Data Fest 2020

Placing Knowledge Graphs In Graph ML

Mikhail Galkin
PhD, Research Scientist
TU Dresden & Fraunhofer IAIS
Dresden, Germany

 @migalkin

 @michael_galkin

 @mgalkin

On the definition of a Knowledge Graph

Given entities E, relations R, KG is a directed multi-relational graph G that comprises triples (s, p, o)

$$\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$$

$$(s, p, o) \in \mathcal{G}$$

* describes entities and relations
* defines a schema
* interrelating arbitrary entities
* various topical domains

"Every RDF / LPG / RDF* graph is a knowledge graph"

"Abstract schema and instances"

On the definition of a Knowledge Graph

Given entities E, relations R, KG is a directed multi-relational graph G that comprises triples (s, p, o)

$$\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$$
$$(s, p, o) \in \mathcal{G}$$

* describes entities and relations
* defines a schema
* interrelating arbitrary entities
* various topical domains

"Every RDF / LPG / RDF* graph is a knowledge graph"

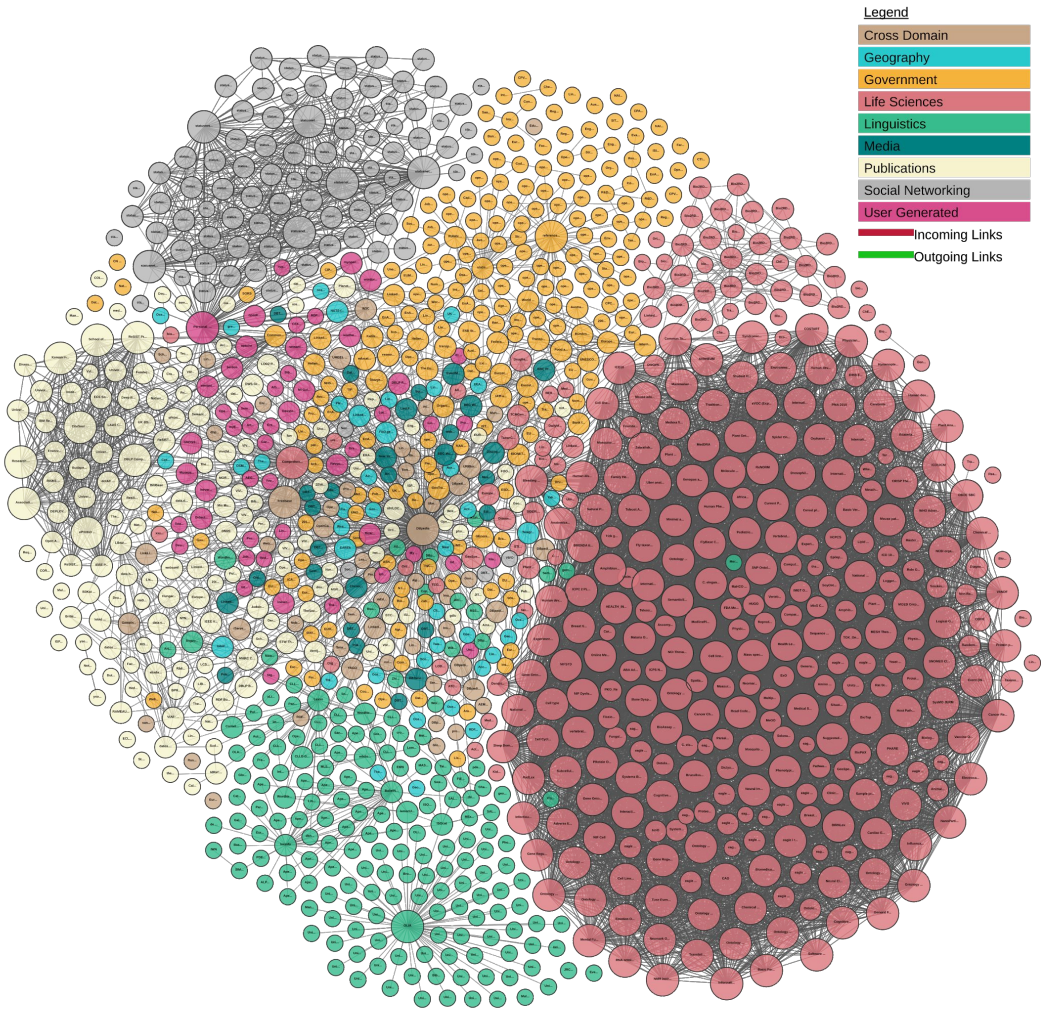
"Abstract schema and instances"

Graph-structured world model

World models?

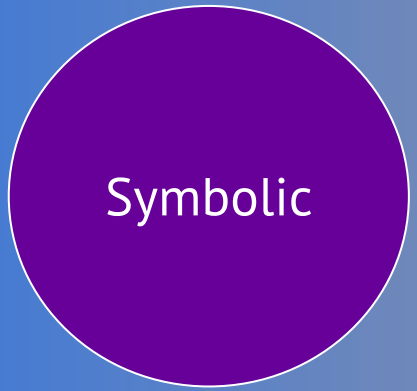
Entities and relations define our **domain of discourse**

How to encode it?

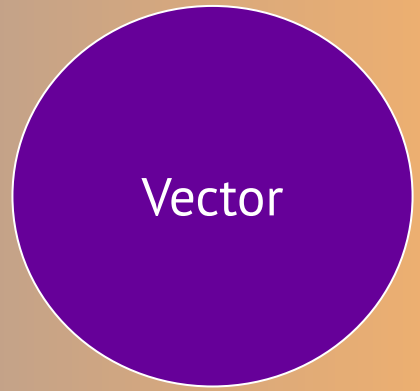


Source: https://lod-cloud.net/

On representation of Knowledge Graphs

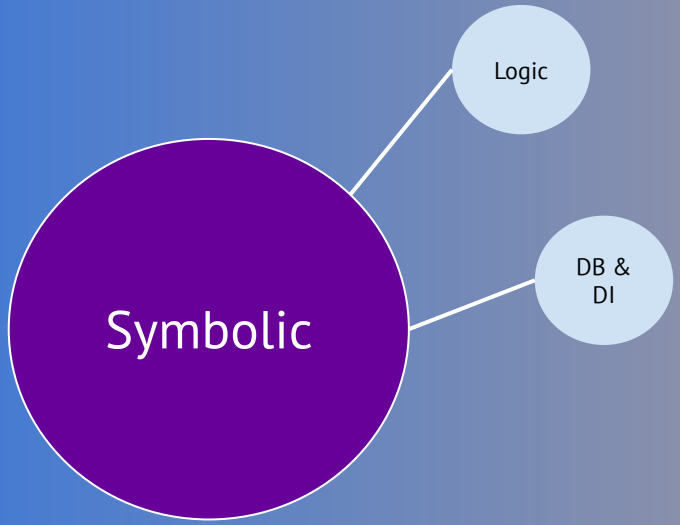


s,p,o
p(s,o)
(s,p,o)



$$s, p, o \in \mathbb{R}^d$$

On representation of Knowledge Graphs

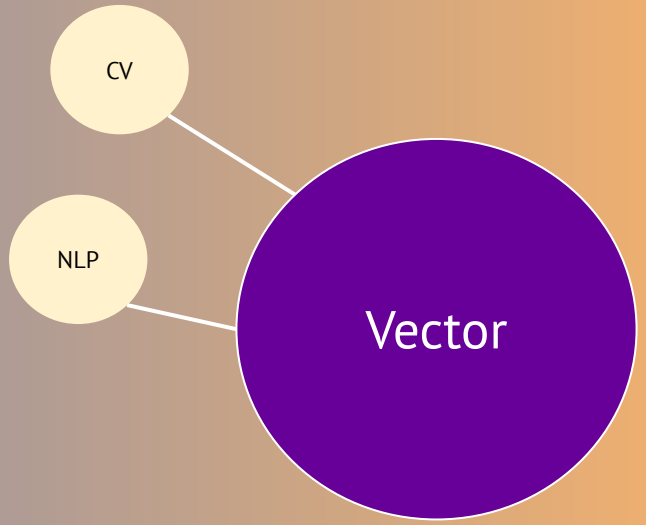


s,p,o
p(s,o)
(s,p,o)

Open-world assumption

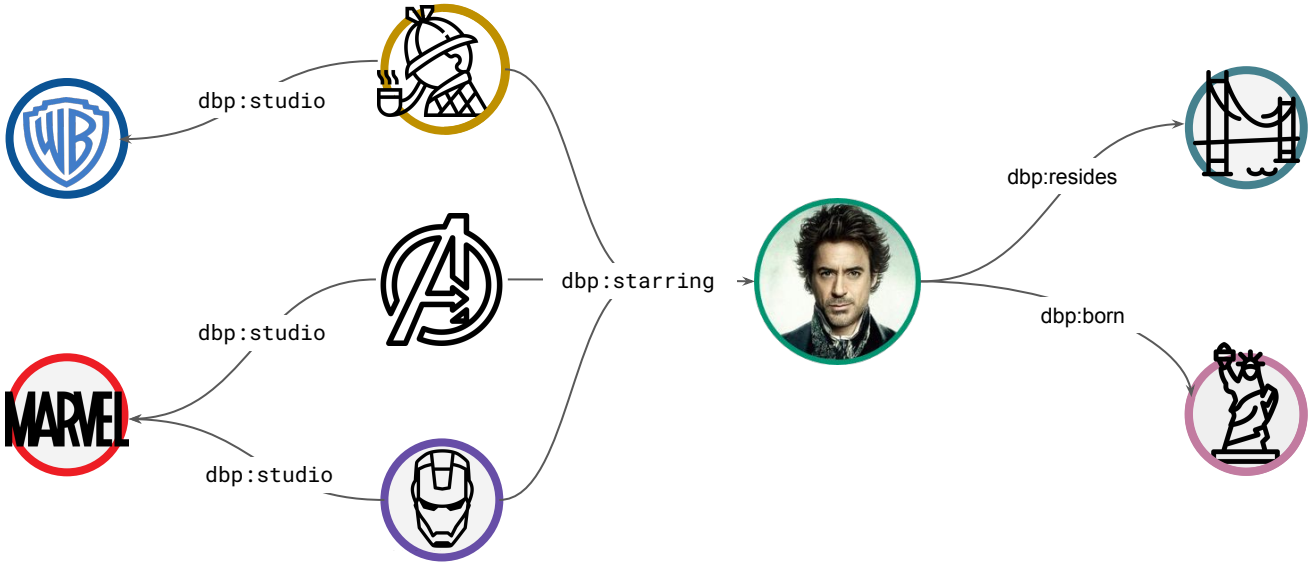
Closed-world assumption

Temporal / evolving



$s, p, o \in \mathbb{R}^d$

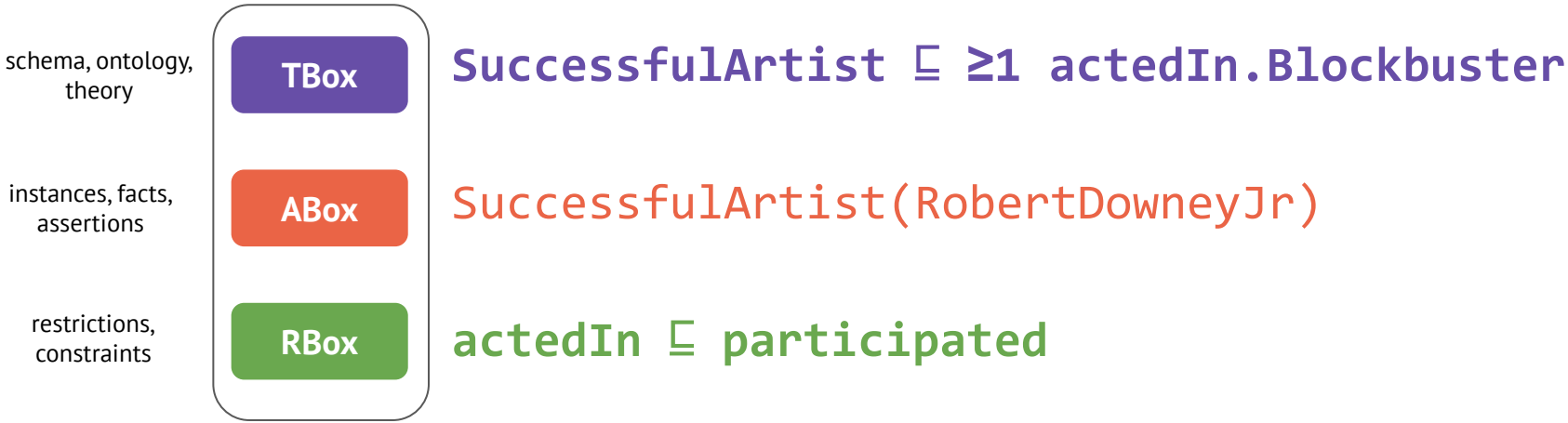
Symbolic: Triples




```
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 .
```

