Transferable Representation Learning for Multi-relational Data

Recent Advances in Transferable Representation Learning (Part III)

Muhao Chen, Kai-Wei Chang, Dan Roth

Feb 2020

AAAI Tutorials

Recent Advances in Transferable Representation Learning

Outline

- Background and Motivation
- Relational embedding learning methods
  - First-order and high-order methods
  - Non-Euclidean methods
- Knowledge association methods
  - Supervised and semi-supervised methods
  - Auxiliary supervision methods
- Cross-domain and interdisciplinary tasks
  - KBP tasks
  - Computational bio-med tasks

Understanding Relations Is Prominent In Practice

QA and Semantic Search:

Google

mazda car that won 24 Hours of Le Mans

All | Images | News | Shopping | Videos | More | Settings | Tools

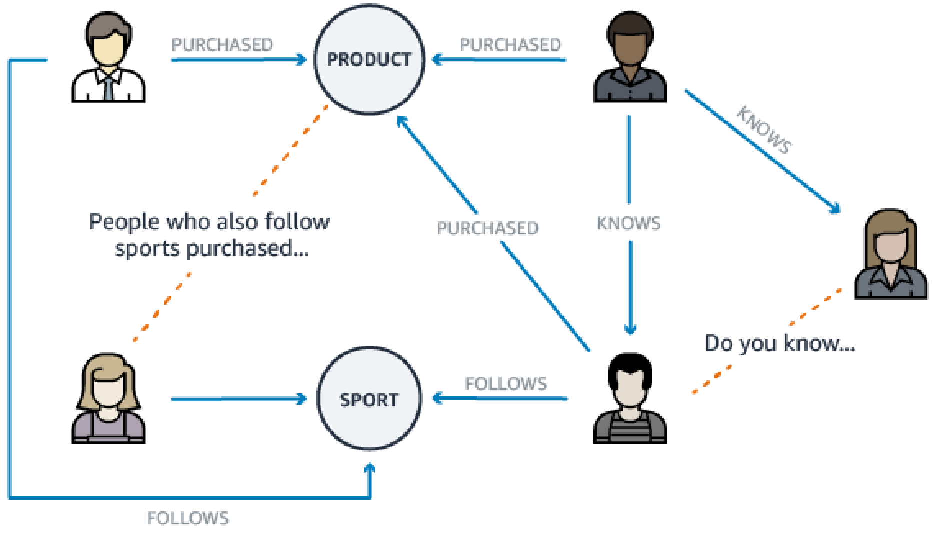
About 34,600,000 results (1.04 seconds)

787B

(?car, produced by, Mazda)
(?car, won, 24 Hours of Le Mans)

