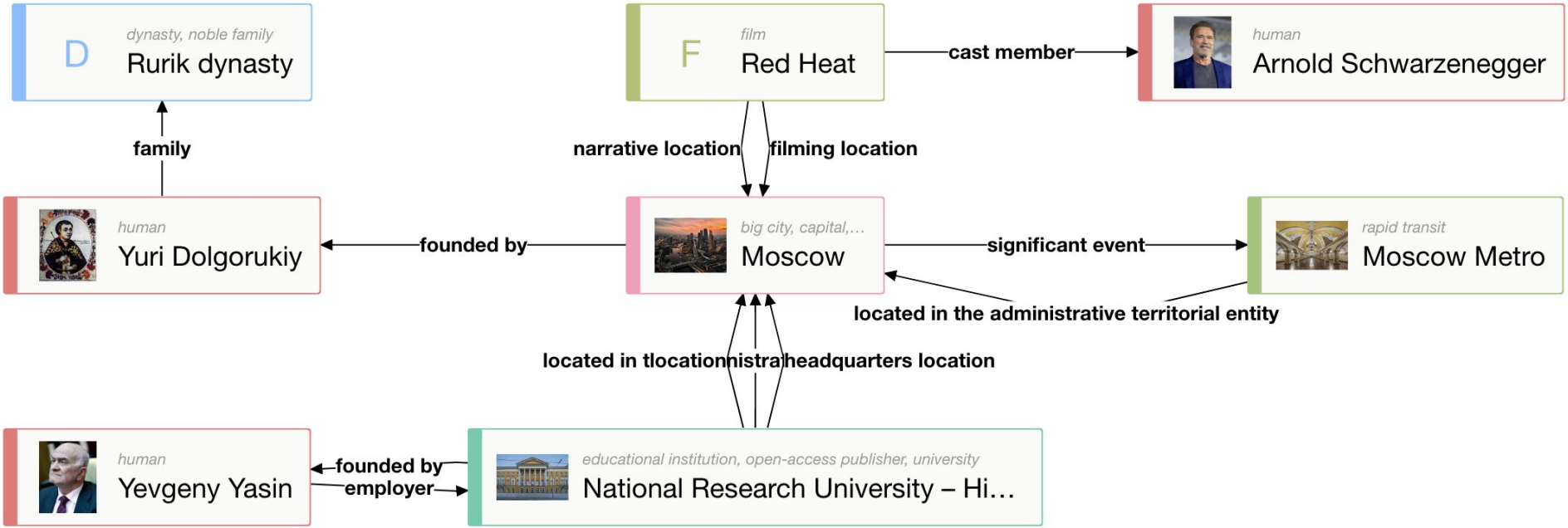
Knowledge Graph - Entities



Knowledge Graph - Relations



Knowledge Graph (real excerpt)