Symbolic: Description Logics

Based on logical formalisms, e.g., Description Logics (DL), RDFS, OWL

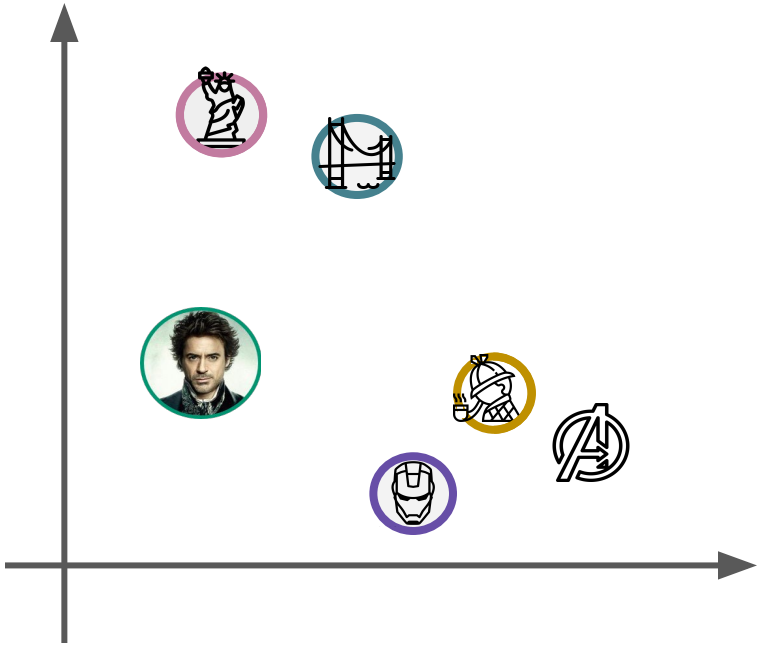


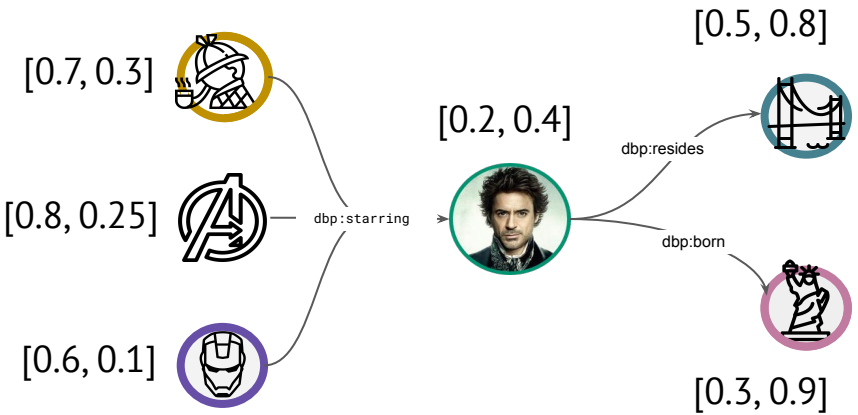
Logically consistent collection of axioms

Vector: Embeddings

$$E \in \mathbb{R}^{N_e \times d}$$

$$R \in \mathbb{R}^{N_r \times d}$$



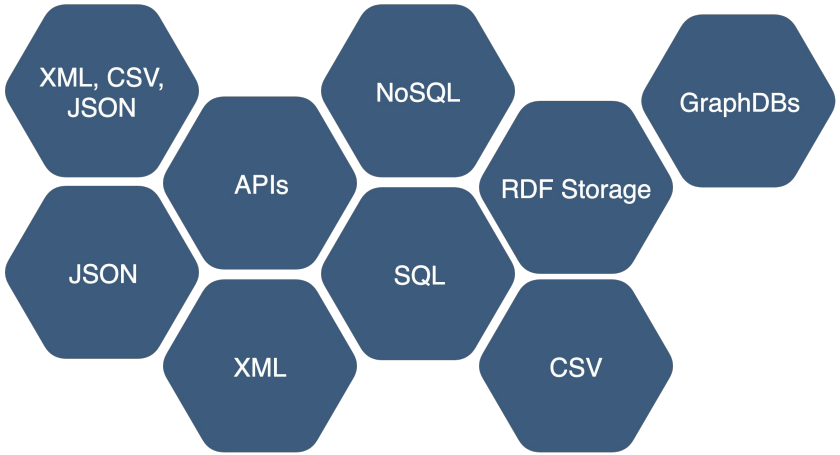


Building KGs

Knowledge Graph

Semantic Data Integration

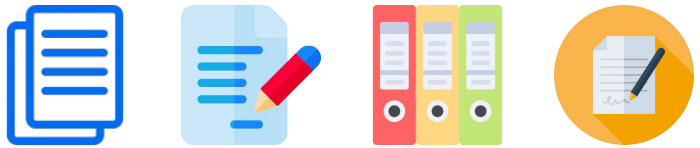
Structured Sources



Knowledge Graph

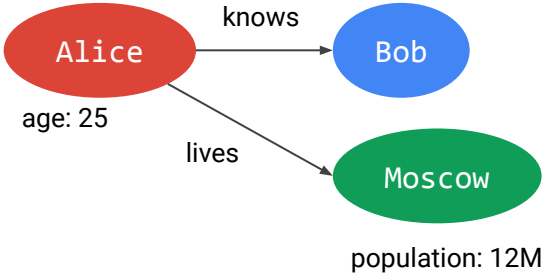
Information Retrieval & NLP

Unstructured Sources



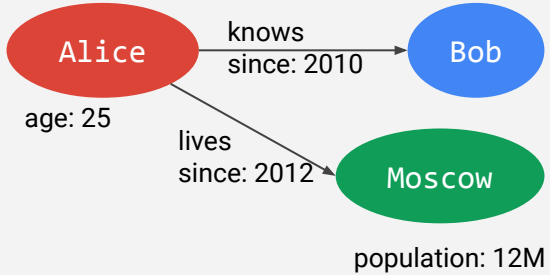
Graph Databases

RDF



- Query language: SPARQL
- Predicate attributes only from RDFS/OWL
- Semantic schema
- Logical reasoning

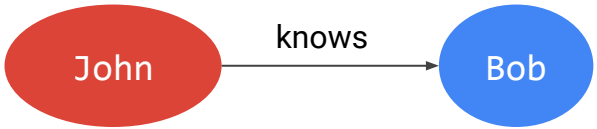
LPG (Labeled Property Graph)



- Query languages: Cypher, Gremlin, GraphQL
- Key-value predicate attributes
- Non-semantic schema
- No reasoning

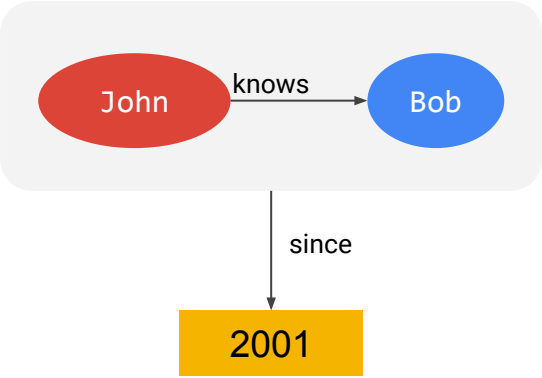
Graph Databases - Queries

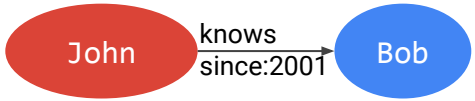
SPARQL	Cypher
SELECT ?s ?friend WHERE { ?s a :Person; :name “John” ; :knows ?friend .}	MATCH (s:Person)-[:knows]-(friend) WHERE s.name = “John” RETURN s, friend ;



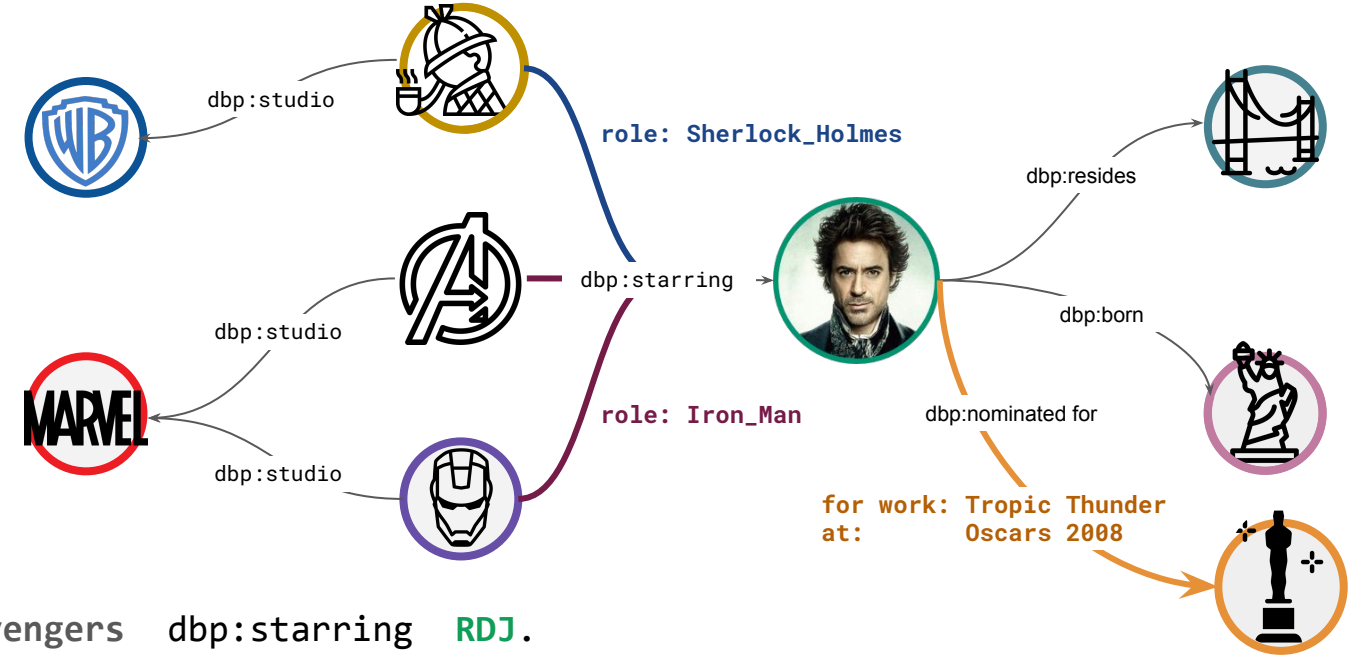
Graph Databases - Queries

SPARQL* (Reification)	Cypher
SELECT ?s WHERE { <<?s :knows :js>> :since 2001 }	MATCH (s:Person)-[:knows {since:2001}] -> (js) RETURN s;



