Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

A review: Knowledge reasoning over knowledge graph

Xiaojun Chen, Shengbin Jia, Yang Xiang*

College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

Published in 2020

Backgrounds

- KGs (large amount of prior knowledge but can also effectively organize data) -> Question-answering systems, search engines, and recommendation systems

- Knowledge reasoning -> identify errors and infer new conclusions (knowledge graph enrichment)

"In short, reasoning is the process of drawing conclusions from existing facts by the rules."

Leading knowledge graphs

Table 2
Examples of world's leading knowledge graphs and their statistics (Paulheim, 2017).

Knowledge graphs	#Entities	#Relations	#Facts
WordNet	0.15M	200,000	4.5M
Freebase	50M	38,000	3B
YAGO	17M	76	150M
DBpedia (En)	4.8M	2800	176M
Wikidata	16M	1673	66M
NELL	2M	425	120M

WordNet:
A lexical database for the English language

Freebase:
Contains data harvested from sources: Wikipedia, NNDB, Fashion Model Directory, MusicBrainz, and data contributed by its users

YAGO:
Extracted from Wikipedia, WordNet, and GeoNames

Dbpedia:
Consistent ontology, including persons, places, music albums, films, video games, organizations, species, and diseases. Can be integrated into AWS.

Wikidata:
multilingual, open, linked, structured knowledge base

NELL:
Never-Ending Language Learning system
A semantic machine learning system that runs 24/7, forever, learning to read the web

Definition: knowledge reasoning over KGs

Definition 1Knowledge reasoning over KGs:. Given a knowledge graph $KG = < E, R, T >$ and the relation path P, where E, T represent the set of entities, R denotes the set of relations, and the edges in R link two nodes to form a triple $(h, r, t) \in T$, generating a triplet that does not exist in the KG $G' = \{(h, r, t) | h \in E, r \in R, t \in T, (h, r, t) \notin G\}$.

Knowledge reasoning based on distributed representation

- Tensor factorization
- Distance model
- Semantic matching model
- Multi-source information

Based on tensor factorization

- **RESCAL**

A Three-Way Model for Collective Learning on Multi-Relational Data

Maximilian Nickel NICKEL@CIP.IFI.LMU.DE
Ludwig-Maximilians-Universität, Munich, Germany

Volker Tresp VOLKER.TRESP@SIEMENS.COM
Siemens AG, Corporate Technology, Munich, Germany

Hans-Peter Kriegel KRIEGEL@DBS.IFI.LMU.DE
Ludwig-Maximilians-Universität, Munich, Germany

computes a three-way factorization of an adjacency tensor that represents the knowledge graph.
Alternatively, it can be interpreted as a *compositional model*, where pairs of entities are represented via the tensor product of their embeddings.

an entity and an attribute value. In order to model dyadic relational data as a tensor, we employ a three-way tensor $\mathcal{X}$, where two modes are identically formed by the concatenated entities of the domain and the third mode holds the relations.[1] Figure 1 provides an illustration of this modelling method. A tensor entry $\mathcal{X}_{ijk} = 1$ denotes the fact that there exists a relation `(i-th entity, k-th predicate, j-th entity)`. Otherwise, for non-existing and unknown relations, the entry is set to zero.