https://wikidata.metaphacts.com/

Available Knowledge Graphs

Knowledge Graphs

Open-domain: Wikidata, DBpedia

Biomed: Drugbank, SNOMED-CT, Bio2RDF

Ontologies

Industry 4.0: RAMI

Finance: FIBO, FRO, XBRL, FinReg

Available Knowledge Graphs

Knowledge Graphs

Open-domain: Wikidata, DBpedia

Biomed: Drugbank, SNOMED-CT, Bio2RDF

Ontologies

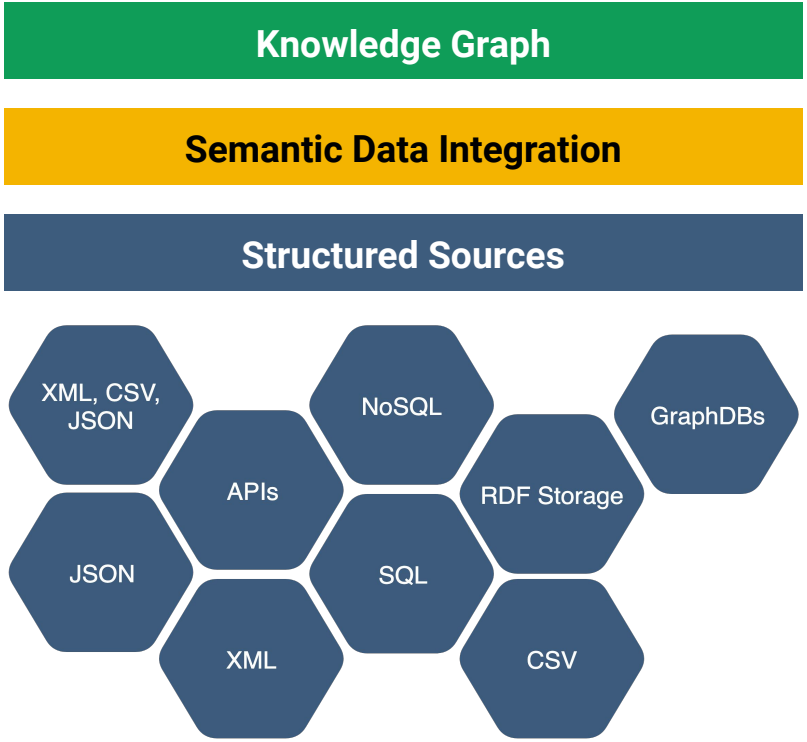
Industry 4.0: RAMI

Finance: FIBO, FRO, XBRL, FinReg

Custom

Enterprise Knowledge Graphs

Building Knowledge Graphs

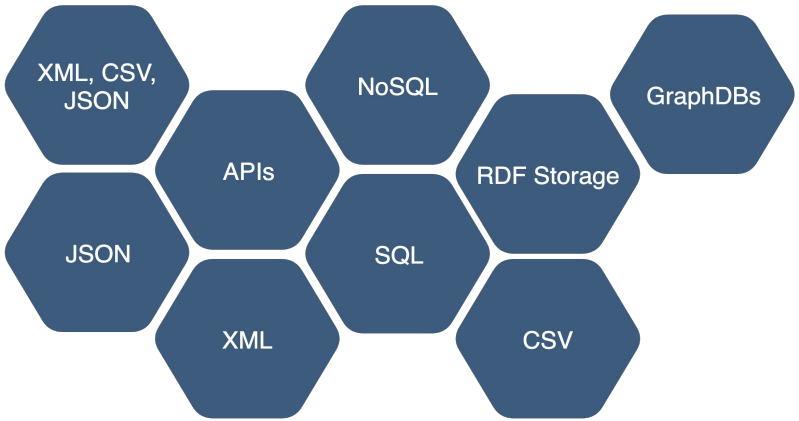


Building Knowledge Graphs

Knowledge Graph

Semantic Data Integration

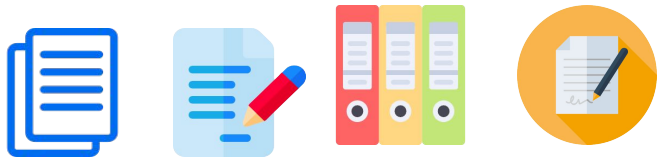
Structured Sources



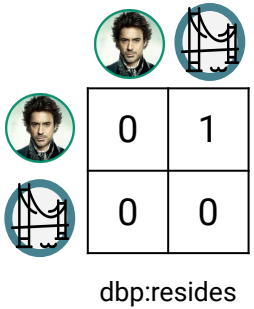
Knowledge Graph

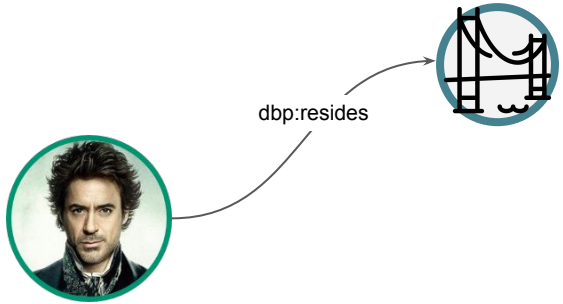
Information Retrieval & NLP

Unstructured Sources

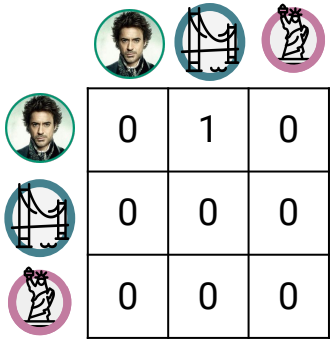


Knowledge Graphs as Tensors

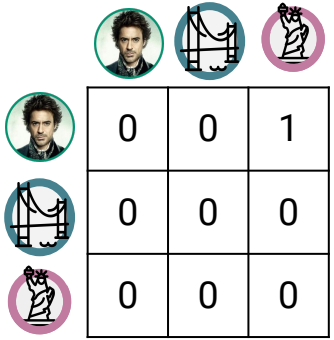




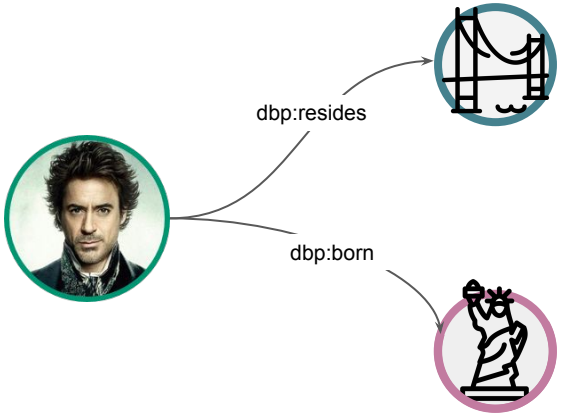
Knowledge Graphs as Tensors



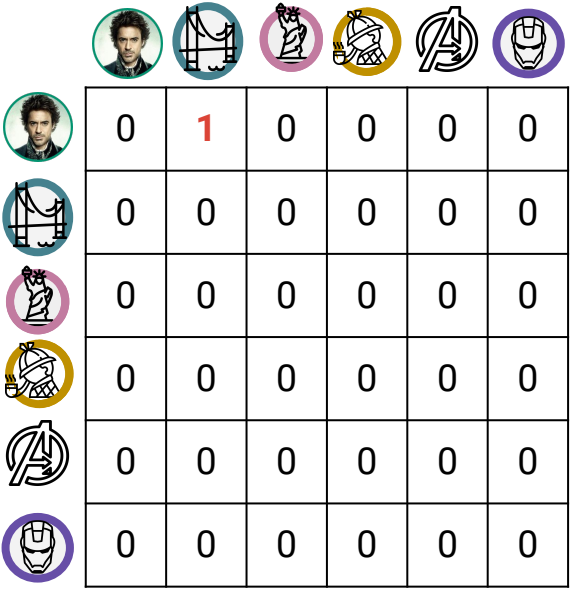
dbp:resides



dbp:born

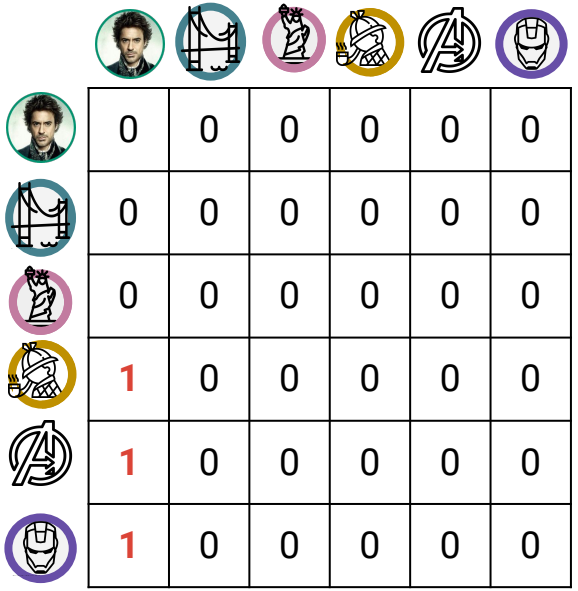


Knowledge Graphs as Tensors



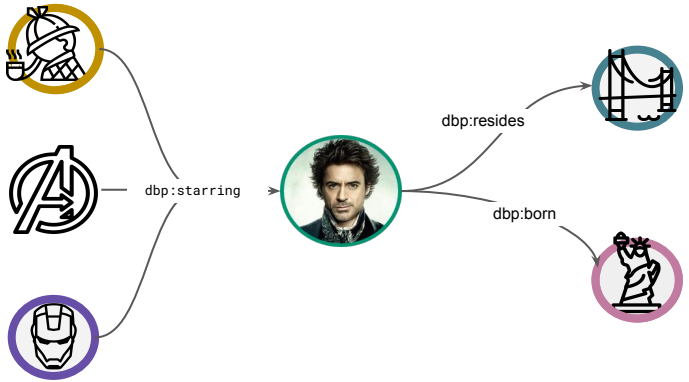
0	1	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

dbp:resides

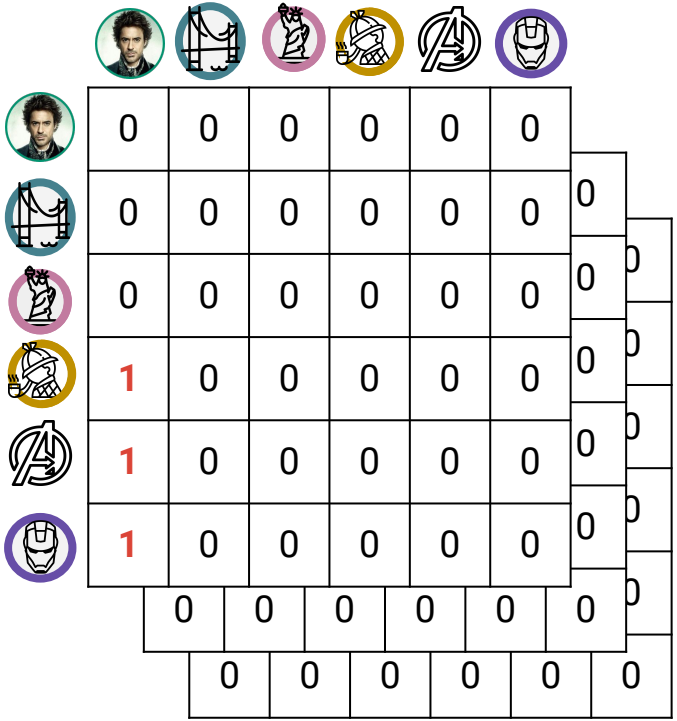


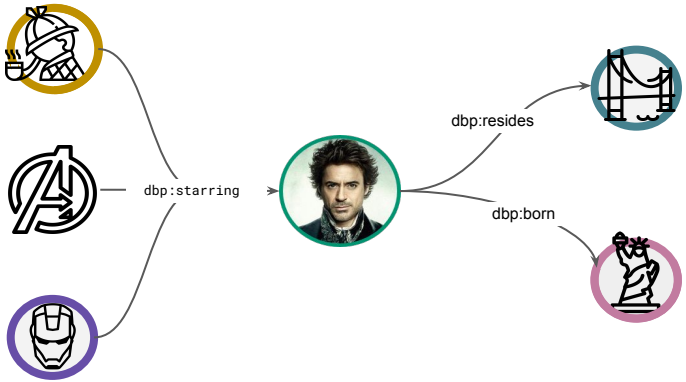
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0

dbp:starring



Knowledge Graphs as Tensors





$$\mathcal{T} : \mathbb{R}^{|E| \times |E| \times |R|}$$

Tensor Factorization

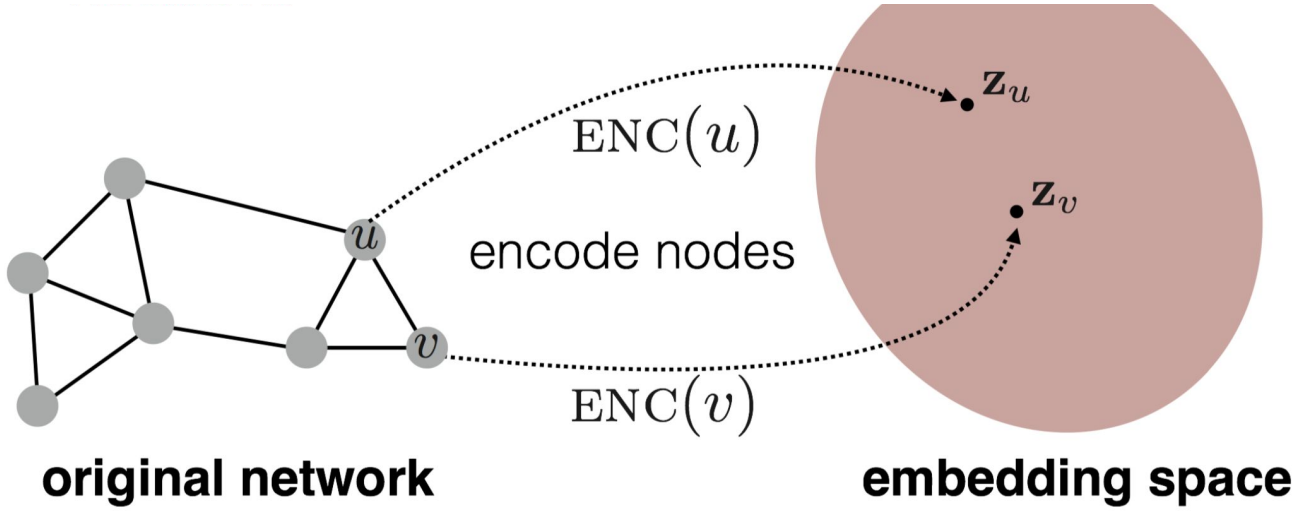
Translation

Convolution

Graph Neural Nets

Knowledge Graph Embeddings

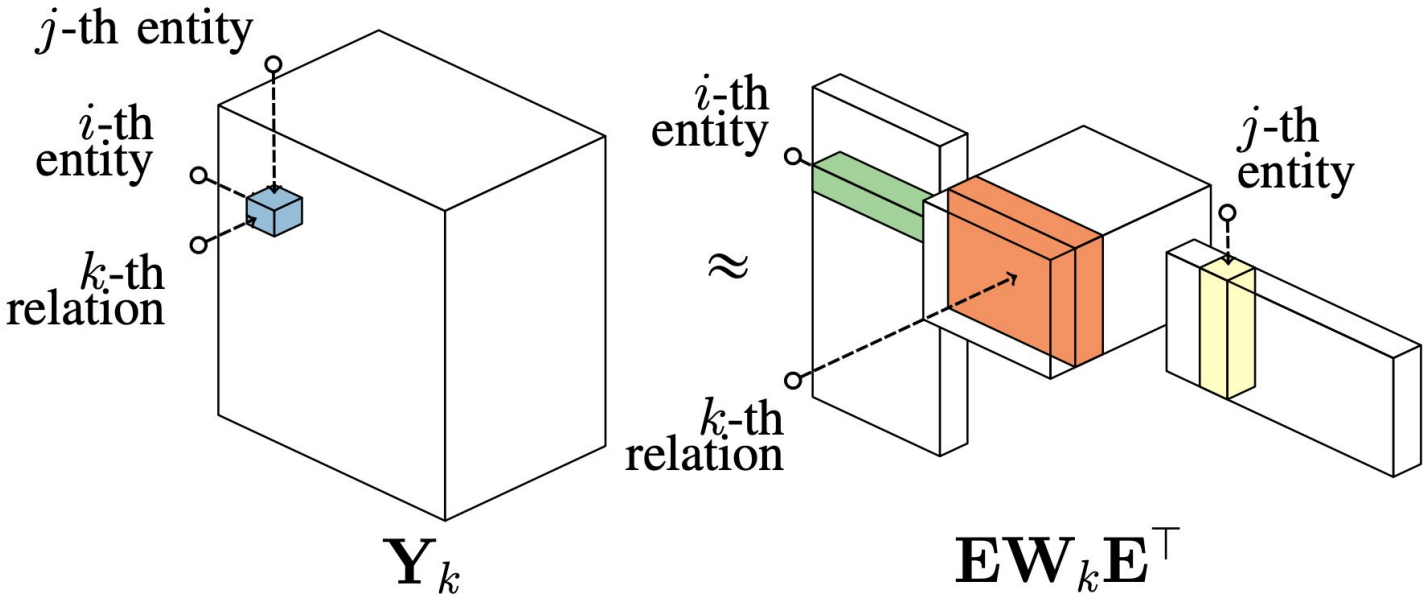
Goal: encode nodes so that **similarity in the embedding space (e.g., dot product)** approximates **similarity in the original network**



