

COMP4222 Machine Learning with Structured Data

Heterogeneous Graphs and Knowledge Graphs

Instructor: Yangqiu Song

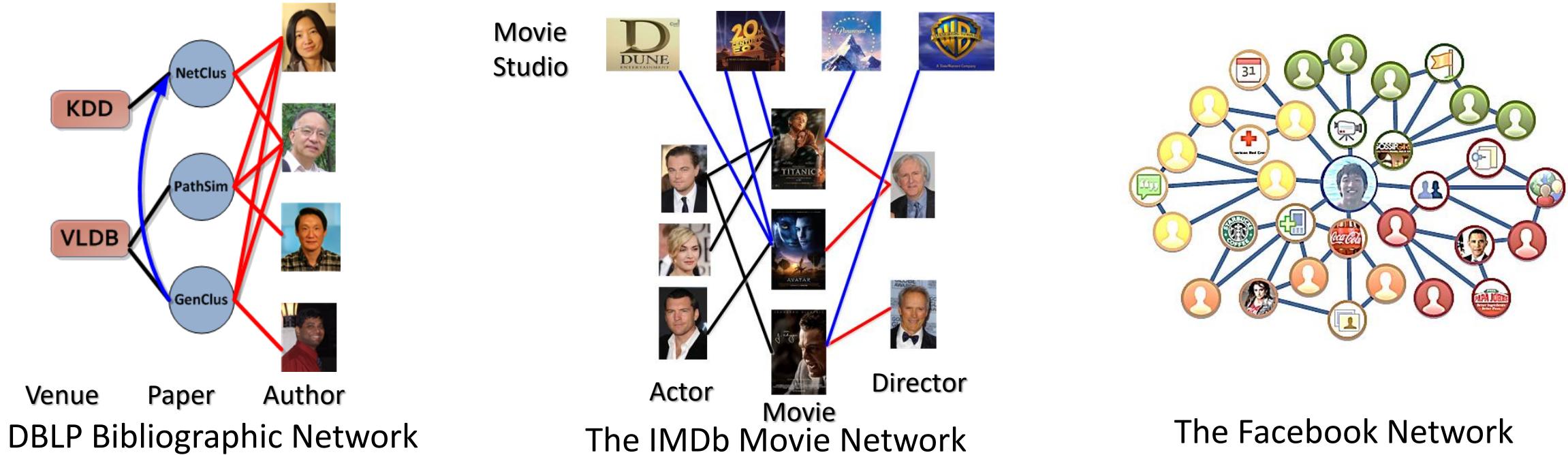
Slides credits: Jure Laskovec, Yizhou Sun, Evgeniy Gabrilovich

What Are Information Networks?

- A network where each **node** represents an **entity** (e.g., user in a social network) and each **link** (e.g., friendship) a **relationship** between entities.
 - Nodes/links may have attributes, labels, and weights.
 - Links may carry rich semantic information.

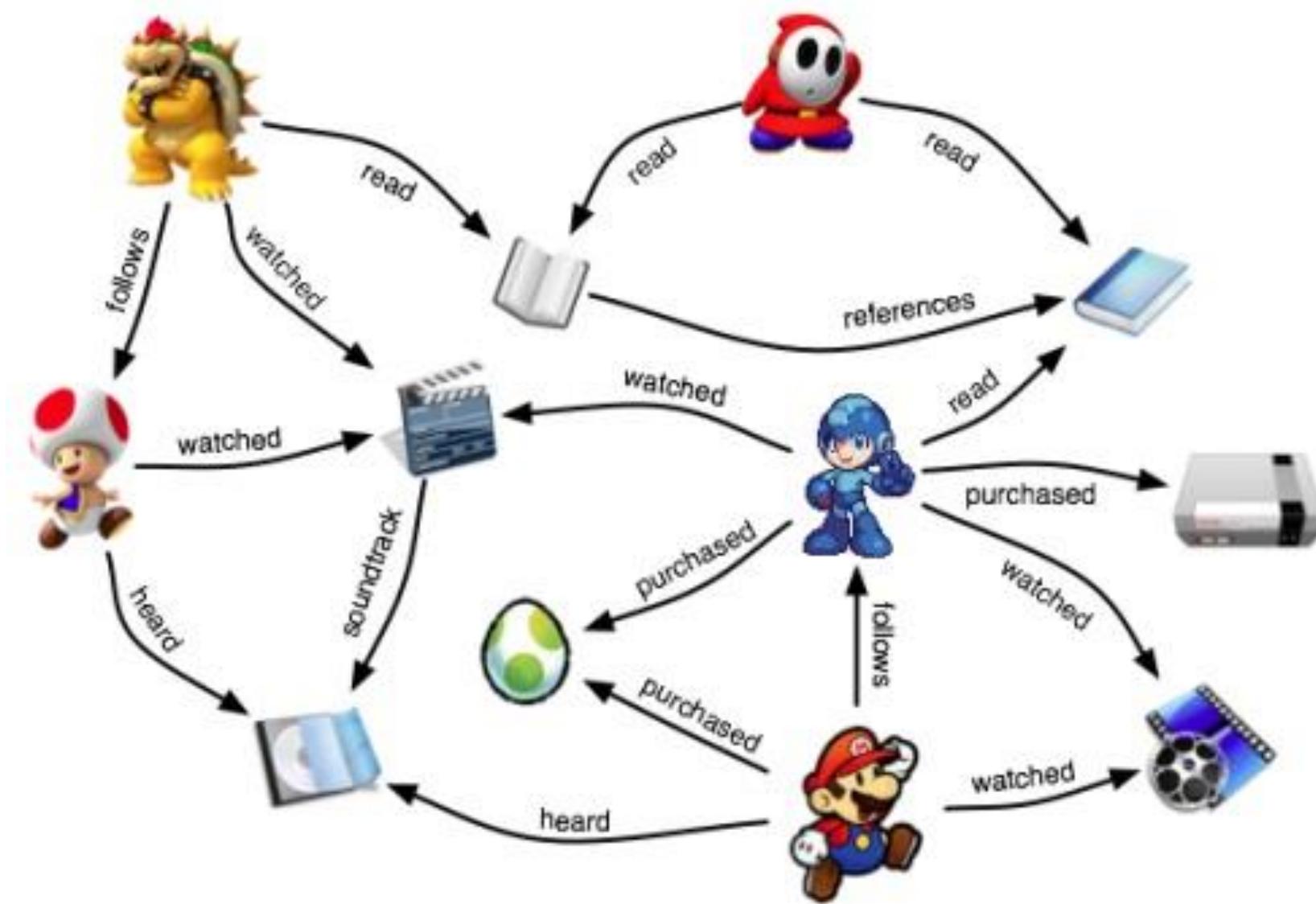


Heterogeneous Information Networks

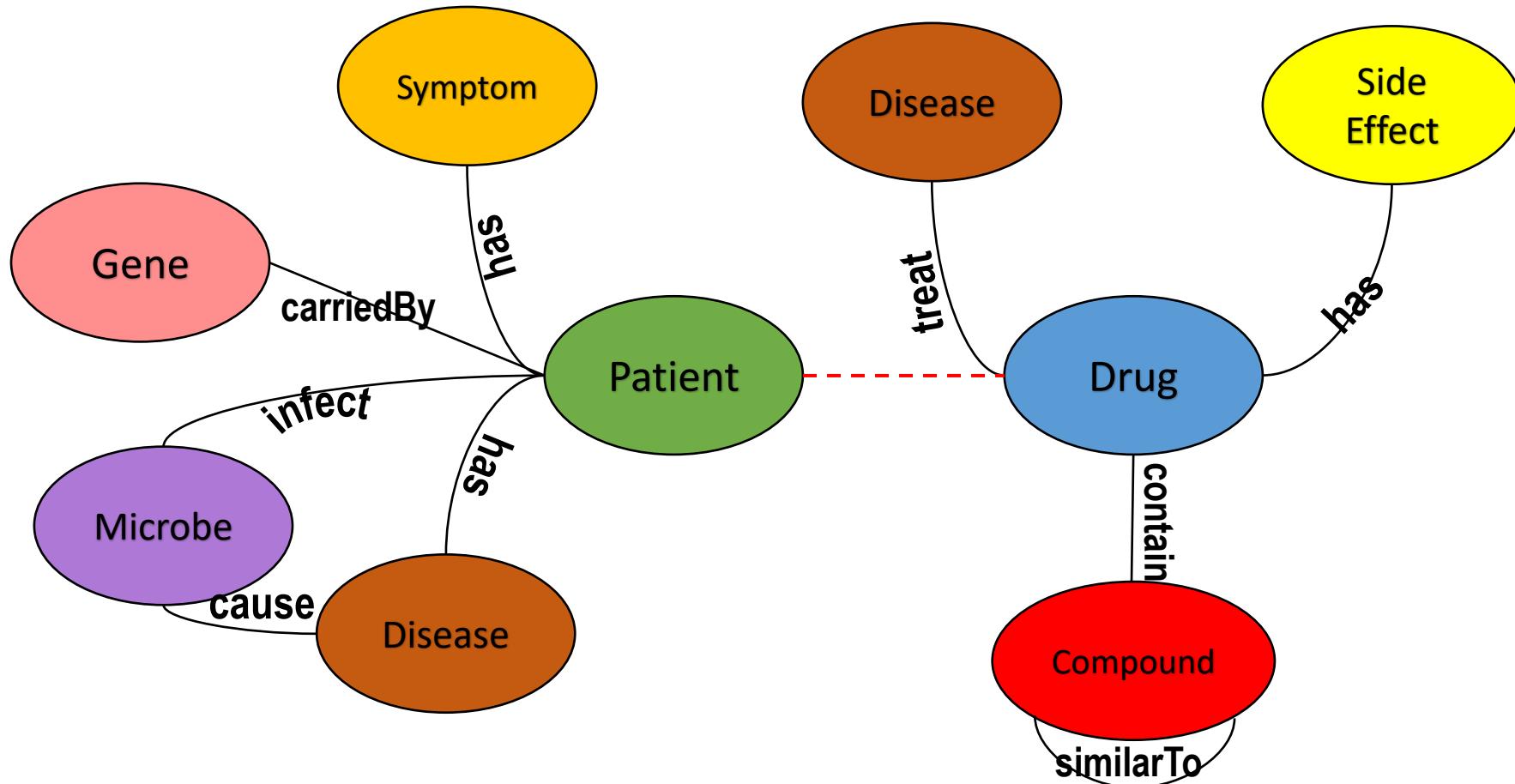


1. Multiple entity types and link types
2. *New problems* are emerging in heterogeneous networks!

We are living in a connected world!



Even in Biomedical Domain



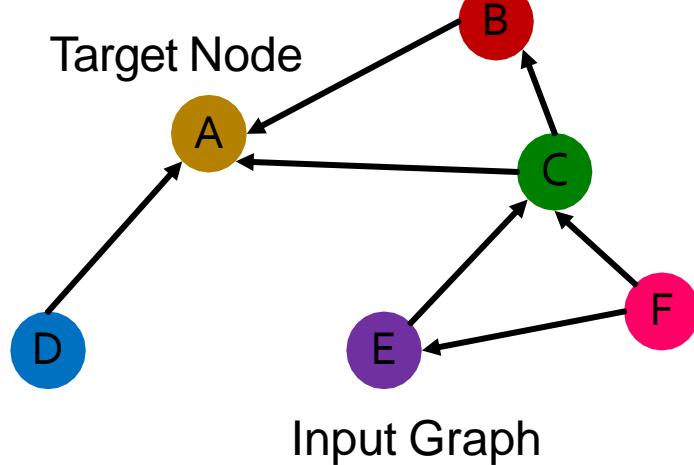
Heterogeneous Graphs

○ A heterogeneous graph is defined as

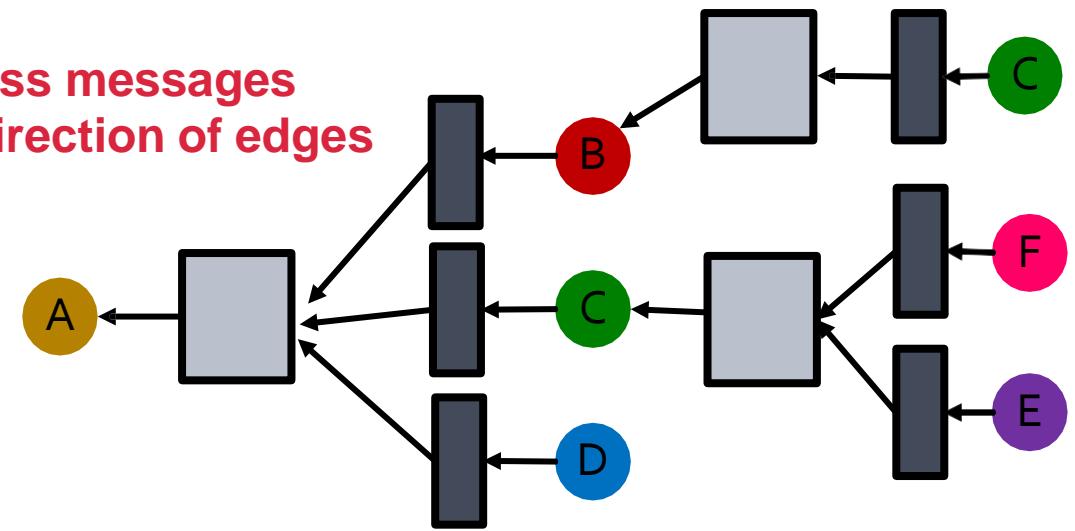
- $G = (V, E, R, T)$
- Nodes with node types $v_i \in V$
- Edges with relation types $(v_i, r, v_j) \in E$
- Node type $T(v_i)$
- Relation type $r \in R$

Relational GCN

- We will extend **GCN** to handle heterogeneous graphs with multiple edge/relation types
- We start with a directed graph with **one** relation
 - How do we run GCN and update the representation of the **target node A** on this graph?

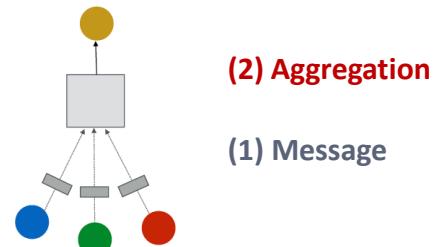


Only pass messages along direction of edges



Recap: Message Passing

- **A Single GNN Layer**
 - **(1) Message:** each node computes a message $\mathbf{m}_u^{(l)} = \text{MSG}^{(l)}\left(\mathbf{h}_u^{(l-1)}\right), u \in N(v) \cup v$
 - **(2) Aggregation:** aggregate messages from neighbors $\mathbf{h}_v^{(l)} = \text{AGG}^{(l)}\left(\left\{\mathbf{m}_u^{(l)}, u \in N(v)\right\}, \mathbf{m}_v^{(l)}\right)$
- **Nonlinearity (activation):** Adds expressiveness
 - Often written as $\sigma(\cdot)$: ReLU(\cdot), Sigmoid(\cdot) , ...
 - Can be added to **message or aggregation**



Recap: GCN

- **Graph Convolutional Networks (GCN)**

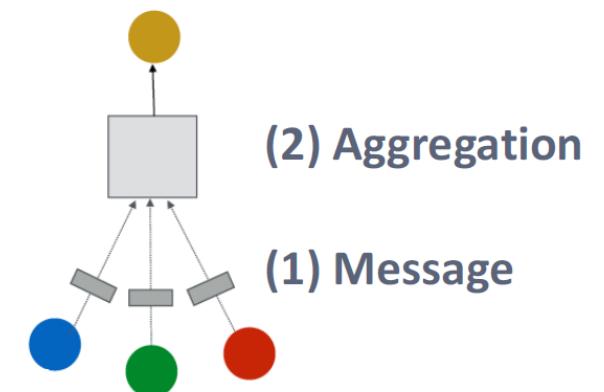
$$\mathbf{h}_v^{(l)} = \sigma \left(\mathbf{W}^{(l)} \sum_{u \in N(v)} \frac{\mathbf{h}_u^{(l-1)}}{|N(v)|} \right)$$

- How to write this as Message + Aggregation?