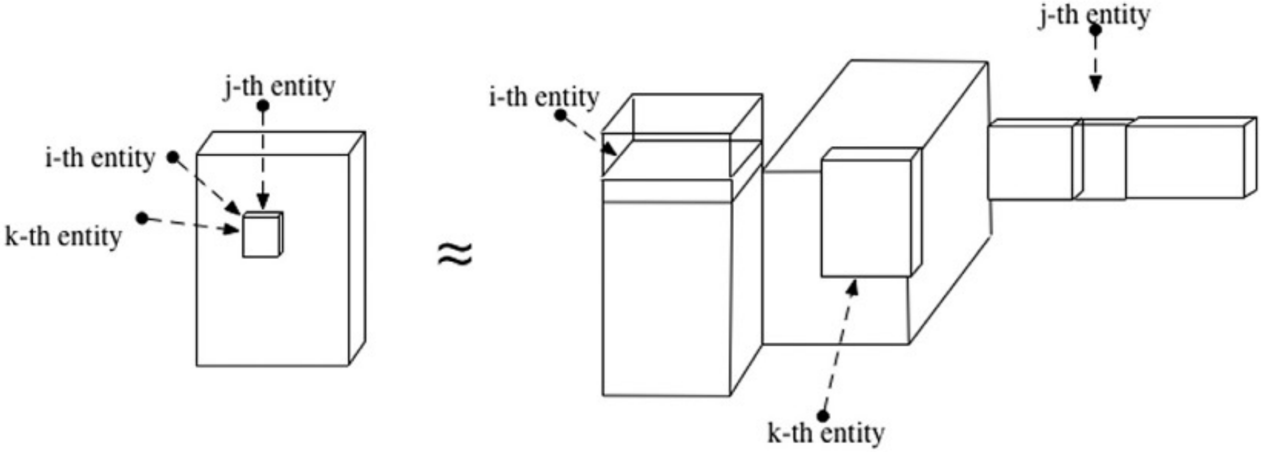


Fig. 3. Simple illustration of RESCAL.

RESCAL for multi-relational data

rank-r factorization

$$\mathcal{X}_k \approx AR_kA^T, \text{ for } k = 1, \ldots, m \qquad (1)$$

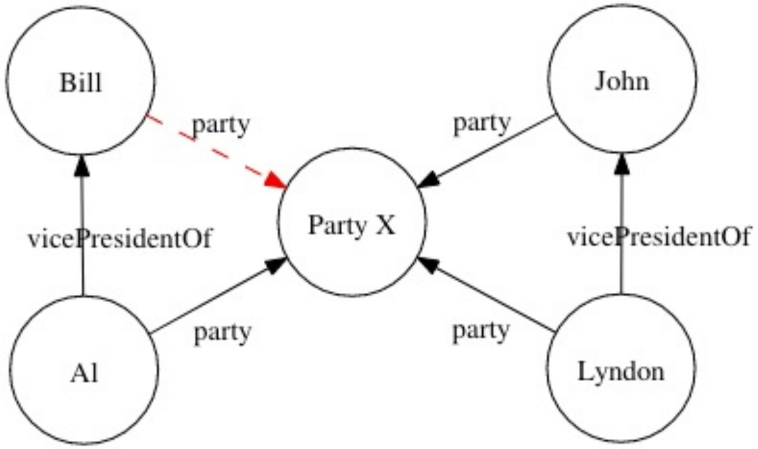


Figure 2: Visualization of a subgraph of the relational graph for the US presidents example. The relation marked red is unknown.

Similar latent-component representations of Al and Lyndon

Bill and John have similar latent-component representations

$$\boldsymbol{a}_{\texttt{Bill}}^T R_{\texttt{party}} \boldsymbol{a}_{\texttt{PartyX}} \qquad \boldsymbol{a}_{\texttt{John}}^T R_{\texttt{party}} \boldsymbol{a}_{\texttt{PartyX}}$$

However, RESCAL can be hard to scale to very large knowledge-graphs because its has a quadratic runtime and memory complexity with regard to the embedding dimension.

Based on tensor factorization

- TRESCAL
  - highly efficient and scalable
- RESCAL-logit
  - improve inference accuracy
- PRESCAL
  - based on paths of tensor factorization is proposed
- Jainet al. (2017) a novel combination of matrix factorization and tensor factorization

Based on distance model

TransE

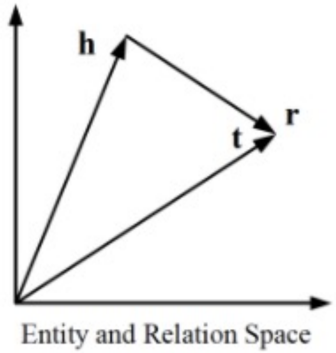


Figure 3: TransE

$$f(h, r, t) = \|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|_{l_1/l_2}$$

Translating Embeddings for Modeling Multi-relational Data

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán
Université de Technologie de Compiègne – CNRS
Heudiasyc UMR 7253
Compiègne, France
{bordesan, nusunier, agarciad}@utc.fr

Jason Weston, Oksana Yakhnenko
Google
111 8th avenue
New York, NY, USA
{jweston, oksana}@google.com

For example if $h_1 = emb(\text{" } Ottawa \text{ "}),\ h_2 = emb(\text{" } Berlin \text{ "}), t_1 = emb(\text{" } Canada \text{ "}), t_2 = (\text{" } Germany \text{ "})$, and finally $r = \text{" } CapilatOf \text{ "}$, then $h_1 + r$ and $h_2 + r$ should approximate t_1 and t_2 respectively.