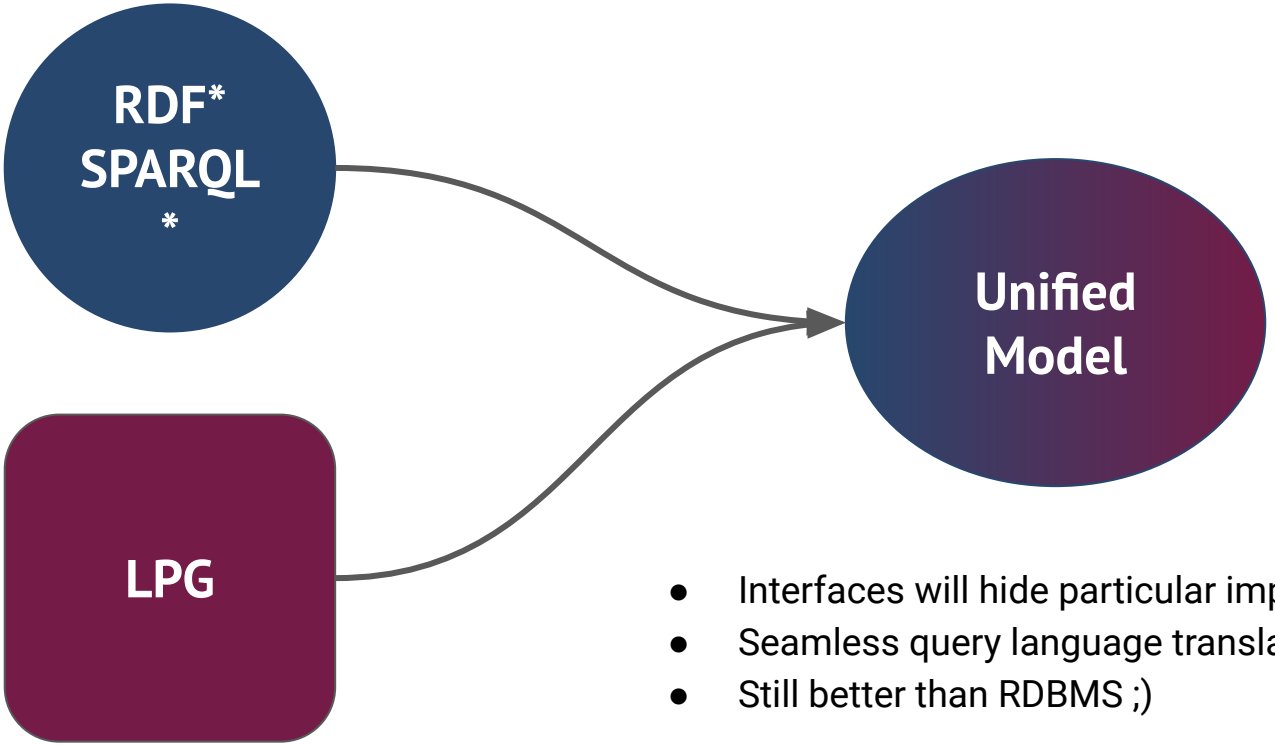


Hyper-Relational KGs: RDF* and SPARQL*



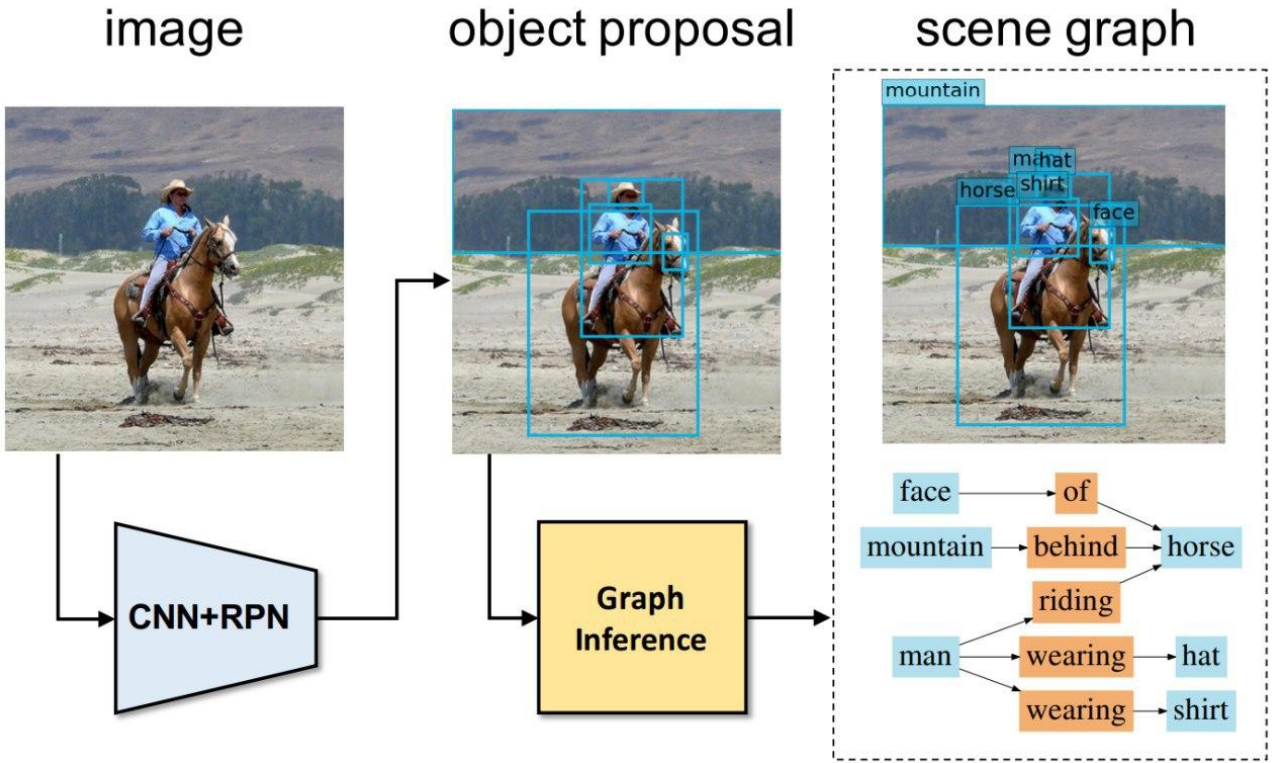

```
The_Avengers  dbp:starring  RDJ.
<< The_Avengers dbp:starring RDJ >> role Iron_Man .
<< RDJ dbp:nominated_for Oscar >> for_work Tropic_Thunder;
                                  at       Oscars_2008 .
```

Graph Databases - Convergence



- Interfaces will hide particular implementations
- Seamless query language translation
- Still better than RDBMS ;)

POV: Computer Vision



Entities:

Identified objects

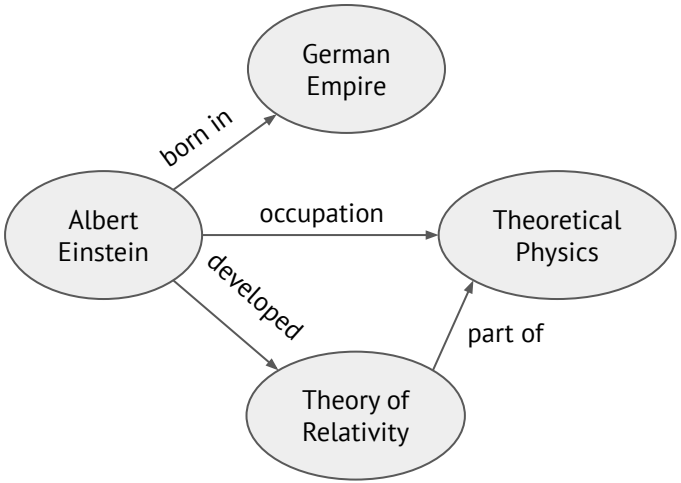
Relations:

Inferred as to objects arrangement

Xu et al. Scene Graph Generation by Iterative Message Passing. CVPR 2017

POV: NLP - Building KGs from texts

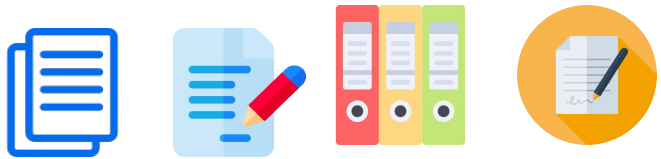
Albert Einstein was a German-born theoretical physicist who developed the theory of relativity.



Knowledge Graph

Information Retrieval

Unstructured Sources



POV: NLP - Named Entity Recognition

Entities in a Knowledge Graph

apple (Q89)

fruit of the apple tree
apples

Apple (Q1754545)

1990 album by Mother Love bone

Apple (Q213710)

UK international record label; imprint of Apple Corps Ltd.
LC 01074 | LC 1074 | Apple Records

Apple Inc. (Q312)

American producer of hardware, software, and services, based in Cupertino, California
Apple Computer, Inc. | Apple Computer | Apple Computer Inc | Apple | Apple Incorporated | Apple Computer Incorporated |

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

POV: NLP - Relation Linking

Relations in a Knowledge Graph

- List of known relations
- Surface forms (synonyms), easily multi-lingual
- Relations constraints
- Relations hierarchy
- Most used types of subjects and objects

wdt:P161

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

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

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

RDJ **has acted** in which movies?

cast member (P161)

actor in the subject production |

starring | film starring | actor | actress | contestant or a play

performer (P175)

actor, musician, band or other performer associated with this role or musical work

artist | musician | played by | portrayed by | recorded by | recording by | dancer | actor | musical artist