Message

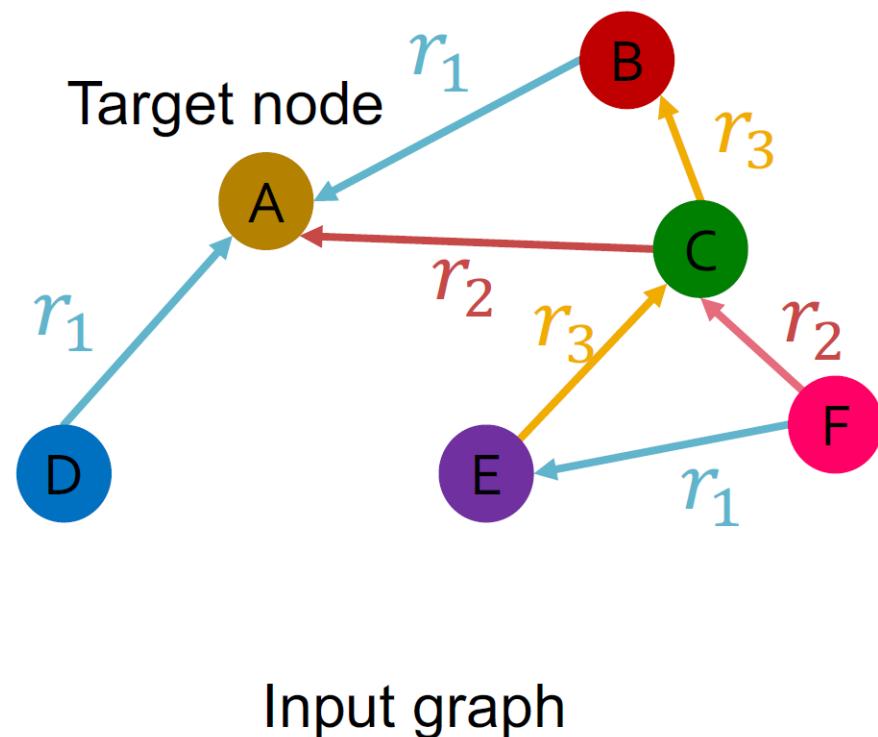
$$\mathbf{h}_v^{(l)} = \sigma \left(\sum_{u \in N(v)} \mathbf{W}^{(l)} \frac{\mathbf{h}_u^{(l-1)}}{|N(v)|} \right)$$

Aggregation



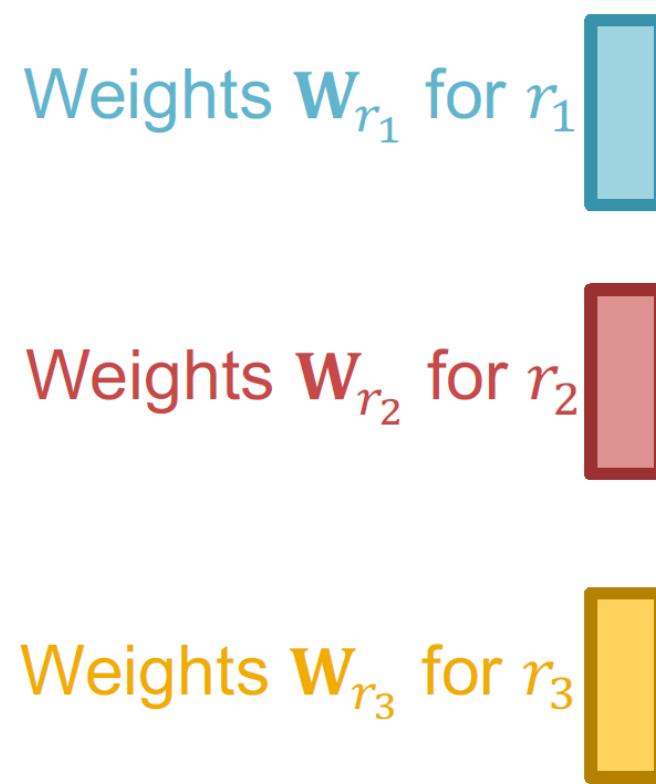
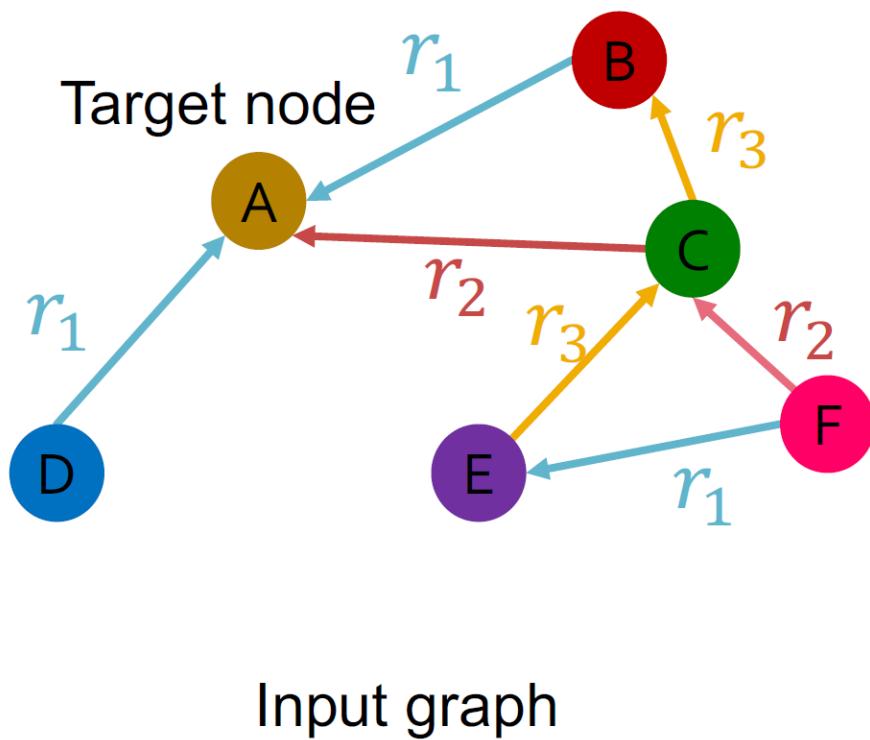
Relational GCN

- What if the graph has **multiple relation types?**



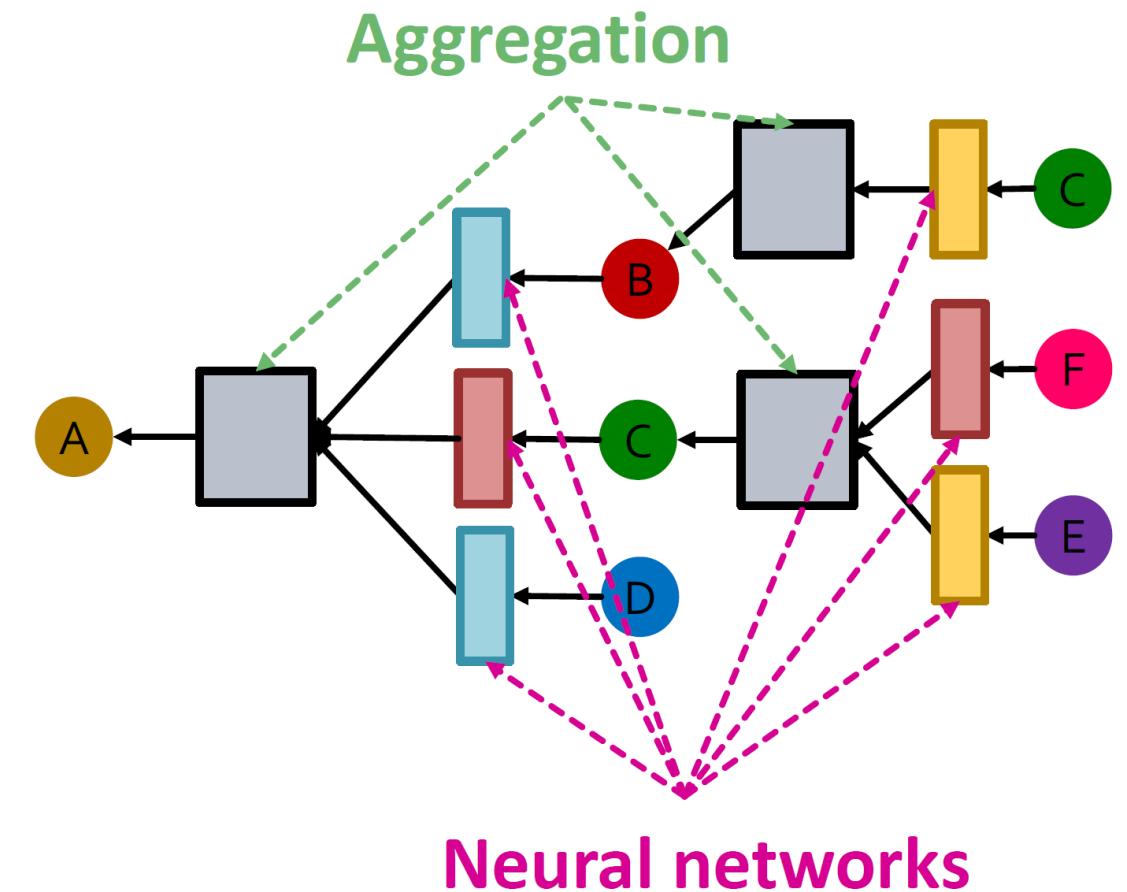
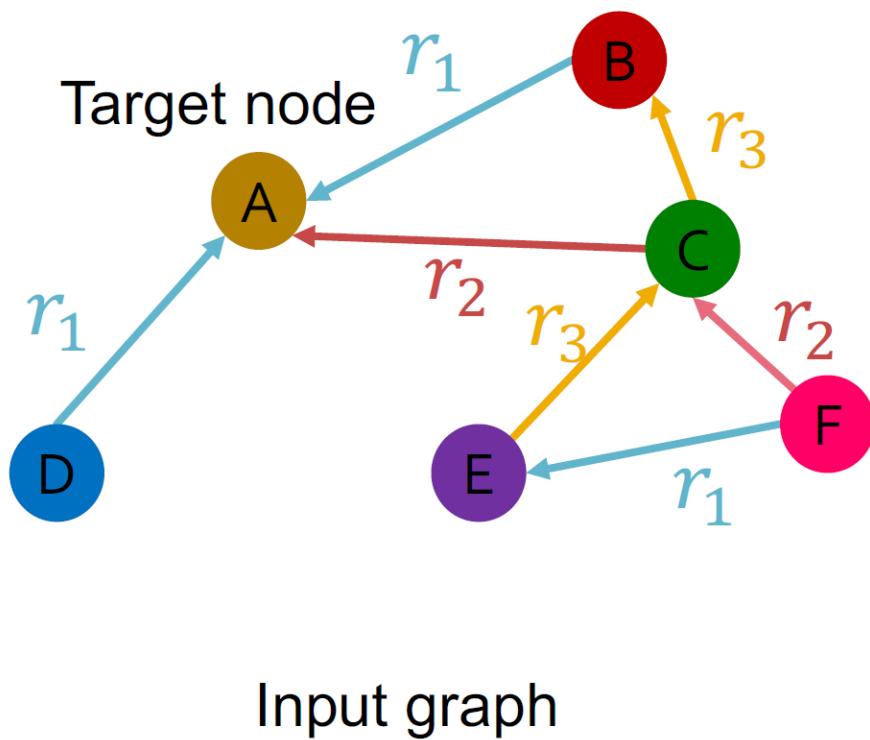
Relational GCN

- Use different neural network weights for different relation types.



Relational GCN

- Use different neural network weights for different relation types.



Relational GCN

- **Relational GCN (RGCN):** $\mathbf{h}_v^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{u \in N_v^r} \frac{1}{c_{v,r}} \mathbf{W}_r^{(l)} \mathbf{h}_u^{(l)} + \mathbf{W}_0^{(l)} \mathbf{h}_v^{(l)} \right)$

• How to write this as Message + Aggregation?

- **Message:**

- Each neighbor of a given relation: $\mathbf{m}_{u,r}^{(l)} = \frac{1}{c_{v,r}} \mathbf{W}_r^{(l)} \mathbf{h}_u^{(l)}$

- Self-loop: $\mathbf{m}_v^{(l)} = \mathbf{W}_0^{(l)} \mathbf{h}_v^{(l)}$

- **Aggregation:**

- Sum over messages from neighbors and self-loop, then apply activation

$$\mathbf{h}_v^{(l+1)} = \sigma \left(\text{Sum} \left(\left\{ \mathbf{m}_{u,r}^{(l)}, u \in N(v) \right\} \cup \left\{ \mathbf{m}_v^{(l)} \right\} \right) \right)$$

RGCN: Scalability

- Each relation has L matrices $\mathbf{W}_r^{(1)}, \mathbf{W}_r^{(2)} \dots \mathbf{W}_r^{(L)}$
 - The size of each $\mathbf{W}_r^{(l)}$ is $d^{(l+1)} \times d^{(l)}$
 - $d^{(l)}$ is the hidden dimension in layer l
- **Rapid # parameters growth w.r.t # relations!**
- **Overfitting becomes an issue**
 - **Two methods to regularize the weights $\mathbf{W}^{(l)}$**
 - **(1)** Use block diagonal matrices
 - **(2)** Basis/Dictionary learning

(1) Block Diagonal Matrices

- Key insight: make the weights sparse!
 - Use block diagonal matrices for $\mathbf{w}_r^{(l)}$

$$W_r =$$


Limitation: only nearby neurons/dimensions can interact through W

- If use B low-dimensional matrices, then # param reduces from $d^{(l+1)} \times d^{(l)}$ to $B \times \frac{d^{(l+1)}}{B} \times \frac{d^{(l)}}{B}$

(2) Basis Learning

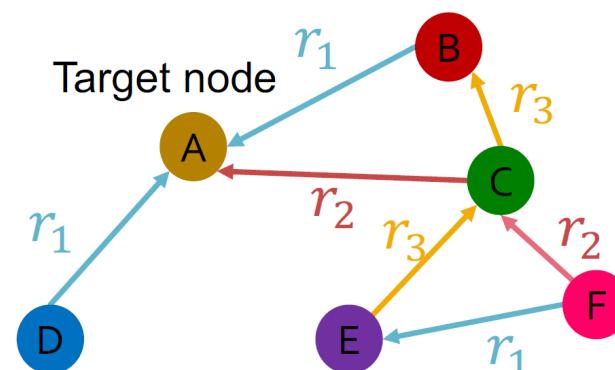
- Key insight: **Share weights** across different relations!
- Represent the matrix of each relation as a **linear combination** of **basis transformations**

$$\mathbf{W}_r = \sum_{b=1}^B a_{rb} \cdot \mathbf{V}_b$$

- \mathbf{V}_b is shared across relations (the basis matrices)
- a_{rb} is the importance weight of matrix \mathbf{V}_b
- Now each relation only needs to learn $\{a_{rb}\}_{b=1}^B$, which is B scalars.

Example: Entity/Node Classification

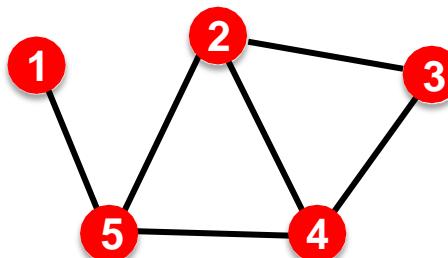
- **Goal:** Predict the label of a given node
- **RGCN** uses the representation of the final layer:
 - If we predict the class of **node A** from **k classes**.
 - Take the **final layer (prediction head)**: $\mathbf{h}_A^{(L)} \in \mathbb{R}^k$, each item in $\mathbf{h}_A^{(L)}$ represents **the probability of that class**.



Input graph

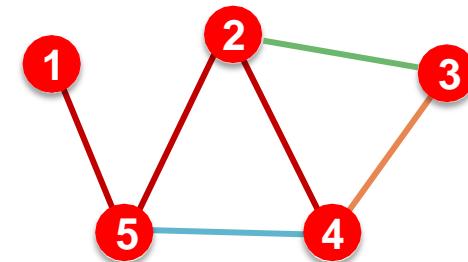
Example: Link Prediction

- Link prediction split:



The original graph

Split



Split Graph with 4 categories of edges

Training message edges for r_1

Training supervision edges for r_1

Validation edges for r_1

Test edges for r_1

.....

Training message edges for r_n

Training supervision edges for r_n

Validation edges for r_n

Test edges for r_n

Training message edges

Training supervision edges

Validation edges

Test edges

Every edge also has a relation type, this is independent of the 4 categories.

In a heterogeneous graph, the homogeneous graphs formed by every single relation also have the 4 splits.

Summary of RGCN

- **Relational GCN**, a graph neural network for **heterogeneous graphs**
- Can perform entity classification as well as link prediction tasks.
- Ideas can easily be extended into RGNN (RGraphSAGE, RGAT, etc.)