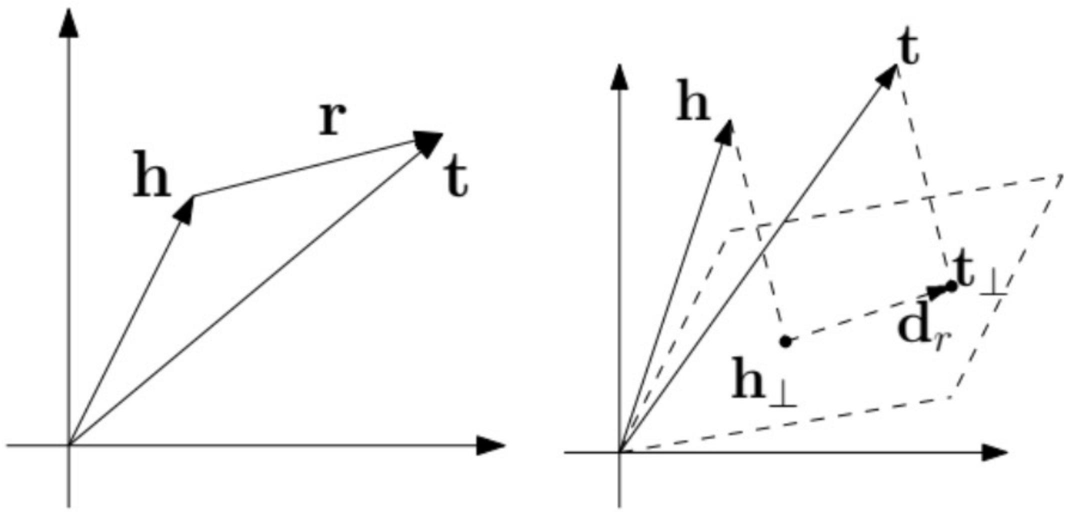
Relationships as translations in the embedding space In this paper, we introduce TransE, an energy-based model for learning low-dimensional embeddings of entities. In TransE, relationships are represented as *translations in the embedding space*: if (h, ℓ, t) holds, then the embedding of the tail entity t should be close to the embedding of the head entity h plus some vector that depends on the relationship ℓ. Our approach relies on a reduced set of parameters as it learns only one low-dimensional vector for each entity and each relationship.

Despite its simplicity and efficiency, TransE cannot deal with One-to-N, N-to-One, and N-to-N relations effectively.
e.g., PresidentOf, TransE might learn indistinguishable representations for Trump and Obama

TransH, TransR

- TransH
- Introducing a relation-specific hyperplane

$$f_r(h, r, t) = \|\boldsymbol{h_r} + \boldsymbol{r} - \boldsymbol{t_r}\|_{l_1/l_2}$$



(a) TransE (b) TransH

- TransR
- Introduces relation-specific spaces
- Projects entity vectors h and t using the space-specific matrix ***Mr***

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

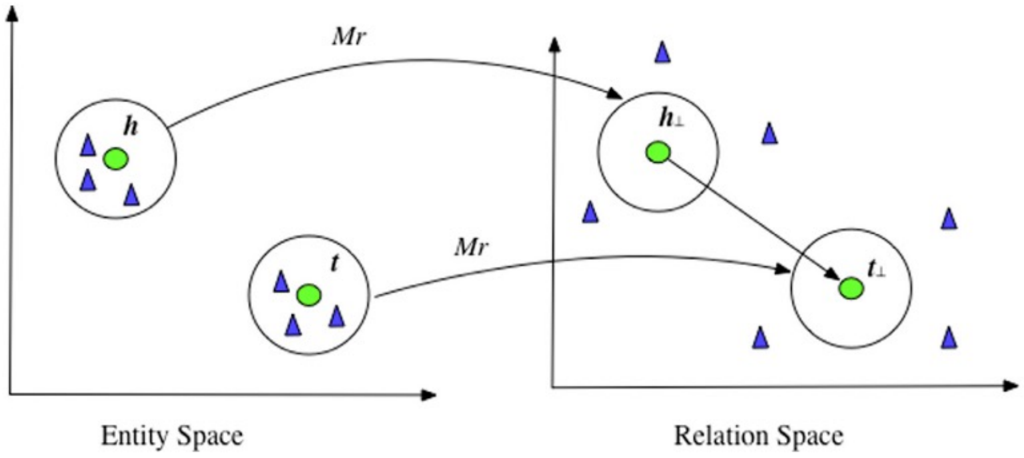


Fig. 5. Simple illustration of TransR (Lin et al., 2015b).