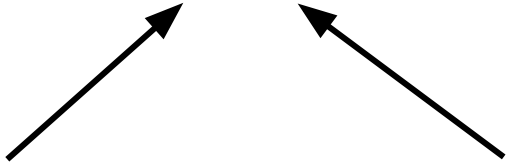
POV: NLP - Question Answering

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
}
```

POV: NLP - Language Modeling

Robert Downey Jr. portrayed **[MASK]** in the Marvel movie in 2008.



Knowledge Graph

(Iron Man, cast member, Robert Downey Jr)
(Iron Man, production company, Marvel)
(Iron Man, released, 2008)
(Robert Downey Jr, character role, Tony Stark)
(Tony Stark, pseudonym, Iron Man)

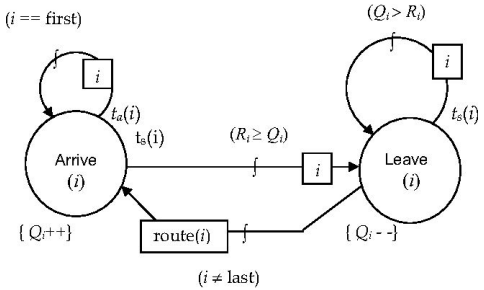
Precise facts | Entities & relations | Explainability

Unstructured Sources

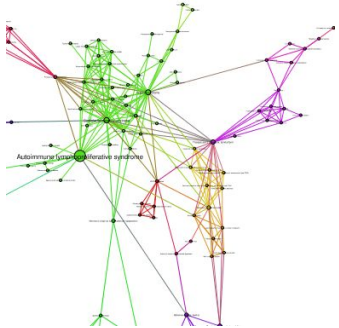
Large-scale text corpora
(Wikipedia, OpenBooks, Reddit, CommonCrawl, etc)

KGs in Graph ML

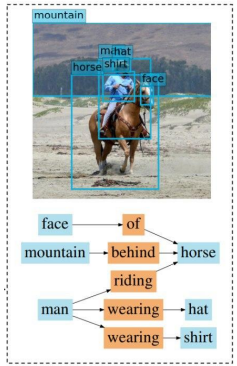
Cell similarity networks



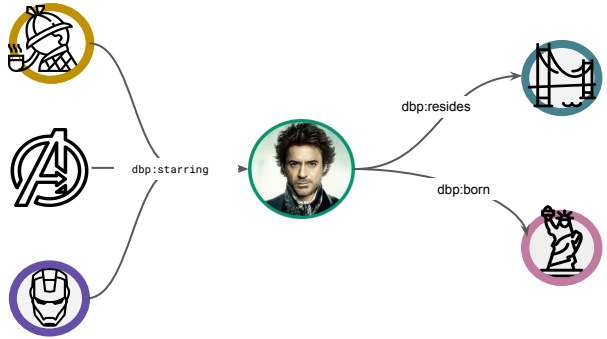
Event graphs



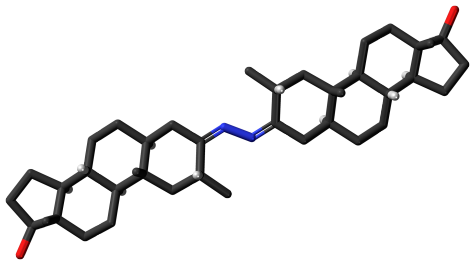
Disease pathways



Scene Graphs

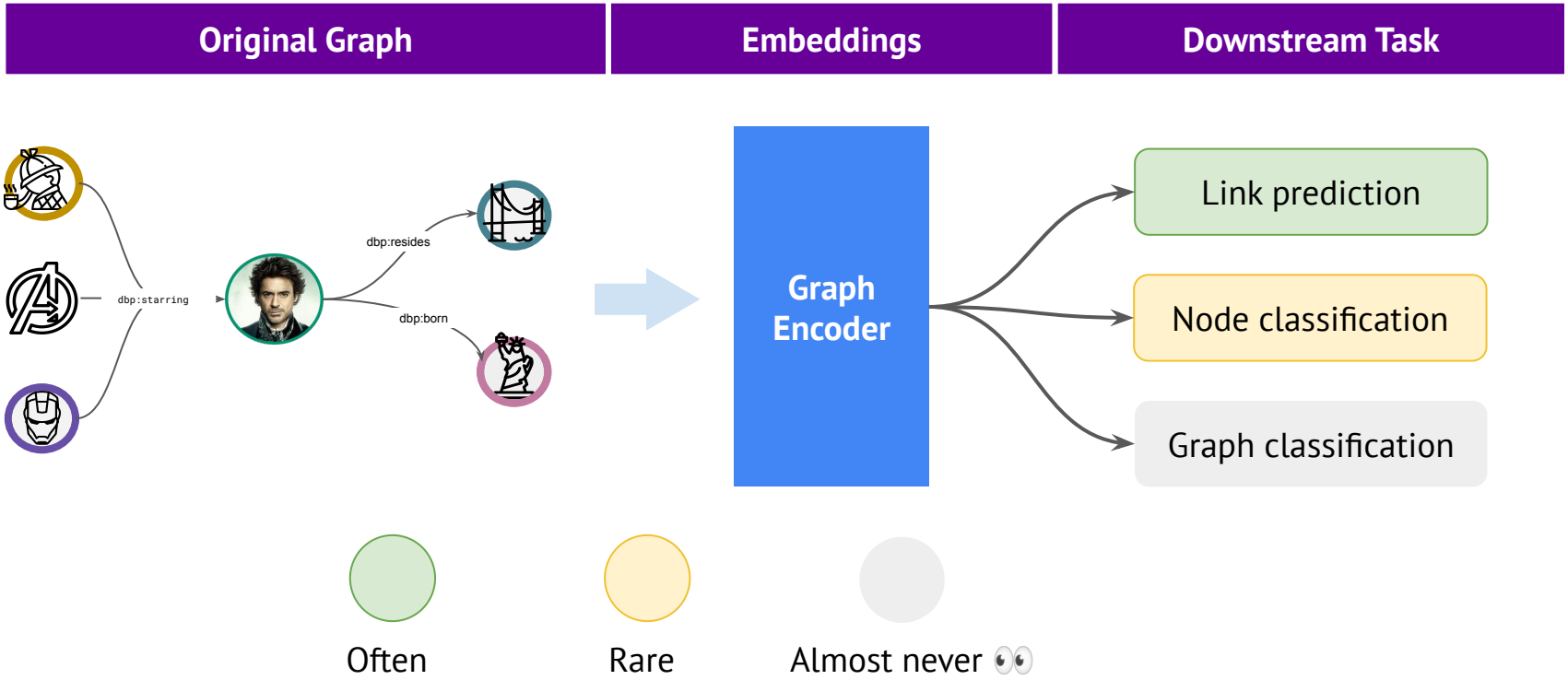


Knowledge Graphs



Molecules

KGs in Graph ML



KG Embeddings: Link Prediction

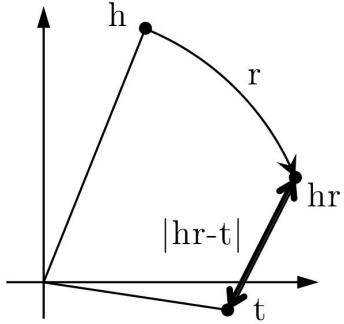
Tensor Factorization

Translation

Convolution

- Rank link probabilities with a score function
- Transductive setup: can only predict known relations among seen entities
- No node features - random initialization of embeddings

$$\text{score}(h, r, t)$$



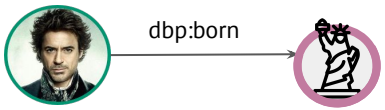
(b) RotatE models **r** as rotation in complex plane.

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	$\text{Re}(\langle \mathbf{h}, \mathbf{r}, \bar{\mathbf{t}} \rangle)$	✓	✓	✓	✗
RotatE	$-\|\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|\|$	✓	✓	✓	✓

Sun et al. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. ICLR 2019

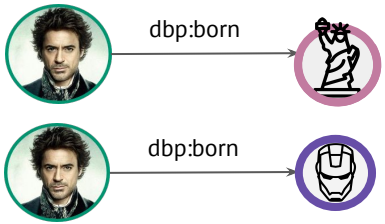
KG Embeddings: Link Prediction

Pointwise



$$\text{score}(h, r, t)$$

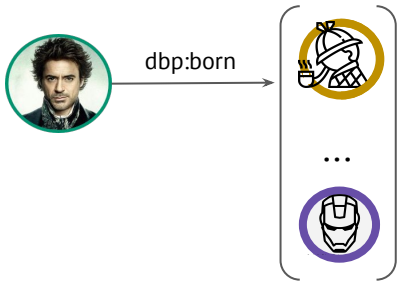
Negative Sampling



$$\text{score}(h, r, t)$$