Knowledge Graphs: Completion with Embeddings

Google Knowledge Graph

- Launched in May 2012
- 570 million entities and 18 billions of relationships

The screenshot shows a Google search results page for the query "albert einstein". The search bar at the top contains the query. Below it, the search results are displayed under the heading "Search" with a result count of "73,300,000 other results (0.36 seconds)". On the left, a sidebar lists search categories: Everything (selected), Images, Maps, Videos, News, Shopping, Books, and More. A red arrow points from the sidebar towards the right-hand knowledge graph panel. The main search results include links to the Albert Einstein Official Site, Wikipedia, Biography.com, and BrainyQuote. To the right of the search results is a detailed knowledge graph panel for Albert Einstein. It features a portrait photo, his birth date (March 14, 1879, Ulm), death date (April 18, 1955, Princeton), and children (Hans Albert Einstein, Eduard Einstein, Lieserl Einstein). It also lists his education (University of Zurich, ETH Zurich) and spouse (Elsa Einstein). Below this, there are sections for books (including "Albert Einstein: Ideas and Opinions" and "Albert Einstein: Out of my later years"), and a "People also search for" section with links to Isaac Newton, Thomas Edison, Stephen Hawking, Benjamin Franklin, and Galileo Galilei. At the bottom right of the panel is a "Report a problem" link.

Surfacing structured results in web search

Google pisa italy

All Maps Images News Videos More Search tools

About 39,200,000 results (0.53 seconds)

Pisa - Wikipedia, the free encyclopedia
<https://en.wikipedia.org/wiki/Pisa> Wikipedia
Pisa is a city in Tuscany, Central Italy, straddling the River Arno just before it empties into the Tyrrhenian Sea. It is the capital city of the Province of Pisa.
Leaning Tower of Pisa · Piazza dei Miracoli · Baptistry · San Francesco (Pisa)

Images for pisa italy Report images



More images for pisa italy

Virtual tour of Pisa Italy - History, facts, top attractions & things to do ...
www.italyguides.it/en/tuscany/pisa ▾
★★★★★ Rating: 4 - 5,520 votes
Travel guide of Pisa Italy. Maps, articles, photos and destination guides about Pisa major attractions.
Leaning Tower of Pisa · Piazza dei Miracoli · Interactive map of Pisa

Pisa, Italy: Tourist Guide to Visiting the Leaning Tower of Pisa and ...
<https://www.discovertuscany.com/pisa/> ▾ Discover Tuscany
Pisa: City of the Leaning Tower and More. ... Buy Tickets for the Leaning Tower. Visit the Leaning Tower in Pisa: do not miss the chance to discover the secrets of one of the most famous monuments in the world.

Pisa 2016: Best of Pisa, Italy Tourism - TripAdvisor
https://www.tripadvisor.com/Tourism-g187899-Pisa_Province_of_Pisa_T... ▾ TripAdvisor
Pisa Tourism: TripAdvisor has 144925 reviews of Pisa Hotels, Attractions, and Restaurants making it your best Pisa resource.
Things to do in Pisa · Pisa Hotels · Restaurants · Pisa Hotel Deals

Travel Guide to Pisa in Tuscany, Italy - Italy Travel - About.com
goitaly.about.com/od/pisa/p/pisa.htm ▾
Pisa, Italy: Pisa is best known for its leaning tower but there is much more to see in this Tuscan town. The area around the cathedral and tower, Piazza dei ...

Pisa - Lonely Planet
<https://www.lonelyplanet.com/italy/tuscany/pisa> ▾ Lonely Planet
Once a maritime power to rival Genoa and Venice, Pisa now draws its fame ... economy since the 1400s, and students from across Italy compete for places in its ...

SIGIR 2016 | July 17-21 2016 – Pisa, Tuscany, Italy
sigir2016/ ▾ Special Interest Group on Information Retrieval
We welcome the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, exactly 30 years since SIGIR 1986, also ...
You've visited this page many times. Last visit: 6/21/16


Pisa
City in Italy

Pisa is a city in central Italy's Tuscany region best known for its iconic Leaning Tower. Already tilting when it was completed in 1372, the 56m white-marble cylinder is the bell tower of the Romanesque, striped-marble cathedral that rises next to it in the Piazza del Miracoli, a grassy, walled square.

Hotels: 3-star averaging \$90, 5-star averaging \$140. View hotels
Getting there: 14 h 10 min flight, around \$2,560. View flights
Weather: 75°F (24°C), Wind NW at 3 mph (5 km/h), 83% Humidity
Local time: Thursday 10:09 PM
Province: Province of Pisa

Points of interest View 15+ more



Leaning Tower of Pisa Piazza dei Miracoli Camposanto Monumentale Santa Maria della Spina Knights' Square

Colleges and Universities



University of Pisa Sant'Anna School of Advanced Studies Scuola Normale Superiore...

More about Pisa Feedback

Augmenting the presentation with relevant facts

Surfacing facts proactively

The image shows two Google search results side-by-side. The left result is for 'italy population' and the right is for 'italy'. A yellow arrow points from the 'italy population' search to the 'italy' search, indicating how one query can lead to another.

Search 1: Italy Population

Google search bar: italy population

Results:

- Italy / Population**
59.83 million
80M
60M
40M
20M
0
1960 1970
- Explore more
- Sources include: World Bank
- Italy Population (2014)**
www.worldometers.info/
Italy Population (LIVE) Th
based on the latest United
population. Italy ranks nur

Search 2: Italy

Google search bar: italy

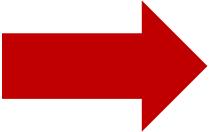
Results:

- Italian Tourism Official Website**
www.italia.it/en/home.html ▾
Italian tourism official website: vacation, art and culture, history, events, nature, lakes, mountains, golf, sci, boating, thermal spas, sports and adventure.
- Italy - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/Italy ▾
Wikipedia
Location of Italy (dark green). – in Europe – in the European Union (light green) – [Legend]. Capital and largest city, Rome - 41°54'N 12°29'E / 41.900°N ...
Sergio Mattarella · History of Italy · Italian language · Demographics of Italy
- In the news**
Chris Wood: "It Will Take A Political Genius To Hold The EU Together", Italy Is The Flash Point
Zero Hedge - 1 hour ago
This Italian issue was discussed in more detail here a few months ago (see GREED & fear ...)
- Germany Vs Italy: Heavyweights Collide at Euro 2016
ABC News - 10 hours ago
- Thomas Muller confident Germany can breach Italy wall in Euro 2016 quarter-final
Daily Mail - 9 hours ago
- More news for italy
- Images for italy**
Report images

Right Panel: Italy Summary

- Flag of Italy (green, red, and blue vertical stripes)
- Map of Italy in Europe
- Italy**
Country in Europe
- Text: Italy, commanding a long Mediterranean coastline, has left a powerful mark on Western culture and cuisine. Its capital, Rome, is home to the Vatican as well as landmark art and ancient ruins. Other major cities include Florence, with Renaissance treasures such as Michelangelo's "David" and its leather and paper artisans; Venice, the sinking city of canals; and Milan, Italy's fashion capital.
- Capital:** Rome
- Currency:** Euro
- Population:** 59.83 million (2013) World Bank
- President:** Sergio Mattarella
- Prime minister:** Matteo Renzi
- Points of interest**
View 45+ more
- Colosseum (Roma Capital)
- Lake Garda
- Leaning Tower of Pisa
- Amalfi Coast
- Lake Como

Exploratory search



Google pisa sightseeing

All Maps Images News Videos More Search tools

Pisa / Points of interest

Leaning Tower of Pisa Ornate 14th-century tower with a tilt		Knights' Square Renaissance square with Medici statue		Palazzo della Carovana Ornate art-filled palace & university	
Piazza dei Miracoli Green space that's home to leaning tower		Pisa Baptistry Celebrated acoustics & Gothic sculptures		San Paolo a Ripa d'Arno Romanesque church from the 11th century	
Santa Maria della Spina Petite church with gothic exterior		Camposanto Monumentale Cloistered cemetery with unique frescoes		San Michele in Borgo Vintage church with medieval artworks	

The Top 10 Things to Do in Pisa - TripAdvisor - Pisa, Italy Attractions ...
https://www.tripadvisor.com/Attractions-g187899-Activities-Pisa_Provinc... ▾ TripAdvisor ▾
Book your tickets online for the top things to do in Pisa, Italy on TripAdvisor. See 27414 traveler reviews and photos of Pisa tourist attractions. Find what to do ...
Piazza dei Miracoli · Torre di Pisa · Duomo Pisa · Gelateria Gentile

One Day in Pisa: What to See and Do | Indiana Jo
indianajo.com/2013/07/one-day-in-pisa.html ▾
Jul 11, 2013 - UPDATED SEPTEMBER 2015: I returned to Pisa in 2015 and I've updated this post to include a few more sights, a local market I stumbled ...

Sights in Pisa, Italy - Lonely Planet
www.lonelyplanet.com/.../Italy/Tuscany/Northwestern-Tuscany/Lonely-Planet/17-sights-in-Pisa, including Leaning Tower, Duomo, and Battistero.

10 things you must see and do during your stay in Pisa - 10things.it
www.10things.it/guide/pisa/top-10/ ▾
10 things you must see and do during your stay in Pisa. The night in Pisa offers interesting attractions for tourists, don't forget that Pisa is a young city thanks to ...

Images for pisa sightseeing Report images

More images for pisa sightseeing

Pisa one day itinerary:miracles square,leaning tower and all Pisa main ...
<https://www.discovertuscany.com/pisa/one-day-itinerary.html> ▾ Discover Tuscany ▾
A short guide to the discovery of Pisa and its main artistic attractions in just one day.


Pisa
City in Italy

Pisa is a city in central Italy's Tuscany region best known for its iconic Leaning Tower. Already tilting when it was completed in 1372, the 56m white-marble cylinder is the bell tower of the Romanesque, striped-marble cathedral that rises next to it in the Piazza dei Miracoli, a grassy, walled square.

Hotels: 3-star averaging \$90, 5-star averaging \$140. [View hotels](#)
Getting there: 14 h 10 min flight, around \$2,560. [View flights](#)
Weather: 75°F (24°C), Wind NW at 3 mph (5 km/h), 83% Humidity
Local time: Thursday 10:07 PM
Province: Province of Pisa

Colleges and Universities

 **UNIVERSITÀ DI PISA**
University of Pisa

 **SANT'ANNA SCHOOL OF ADVANCED STUDIES**
Sant'Anna School of Advance...

 **SCUOLA NORMALE SUPERIORE**
Scuola Normale Superiore...

Connecting people, places, and things

The diagram illustrates the interconnected nature of social media data, specifically focusing on the Harvard University Facebook page and its connections to other users and locations.

Harvard University (Top Center):
A Facebook page for Harvard University, featuring a red shield logo with "VERITAS" and a building in the background. It has 2,703,624 likes, 47,280 people talking about it, and 430,473 people were here.

People who like Harvard University (Bottom Left):
A Facebook page showing profiles of users who like Harvard University. Examples include Paul McDonald (Engineer at Facebook), Ekaterina Skorobogatova (Works at Facebook), Gary Johnson (Corporate Development at Facebook), and Greg Marra (Product Manager at Facebook).

People who visited Harvard University (Bottom Right):
A Facebook page showing profiles of users who have visited Harvard University. Examples include Florence Trouche (Global Client Partner at Facebook), Andrew Tulloch (Machine Learning at Facebook), Joseph Barillari (Software Engineer at Facebook), and Sheryl Sandberg (Chief Operating Officer at Facebook).

Third Page (Partially Visible):
A fourth Facebook page is partially visible on the right, showing a profile picture of a person and some text.

Connecting people, places and things

Structured search within the graph

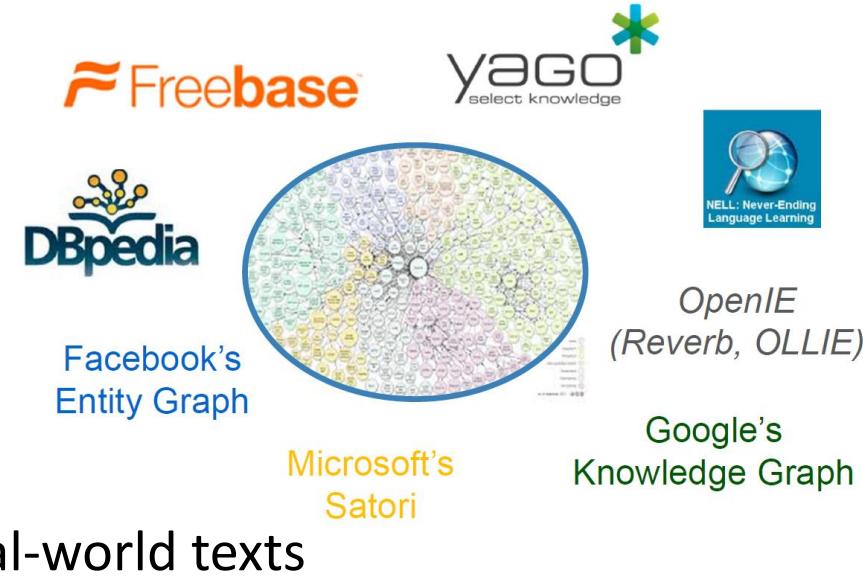
The screenshot shows a Facebook search results page with the query "People who like Harvard University and Basketball and work at Facebook". The results list five profiles:

- Mike Vernal** (VP Engineering at Facebook): Likes Harvard University, Harvard Crimson and F@ceb00k Su... Studied Computer Science at Harvard University '02
10 mutual friends including Keith Adams and Philip Bohannon
- Jared Morgenstern** (Product Manager / Ninja - Games ...): Likes Harvard University, F@ceb00k Summer Basketball Leag... Studied Computer Science at Harvard University
5 mutual friends including Clodagh Chloe Takeuchi and Pierre ...
- Florin Ratiu** (Software Engineer at Facebook): Likes Harvard School of Public Health, Stanford 6th Man and ba... Studied Management Science and Engineering at Stanford Univ...
3 mutual friends including Alexey Spiridonov and Serkan Plantino
- Ning Zhang (张宁)** (Software Engineer at Facebook): Likes Harvard University, Basketball and 314 others
Studied Computer Science at University of Waterloo '06
5 mutual friends including Tudor Bosman and Ves Stoyanov
- Zhongyuan Xu (徐重远)** (Software Engineer at Facebook): Likes Harvard University, Basketball and 364 others
Studied at Stony Brook University
1 mutual friend: Ledell Wu

The page took 2147 ms to load.

Knowledge Graph Summary

- WordNet
 - Expert annotation
 - Linguistic
- Cyc
 - Expert annotation
 - Formalized logical reasoning
- ConceptNet (MIT)
 - Crowdsourcing
 - Context-based inferences over real-world texts
- KnowItAll (UW), NELL (CMU)
 - Information Extraction
 - Triplets (entities and relations)
- Google Knowledge Graph
 - Combination
 - Triplets (entities and relations)
- Amazon Product Graph
 - More products and attributes



Knowledge Graph Datasets

- **Publicly available KGs:**
 - FreeBase, Wikidata, DBpedia, YAGO, NELL, etc.
- **Common characteristics:**
 - **Massive**: millions of nodes and edges
 - **Incomplete**: many true edges are missing

Given a massive KG,
enumerating all the
possible facts is
intractable!



Can we predict plausible
BUT missing links?

Example: Freebase

○ Freebase

- ~50 million **entities**
- ~38K **relation types**
- ~3 billion **facts/triples**



93.8% of persons from Freebase have no place of birth and 78.5% have no nationality!