- TransE, TransH, TransR -> each relation has only one semantics

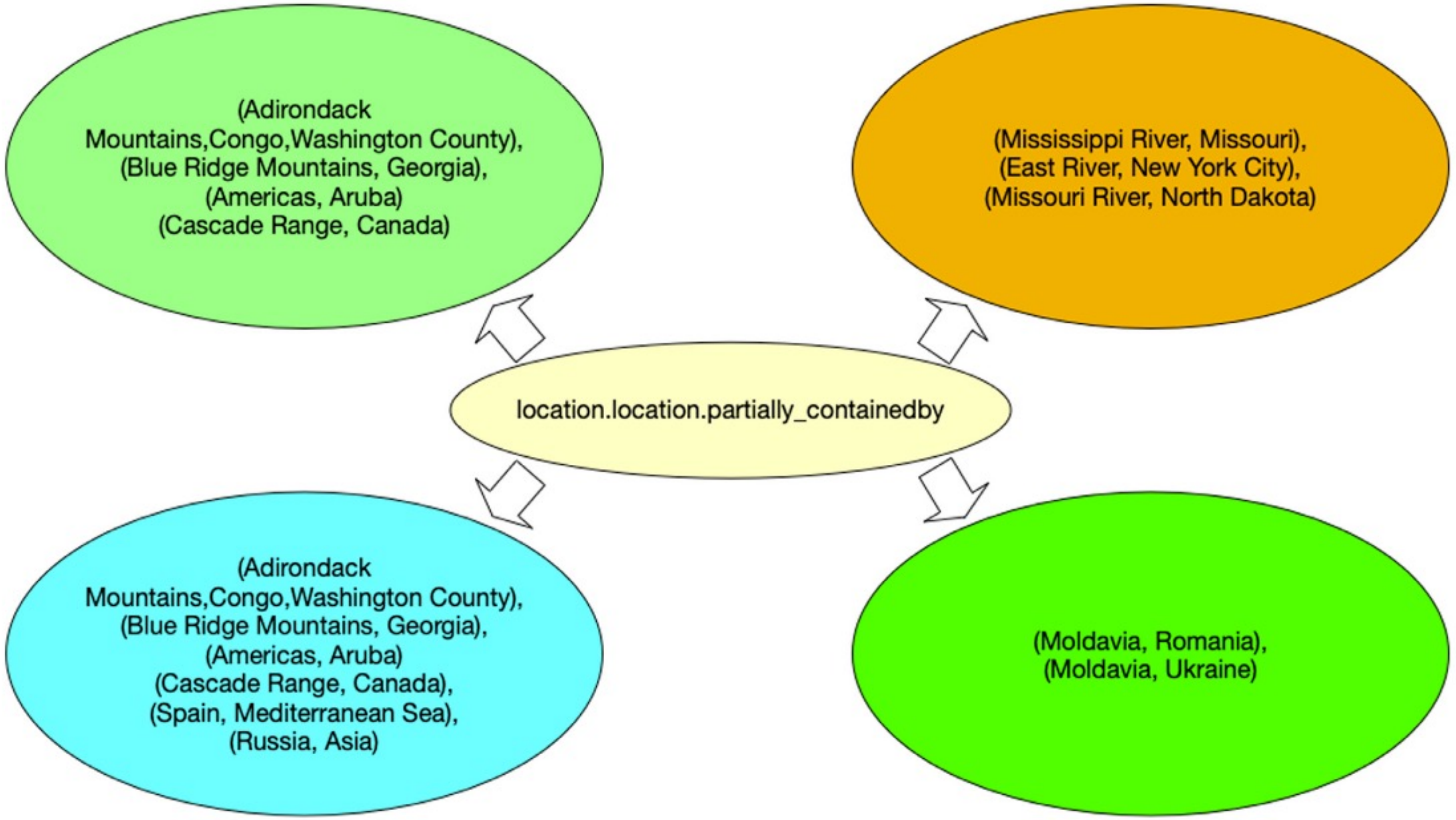


Fig. 6. Multiple types of entities of relation location (the relation *HasPart* has at least two latent semantics: composition related as (*Table, HasPart, Leg*) and location related as (*Atlantic, HasPart, NewYorkBay*)) (Ji et al., 2015) .