Tensor Factorization

KGE - RESCAL

Goal - factorize a sparse 3D tensor to dense E and R



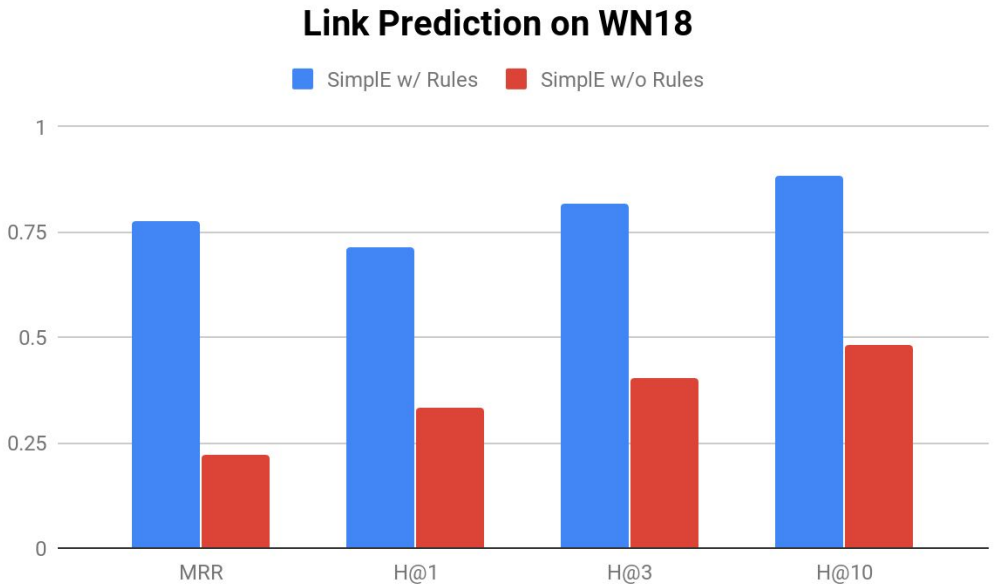
Nickel et al. A review of relational machine learning for knowledge graphs. IEEE. 2015

Tensor Factorization

KGE - Expressivity & Rules

TF can be enriched with logical rules and can learn rules

- Symmetric
- Inverse
- Anti-symmetric
- Composition



Kazemi et al. SimplE Embedding for Link Prediction in Knowledge Graphs. NIPS 2018

Tensor Factorization

KGE - Expressivity & Rules

TF can be enriched with logical rules and can learn rules

Model	Score Function	Symmetry	Antisymmetry	Inversion	Composition
SE	$-\|\|\boldsymbol{W}_{r,1}\mathbf{h}-\boldsymbol{W}_{r,2}\mathbf{t}\|\|$	✗	✗	✗	✗
TransE	$-\|\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|\|$	✗	✓	✓	✓
TransX	$-\|\|g_{r,1}(\mathbf{h})+\mathbf{r}-g_{r,2}(\mathbf{t})\|\|$	✓	✓	✗	✗
DistMult	$\langle\mathbf{h},\mathbf{r},\mathbf{t}\rangle$	✓	✗	✗	✗
ComplEx	$\mathrm{Re}(\langle\mathbf{h},\mathbf{r},\bar{\mathbf{t}}\rangle)$	✓	✓	✓	✗
RotatE	$-\|\|\mathbf{h}\circ\mathbf{r}-\mathbf{t}\|\|$	✓	✓	✓	✓

Table 2: The pattern modeling and inference abilities of several models.

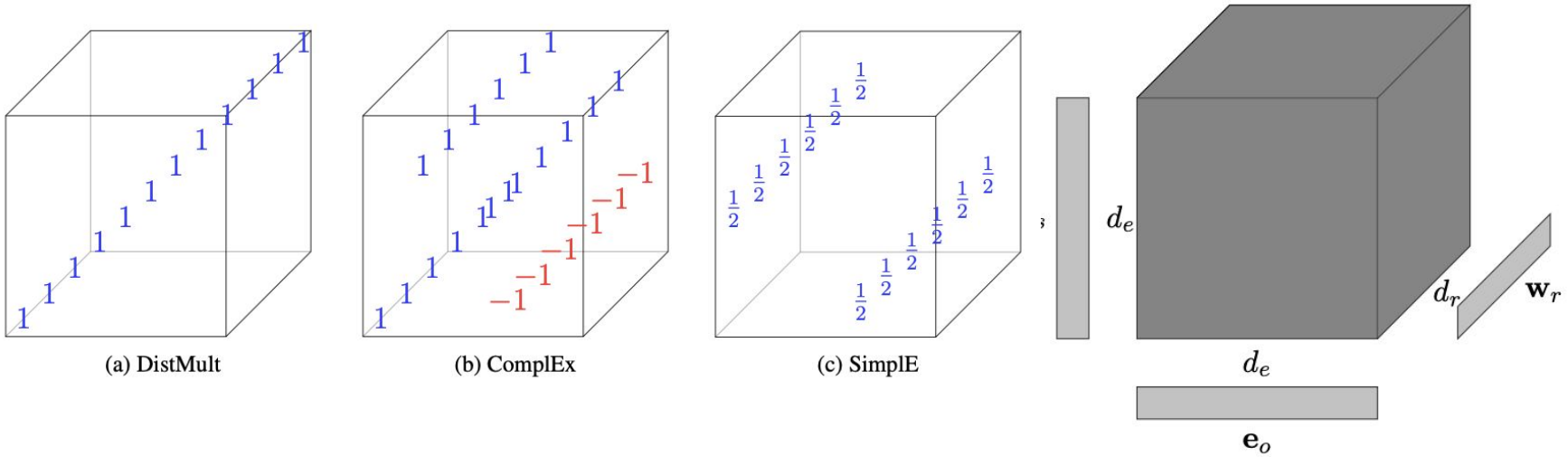
Sun et al. Rotate: Knowledge graph embedding by relational rotation in complex space. ICLR 2019

Tensor Factorization

KGE - TuckER

Goal - factorize a sparse 3D tensor to dense core W, entities E and relations R

$$\phi(e_s, r, e_o) = \mathcal{W} \times_1 \mathbf{e}_s \times_2 \mathbf{w}_r \times_3 \mathbf{e}_o$$



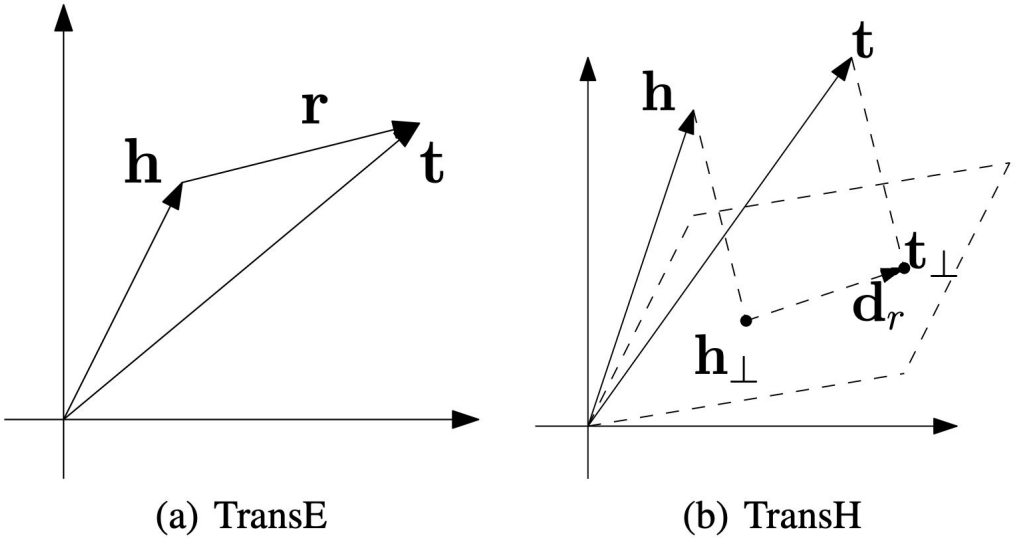
Tensor Factorization

Translation

KGE - TransE

Translate entities and relations into one embedding space

$$h + r \approx t \quad \text{Moscow} + \text{capitalOf} \approx \text{Russia}$$



(a) TransE (b) TransH

Wang et al. Knowledge Graph Embedding by Translating on Hyperplanes. AAAI 2014

Tensor Factorization

Translation

KGE - TransE

LOTS of models

TABLE 9
Knowledge graph embedding using margin-based ranking loss.