○ Datasets: FB15k/FB15k-237

- A **complete** subset of Freebase, used by researchers to learn KG models

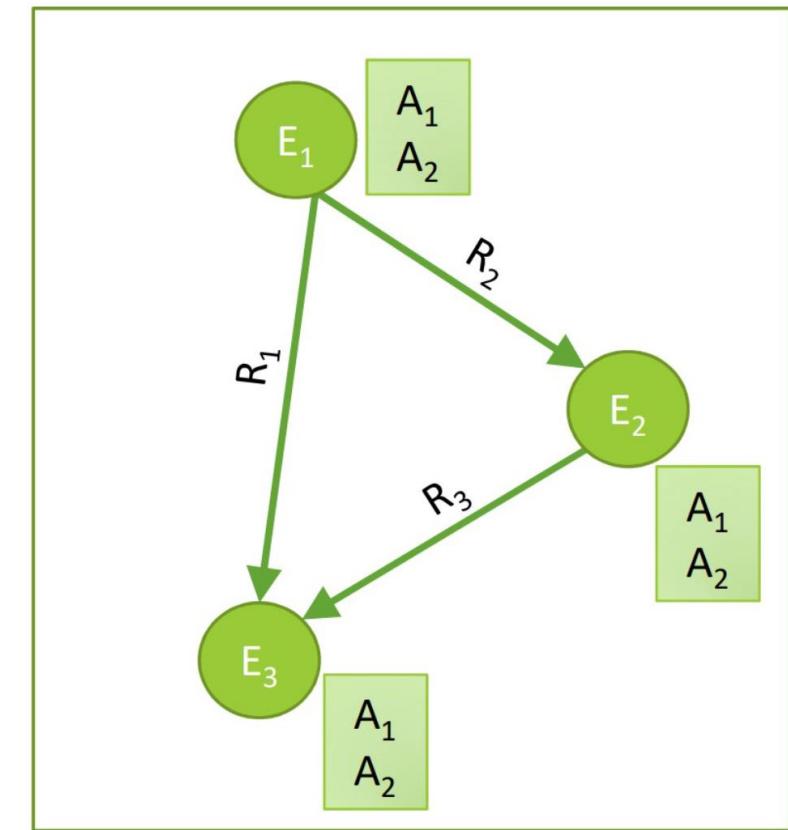
Dataset	Entities	Relations	Total Edges
FB15k	14,951	1,345	592,213
FB15k-237	14,505	237	310,079

¹ Paulheim, Heiko. "Knowledge graph refinement: A survey of approaches and evaluation methods." *Semantic web* 8.3 (2017): 489-508.

²Min, Bonan, et al. "Distant supervision for relation extraction with an incomplete knowledge base." *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2013.

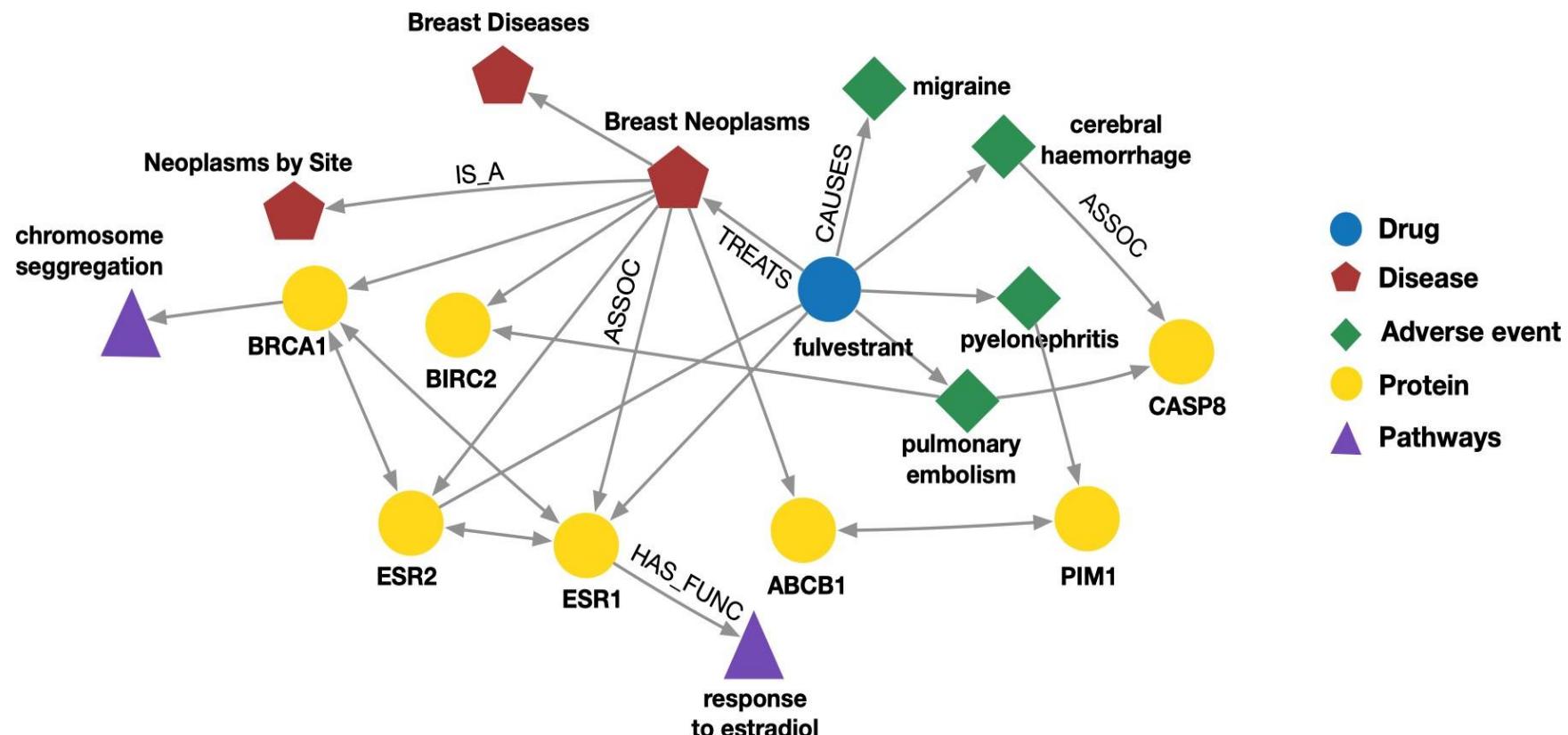
Knowledge Graph Representation

- Knowledge in graph form:
 - Capture entities, types, and relationships
- Nodes are **entities**
- Nodes are labeled with their **types**
- Edges between two nodes capture **relationships** between entities
- Some entities also have **attributes**
- **KG is an example of a heterogeneous graph**



Bio Knowledge Graphs

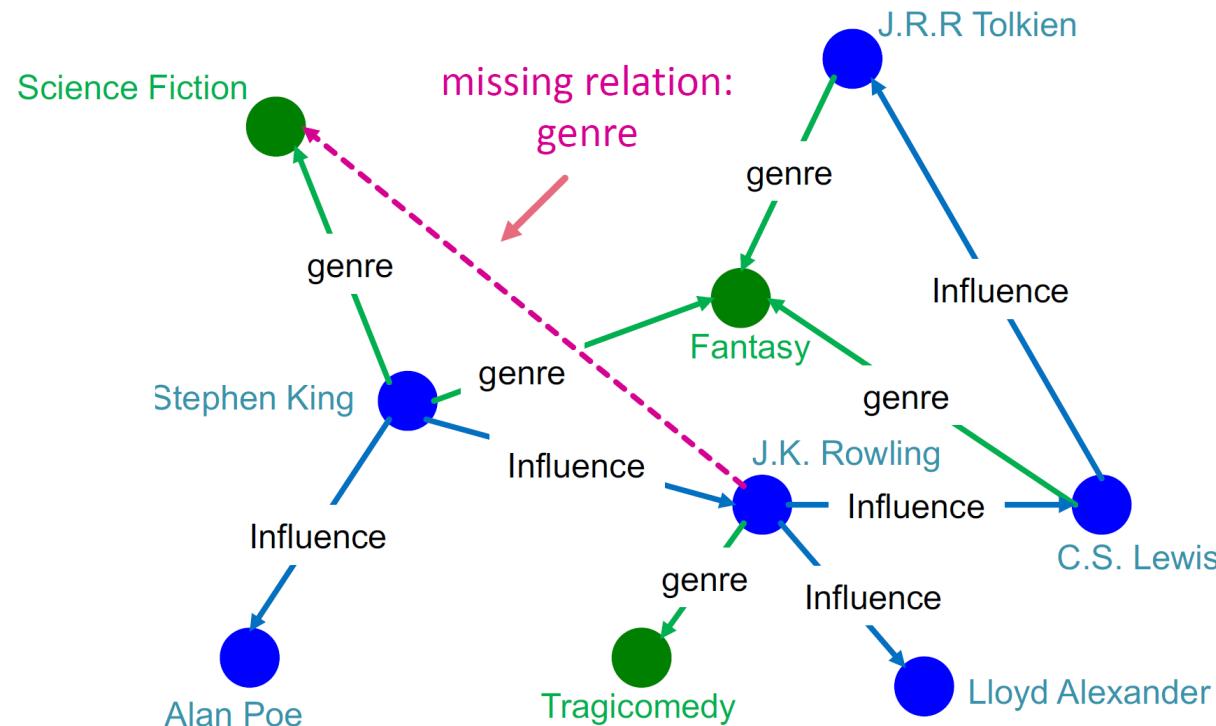
- **Node types:** drug, disease, adverse event, protein, pathways
- **Relation types:** has_func, causes, assoc, treats, is_a



KG Completion Task

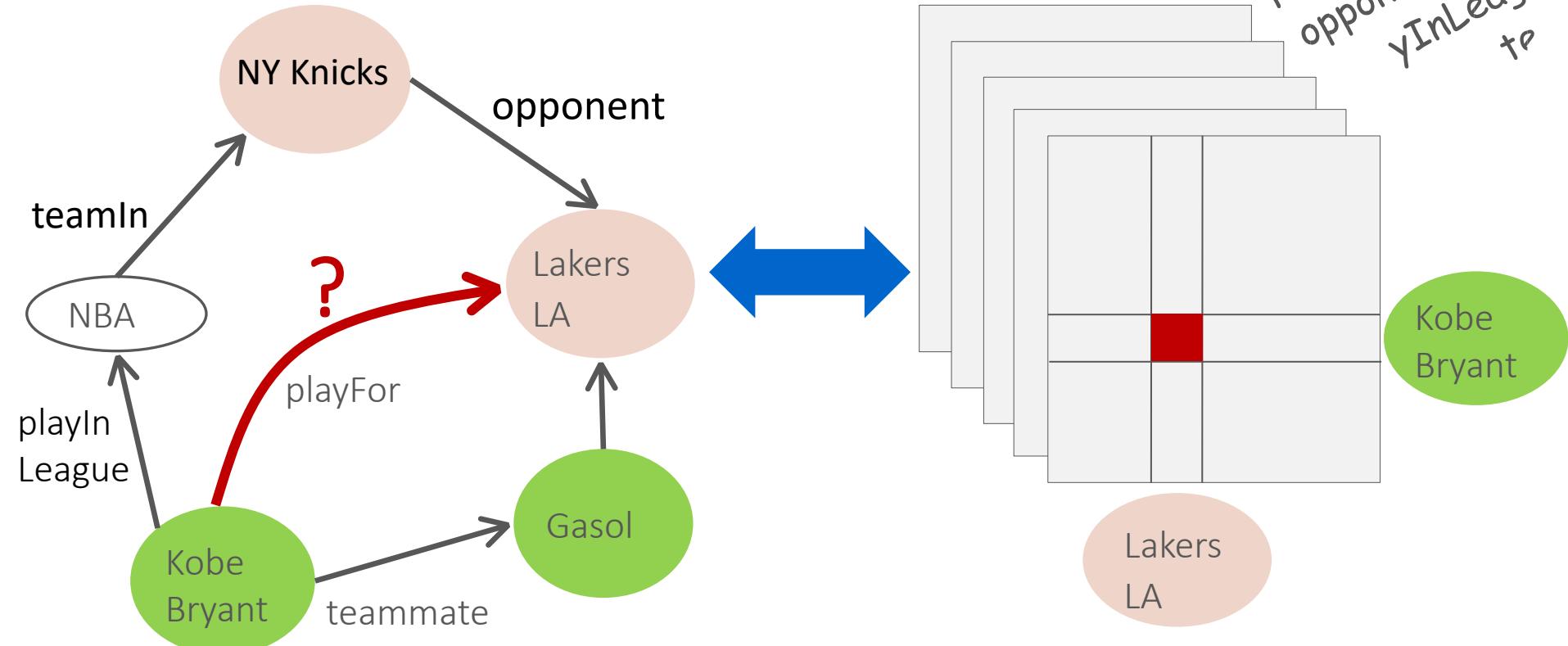
- Given an enormous KG, can we complete the KG?
- For a given (**head**, **relation**), we predict missing **tails**.
 - (Note this is slightly different from link prediction task)

Example task: predict the tail “Science Fiction” for (“J.K. Rowling”, “genre”)

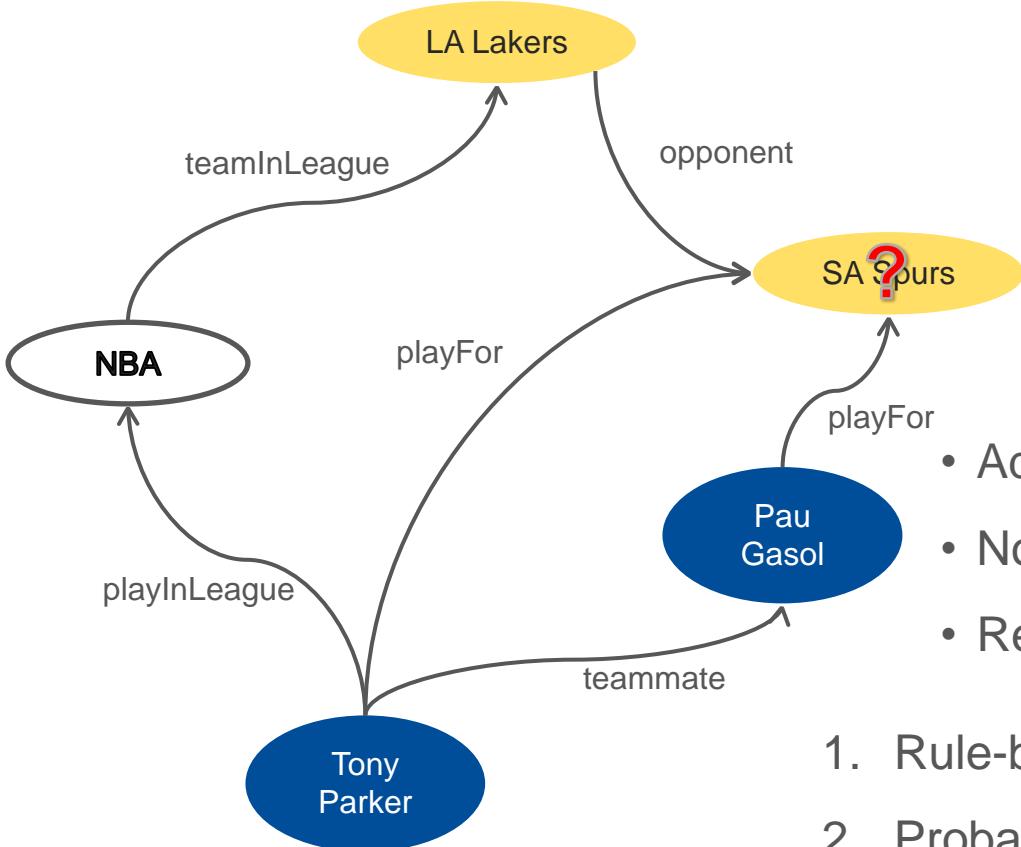


Knowledge Graph Completion as Link Prediction

Compute $\Pr(x_{ijk} = 1 \mid \text{Graph}) :$



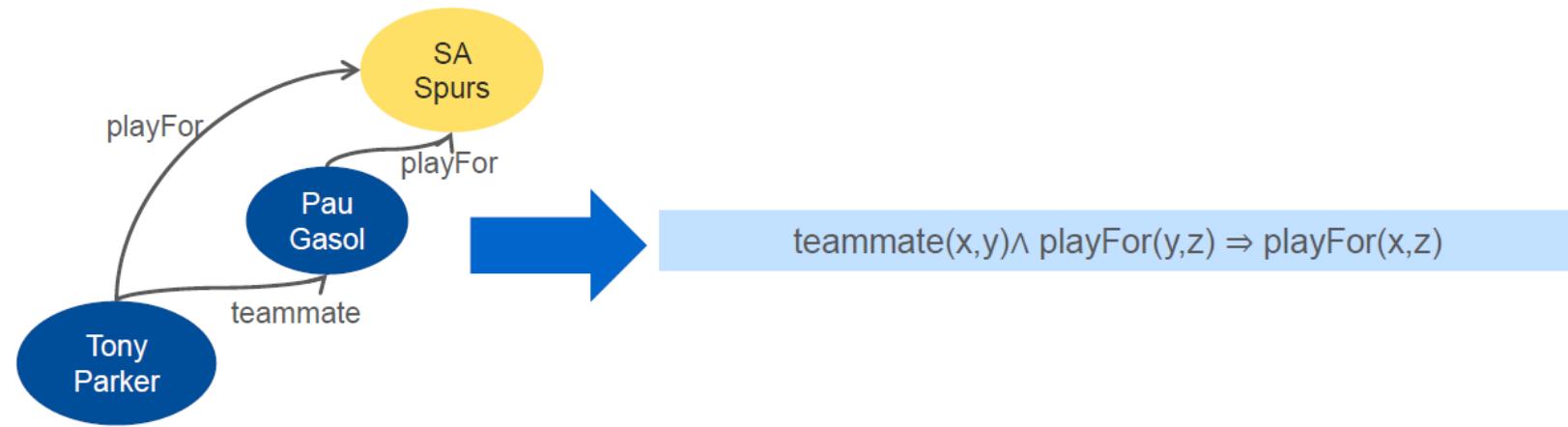
Link Prediction



- 1. Rule-based methods
 - 2. Probabilistic models
 - 3. Factorization models
 - 4. Embedding models
- Add knowledge from existing graph
 - No external source
 - Reasoning within the graph