Understanding Relations Is Prominent In Practice

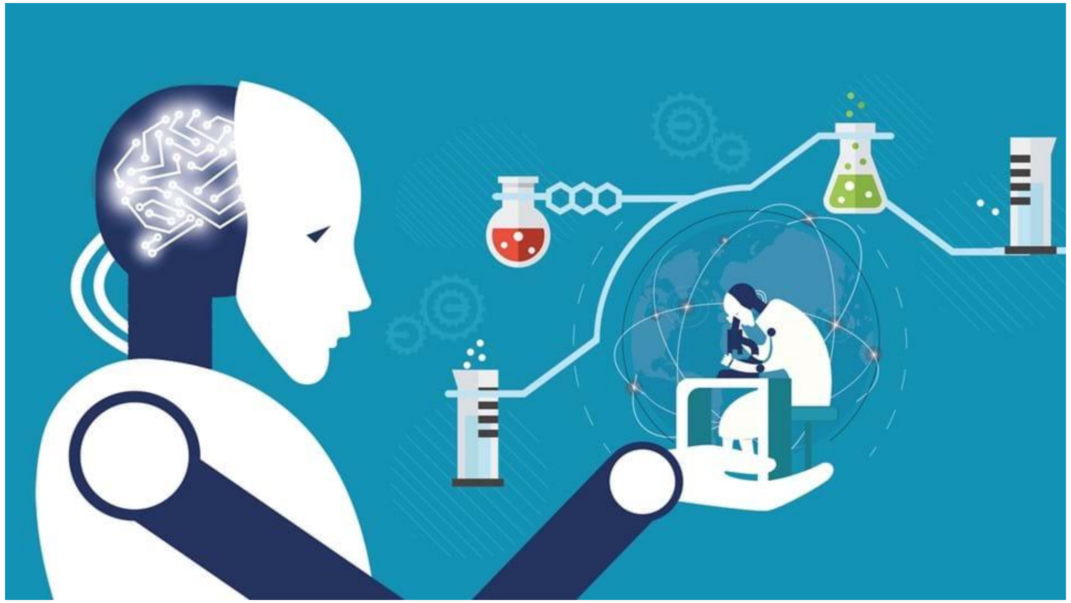
E-Commerce



Co-purchase relations of products

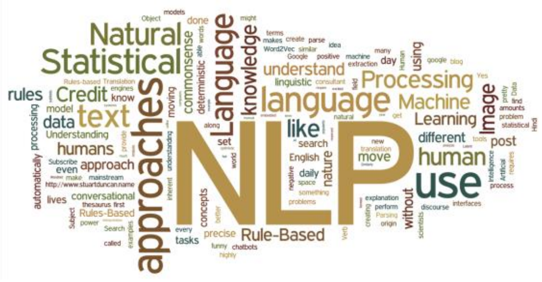
Social relations of users

Computational Biology Research



Interactions of molecules and biomolecules.

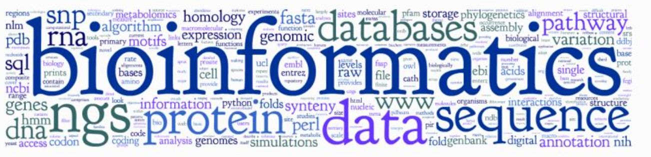
Understanding Relations Is Prominent In Practice



- QA
- Discourse relation detection
- Dialogue state tracking
- Event prediction
- Narrative cloze
- Entity/event typing and linking

Semantic Web

- Semantic search
- Relational rule mining
- Ontology population
- Ontology matching and knowledge integration



Medical INFORMATICS

- Interaction prediction of biomolecules
- Mutation effect estimation
- Non-coding RNA alignment
- Drug discovery
- Polypharmacy side effect detection