GE Algorithm	Energy Function $\mathbf{f_r(h,t)}$
TransE [91]	$\|h+r-t\|_{l1}$
TKRL [53]	$\|M_{rh}h+r-M_{rt}t\|$
TransR [15]	$\|hM_r+r-tM_r\|_2^2$
CTransR [15]	$\|hM_r+r_c-tM_r\|_2^2+\alpha\|r_c-r\|_2^2$
TransH [14]	$\|(h-w_r^Thw_r)+d_r-(t-w_r^Ttw_r)\|_2^2$
SePLi [39]	$\frac{1}{2}\|W_ie_{ih}+b_i-e_{it}\|^2$
TransD [125]	$\|M_{rh}h+r-M_{rt}t\|_2^2$
TranSparse [126]	$\|M_r^h(\theta_r^h)h+r-M_r^t(\theta_r^t)t\|_{l_1/2}^2$
m-TransH [127]	$\|\sum_{\rho\in\mathcal{M}(R_r)}a_r(\rho)\mathbb{P}_{n_r}(t(\rho))+b_r\|^2, t\in\mathcal{N}^{\mathcal{M}(R_r)}$
DKRL [128]	$\|h_d+r-t_d\|+\|h_d+r-t_s\|+\|h_s+r-t_d\|$
ManifoldE [129]	Sphere: $\|\varphi(h)+\varphi(r)-\varphi(t)\|^2$ Hyperplane: $(\varphi(h)+\varphi(r_{head}))^T(\varphi(t)+\varphi(r_{tail}))$ φ is the mapping function to Hilbert space
TransA [130]	$\|h+r-t\|$
puTransE [43]	$\|h+r-t\|$
KGE-LDA [60]	$\|h+r-t\|_{l1}$
SE [90]	$\|R_uh-R_ut\|_{l1}$
SME [92] linear	$(W_{u1}r+W_{u2}h+b_u)^T(W_{v1}r+W_{v2}t+b_v)$
SME [92] bilinear	$(W_{u1}r+W_{u2}h+b_u)^T(W_{v1}r+W_{v2}t+b_v)$
SSP [59]	$-\lambda\|e-s^Tes\|_2^2+\|e\|_2^2, S(s_h,s_t)=\frac{s_h+s_t}{\|s_h+s_t\|_2^2}$
NTN [131]	$u_r^T\tanh(h^TW_rt+W_{rh}h+W_{rt}t+b_r)$
HOLE [132]	$r^T(h\star t)$, where $\star$ is circular correlation
MTransE [133]	$\|h+r-t\|_{l1}$

Cai et al. A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications. IEEE TKDE 2017

Tensor Factorization

Translation

KGE - LogicENN

$$\min_{\theta} \sum_{(h,r,t)\in\mathcal{S}} \alpha^r_{h,t} \log(1+\exp(-y^r_{h,t} f^r_{h,t})) + \lambda \sum_{i=1}^{l} \frac{\mathcal{R}_i}{N_i}$$

subject to $\|h\| = 1$ and $\|t\| = 1$.

Rule	Definition $\forall \mathtt{h}, \mathtt{t}, \mathtt{s} \in \mathcal{E} : \ldots$	Formulation based on score function	Formulation based on NN	Equivalent regularization form (Denoted as $\mathcal{R}_i$ in Equation (2))
Equivalence	$(\mathtt{h}, r_1, \mathtt{t}) \Leftrightarrow (\mathtt{h}, r_2, \mathtt{t})$	$f^{r_1}_{h,t} = f^{r_2}_{h,t} + \xi_{h,t}$	$\Phi^T_{h,t}(\boldsymbol{\beta}^{r_1} - \boldsymbol{\beta}^{r_2}) = \xi_{h,t}$	$\max(\|\boldsymbol{\beta}^{r_1} - \boldsymbol{\beta}^{r_2}\|_1 - \xi_{\text{Eq}}, 0)$
Symmetric	$(\mathtt{h}, r, \mathtt{t}) \Leftrightarrow (\mathtt{t}, r, \mathtt{h})$	$f^r_{h,t} = f^r_{t,h} + \xi_{h,t}$	$(\Phi_{h,t} - \Phi_{t,h})^T \boldsymbol{\beta}^r = \xi_{h,t}$	$\max(\|(\Phi_{h,t} - \Phi_{t,h})^T \boldsymbol{\beta}^r\| - \xi_{\text{Sy}}, 0)$
Asymmetric	$(\mathtt{h}, r, \mathtt{t}) \Rightarrow \neg(\mathtt{t}, r, \mathtt{h})$	$f^r_{h,t} = f^r_{t,h} + \mathcal{M}_{h,t}$	$(\Phi_{h,t} - \Phi_{t,h})^T \boldsymbol{\beta}^r = \mathcal{M}$	NC
Negation	$(\mathtt{h}, r_1, \mathtt{t}) \Leftrightarrow \neg(\mathtt{h}, r_2, \mathtt{t})$	$f^{r_1}_{h,t} = \mathcal{M} - f^{r_2}_{h,t} + \xi_{h,t}$	$\Phi^T_{h,t}(\boldsymbol{\beta}^{r_1} + \boldsymbol{\beta}^{r_2}) = \mathcal{M} + \xi_{h,t}$	NC
Implication	$(\mathtt{h}, r_1, \mathtt{t}) \Rightarrow (\mathtt{h}, r_2, \mathtt{t})$	$f^{r_1}_{h,t} \leq f^{r_2}_{h,t}$	$\Phi^T_{h,t}(\boldsymbol{\beta}^{r_1} - \boldsymbol{\beta}^{r_2}) \leq 0$	$\max(\sum_i (\boldsymbol{\beta}^{r_1}_i - \boldsymbol{\beta}^{r_2}_i) + \xi_{\text{Im}}, 0)$
Inverse	$(\mathtt{h}, r_1, \mathtt{t}) \Rightarrow (\mathtt{t}, r_2, \mathtt{h})$	$f^{r_1}_{h,t} \leq f^{r_2}_{t,h}$	$\Phi^T_{h,t}\boldsymbol{\beta}^{r_1} - \Phi^T_{t,h}\boldsymbol{\beta}^{r_2} \leq 0$	$\max(\Phi^T_{h,t}\boldsymbol{\beta}^{r_1} - \Phi^T_{t,h}\boldsymbol{\beta}^{r_2} + \xi_{\text{In}}, 0)$
Reflexivity	$(\mathtt{h}, r, \mathtt{h})$	$f^r_{h,h} = \mathcal{M} - \xi_{h,h}$	$\Phi^T_{h,h}\boldsymbol{\beta}^r = \mathcal{M} - \xi_{h,h}$	NC
Irreflexive	$\neg(\mathtt{h}, r, \mathtt{h})$	$f^r_{h,h} = \xi_{h,h}$	$\Phi^T_{h,h}\boldsymbol{\beta}^r = \xi_{h,h}$	NC
Transitivity	$(\mathtt{h}, r, \mathtt{t}) \wedge (\mathtt{t}, r, \mathtt{s}) \Rightarrow (\mathtt{h}, r, \mathtt{s})$	$\sigma(f^r_{h,s}) \geq \sigma(f^r_{h,t}) \times \sigma(f^r_{t,s})$	$\sigma(\Phi_{h,t}\boldsymbol{\beta}^r) \times \sigma(\Phi_{t,s}\boldsymbol{\beta}^r) - \sigma(\Phi^T_{h,s}\boldsymbol{\beta}^r) \leq 0$	$\max(\sigma(\Phi_{h,t}\boldsymbol{\beta}^r) \times \sigma(\Phi_{t,s}\boldsymbol{\beta}^r) - \sigma(\Phi^T_{h,s}\boldsymbol{\beta}^r) + \xi_{\text{Tr}}, 0)$
Composition	$(\mathtt{h}, r_1, \mathtt{t}) \wedge (\mathtt{t}, r_2, \mathtt{s}) \Rightarrow (\mathtt{h}, r_3, \mathtt{s})$	$\sigma(f^{r_1}_{h,s}) \geq \sigma(f^{r_2}_{h,t}) \times \sigma(f^{r_3}_{t,s})$	$\sigma(\Phi_{h,t}\boldsymbol{\beta}^{r_1}) \times \sigma(\Phi_{t,s}\boldsymbol{\beta}^{r_2}) - \sigma(\Phi^T_{h,s}\boldsymbol{\beta}^{r_3}) \leq 0$	$\max(\sigma(\Phi_{h,t}\boldsymbol{\beta}^{r_1}) \times \sigma(\Phi_{t,s}\boldsymbol{\beta}^{r_2}) - \sigma(\Phi^T_{h,s}\boldsymbol{\beta}^{r_3}) + \xi_{\text{Co}}, 0)$

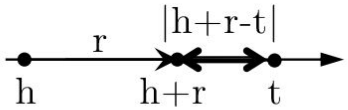
Table 1: Formulation and representation of rules (NC: Not considered for implementation).

Nayyeri et al. LogicENN: A Neural Based Knowledge Graphs Embedding Model with Logical Rules. arxiv:1908.07141

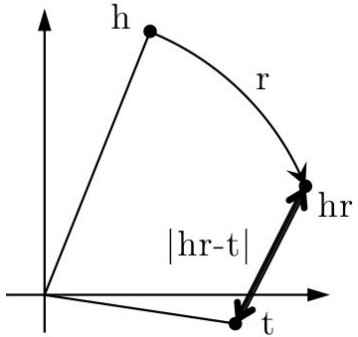
Tensor Factorization

Translation

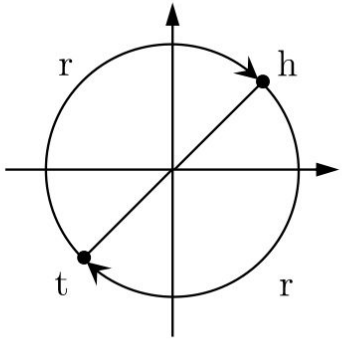
KGE - RotatE



(a) TransE models **r** as translation in real line.



(b) RotatE models **r** as rotation in complex plane.



(c) RotatE: an example of modeling symmetric relations **r** with $r_i = -1$

Figure 1: Illustrations of TransE and RotatE with only 1 dimension of embedding.

Idea:
Entities are vectors in **complex space**

Relations: rotations in **complex space**

Score function: $d_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \qquad |r_i| = 1$

Loss & Optimization: $L = -\log \sigma(\gamma - d_r(\mathbf{h}, \mathbf{t})) - \sum_{i=1}^{n} \frac{1}{k} \log \sigma(d_r(\mathbf{h}'_i, \mathbf{t}'_i) - \gamma),$

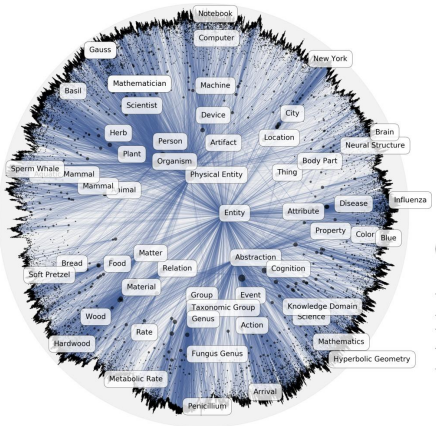
Sun et al. RotatE: Knowledge Graph Embedding By Relational Rotation In Complex Space. ICLR 2019

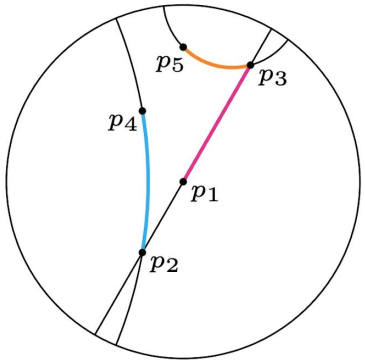
Tensor Factorization

Translation

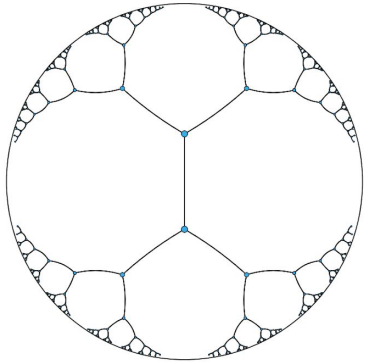
KGE - Hyperbolic

Goal: embed hierarchical structures into an n-dimensional Poincaré ball.