First Order Inductive Learner

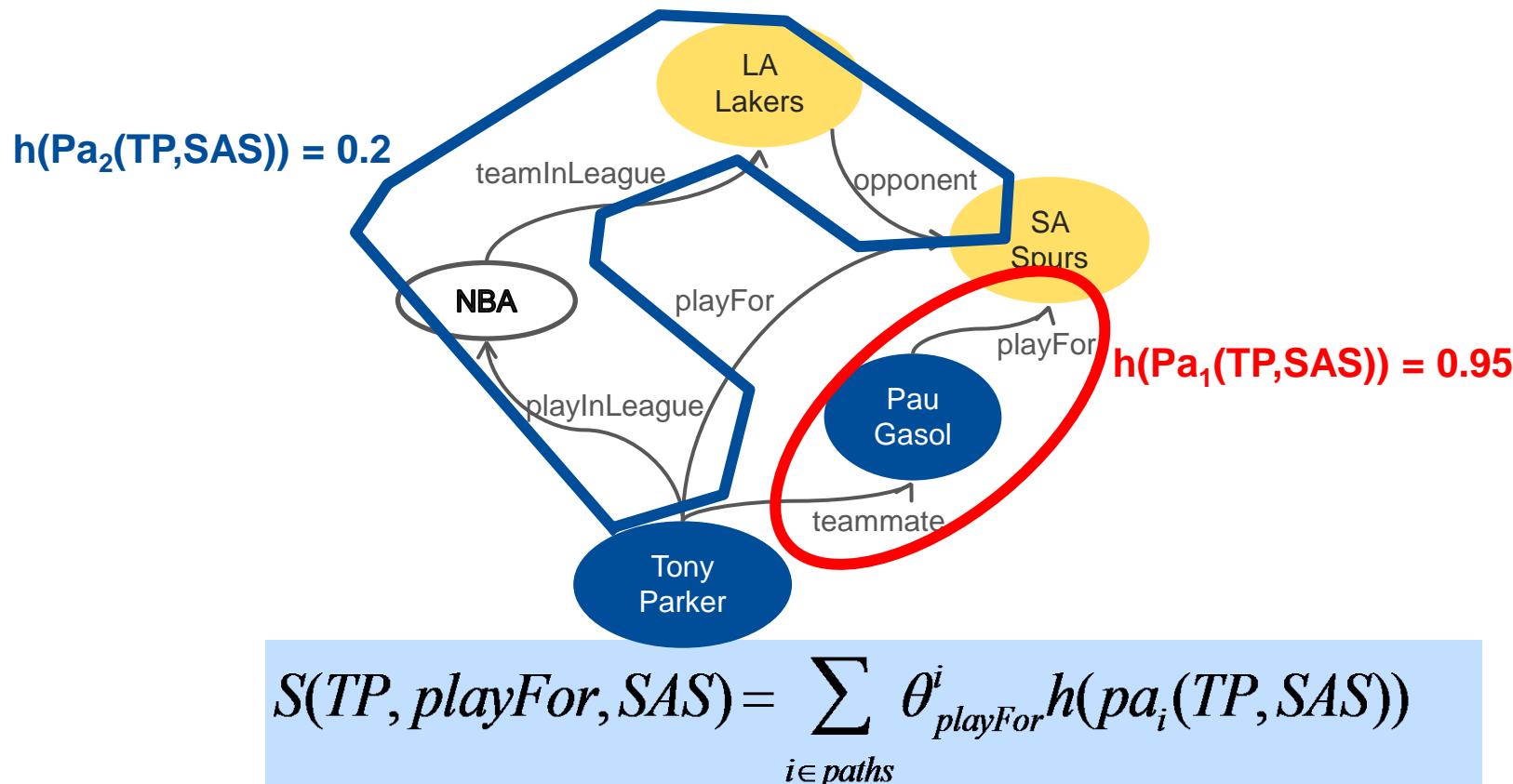
- FOIL learns function-free Horn clauses:
 - given positive negative examples of a concept
 - a set of background-knowledge predicates
 - FOIL inductively generates a logical rule for the concept that cover all + and no -



- Computationally expensive: huge search space large, costly Horn clauses
- Must add constraints → high precision but low recall
- Inductive Logic Programming: deterministic and potentially problematic

Path Ranking Algorithm [Lao et al., 11]

- Random walks on the graph are used to sample paths
- Paths are weighted with probability of reaching target from source
- Paths are used as ranking experts in a scoring function



Link Prediction with Scoring Functions

- A scoring function alone does not grant a decision $S(TP, playFor, SAS) > \theta$
- **Thresholding:** determine a threshold ϑ
 $(TP, playFor, SAS)$ is *True* iff

- **Ranking**
 - The most likely relation between Kobe Bryant and LA Lakers is:

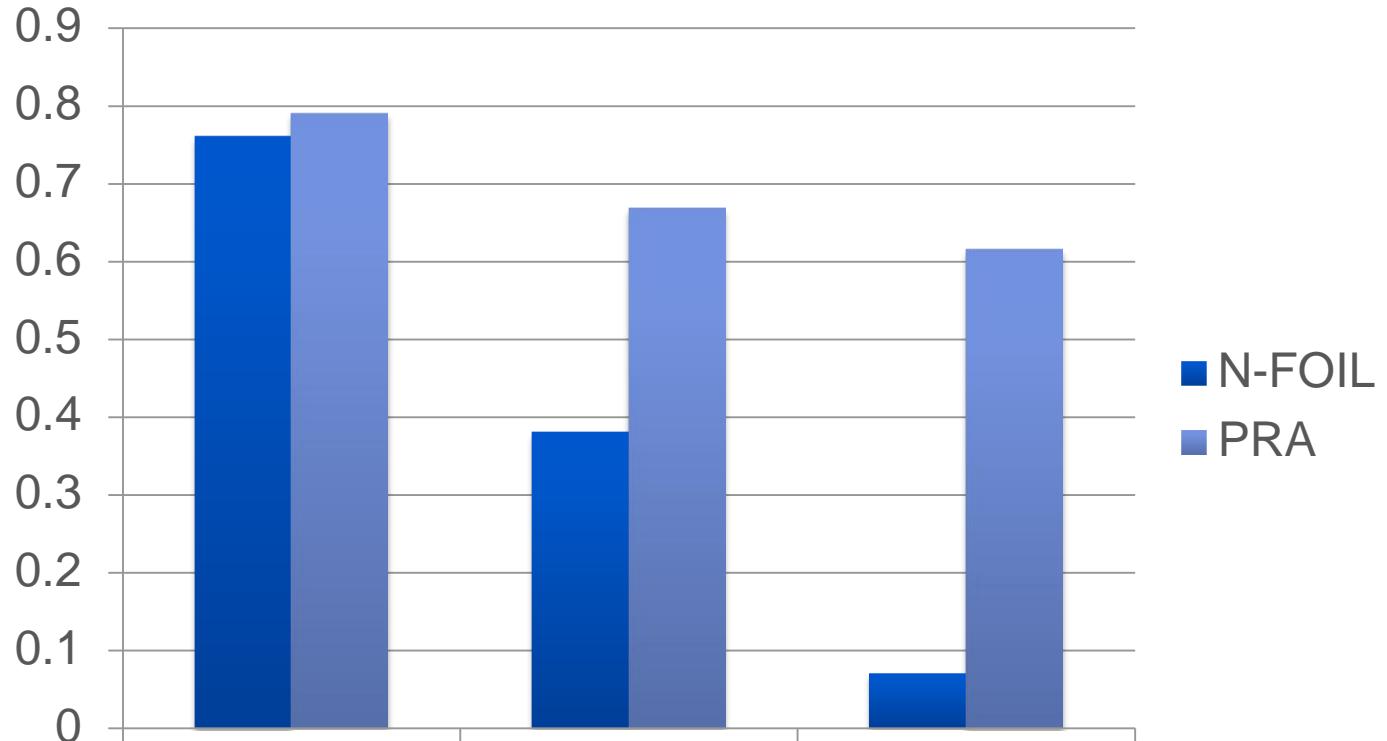
$$rel = \operatorname{argmax}_{r' \in rels} S(TP, r', SAS)$$

- The most likely team for Kobe Bryant is:

$$obj = \operatorname{argmax}_{e' \in ents} S(TP, playFor, e')$$

- **As prior** for extraction models (cf. Knowledge Vault [Dong et al., 2014])
- **No calibration of scores** like probabilities

Random Walks Boost Precision

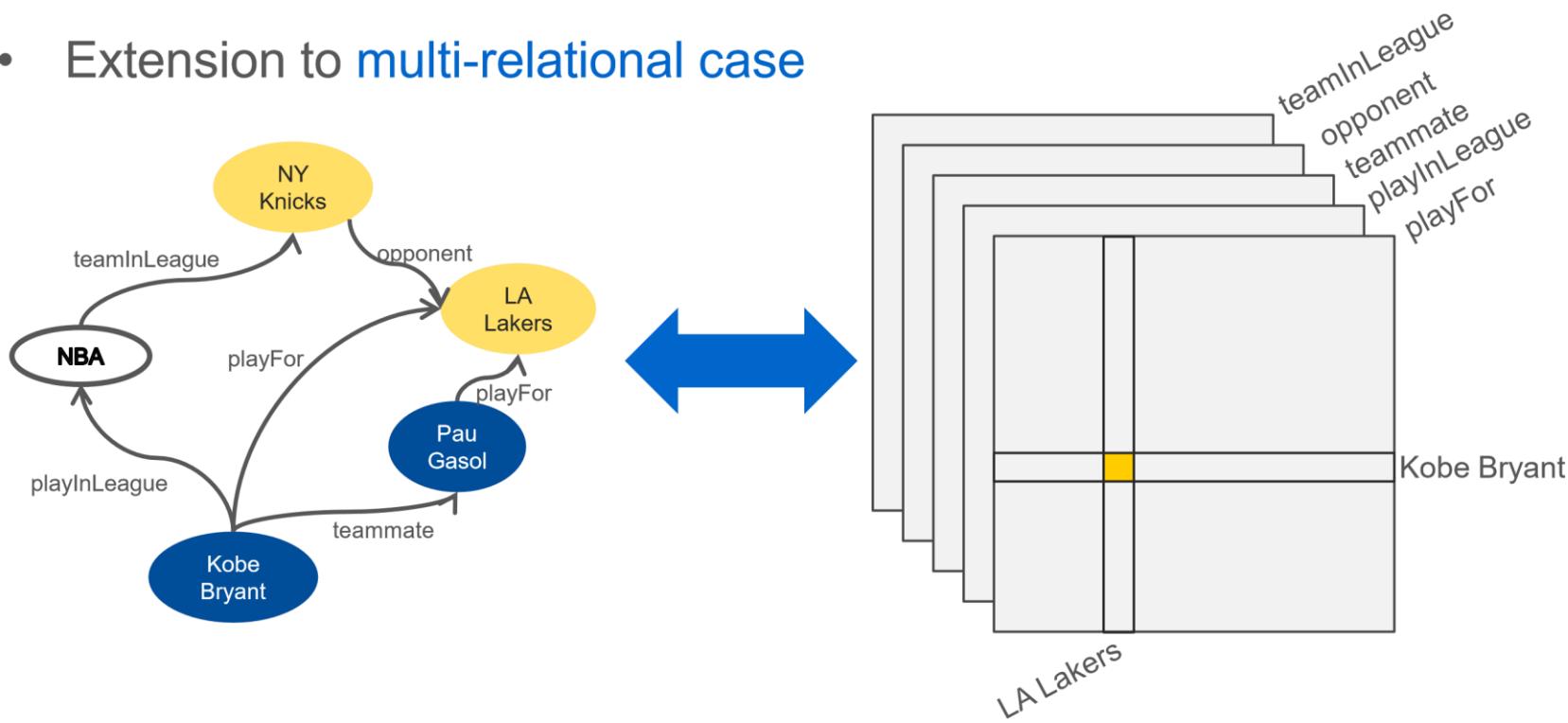


Precision of generalized facts for three cutoffs [Lao et al., 2011]

Factorization methods

- Matrix factorization is successful: collaborative filtering, recommendation, etc.

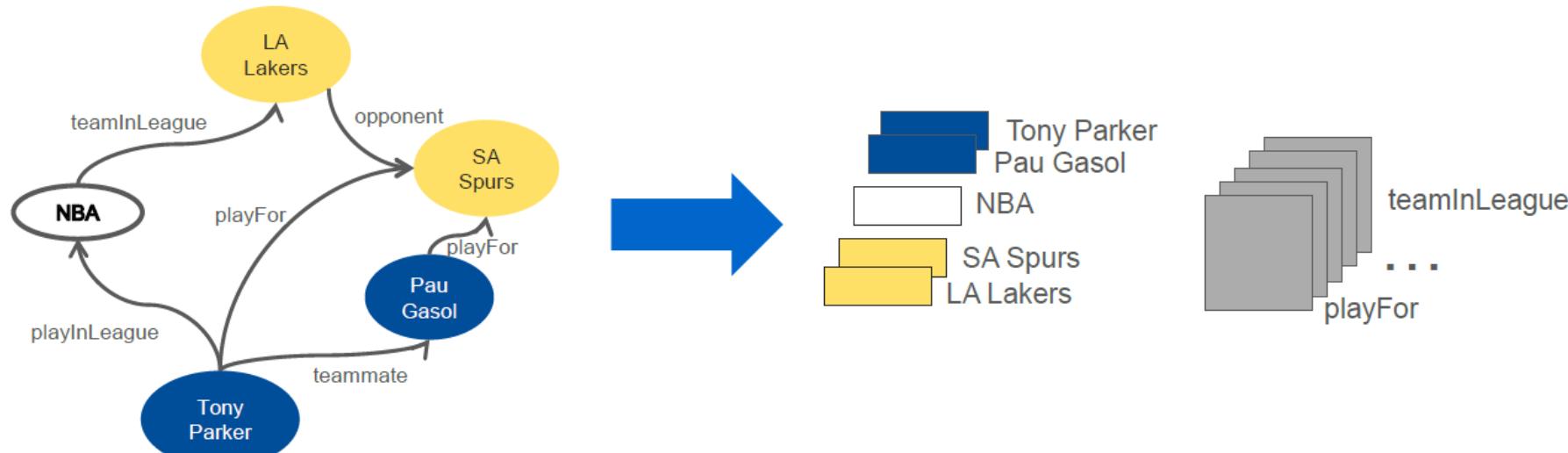
- Extension to multi-relational case



- Collective matrix factorization or tensor factorization

Embedding models

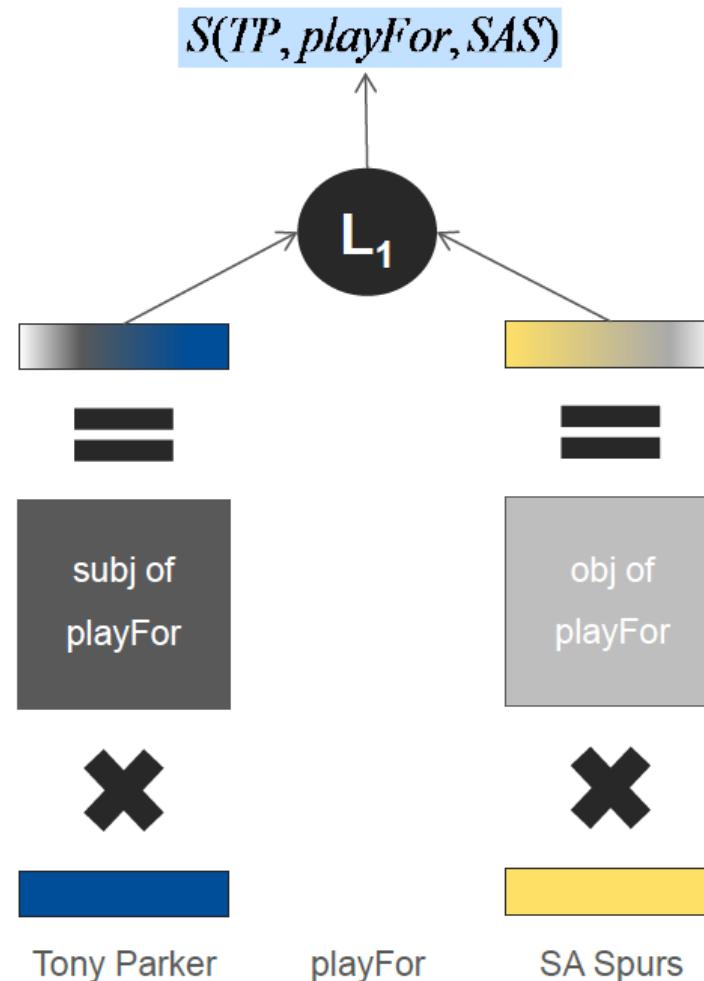
- Related to Deep Learning methods
- Entities are **vectors** (low-dimensional sparse)
- Relation types are **operators** on these vectors