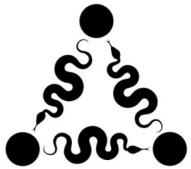
$$\text{score}(h, r, t')$$

1-N scoring



$$\text{score}(h, r) \cdot E^T$$

KG Embeddings: PyKEEN 1.0



PyKEEN

https://github.com/pykeen/pykeen

- → **PyTorch** 😍
- → 13 datasets + your own graphs
- → 23 KG embedding models and counting

- → 7 losses
- → 6 optimizers
- → 6 metrics
- → 5 regularizers
- → 2 training loops
- → 2 negative samplers
- → Tracking in MLFlow, WANDB

📈 Benchmarked!

Ali et al. Bringing Light Into the Dark: A Large-scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework. arxiv:2006.13365

GNN Encoders for KGs

GCN
GraphSAGE
GAT
GIN

GNN Encoders for KGs

- 👉 Node features are absent
- 👉 Multi-relational (100-1000 edge types)

- 👉 GCN, GAT, GIN, etc do not explicitly model edges
- 👉 GCN, GAT, GIN, etc subsume homogeneous graphs

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

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

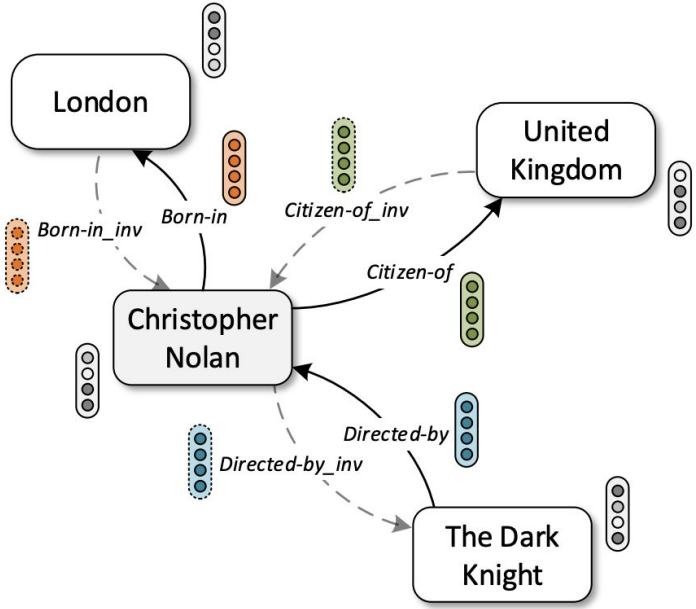
R-GCN [1]: a whole matrix W per relation

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

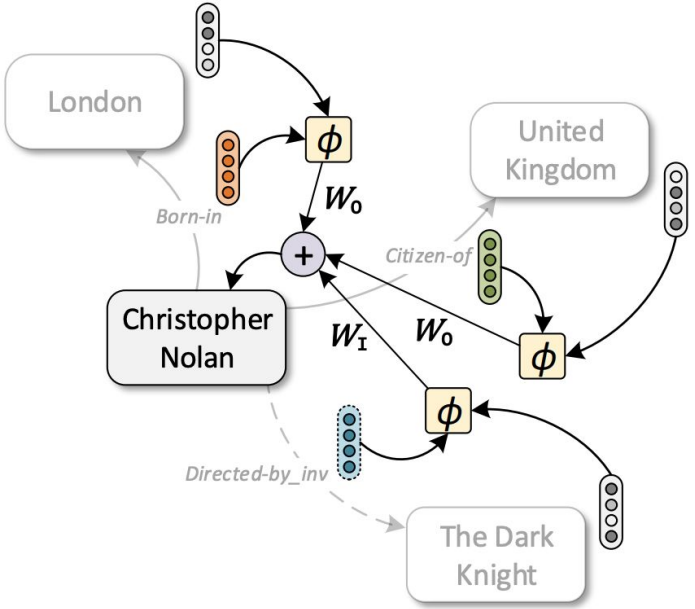
CompGCN [2]: a vector z_r per relation +
composition of (h,r) +
only 3 different W: input/output/self-loop

[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

Multirelational GNN Encoders for KGs



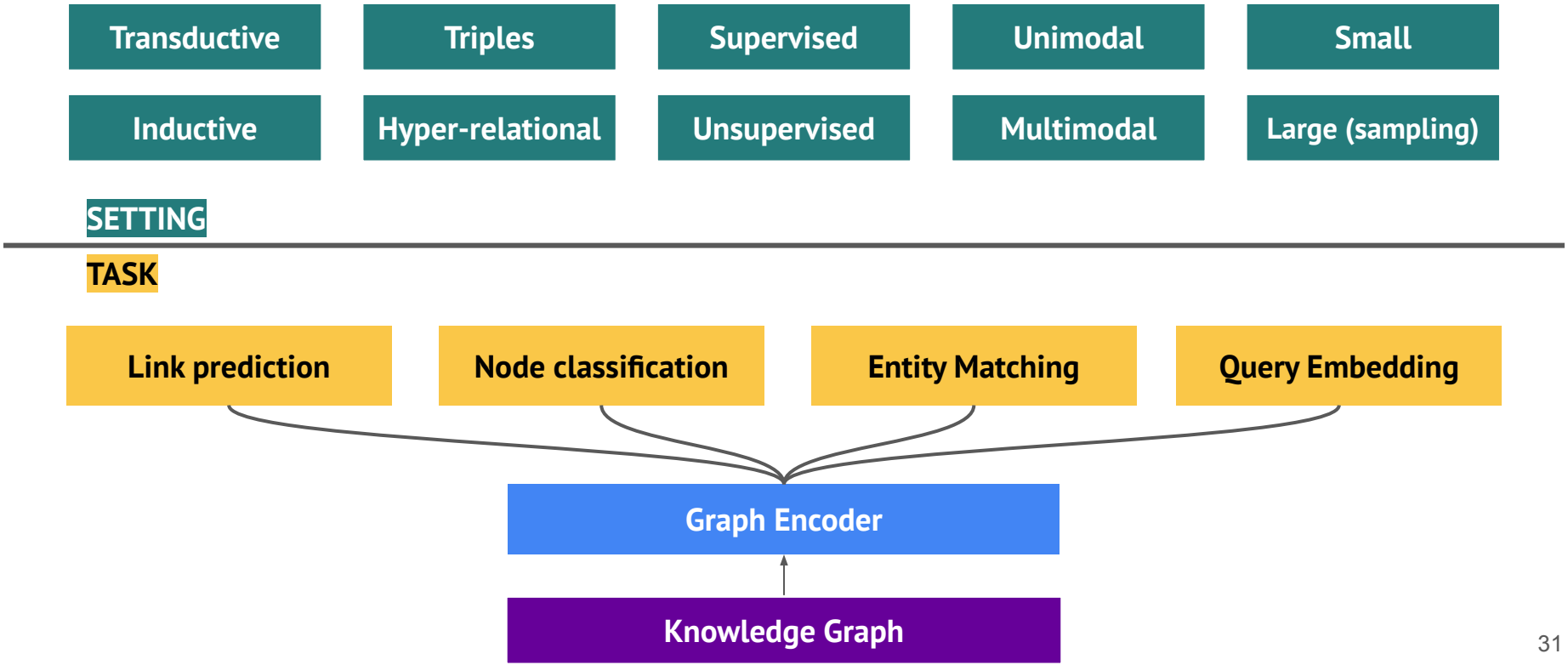
Relational Graph with Embeddings



CompGCN Update

Vashishth et al. Composition-Based Multi-Relational Graph Convolutional Networks. ICLR 2020

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



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

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

SETTING

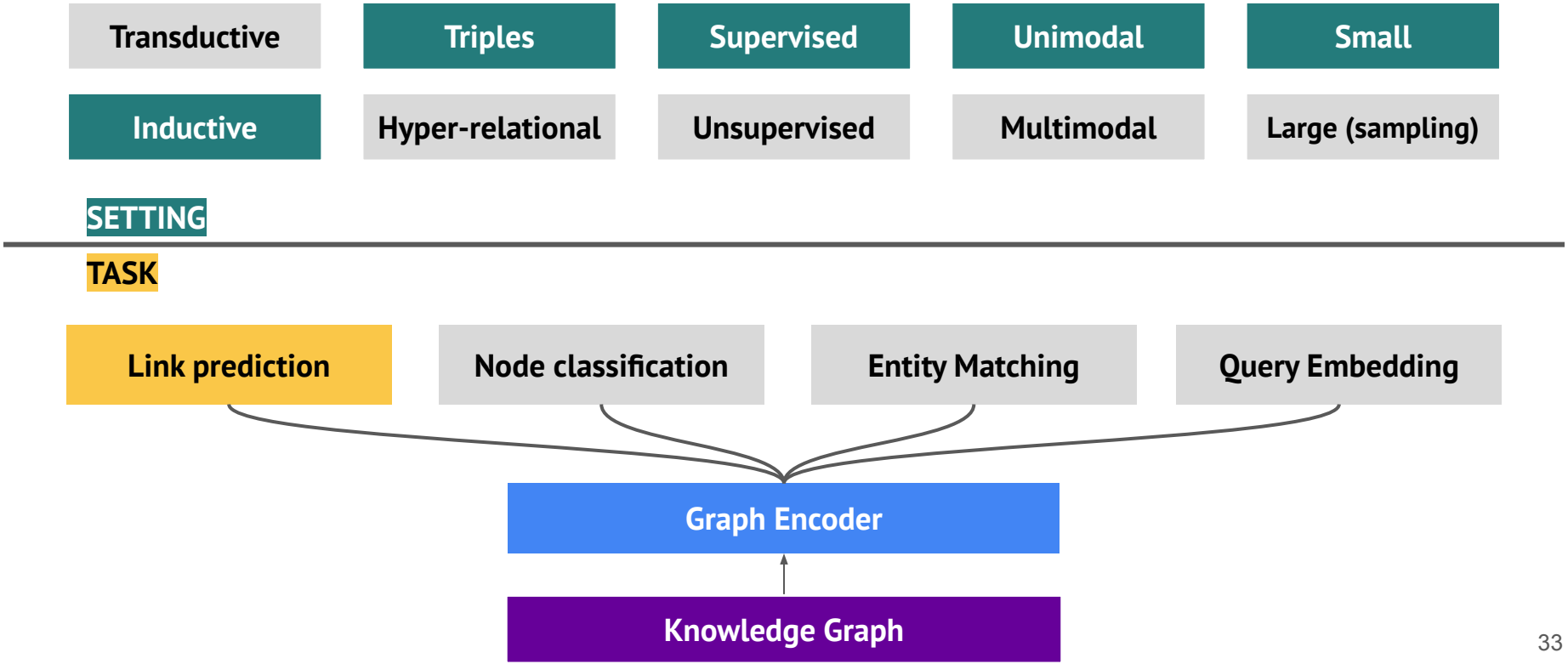
TASK

Link prediction | Node classification | Entity Matching | Query Embedding

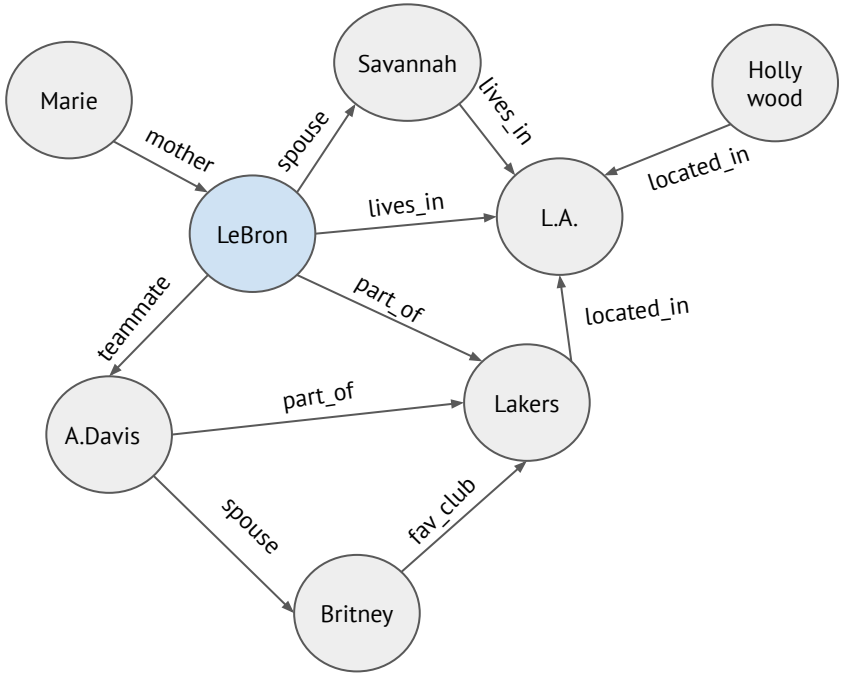
Graph Encoder

Knowledge Graph