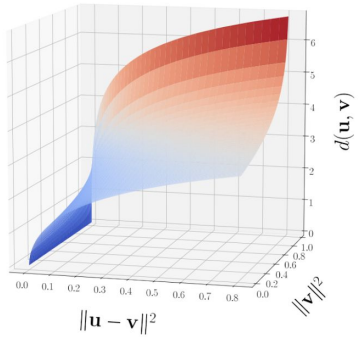




(a) Geodesics of the Poincaré disk



(b) Embedding of a tree in $\mathcal{B}^2$



(c) Growth of Poincaré distance

Figure 1: (a) Due to the negative curvature of $\mathcal{B}$, the distance of points increases exponentially (relative to their Euclidean distance) the closer they are to the boundary. (c) Growth of the Poincaré distance $d(\boldsymbol{u}, \boldsymbol{v})$ relative to the Euclidean distance and the norm of $\boldsymbol{v}$ (for fixed $\|\boldsymbol{u}\| = 0.9$). (b) Embedding of a regular tree in $\mathcal{B}^2$ such that all connected nodes are spaced equally far apart (i.e., all black line segments have identical hyperbolic length).

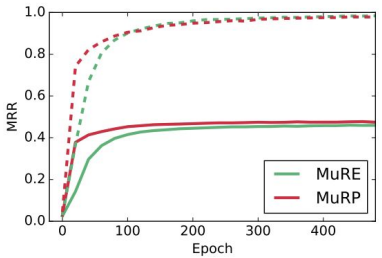
Nickel et al. Poincaré Embeddings for Learning Hierarchical Representations. NIPS 2017

Tensor Factorization

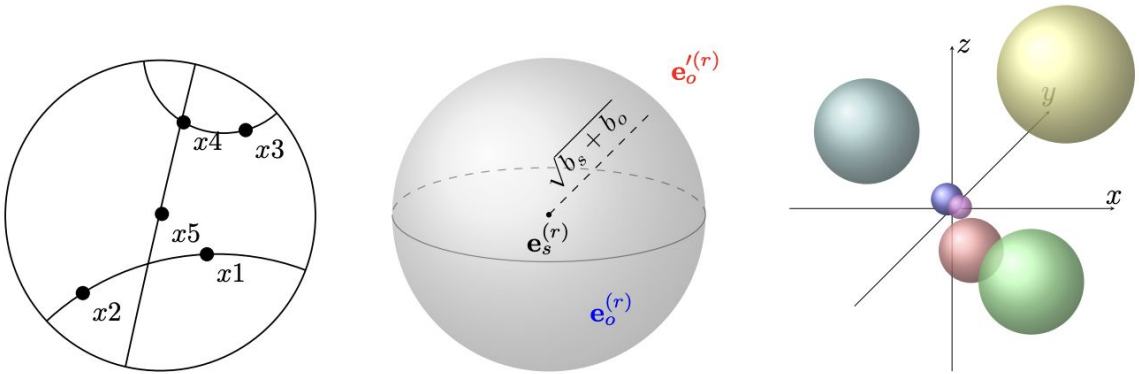
KGE - Hyperbolic - MuRP

Goal: embed hierarchical structures into an n-dimensional Poincaré ball.

Translation



(b) MRR covergence rate per epoch.



(a) Poincaré disk geodesics. (b) Model decision boundary. (c) Spheres of influence.

Figure 1: (a) Geodesics in the Poincaré disk, indicating the shortest paths between pairs of points. (b) The model predicts the triple (e_s, r, e_o) as true and (e_s, r, e'_o) as false. (c) Each entity embedding has a *sphere of influence*, whose radius is determined by the entity-specific bias.

Balazevic et al. Multi-relational Poincaré Graph Embeddings. NeurIPS 2019

Tensor Factorization

Translation

KGE - Hyperbolic - MuRP

Goal: embed hierarchical structures into an n-dimensional Poincaré ball.

	WN18RR				FB15k-237			
	MRR	Hits@10	Hits@3	Hits@1	MRR	Hits@10	Hits@3	Hits@1
TransE [6]	.226	.501	–	–	.294	.465	–	–
DistMult [37]	.430	.490	.440	.390	.241	.419	.263	.155
ComplEx [34]	.440	.510	.460	.410	.247	.428	.275	.158
Neural LP [38]	–	–	–	–	.250	.408	–	–
MINERVA [9]	–	–	–	–	–	.456	–	–
ConvE [11]	.430	.520	.440	.400	.325	.501	.356	.237
M-Walk [29]	.437	–	.445	.414	–	–	–	–
TuckER [2]	.470	.526	.482	**.443**	**.358**	**.544**	**.394**	**.266**
RotatE [30]	–	–	–	–	.297	.480	.328	.205
MuRE $d = 40$	.459	.528	.474	.429	.315	.493	.346	.227
MuRE $d = 200$	.475	.554	.487	.436	**.336**	**.521**	**.370**	**.245**
MuRP $d = 40$	**.477**	**.555**	**.489**	.438	.324	.506	.356	.235
MuRP $d = 200$	**.481**	**.566**	**.495**	**.440**	.335	.518	.367	.243