Embeddings trained to define a **similarity score** on triples such that

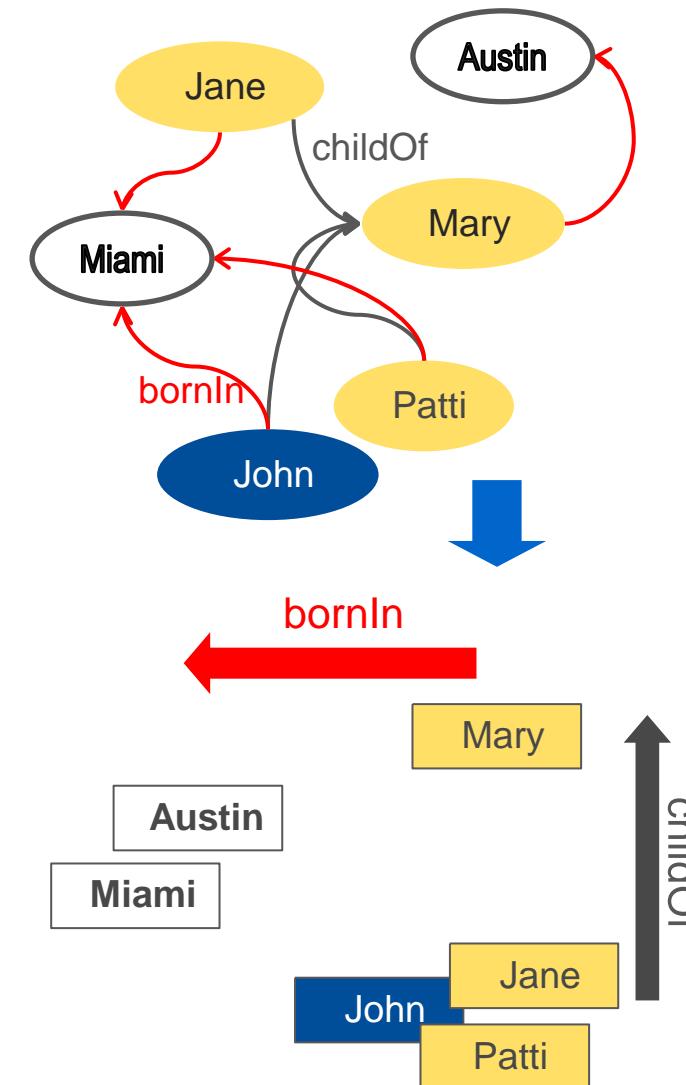
$$S(KB, playFor, LAL) > S(KB, playFor, NYK)$$

Structured Embeddings [Bordes et al., 11]



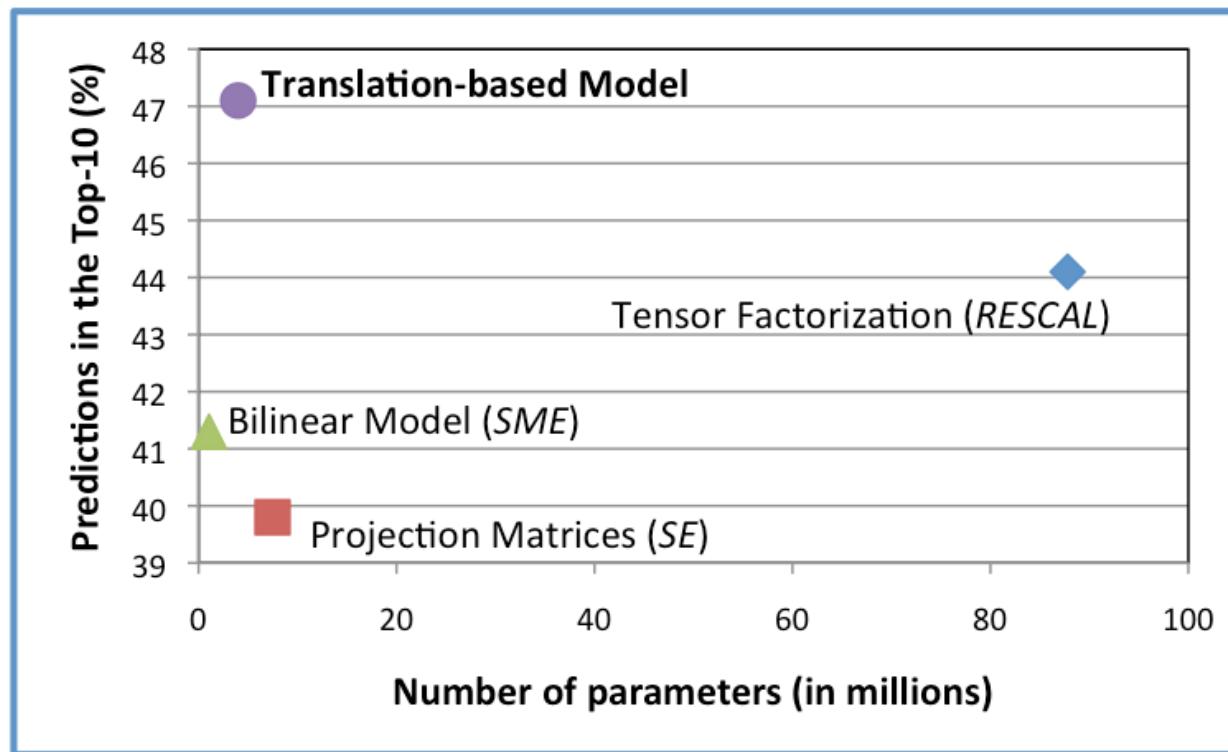
$$S(TP, playFor, SAS) = -\left\| M_{playFor}^{sub} e_{TP} - M_{playFor}^{obj} e_{SAS} \right\|_1$$

Translating Embeddings [Bordes et al. 13]



$$S(john, bornIn, miami) = -\left\| e_{john} + e_{bornIn} - e_{miami} \right\|_2$$

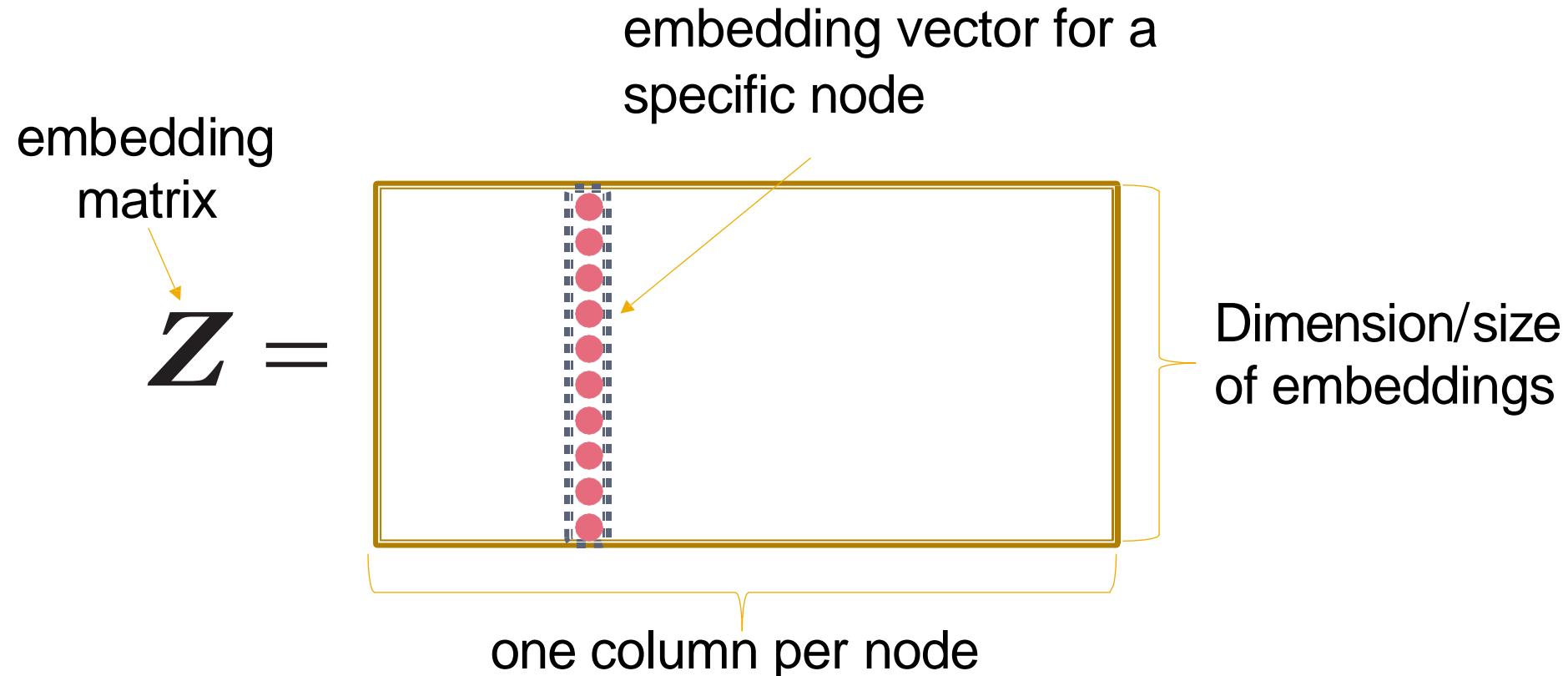
The Simpler, The Better



Ranking object entities on a subset of Freebase [Bordes et al. 13]

Recap: Shallow Encoding

- Simplest encoding approach: **encoder is just an embedding-lookup**



Formal KG Representation

- Edges in KG are represented as **triples** (h, r, t)
 - head (h) has relation (r) with tail (t)
- **Key Idea:**
 - Model entities and relations in the embedding/vector space \mathbb{R}^d .
 - Associate entities and relations with **shallow embeddings**
 - **Note we do not learn a GNN here!**
 - Given a true triple (h, r, t) , the goal is that the **embedding of (h, r) should be close to the embedding of t .**

TransE

embedding vectors will appear in boldface

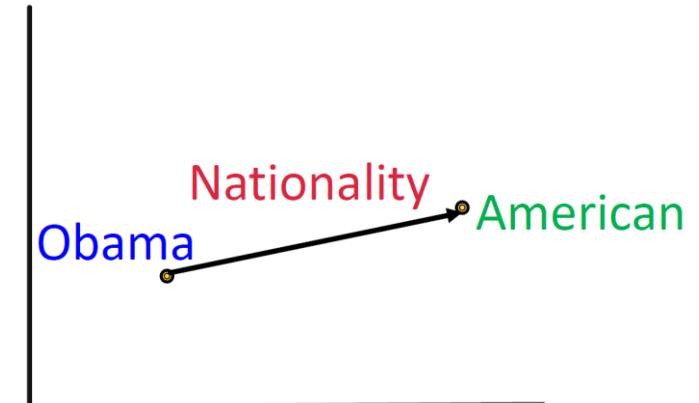
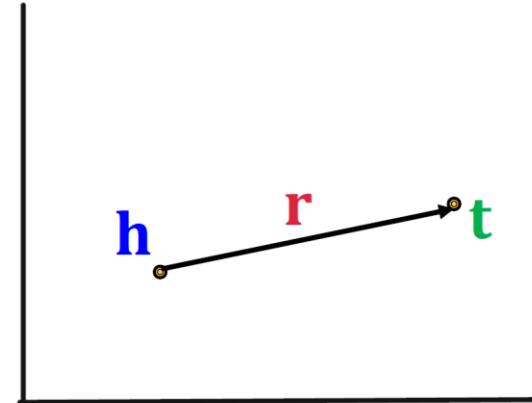
- **Translation Intuition:**

- For a triple (h, r, t) , $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$,

- $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ if the given fact is true
- $\mathbf{h} + \mathbf{r} \neq \mathbf{t}$ else
- $f_r(h, t) = -d_r(h, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$

- **Loss function:**
$$L(h, r, t) = \max(0, \gamma + f_{r,neg} - f_{r,pos}) \\ = \max(0, \gamma + d_{r,pos} - d_{r,neg})$$

- In learning, we minimize the loss function for all pairs of triples $\min L(h, r, t)$



TransE Algorithm

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L , margin γ , embeddings dim. k .

- 1: **initialize** $\ell \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$
- 2: $\ell \leftarrow \ell / \|\ell\|$ for each $\ell \in L$
- 3: $e \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$
- 4: **loop**
- 5: $e \leftarrow e / \|e\|$ for each entity $e \in E$
- 6: $S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b
- 7: $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
- 8: **for** $(h, \ell, t) \in S_{batch}$ **do**
- 9: $(h', \ell, t') \leftarrow \text{sample}(S'_{(h, \ell, t)})$ // sample a corrupted triplet
- 10: $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$
- 11: **end for**
- 12: Update embeddings w.r.t.
- 13: **end loop**

Entities and relations are initialized uniformly, and normalized

Negative sampling with triplet that does not appear in the KG

$$\sum_{((h, \ell, t), (h', \ell, t')) \in T_{batch}} \nabla [\gamma + d(\mathbf{h} + \ell, \mathbf{t}) - d(\mathbf{h}' + \ell, \mathbf{t}')]_+$$

positive sample negative sample

d represents distance (negative of score)

Contrastive loss: favors lower distance (or higher score) for valid triplets, high distance (or lower score) for corrupted ones

Connectivity Patterns in KG

- **Relations in a heterogeneous KG have different properties**
 - Example
 - **Symmetry:** If the edge $(h, \text{"Roommate"}, t)$ exists in KG, then the edge $(t, \text{"Roommate"}, h)$ should also exist.
 - **Inverse relation:** If the edge $(h, \text{"Advisor"}, t)$ exists in KG, then the edge $(t, \text{"Advisee"}, h)$ should also exist.
- Can we **categorize** these relation patterns?
- Are KG embedding methods (e.g., **TransE**) expressive enough to model these patterns?

Relation Patterns

- **Symmetric (Antisymmetric) Relations:**

- **Example:**

- Symmetric: Family, Roommate
 - Antisymmetric: Hyponym

$$r(h, t) \Rightarrow r(t, h) \quad (r(h, t) \Rightarrow \neg r(t, h)) \quad \forall h, t$$

- **Inverse Relations:**

$$r_2(h, t) \Rightarrow r_1(t, h)$$

- **Example :** (Advisor, Advisee)

- **Composition (Transitive) Relations:** $r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z) \quad \forall x, y, z$

- **Example:** My mother's husband is my father.

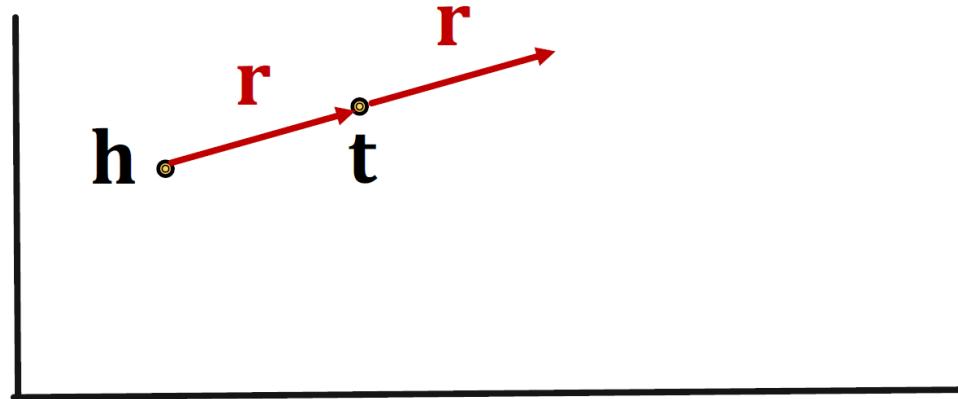
- **1-to-N relations:**

- $r(h, t_1), r(h, t_2), \dots, r(h, t_n)$ are all True.