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



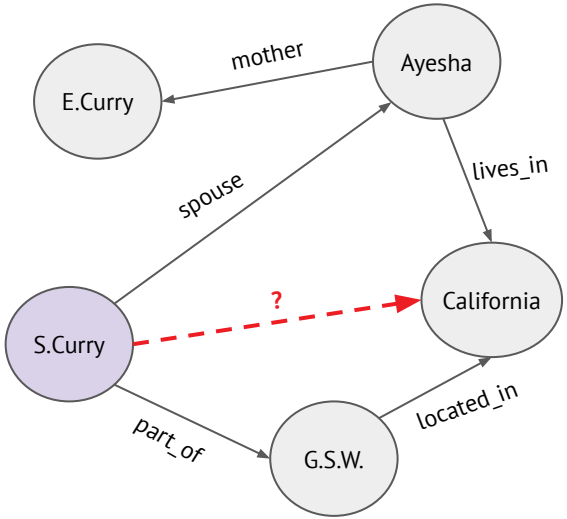
Inductive Link Prediction in KGs



Training graph

Inductive inference

- Unseen nodes
- Known relations



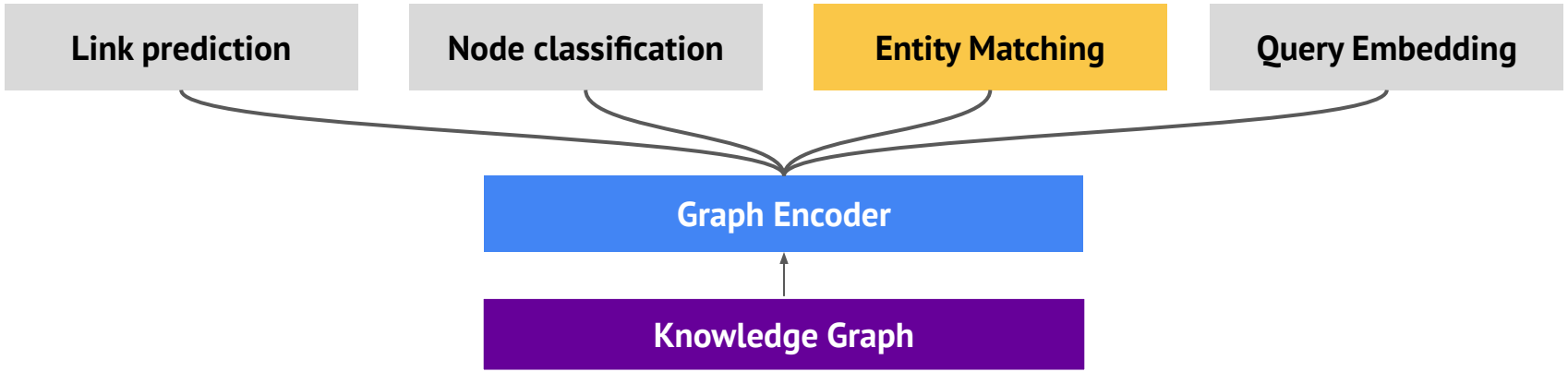
Teru et al. Inductive Relation Prediction by Subgraph Reasoning. ICML 2020

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

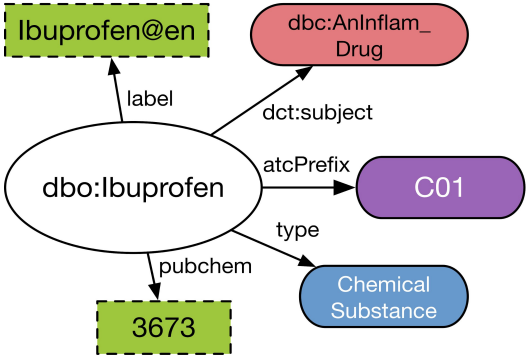
Transductive	Triples	Supervised	Unimodal	Small
Inductive	Hyper-relational	Unsupervised	Multimodal	Large (sampling)

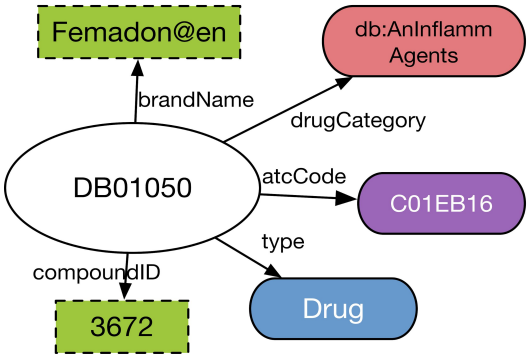
SETTING

TASK



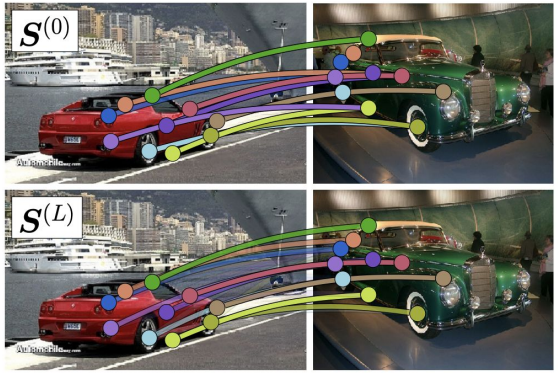
KGs in Graph ML: Entity Matching





The same entity in two KGs
DBpedia vs DrugBank

(a) Motorbike



(b) Car

Similar objects

Fey et al. Deep Graph Matching Consensus. ICLR 2020

KGs in Graph ML: Entity Matching

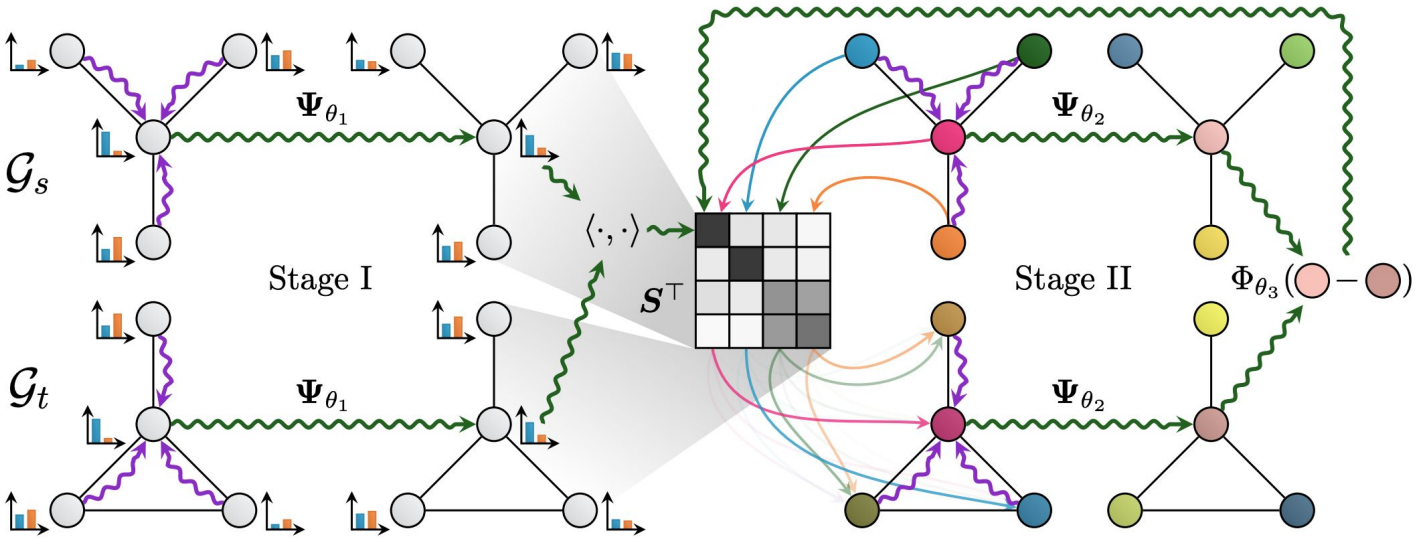
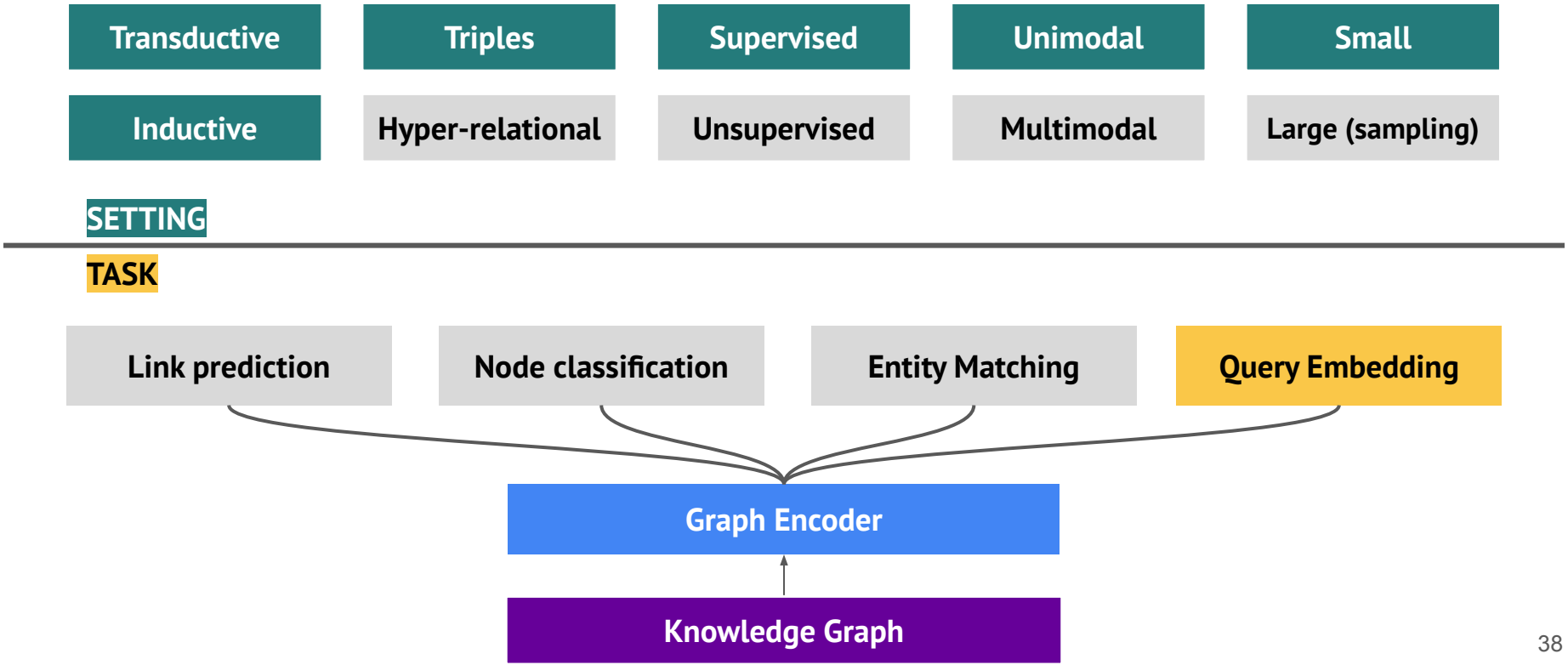


Figure 1: High-level illustration of our two-stage neighborhood consensus architecture. Node features are first locally matched based on a graph neural network Ψ_{θ_1}, before their correspondence scores get iteratively refined based on neighborhood consensus. Here, an injective node coloring of $\mathcal{G}_s$ is transferred to $\mathcal{G}_t$ via $\boldsymbol{S}$, and distributed by Ψ_{θ_2} on both graphs. Updates on $\boldsymbol{S}$ are performed by a neural network Φ_{θ_3} based on pair-wise color differences.

Fey et al. Deep Graph Matching Consensus. ICLR 2020

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



KGs in Graph ML: Query Embedding

Where did Canadian citizens with Turing Award graduate?

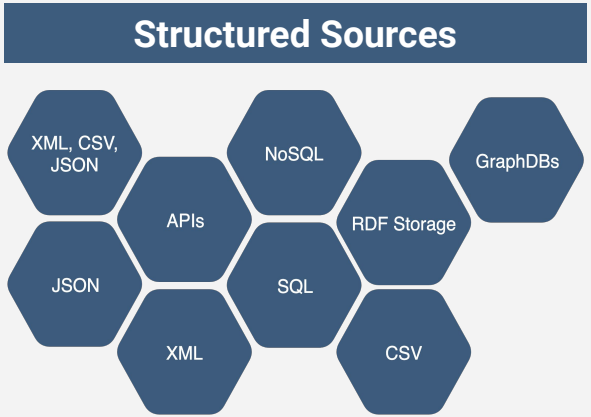
```
SELECT ?y WHERE {
     ?x :win       :TuringAward .
     ?x :citizen   :Canada .
     ?x :graduate  ?y . }
```

KGs in Graph ML: Query Embedding

Where did Canadian citizens with Turing Award graduate?