Multi-relational Data

Knowledge Graphs

Baidu知识图谱

Google KNOWLEDGE GRAPH

DBpedia

WIKIDATA

yago select knowledge

NELL

Freebase

Bio-med Ontologies /Data Banks

MeSH

SKEMPI

neXtprot

STRING

PubMed

GENEONTOLOGY Unifying Biology

RCSB PDB PROTEIN DATA BANK

Product & E-Commerse Graphs

product graph

ebay

Alibaba.com

Common Sense and Semantic Graphs

WORDNET

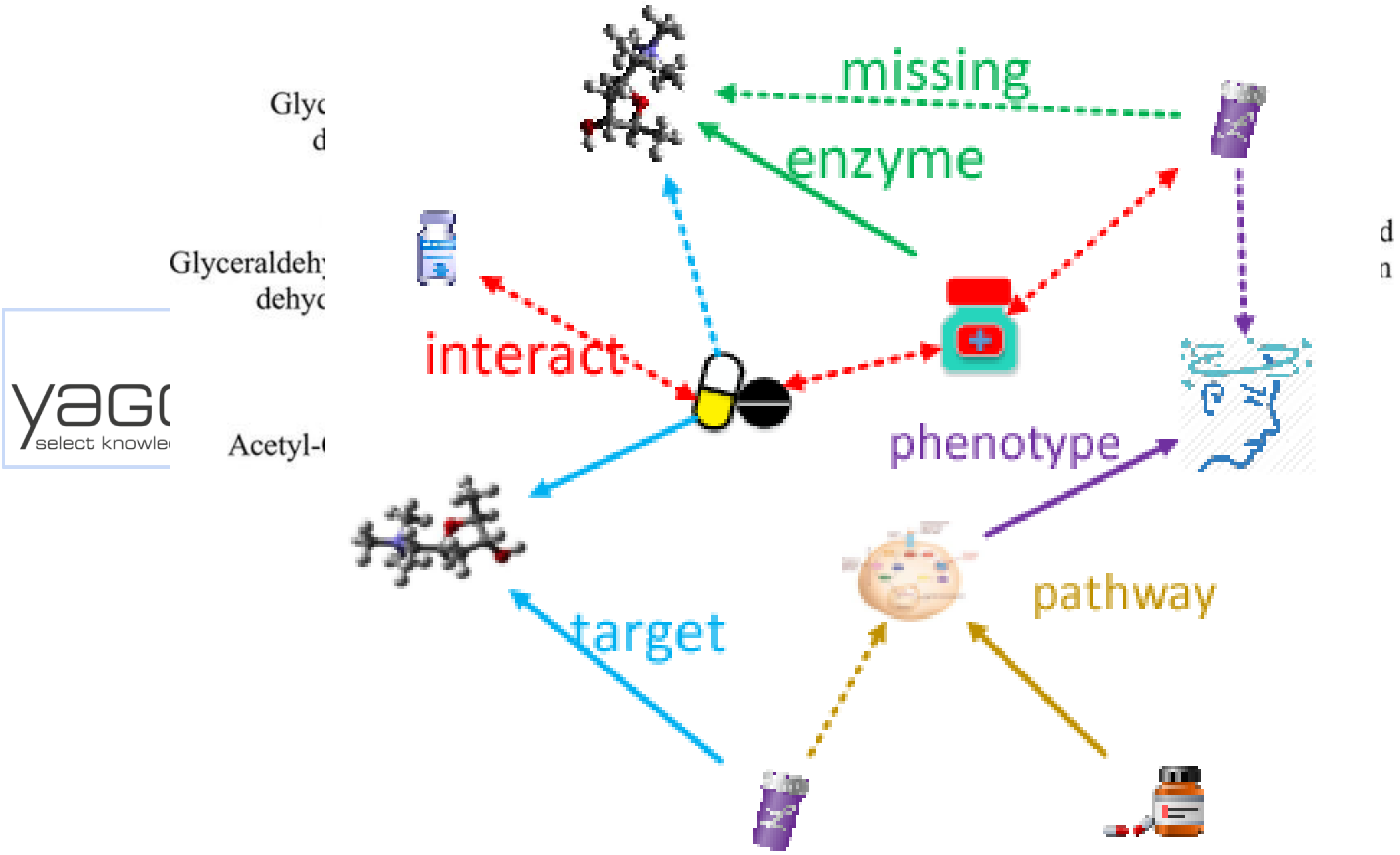
ConceptNet An open, multilingual knowledge graph