Balazevic et al. Multi-relational Poincaré Graph Embeddings. NeurIPS 2019

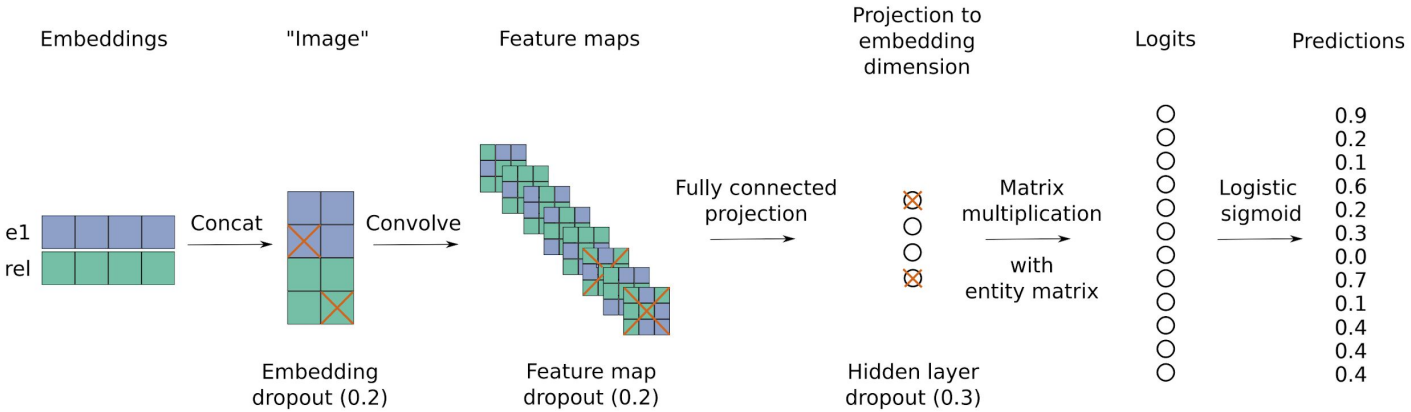
Tensor Factorization

Translation

Convolution

KGE - ConvE

Goal: CNNs for predicting a probability of the object



Minervini et al. Convolutional 2D Knowledge Graph Embeddings. AAAI 2018

Tensor Factorization

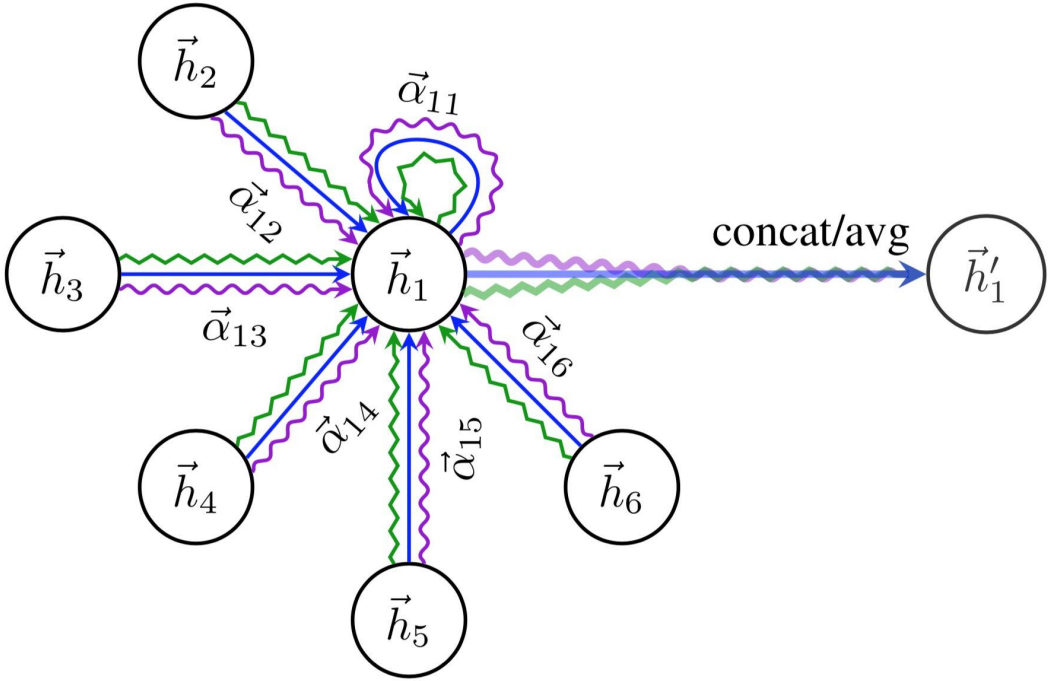
Translation

Convolution

Graph Neural Nets

KGE - Graph Networks

Goal: leverage topological graph characteristics



Veličković et al. Graph Attention Networks. ICLR 2018

Tensor Factorization

Translation

Convolution

Graph Neural Nets

KGE - Graph Networks

Goal: leverage topological graph characteristics

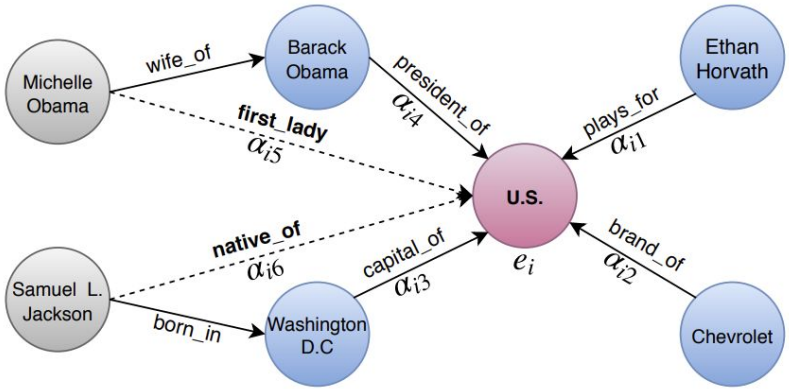


Figure 2: This figure shows the aggregation process of our graph attentional layer. α_{ij} represents relative attention values of the edge. The dashed lines represent an *auxiliary* edge from a n-hop neighbors, in this case $n = 2$.

Nathani et al. Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs. ACL 2019

Question Answering over KGs

How many Marvel movies was Robert Downey Jr. casted in?

KGQA

How many **Marvel** **movies** was **Robert Downey Jr.** **casted** in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

KGQA

How many Marvel movies was Robert Downey Jr. casted in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

KGQA

How many Marvel movies was Robert Downey Jr. casted in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

KGQA

How many Marvel movies was Robert Downey Jr. casted in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

Entity Linking

Who is the CEO of **Apple**?

Apple belongs to which genus?

movie character

Downey played **Iron Man** in which year?

Who is the alter ego of **Iron man**?

comic character

Relation Linking

dbo:starring

Name all the movies in which **Robert Downey Jr** **Acted**?

Which movies have **RDJ**?

Flicks where I can see **Robert DJ**?

Find me all the films **casting** **Rober Downey Jr** ?

List all the movies **starring** **Robert Downey Junior**?

RDJ **has acted** in which movies?

Relation Linking - Implicit Predicates

Name all the movies in which **Robert Downey Jr** **Acted**?

Which movies have **RDJ**?

Flicks where I can see **Robert DJ**?

Find me all the films **casting** **Rober Downey Jr** ?

List all the movies **starring** **Robert Downey Junior**?

RDJ **has acted** in which movies?

LC-QuAD 2.0

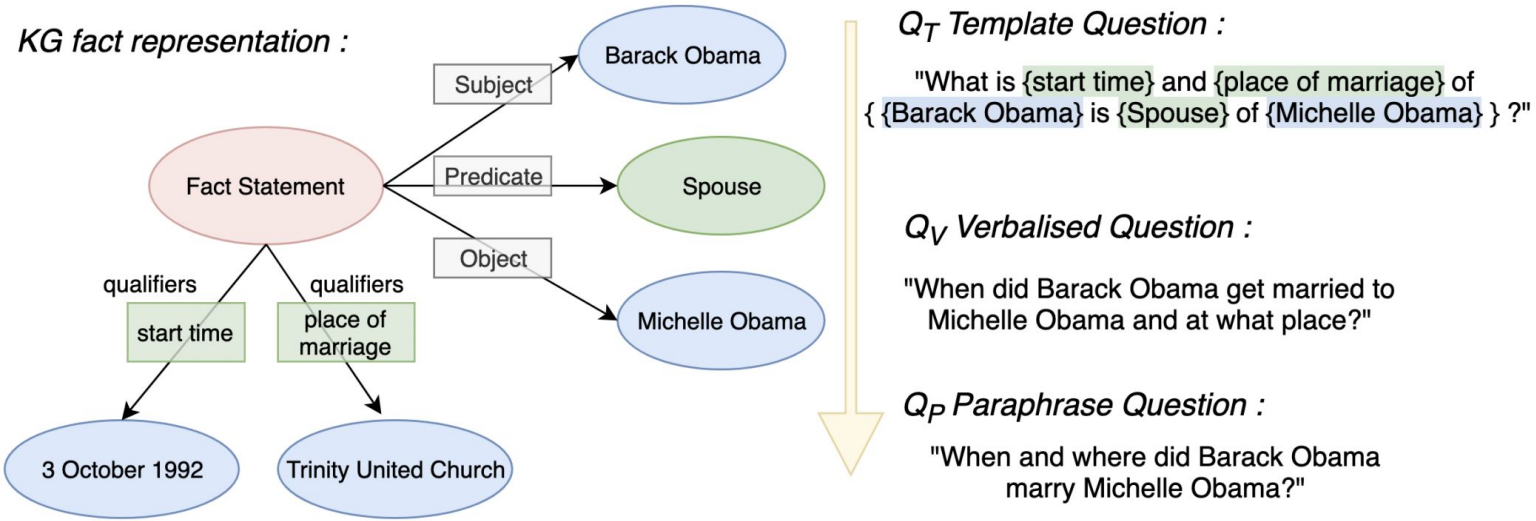


Fig. 2. (left) Representation of a fact with its Qualifiers. (right) Translation of a KG-fact to a verbalised question and then paraphrased question.

Dubey et al. LC-QuAD 2.0: A Large Dataset for Complex Question Answering over Wikidata and DBpedia. ISWC 2019

Knowledge Graphs from Text

QUESTION What is Albert Einstein famous for?

WEB INFORMATION

DOCUMENT 1

Albert Einstein, a German theoretical physicist, published the theory of relativity.

The theory of relativity is one of the two pillars of modern physics.

He won the physics Nobel Prize.

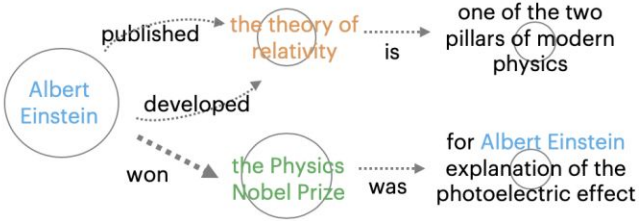
DOCUMENT 2

Albert Einstein (March 14, 1879 to April 18, 1955) developed the theory of relativity.

He won the Nobel Prize.

The great prize was for his explanation of the photoelectric effect.

GRAPH CONSTRUCTION



LINEARIZATION

<sub> Albert Einstein <obj> the theory of relativity <pred> published <s> developed <obj> the Physics Nobel Prize <s> won

<sub> the theory of relativity <obj> one of the two pillars of modern physics <pred> is

<sub> the Physics Nobel Prize <obj> for his explanation of the photoelectric effect <pred> was

Fan et al. Using Local Knowledge Graph Construction to Scale Seq2Seq Models to Multi-Document Inputs. EMNLP 2019

Knowledge Graphs from Text

QUESTION What is Albert Einstein famous for?

WEB INFORMATION

DOCUMENT 1

Albert Einstein, a German theoretical physicist, published the theory of relativity.

The theory of relativity is one of the two pillars of modern physics.

He won the physics Nobel Prize.

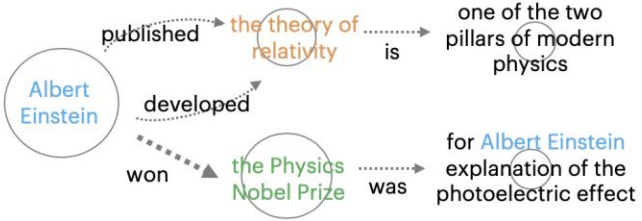
DOCUMENT 2

Albert Einstein (March 14, 1879 to April 18, 1955) developed the theory of relativity.

He won the Nobel Prize.

The great prize was for his explanation of the photoelectric effect.

GRAPH CONSTRUCTION



LINEARIZATION

<sub> Albert Einstein <obj> the theory of relativity <pred> published <s> developed <obj> the Physics Nobel Prize <s> won

<sub> the theory of relativity <obj> one of the two pillars of modern physics <pred> is

<sub> the Physics Nobel Prize <obj> for his explanation of the photoelectric effect <pred> was

GRAPH CONSTRUCTION STEPS

QUERY: Can someone finally explain the theory of general relativity?