```
SELECT ?y WHERE {
     ?x :win        :TuringAward .
     ?x :citizen    :Canada .
     ?x :graduate   ?y . }
```

query



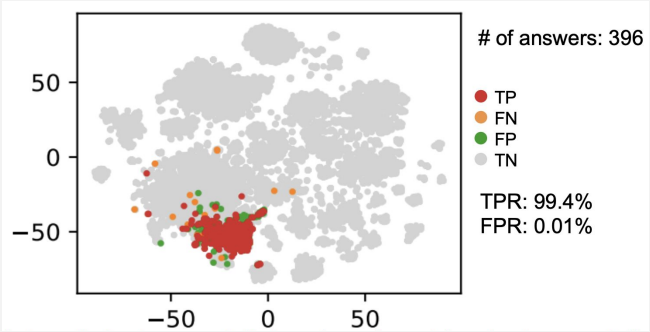
KGs are sparse and incomplete

KGs in Graph ML: Query Embedding

Where did Canadian citizens with Turing Award graduate?

```
SELECT ?y WHERE {
     ?x :win        :TuringAward .
     ?x :citizen    :Canada .
     ?x :graduate   ?y . }
```

embed

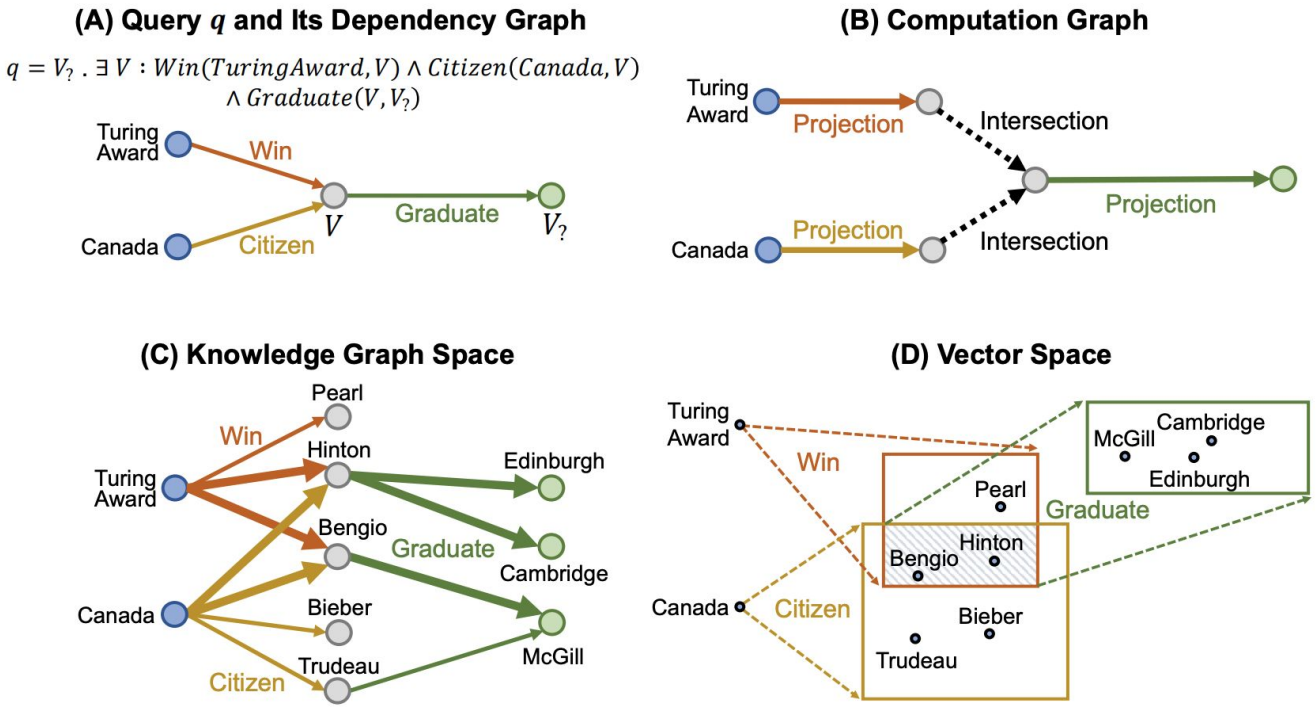


Execution in a vector space

Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
Daza et al. Message Passing Query Embedding. GRL @ ICML 2020
cs224w. snap.stanford.edu

KGs in Graph ML: Query Embedding

Subset of SPARQL - EPFO queries: Conjunctive + disjunction



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

Data Fest 2020

Thanks!

@migalkin

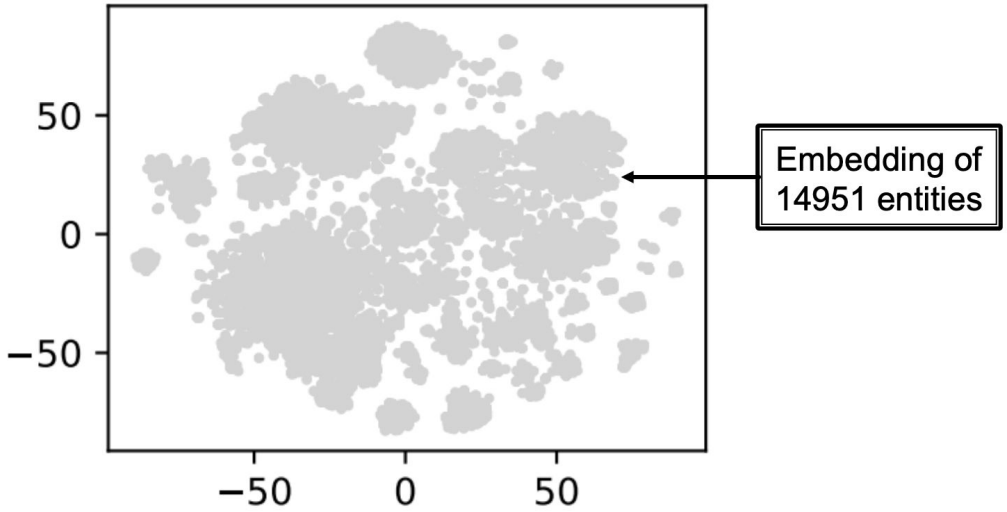
@michael_galkin

@mgalkin

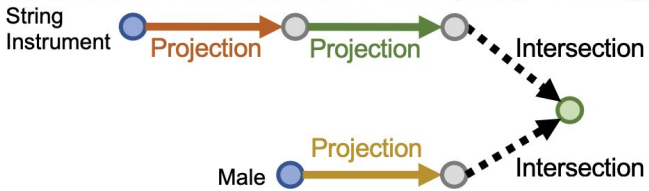
mikhail.galkin@tu-dresden.de

https://t.me/graphML

KGs in Graph ML: Query Embedding

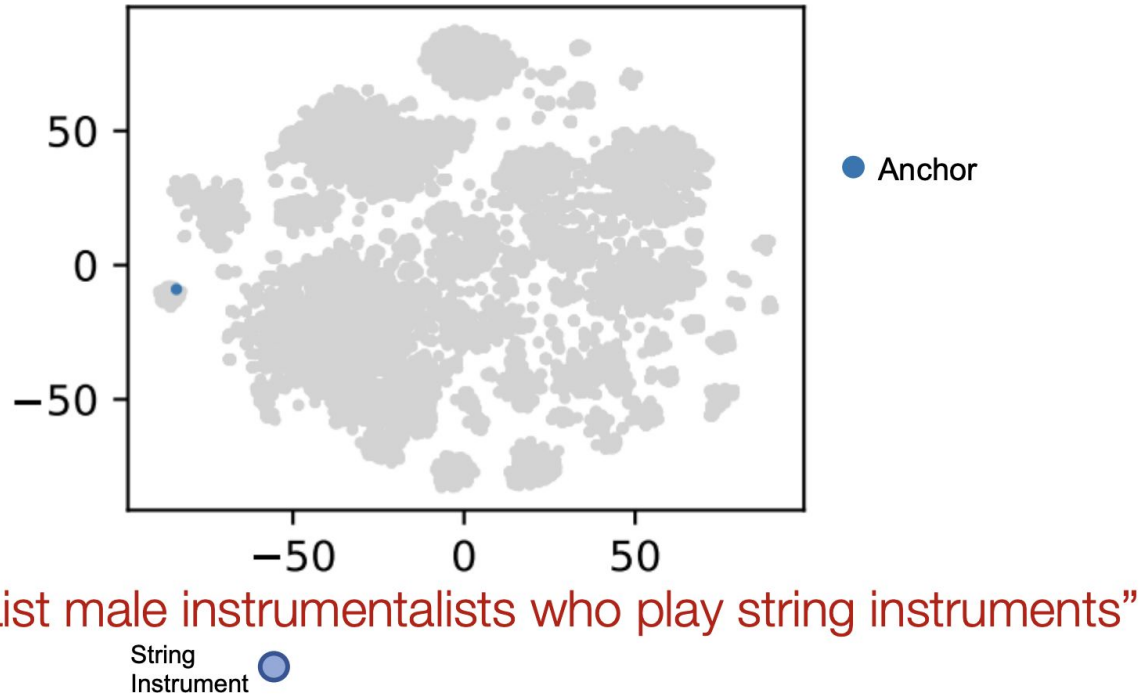


"List male instrumentalists who play string instruments"



Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
cs224w. snap.stanford.edu

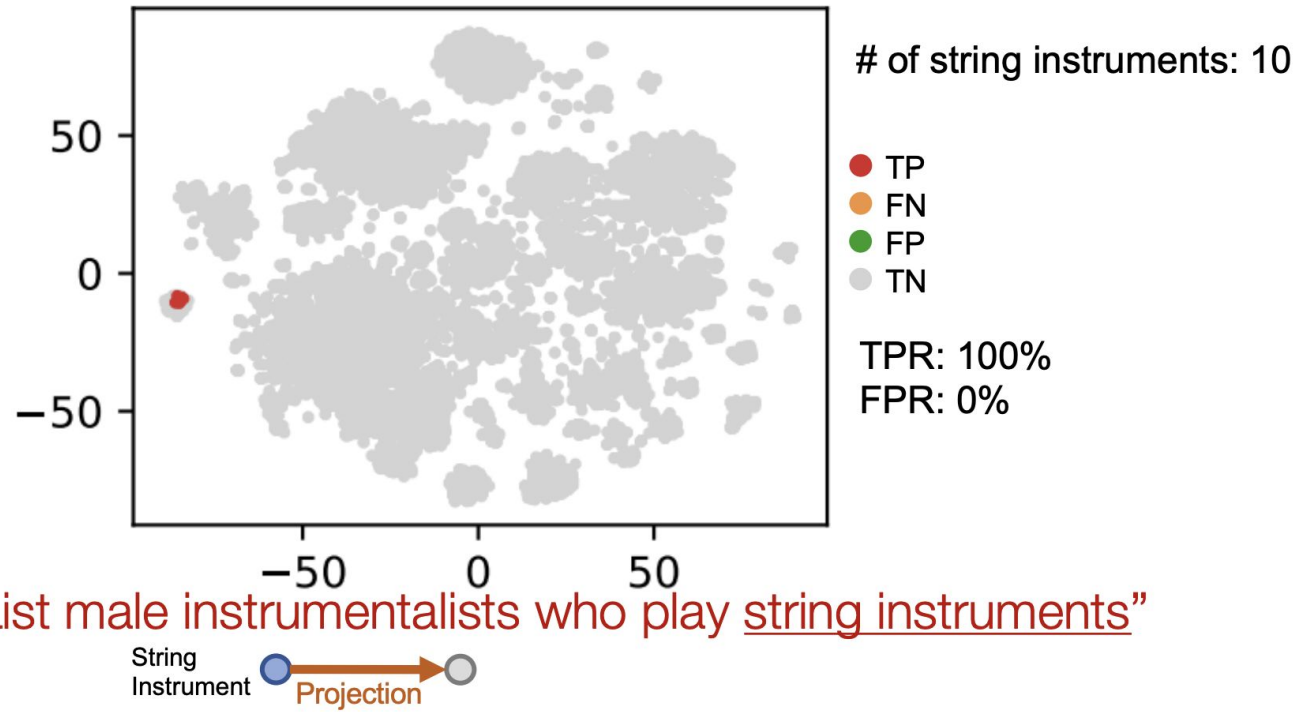
KGs in Graph ML: Query Embedding



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

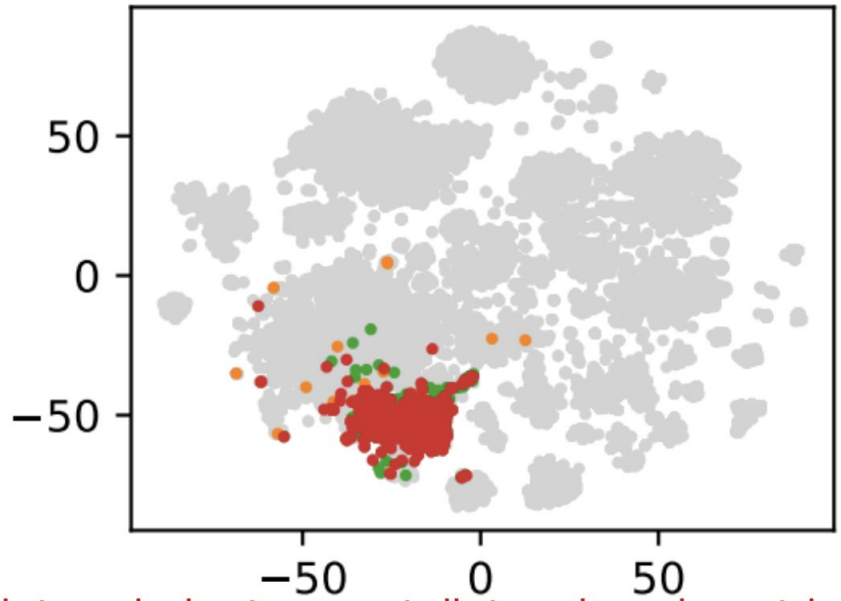
cs224w. snap.stanford.edu

KGs in Graph ML: Query Embedding



Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
cs224w. snap.stanford.edu

KGs in Graph ML: Query Embedding



# of instrumentalists: 472

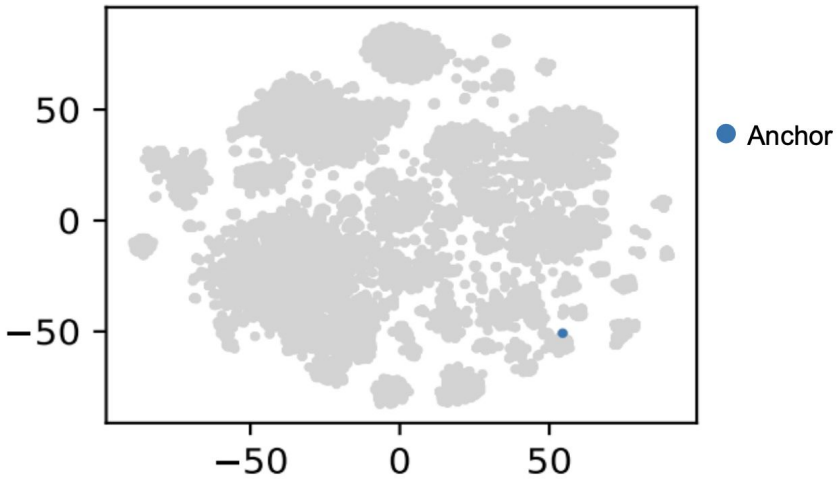
TP
FN
FP
TN

TPR: 98.4%
FPR: 0.01%

"List male instrumentalists who play string instruments"

String Instrument → Projection → Projection

KGs in Graph ML: Query Embedding

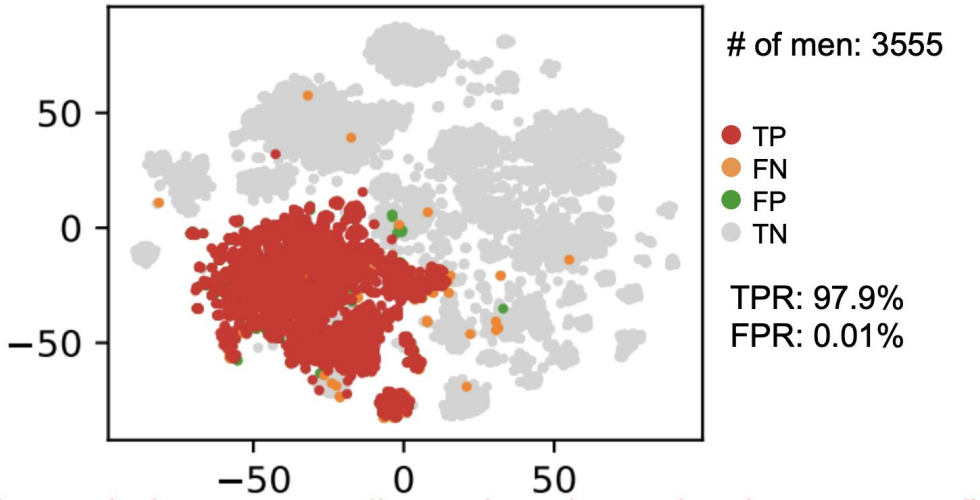


“List male instrumentalists who play string instruments”

Male

Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
cs224w. snap.stanford.edu

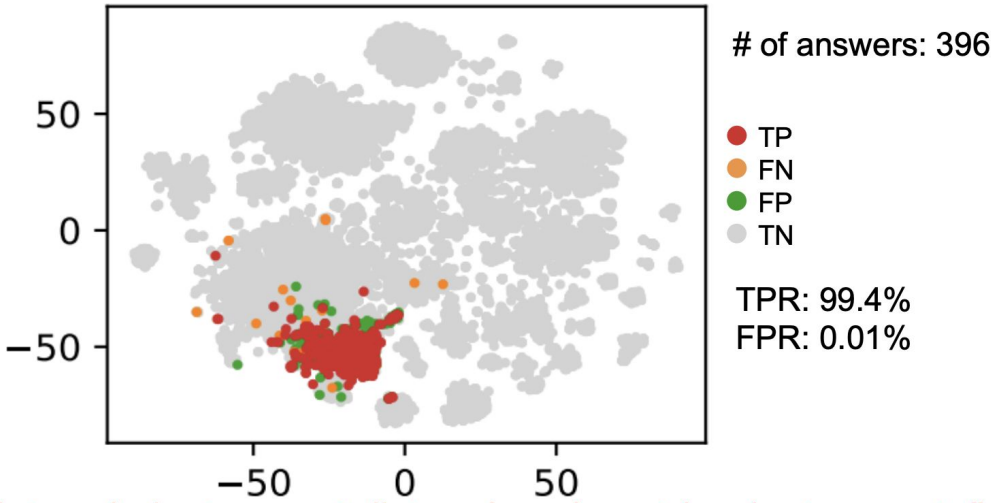
KGs in Graph ML: Query Embedding



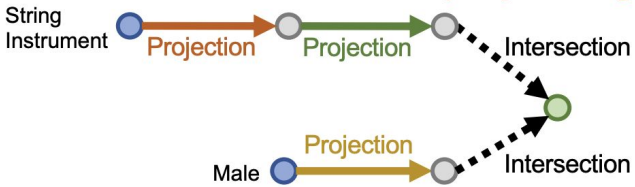
"List male instrumentalists who play string instruments"

Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
cs224w. snap.stanford.edu

KGs in Graph ML: Query Embedding

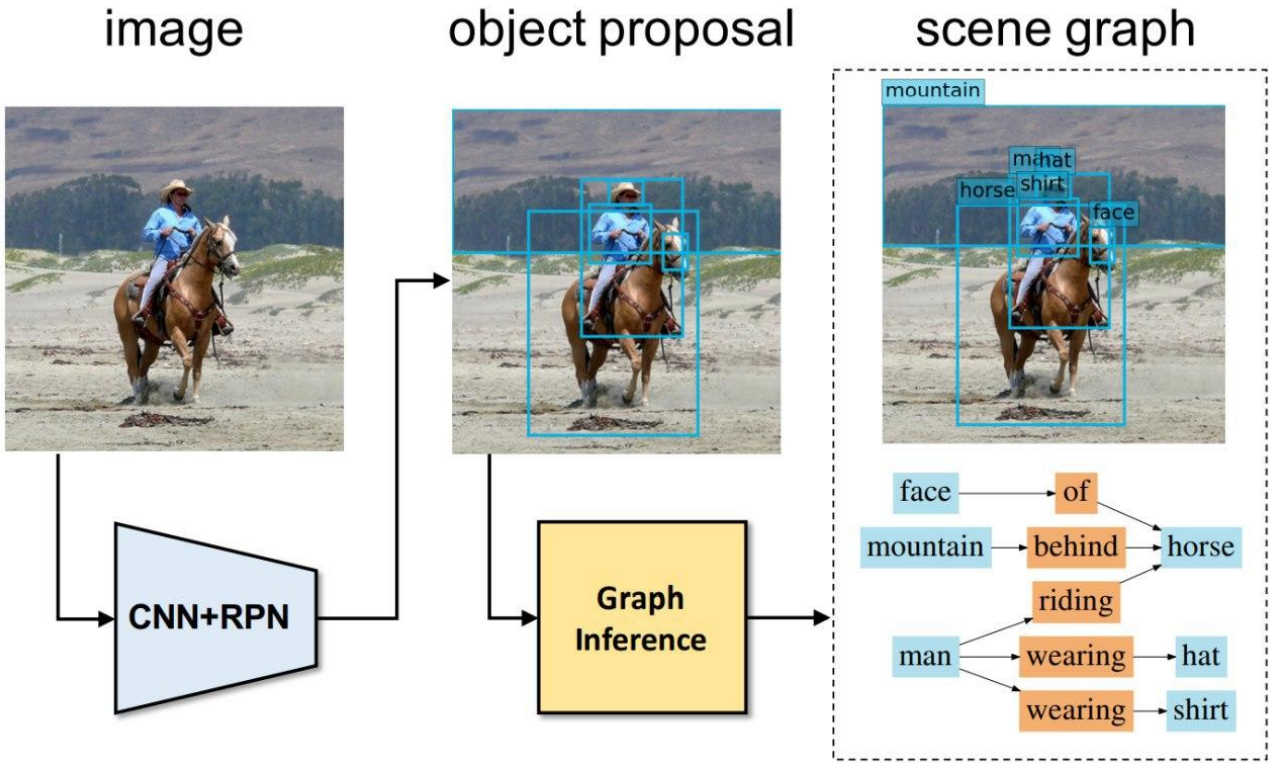


"List male instrumentalists who play string instruments"



Ren et al. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. ICLR 2020