FrameNet

VERBNET

Wiktionary The free dictionary

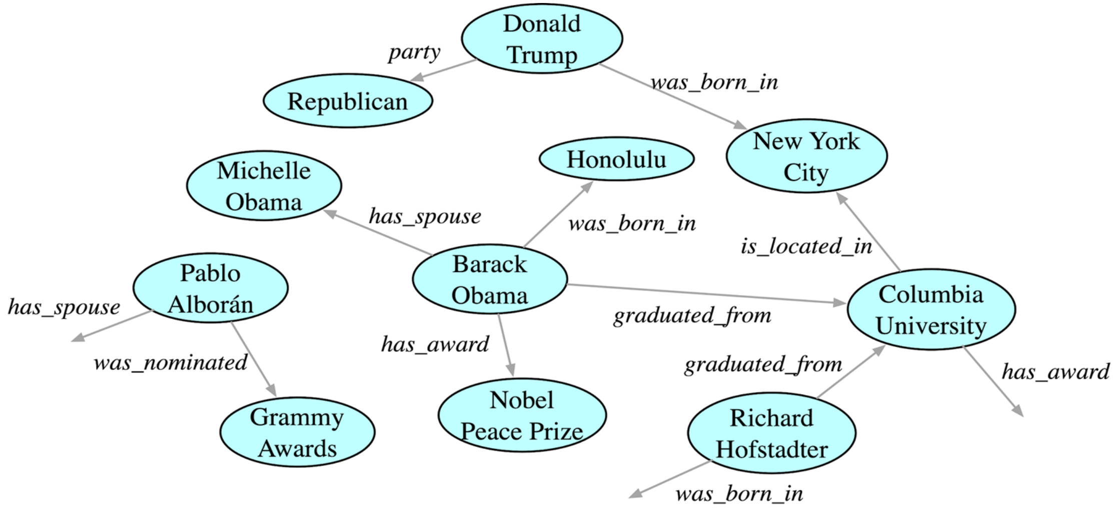
Multi-relational Data



Multi-relational Data

A multi-relational dataset is formally defined as an edge-typed graph ***G***

- ***E***: the vocabulary of nodes (representing entities, objects or concepts)
- ***R***: the vocabulary of relations
- $T=(h, r, t) \in \mathbf{G}$ s.t. $h, t \in E$ and $r \in R$: a **triple** representing the fact of a relation *r* between two entities *h* and *t*



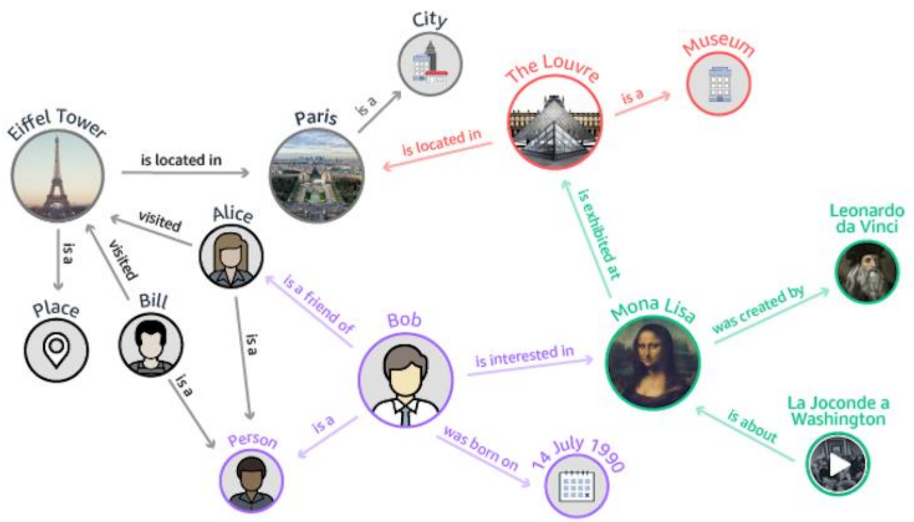
G

Produced By

787B

$T=(h, r, t)$

Why **Representation Learning?**



Downsides of symbolic knowledge representations

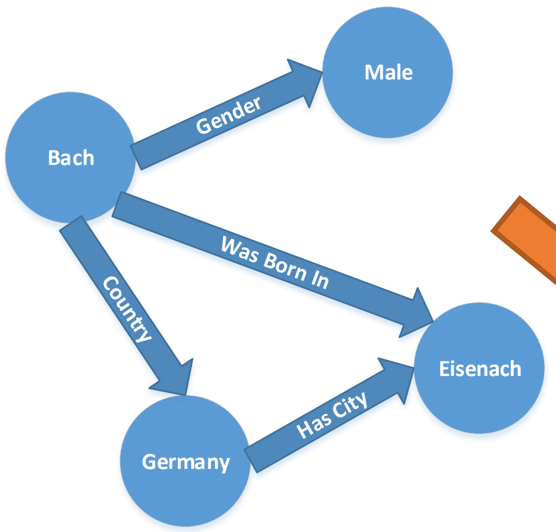
- The data are usually **sparse**
- Not easily supporting machine **inference**

A plausible representation should

- Be **quantifiable**
- Support the inference of **missing knowledge**

Why Representation Learning?