DOCUMENT SENTENCES with GRAPH OPERATIONS

1. Albert Einstein, a German theoretical physicist, published the theory of relativity.

ADDED TO GRAPH

2. The theory of relativity is one of the two pillars of modern physics.

MERGE OPERATION:
theory of relativity
EXISTS AS A NODE
NODE WEIGHT + 1

3. He won the physics Nobel Prize.

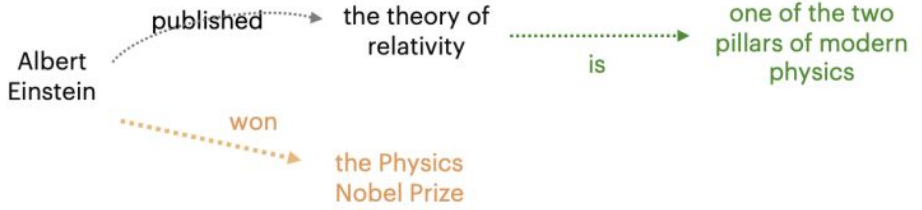
COREFERENCE:
he and *Albert Einstein*

MERGE OPERATION:
Albert Einstein
EXISTS AS A NODE
NODE WEIGHT + 1

4. Puppies are very cute.

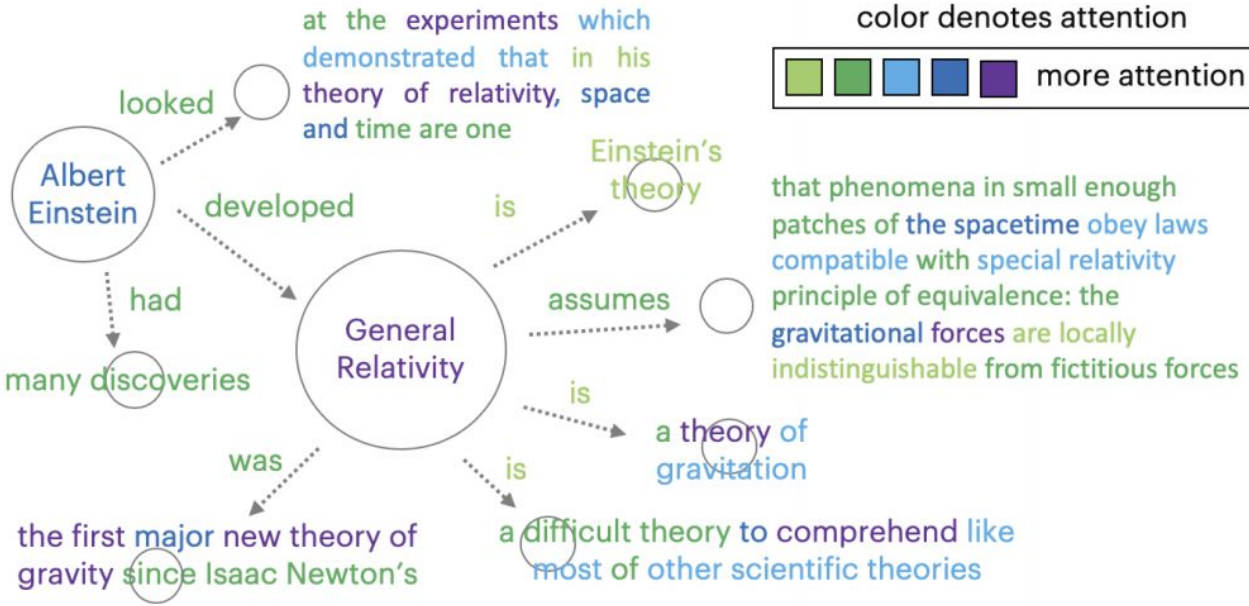
FILTER OPERATION:
low TF-IDF overlap with query
NOT ADDED TO GRAPH

CONSTRUCTED GRAPH



Fan et al. Using Local Knowledge Graph Construction to Scale Seq2Seq Models to Multi-Document Inputs. EMNLP 2019

Knowledge Graphs from Text



Question: *Can someone finally explain the theory of general relativity?*

Generated Answer: It's a bit complicated to explain. General relativity is a theory of Albert Einstein. Einstein developed this theory. Relativity is measure of how fast something moves relative to the reference frame. In theory, the speed of light is constant. If you are traveling at a constant speed, and then you move at a speed relative to an observer, there is a constant rate for that object to move. General relativity is observed in the experiments of Albert Einstein. It is about space, and space, and gravity.

Fan et al. Using Local Knowledge Graph Construction to Scale Seq2Seq Models to Multi-Document Inputs. EMNLP 2019

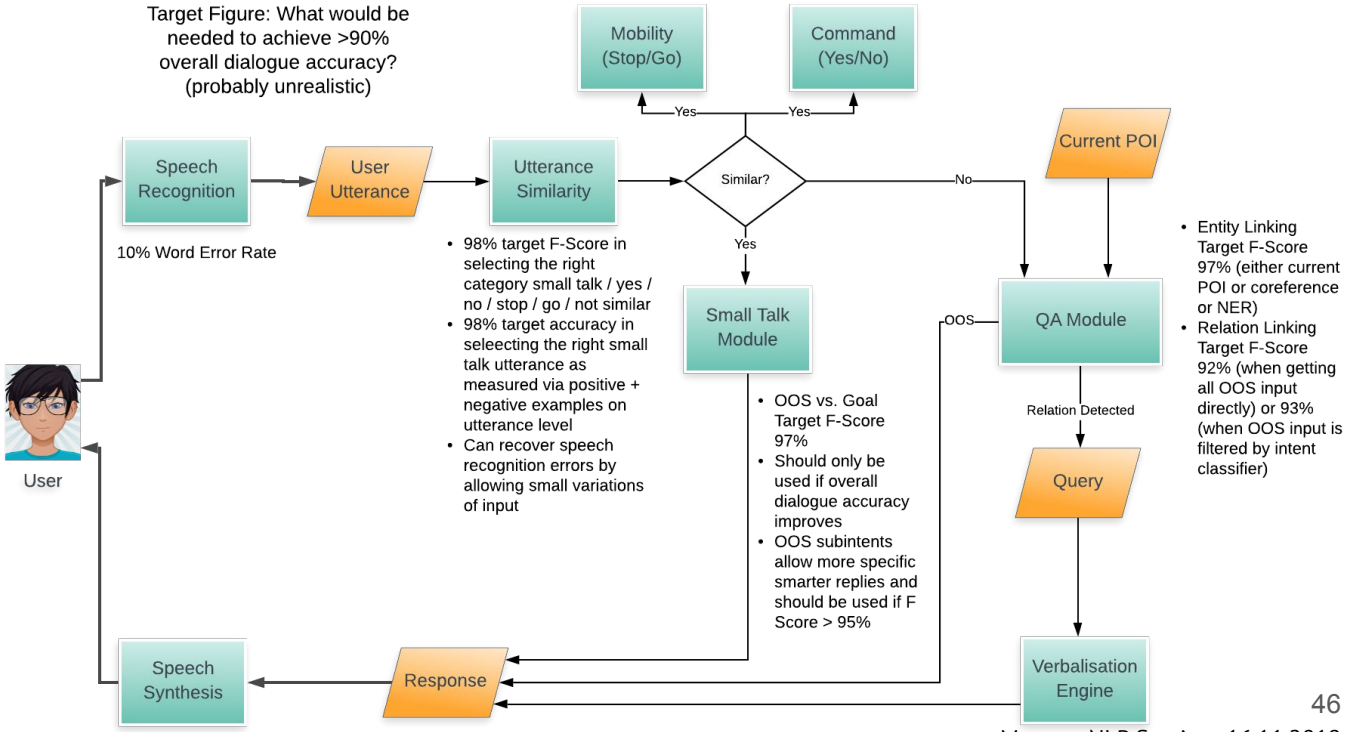
Raw text-based dialogue systems

- Mostly sequence to sequence over textual inputs
- Some basic causal NL inference (s1 follows s2)
- Effective on simple utterances over short paragraphs
  - No memory (with exceptions)
  - No format and justifiable knowledge

KGs & Conversational AI

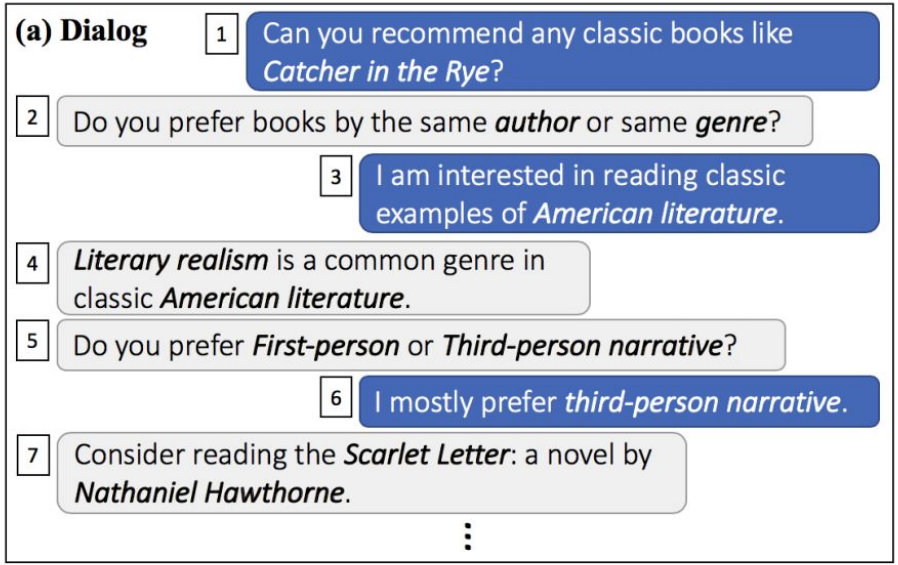
Knowledge-driven in-car dialogue system (EN/DE)