cs224w. snap.stanford.edu

POV: Computer Vision



Entities:

Identified objects

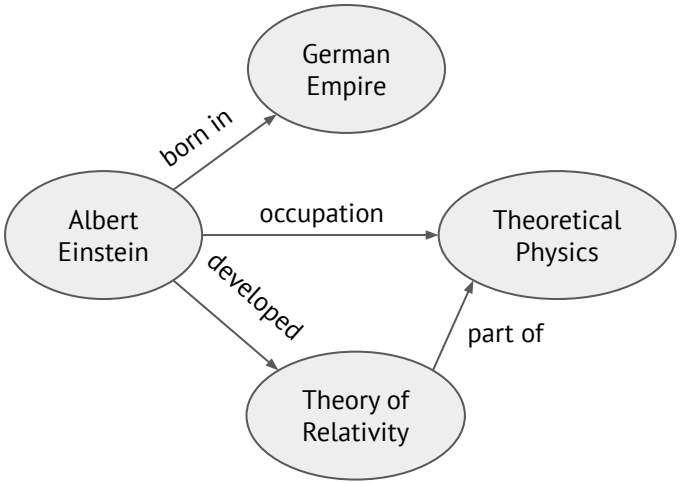
Relations:

Inferred as to objects arrangement

Xu et al. Scene Graph Generation by Iterative Message Passing. CVPR 2017

POV: NLP - Building KGs from texts

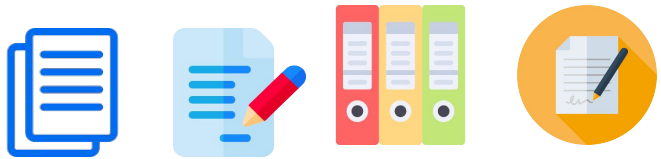
Albert Einstein was a German-born theoretical physicist who developed the theory of relativity.



Knowledge Graph

Information Retrieval

Unstructured Sources



POV: NLP - Named Entity Recognition

Entities in a Knowledge Graph

apple (Q89)

fruit of the apple tree
apples

Apple (Q1754545)

1990 album by Mother Love bone

Apple (Q213710)

UK international record label; imprint of Apple Corps Ltd.
LC 01074 | LC 1074 | Apple Records

Apple Inc. (Q312)

American producer of hardware, software, and services, based in Cupertino, California
Apple Computer, Inc. | Apple Computer | Apple Computer Inc | Apple | Apple Incorporated | Apple Computer Incorporated |

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

POV: NLP - Relation Linking

Relations in a Knowledge Graph

- List of known relations
- Surface forms (synonyms), easily multi-lingual
- Relations constraints
- Relations hierarchy
- Most used types of subjects and objects

wdt:P161

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

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

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

RDJ **has acted** in which movies?

cast member (P161)

actor in the subject production |

starring | film starring | actor | actress | contestant or a play

performer (P175)

actor, musician, band or other performer associated with this role or musical work

artist | musician | played by | portrayed by | recorded by | recording by | dancer | actor | musical artist

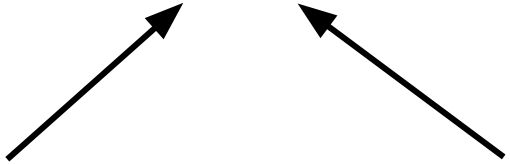
POV: NLP - Question Answering

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
}
```

POV: NLP - Language Modeling

Robert Downey Jr. portrayed **[MASK]** in the Marvel movie in 2008.



Knowledge Graph

(Iron Man, cast member, Robert Downey Jr)
(Iron Man, production company, Marvel)
(Iron Man, released, 2008)
(Robert Downey Jr, character role, Tony Stark)
(Tony Stark, pseudonym, Iron Man)

Precise facts | Entities & relations | Explainability

Unstructured Sources

Large-scale text corpora
(Wikipedia, OpenBooks, Reddit, CommonCrawl, etc)