Latent representations/embeddings are more inferable



Multi-relational data

Encode

Paris (0.036, -0.12, ..., 0.323)
capital (0.102, 0.671, …, -0.101)
France (0.138, 0.551, …, 0.222)
…

Embeddings

Capture

Similarity of entities

- Mistaken ≈ Wrong
- Feline ≈ Cat
- Los Angeles ≈ Hollywood
- …

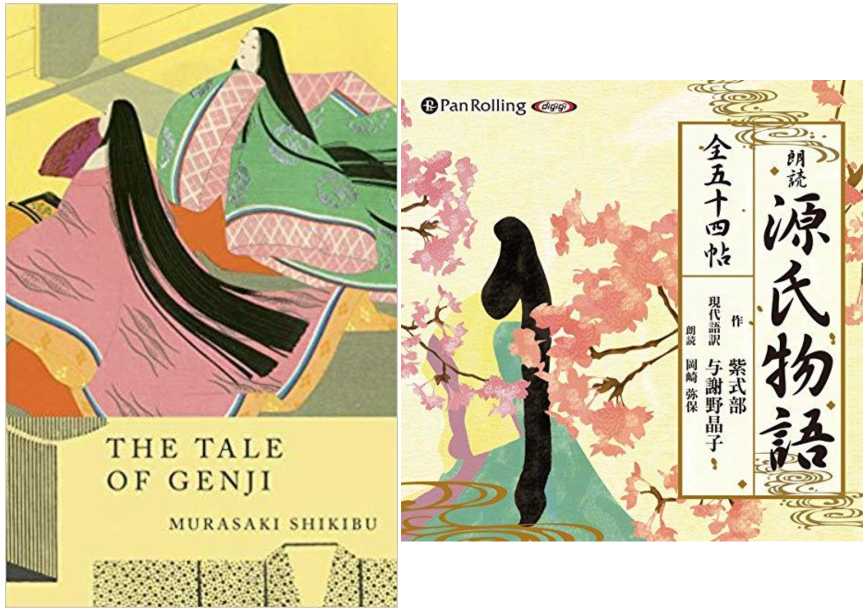
Enable

Relational inference as vector algebra (e.g. a translation)

- France – Paris≈ capital
- USD – US ≈ currency
- Bach – German ≈ nationality
- …

Why **Transferable** Representation Learning?

Different data can possess complementary knowledge



(The Tale of Genji, *Genre*, **?e**)