- **Example:** r is "StudentsOf"

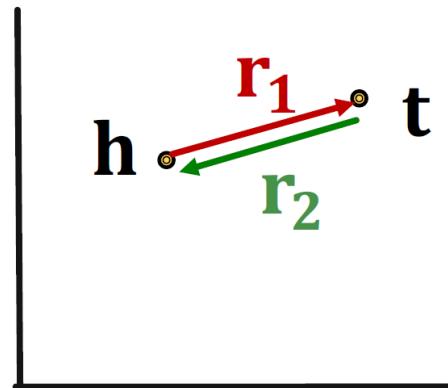
Antisymmetric Relations in TransE

- **Antisymmetric Relations:**
 - $r(h, t) \Rightarrow \neg r(t, h) \forall h, t$
- **Example:** Hyponym
- **TransE** can model antisymmetric relations ✓
 - $\mathbf{h} + \mathbf{r} = \mathbf{t}$, but $\mathbf{t} + \mathbf{r} \neq \mathbf{h}$



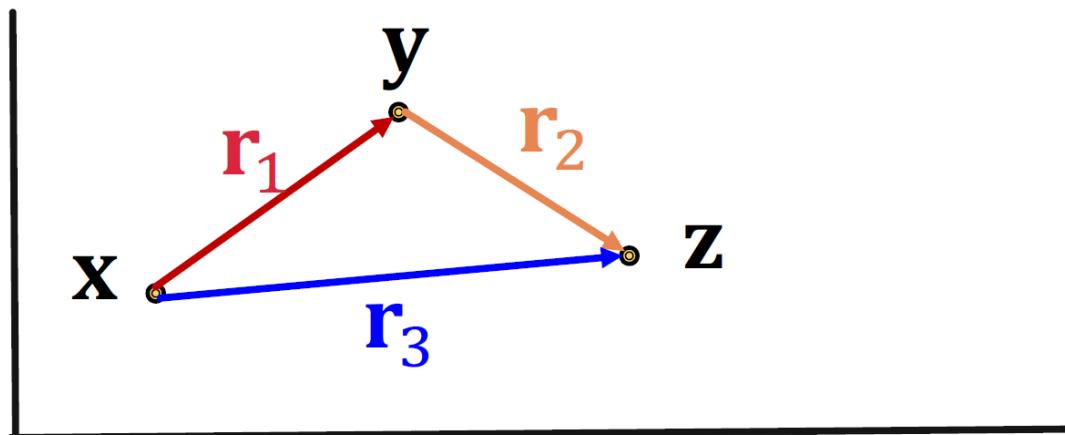
Inverse Relations in TransE

- **Inverse Relations:**
 - $r_1(h, t) \Rightarrow r_2(t, h)$
 - **Example :** (Advisor, Advisee)
- **TransE** can model inverse relations ✓
 - $\mathbf{h} + \mathbf{r}_2 = \mathbf{t}$, we can set $\mathbf{r}_1 = -\mathbf{r}_2$



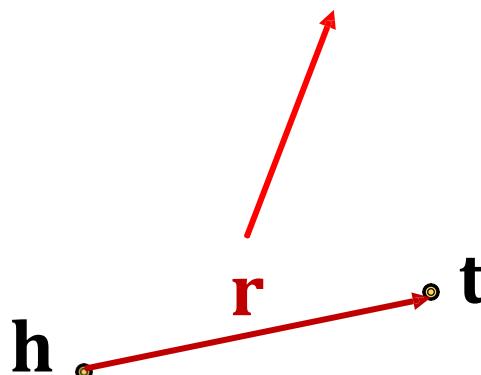
Composition in TransE

- **Composition (Transitive) Relations:**
 - $r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z) \quad \forall x, y, z$
 - **Example:** My mother's husband is my father.
- **TransE** can model composition relations ✓
 - $\mathbf{r}_1 = \mathbf{r}_2 + \mathbf{r}_3$



Limitation: Symmetric Relations

- **Symmetric Relations:**
 - $r(h, t) \Rightarrow r(t, h) \quad \forall h, t$
- **Example:** Family, Roommate
- **TransE cannot** model symmetric relations ✗
 - only if $\mathbf{r} = 0, \mathbf{h} = \mathbf{t}$



For all h, t that satisfy $r(h, t)$, $r(t, h)$ is also True, which means $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| = 0$ and $\|\mathbf{t} + \mathbf{r} - \mathbf{h}\| = 0$. Then $\mathbf{r} = 0$ and $\mathbf{h} = \mathbf{t}$, however h and t are two different entities and should be mapped to different locations.

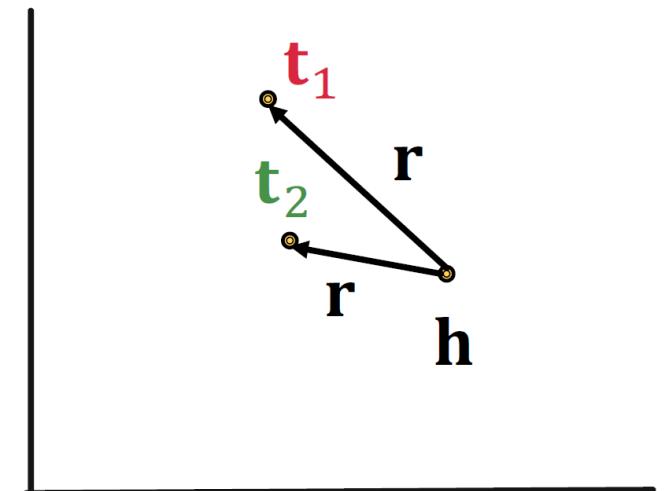
Limitation: 1-to-N Relations

- **1-to-N Relations:**
 - **Example:** (h, r, t_1) and (h, r, t_2) both exist in the knowledge graph, e.g., r is “StudentsOf”
- **TransE cannot** model 1-to-N relations ✗
 - t_1 and t_2 will map to the same vector, although they are different entities

$$t_1 = h + r = t_2$$

$$t_1 \neq t_2$$

contradictory!



TransR

- TransE models translation of any relation in the **same** embedding space.
- Can we design a new space for each relation and do translation in **relation-specific space**?
- TransR: model **entities** as vectors in the entity space \mathbb{R}^d and model each **relation** as vector in relation space $\mathbf{r} \in \mathbb{R}^k$ with $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ as the projection matrix.

TransR

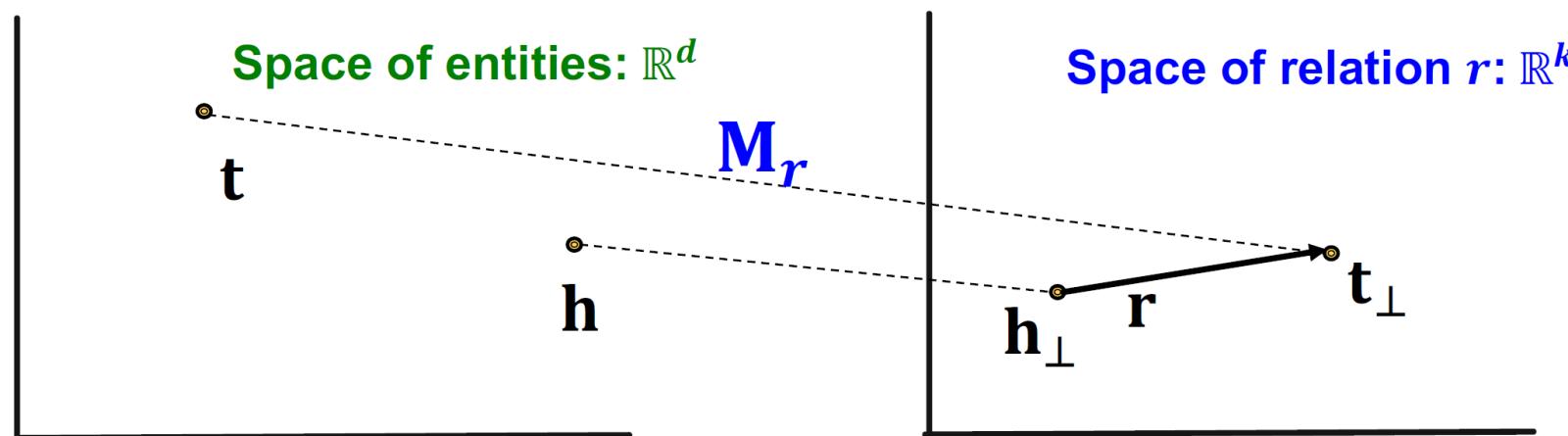
- TransR: model **entities** as vectors in the entity space \mathbb{R}^d and model each **relation** as vector in relation space $\mathbf{r} \in \mathbb{R}^k$ with $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ as the projection matrix.

$$\mathbf{h}_{\perp} = \mathbf{M}_r \mathbf{h}, \quad \mathbf{t}_{\perp} = \mathbf{M}_r \mathbf{t}$$

- Score function:

$$f_r(h, t) = -||\mathbf{h}_{\perp} + \mathbf{r} - \mathbf{t}_{\perp}||$$

Use \mathbf{M}_r to project from entity space \mathbb{R}^d to relation space \mathbb{R}^k



Symmetric Relations in TransR

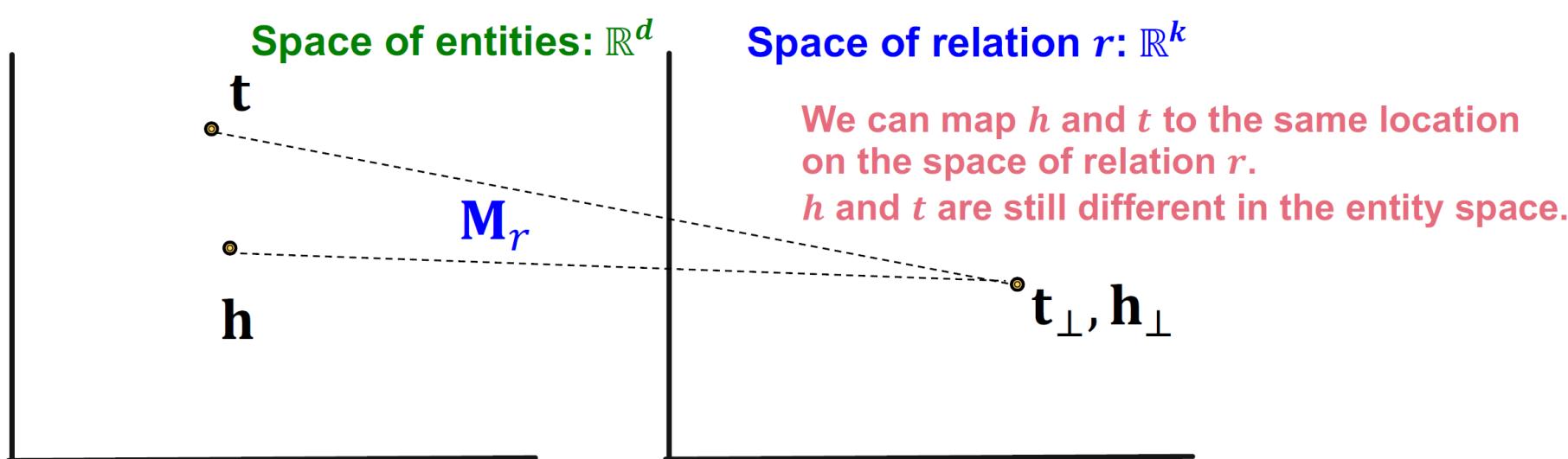
- **Symmetric Relations:**

- $r(h, t) \Rightarrow r(t, h) \forall h, t$

- **Example:** Family, Roommate

- **TransR** can model symmetric relations

$$\mathbf{r} = 0, \mathbf{h}_\perp = \mathbf{M}_r \mathbf{h} = \mathbf{M}_r \mathbf{t} = \mathbf{t}_\perp \checkmark$$



Antisymmetric Relations in TransR

- **Antisymmetric Relations:**

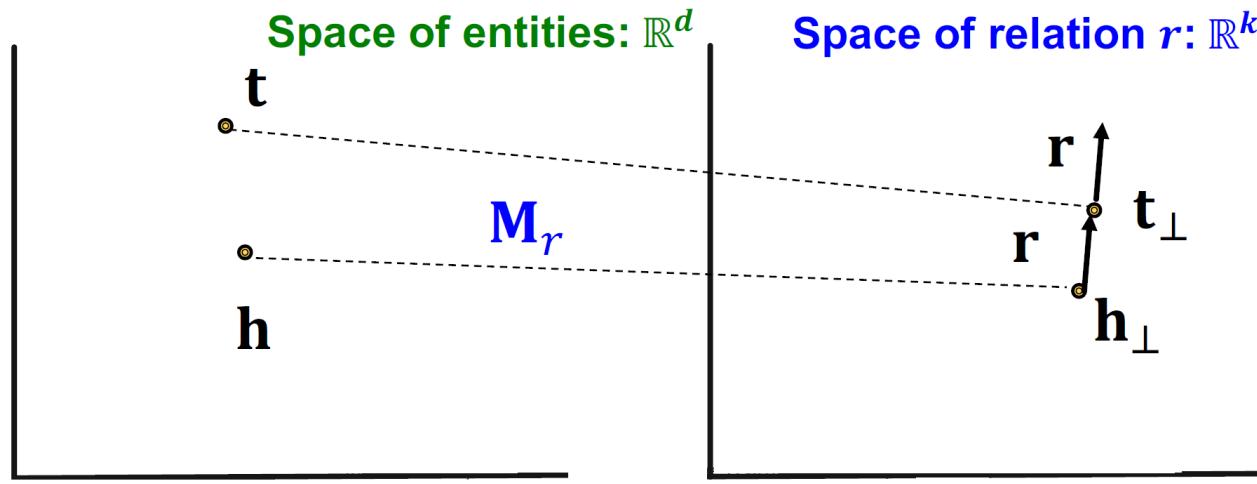
- $r(h, t) \Rightarrow \neg r(t, h) \forall h, t$

- **Example:** Hypernym

- **TransR** can model antisymmetric relations

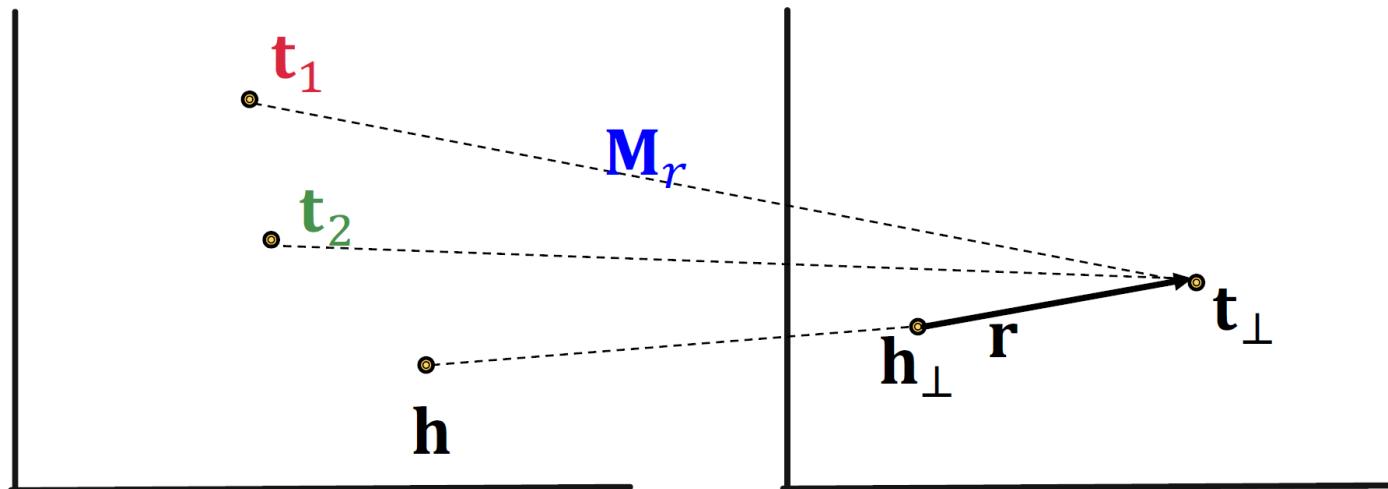
$$\mathbf{r} \neq 0, \mathbf{M}_r \mathbf{h} + \mathbf{r} = \mathbf{M}_r \mathbf{t},$$

Then $\mathbf{M}_r \mathbf{t} + \mathbf{r} \neq \mathbf{M}_r \mathbf{h}$ ✓



1-to-N Relations in TransR

- **1-to-N Relations:**
 - **Example:** (h, r, t_1) and (h, r, t_2) both exist in the knowledge graph, e.g., r is “StudentsOf”
- **TransR** can model 1-to-N relations ✓
 - We can learn \mathbf{M}_r so that $\mathbf{t}_\perp = \mathbf{M}_r \mathbf{t}_1 = \mathbf{M}_r \mathbf{t}_2$
 - Note that \mathbf{t}_1 does not need to be equal to \mathbf{t}_2 !