Full DBpedia 2019 (wikidata branch)
- > 50M entities
- > 4B triples



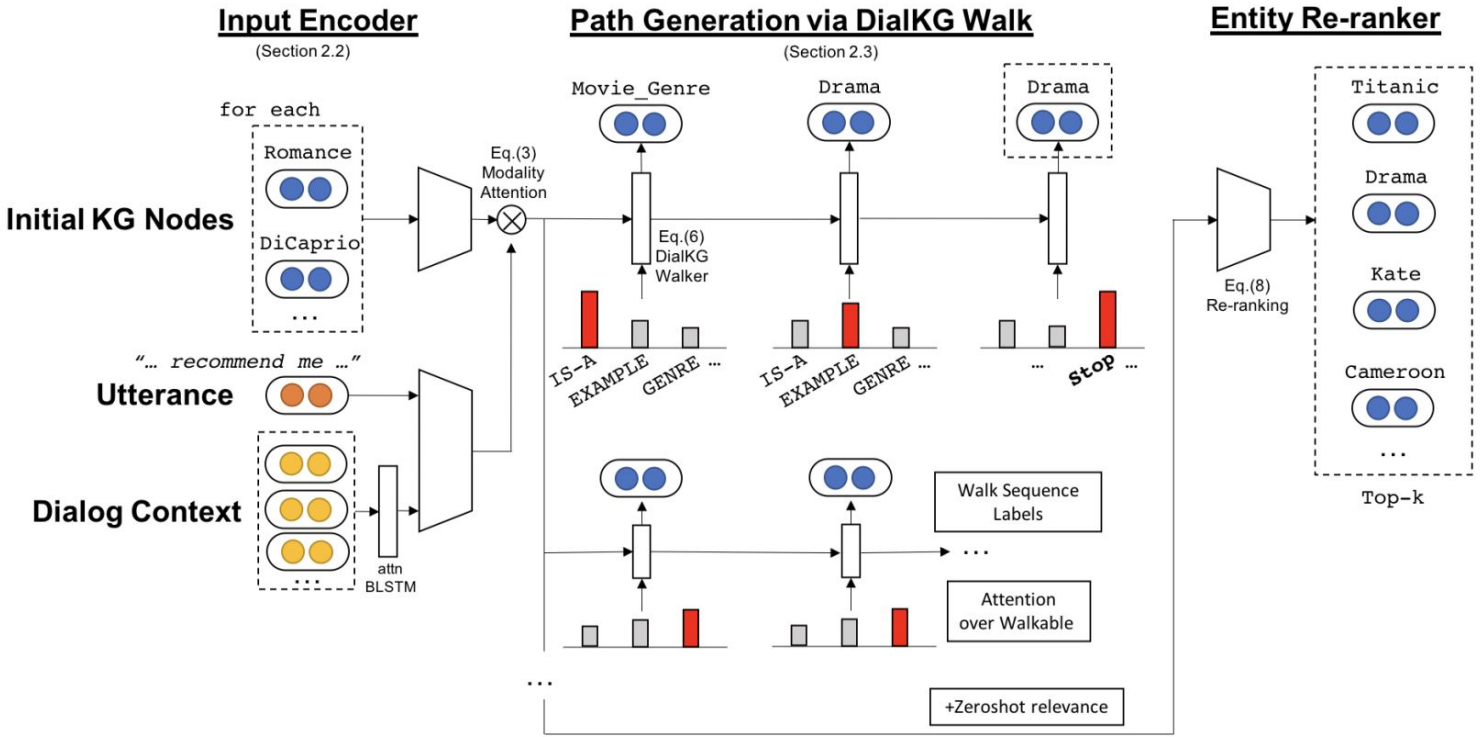
KGs & Conversational AI

Challenge: incorporate graph-based reasoning into dialogue systems over Knowledge Graphs



Moon et al. OpenDialKG: Explainable Conversational Reasoning with Attention-based Walks over Knowledge Graphs. ACL 2019

KGs & Conversational AI



Moon et al. OpenDialKG: Explainable Conversational Reasoning with Attention-based Walks over Knowledge Graphs. ACL 2019

Complex Sequential Question Answering

USER: What is the weapon used by HMS Regulus ?

SPEAKER: Torpedo 61

USER: Who is the manufacturer of that ranged weapon ?

SPEAKER: FFV

USER: Is that social group located in Seeley Union Elementary School District and Succivo ?

SPEAKER: NO

USER: Is Swedish Navy the facility operator of Torpedo 61 ?

SPEAKER: YES

USER: Which ranged weapons are used as weapons by atmost 140 ship types and aircraft power classes ?

SPEAKER: Mark 44, Mark 43, Torpedo 613

USER: Which ship types are the etymology of atmost 1 social group ?

SPEAKER: HMS Spica, HMS Invincible, HMS Queen Elizabeth

USER: How many military branches or organizations are the creator of the design for Mark 44 ?

SPEAKER: 2

USER: What are the port of registry of atleast 1 ship type ?

SPEAKER: Riga, Marseille, Bremerhaven

USER: Which ranged weapons are used as weapons by atleast 17 ship types or aircraft classes ?

SPEAKER: Torpedo 45, Mark 48 torpedo, Torpedo 61

Complex Sequential Question Answering

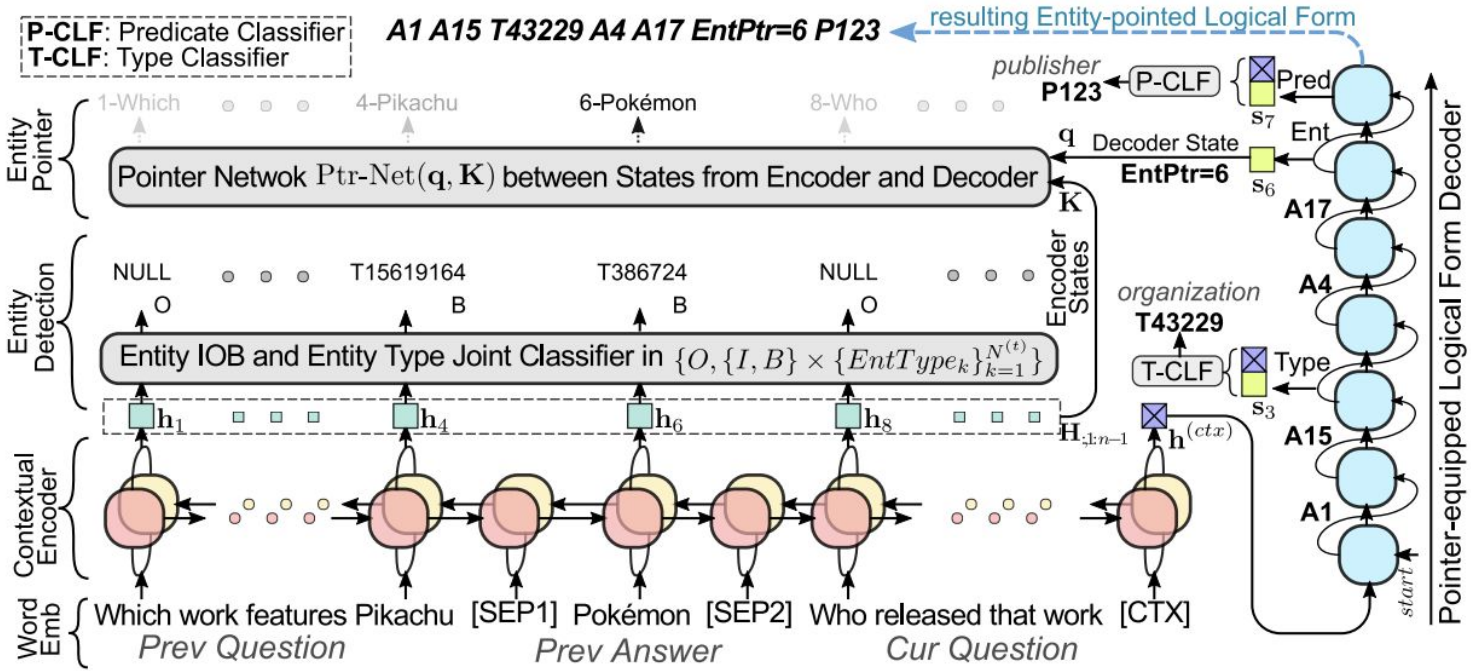


Figure 1: Proposed **M**ulti-**t**ask **S**emantic **P**arsing (MaSP) model. Note that P* and T* are predicate and entity type ids in Wikidata where entity type id originally starts with Q but is replaced with T for clear demonstration.

Shen et al. Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base. EMNLP 2019

Complex Sequential Question Answering

Methods		HRED+KVmem	D2A (Baseline)	MaSP (Ours)	Δ
Question Type	**#Example**	F1 Score	F1 Score	F1 Score	
Overall	203k	9.39%	66.70%	79.26%	+12.56%
Clarification	9k	16.35%	35.53%	80.79%	+45.26%
Comparative Reasoning (All)	15k	2.96%	48.85%	68.90%	+20.05%
Logical Reasoning (All)	22k	8.33%	67.31%	69.04%	+1.73%
Quantitative Reasoning (All)	9k	0.96%	56.41%	73.75%	+17.34%
Simple Question (Coreferenced)	55k	7.26%	57.69%	76.47%	+18.78%
Simple Question (Direct)	82k	13.64%	78.42%	85.18%	+6.76%
Simple Question (Ellipsis)	10k	9.95%	81.14%	83.73%	+2.59%
Question Type	**#Example**	Accuracy	Accuracy	Accuracy	
Verification (Boolean)	27k	21.04%	45.05%	60.63%	+15.58%
Quantitative Reasoning (Count)	24k	12.13%	40.94%	43.39%	+2.45%
Comparative Reasoning (Count)	15k	8.67%	17.78%	22.26%	+4.48%

Table 2: Comparisons with baselines on CSQA. The last column consists of differences between MaSP and D2A.

Shen et al. Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base. EMNLP 2019

So why you need graphs?

How many children does Berlin Hbf have?

Implicit or explicit constraints on produced answers

So why you need graphs?

How many children does Berlin Hbf have?

Train stations don’t have kids

Implicit or explicit constraints on produced answers

- reduce candidates space
- help to fight the mushroom effect
- **ontologies help**

So why you need graphs?

How many children does Berlin Hbf have?

Train stations don't have kids

What is the busiest train station in Germany?

Implicit or explicit constraints on produced answers

- reduce candidates space
- help to fight the mushroom effect
- **ontologies help**

Complex QA via (sub)graphs aggregations

So why you need graphs?

How many children does Berlin Hbf have?

Train stations don't have kids

What is the busiest train station in Germany?

Hamburg Hbf

Implicit or explicit constraints on produced answers

- reduce candidates space
- help to fight the mushroom effect
- **ontologies help**

Complex QA via (sub)graphs aggregations

```
select ?station ?visits where {
 ?station wdt:P31 wd:Q18543139 .      # central stations
 ?station wdt:P17 wd:Q183 .           # in Germany
 ?station wdt:P1373 ?visits .         # daily visits
} ORDER BY DESC(?visits) LIMIT 1      # sort
```

So why you need graphs?

Takeaway 1

Graphs significantly improve reasoning compared to sole natural language inference

So why you need graphs?

Takeaway 1

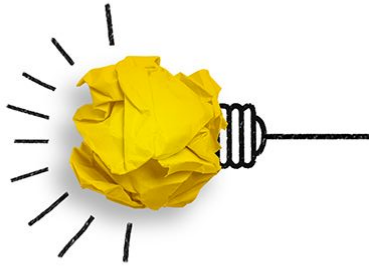
Graphs significantly improve reasoning compared to sole natural language inference

Takeaway 2

Reasoning outcomes are explainable and traceable

Enriched knowledge representations

KGs at scale and robust querying



Graph-based reasoning for complex QA

Possible directions

mikhail.galkin@iais.fraunhofer.de

Self-learning and knowledge extraction from dialogues

NLG from graphs

Commonsense reasoning