mountain-state relation,
but also represents the regional-country relation

TransD

- TransD: based on a dynamic matrix

- Utilizes two vectors to represent an entity or relation
  - First one represents the meaning of entity
  - Second one is used for constructing a mapping matrix

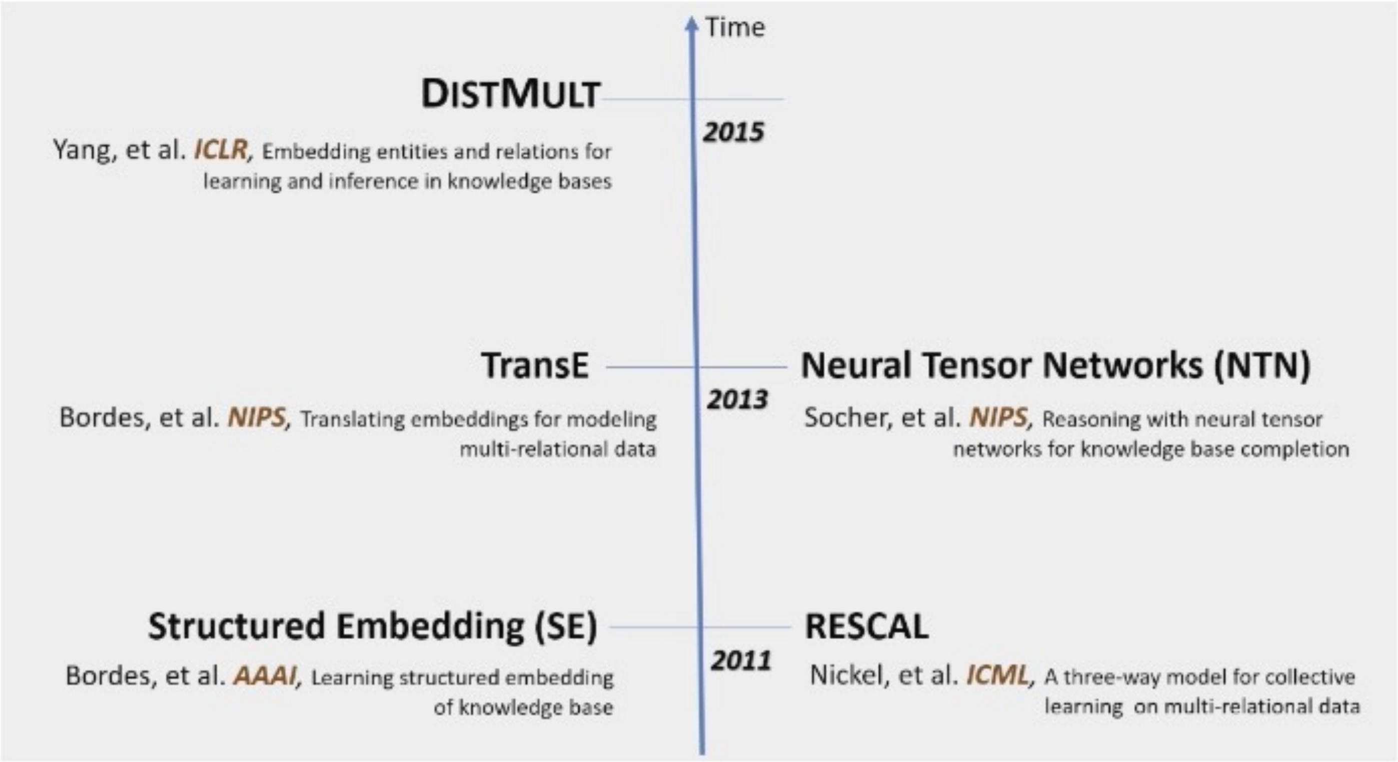
- Compared to TransR, TransD is less complicated and has no matrix-vector multiplication operations

And there are more…

- mTransH
- TransM
- cTransR
- PTransE
- TranSparse – heterogeneity, imbalance
- TransA
- TransAH
- TransG
- KG2E
- t-TransE
- TAE-TransE
- Know-Evolve
- MLN-based approach
- HyTE