Inverse Relations in TransR

- **Inverse Relations:**

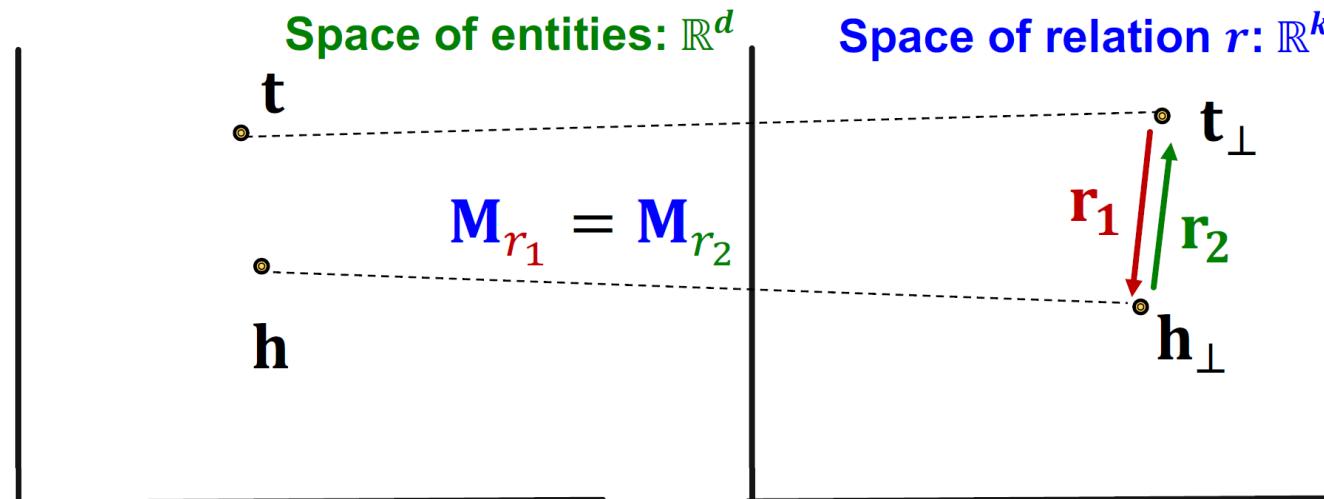
- $r_1(h, t) \Rightarrow r_2(t, h)$

- **Example :** (Advisor, Advisee)

- **TransR** can model inverse relations

$$\mathbf{r}_2 = -\mathbf{r}_1, \mathbf{M}_{r_1} = \mathbf{M}_{r_2}$$

$$\mathbf{M}_{r_1} \mathbf{t} + \mathbf{r}_1 = \mathbf{M}_{r_1} \mathbf{h} \quad \mathbf{M}_{r_2} \mathbf{h} + \mathbf{r}_2 = \mathbf{M}_{r_2} \mathbf{t}$$



Composition Relations in TransR

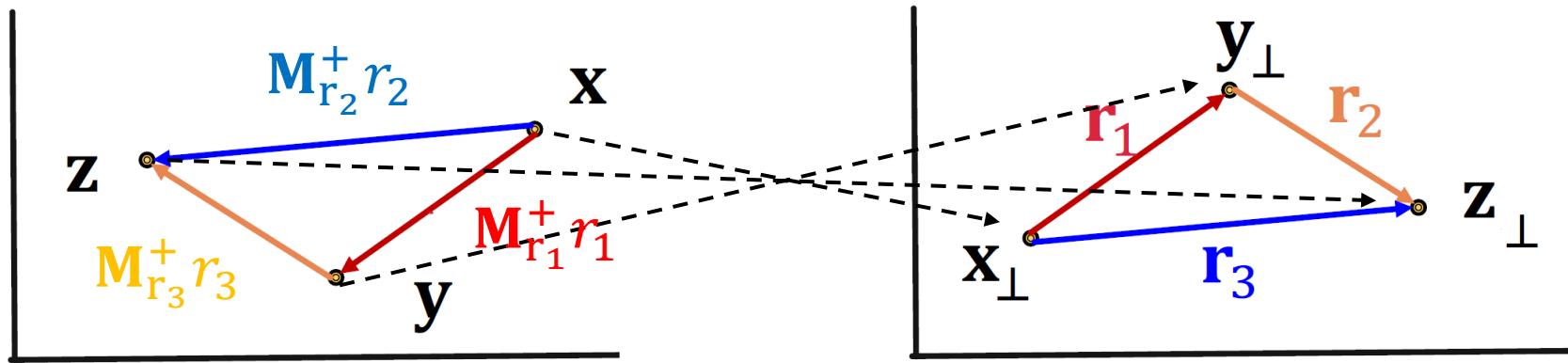
- **Composition (Transitive) Relations:**

- $r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z) \quad \forall x, y, z$

- **Example:** My mother's husband is my father.

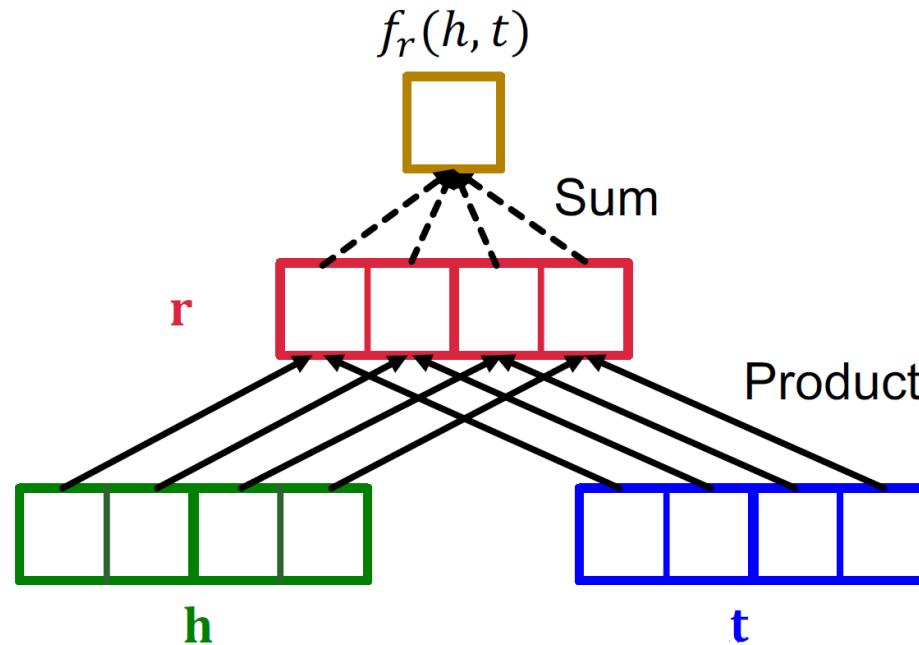
- **TransR** can model composition relations ✓

$\mathbf{M}_{r_i}^+ = \mathbf{M}_{r_i}^T (\mathbf{M}_{r_i} \mathbf{M}_{r_i}^T)^{-1}$ is the Moore–Penrose inverse (pseudoinverse) of \mathbf{M}_{r_i} where $\mathbf{M}_{r_i} \mathbf{M}_{r_i}^T$ is invertible (full rank); so that $\mathbf{M}_{r_i} \mathbf{M}_{r_i}^+ = I$



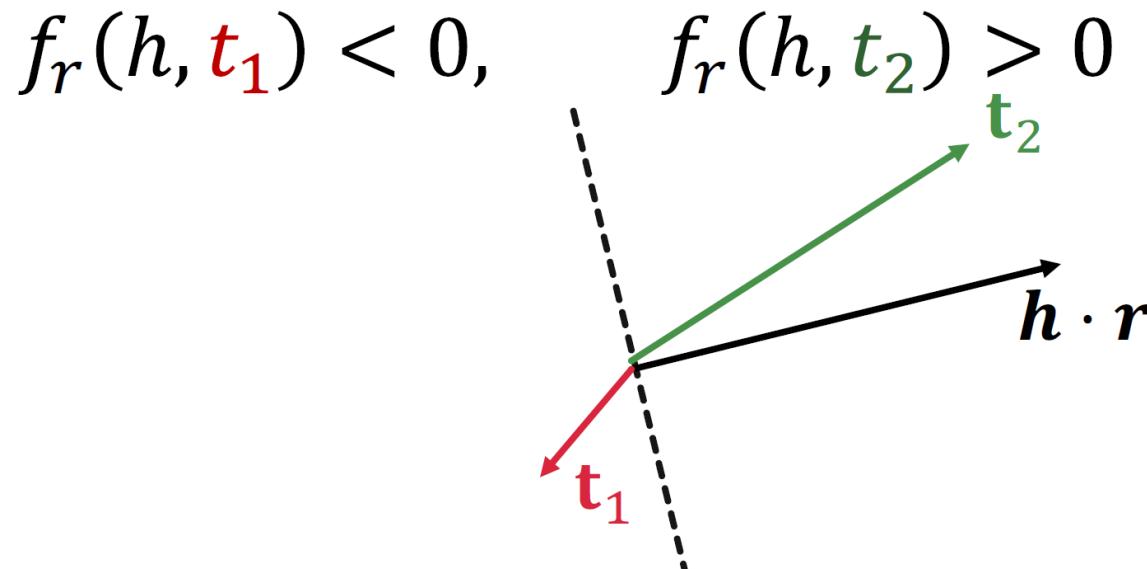
DistMult

- **So far:** The scoring function $f_r(h, t)$ is **negative of L1 / L2 distance** in **TransE** and **TransR**
- Another line of KG embeddings adopt **bilinear modelling**
- **DistMult:** Entities and relations using vectors in \mathbb{R}^k
- **Score funcMon:** $f_r(h, t) = \langle h, r, t \rangle = \sum_i h_i \cdot r_i \cdot t_i$
- $h, r, t \in \mathbb{R}^k$



DistMult

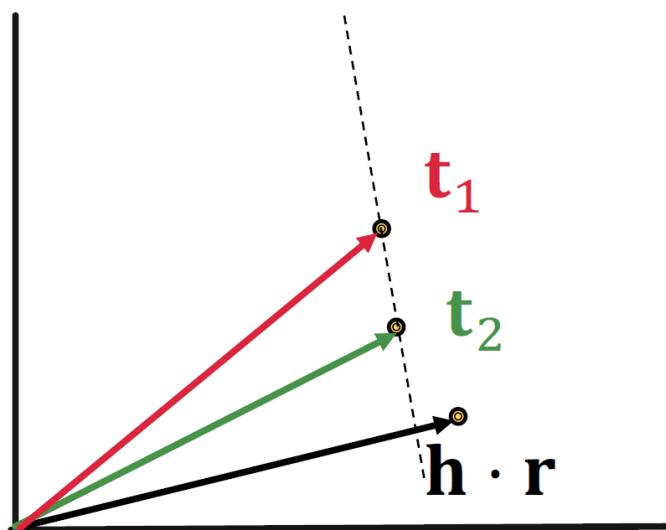
- **Intuition of the score function:** Can be viewed as a **dot product** between $\mathbf{h} \cdot \mathbf{r}$ and \mathbf{t}
- **Example:**



1-to-N Relations in DistMult

- **1-to-N Relations:**
 - **Example:** (h, r, t_1) and (h, r, t_2) both exist in the knowledge graph, e.g., r is “StudentsOf”
- **DistMult** can model 1-to-N relations ✓

$$\langle h, r, t_1 \rangle = \langle h, r, t_2 \rangle$$



Symmetric Relations in DistMult

- **Symmetric Relations:**
 - $r(h, t) \Rightarrow r(t, h) \forall h, t$
- **Example:** Family, Roommate
- **DistMult** can naturally model symmetric relations

$$f_r(h, t) = \langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i = \\ \langle \mathbf{t}, \mathbf{r}, \mathbf{h} \rangle = f_r(t, h)$$

Limitation: Inverse Relations

- **Inverse Relations:**
 - $r_1(h, t) \Rightarrow r_2(t, h)$
 - **Example :** (Advisor, Advisee)
- **DistMult cannot** model inverse relations
 - If it does model inverse relations:
$$f_{r_2}(h, t) = \langle \mathbf{h}, \mathbf{r}_2, \mathbf{t} \rangle = \langle \mathbf{t}, \mathbf{r}_1, \mathbf{h} \rangle = f_{r_1}(t, h)$$
 - This means $\mathbf{r}_1 = \mathbf{r}_2$
 - But semantically this does not make sense: **The embedding of “Advisor” should not be the same with “Advisee”.**

Limitation: Composition Relations

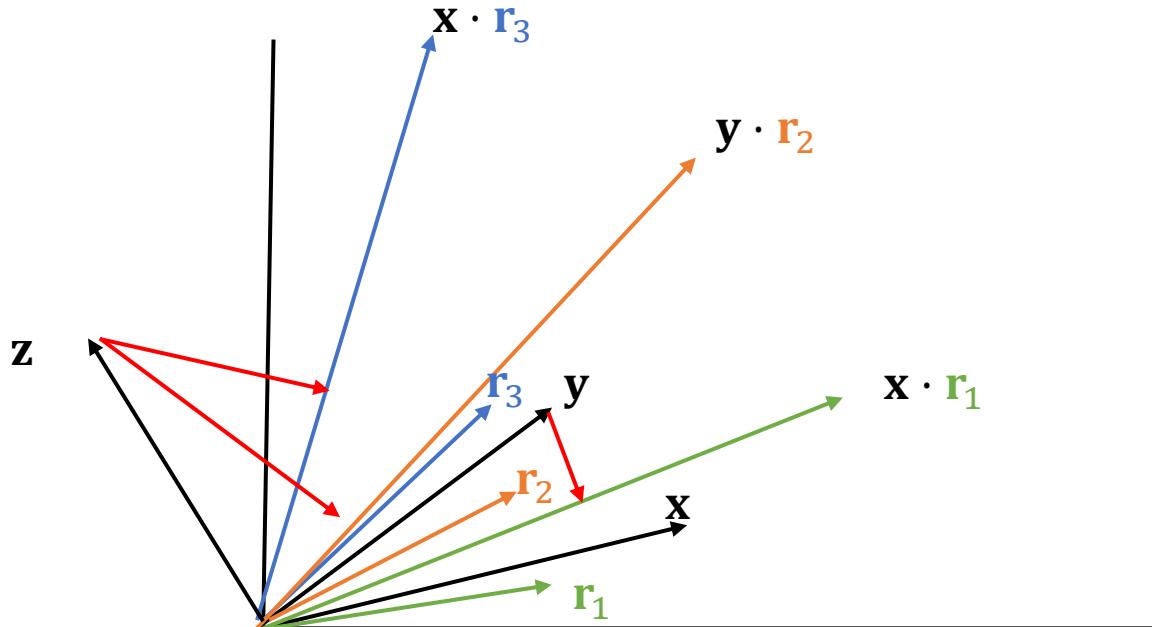
- **Composition (Transitive) Relations:**

- $r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z) \quad \forall x, y, z$

- **Example:** My mother's husband is my father.

- **DistMult cannot** model composition relations

- **Intuition:** **DistMult** defines a hyperplane for each (head, relation), the union of the hyperplanes induced by multihops of relations, e.g., (r_1, r_2) , cannot be expressed using a single hyperplane.



Some Other KG Embedding Algorithms

Model	Sym.	Antisym.	Inv.	Compos.	1-to-N
TransE	✗	✓	✓	✓	✗
TransR	✓	✓	✓	✓	✓
DistMult	✓	✗	✗	✗	✓
ComplEx	✓	✓	✓	✗	✓
RotatE	✓	✓	✓	✓	✓

KG Embedding in Practice

- Different KGs may have drastically **different relation patterns!**
- There is not a general embedding that works for all KGs, use the **table** to select models
- Try **TransE** for a quick run if the target KG does not have much symmetric relations
- Then use more expressive models, e.g., **ComplEx**, **RotatE** (**TransE** in Complex space)

Summary

- Link prediction / Graph completion is one of the prominent tasks on knowledge graphs
- Introduce **TransE** / **TransR** / **DistMult** models with different embedding space and expressiveness