Based on semantic matching model

- **DistMult**
- **ComplEx**



Milestones for KG embeddings

DistMult and ComplEx

https://adi-sharma.github.io/files/acl18_kg_geometry_paper.pdf

DistMult (Yang et al., 2014) models entities and relations as vectors in $\mathbb{R}^d$. It uses an entry-wise product ($\odot$) to measure compatibility between head and tail entities, while using logistic loss for training the model.

$$\sigma_{DistMult}(h, r, t) = \mathbf{r}^\top(\mathbf{h} \odot \mathbf{t}) \tag{3}$$

Since the entry-wise product in (3) is symmetric, DistMult is not suitable for asymmetric and anti-symmetric relations.

ComplEx (Trouillon et al., 2016) represents entities and relations as vectors in $\mathbb{C}^d$. The compatibility of entity pairs is measured using entry-wise product between head and complex conjugate of tail entity vectors.

$$\sigma_{ComplEx}(h, r, t) = \mathbf{Re}(\mathbf{r}^\top(\mathbf{h} \odot \bar{\mathbf{t}})) \tag{5}$$

In contrast to (3), using complex vectors in (5) allows ComplEx to handle symmetric, asymmetric and anti-symmetric relations using the same score function. Similar to DistMult, logistic loss is used for training the model.

- Semantic matching energy (SME)
- Latent factor model
- HOLE (holographic embeddings)

Based on multi-source information

Injecting logic rules into embeddings for inference

KALE, a novel method that learns entity and relation for reasoning by jointly modelling knowledge and logic. KALE consists of three key components: triple modelling module, rule modelling module, and joint learning module.

pLogicNet,

IterE

TEKE – making use of rich context information in a text corpus

TKRL

MKRL

…

OpenBioLink

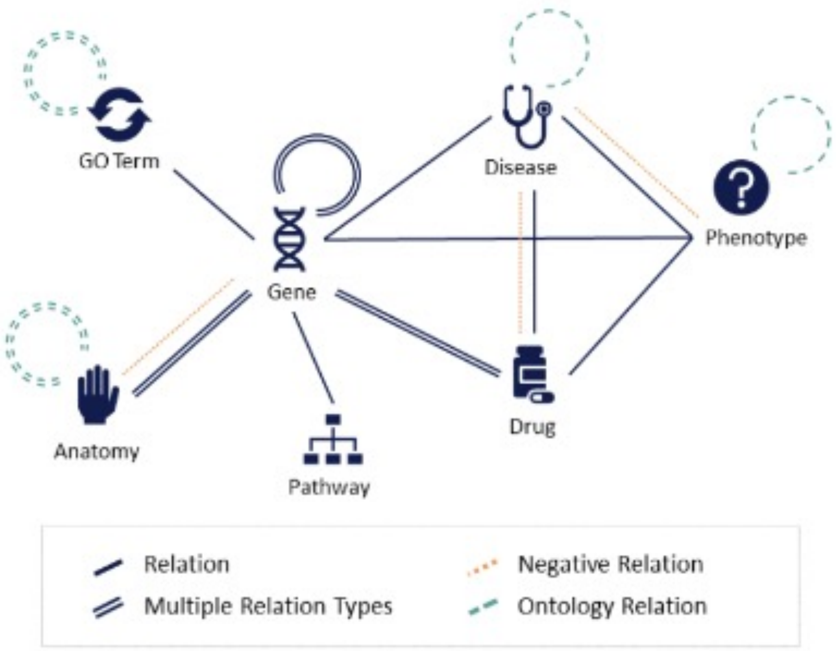


Fig. 1. An overview of the OpenBioLink benchmark graph.

Baseline results

	Model	MRR	h@1	h@10
Latent	RESCAL	**.320**	.212	.544
	TransE	.280	.175	.500
	DistMult	.300	.193	.521
	ComplEx	.319	.211	**.547**
	ConvE	.288	.186	.510
	RotatE	.286	.180	.511
Interpretable	AnyBURL (Maximum)	.277	.192	.457
	AnyBURL (Noisy-OR)	.159	.098	.295
	SAFRAN*	.306	**.214**	.501

$$\mathrm{MRR} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\mathrm{rank}_i} \qquad \mathrm{Hits@k} = \frac{|\{t \in \mathcal{K}_{test} \mid rank(t) \leq k\}|}{|\mathcal{K}_{test}|}$$