DBpedia

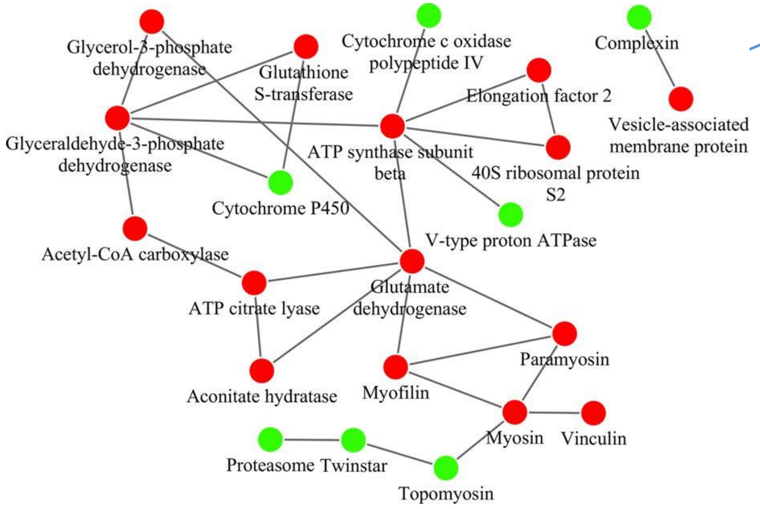
Novel

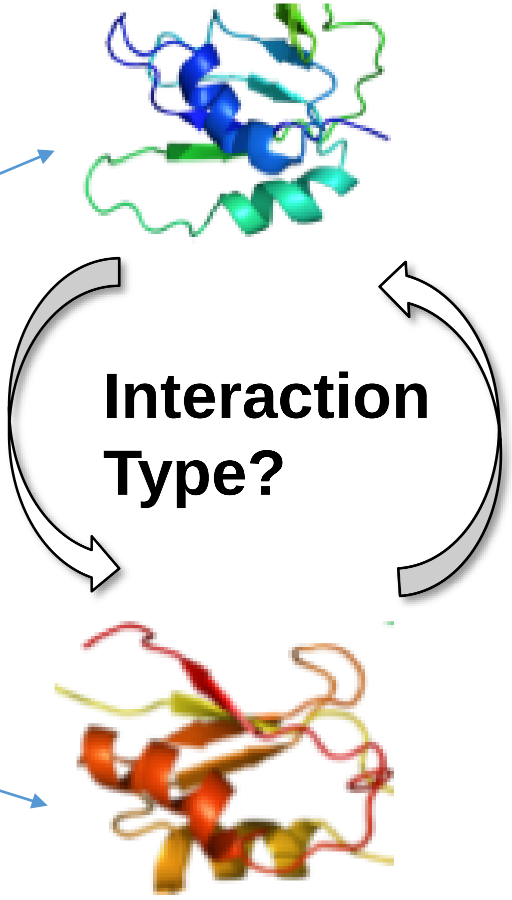
ヂビペディア

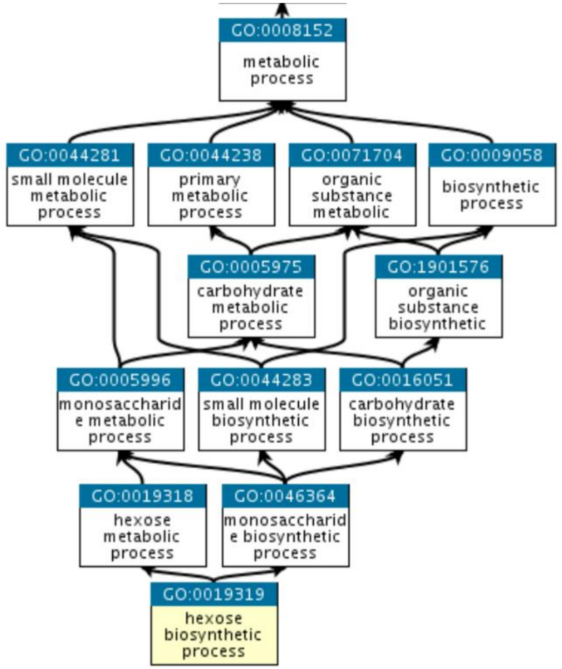
Monogatari (story)
Love story
Royal family story
Realistic novel
Ancient literature

Why **Transferable** Representation Learning

Different data can possess complementary knowledge







GENEONTOLOGY
Unifying Biology

Inferring missing knowledge for
a proteomic data bank

Key Research Questions

Interchangeable knowledge in many scenarios

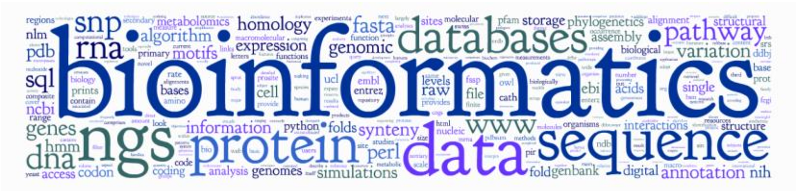
- ☐ Multiple language-specific KGs
- ☐ Multiple knowledge bases
- ☐ Instance KGs and concept ontologies
- ☐ Protein-protein interaction (PPI) data, gene ontologies and cell clusters
- ☐ Drug-drug interaction data, disease ontologies and PPI data
- ☐ Social networks and product graphs
- ☐ …

1. How to capture the association of knowledge with representation learning?
2. How to leverage knowledge transfer to populate missing knowledge?

A General Methodology to Benefit A Wide Range of Tasks

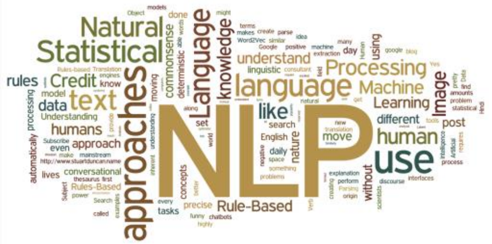
Knowledge Base

Knowledge alignment
Mono/Cross-lingual KG completion
Ontology population
Zero-shot entity matching



Protein-protein interaction prediction
Protein binding affinity estimation
Single cell RNA-sequence imputation
Gene Ontology term assignment

Transferable representation learning



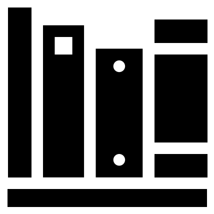
Semantic search
Entity typing
Paraphrase identification
Sub-article relation extraction

Medical INFORMATICS

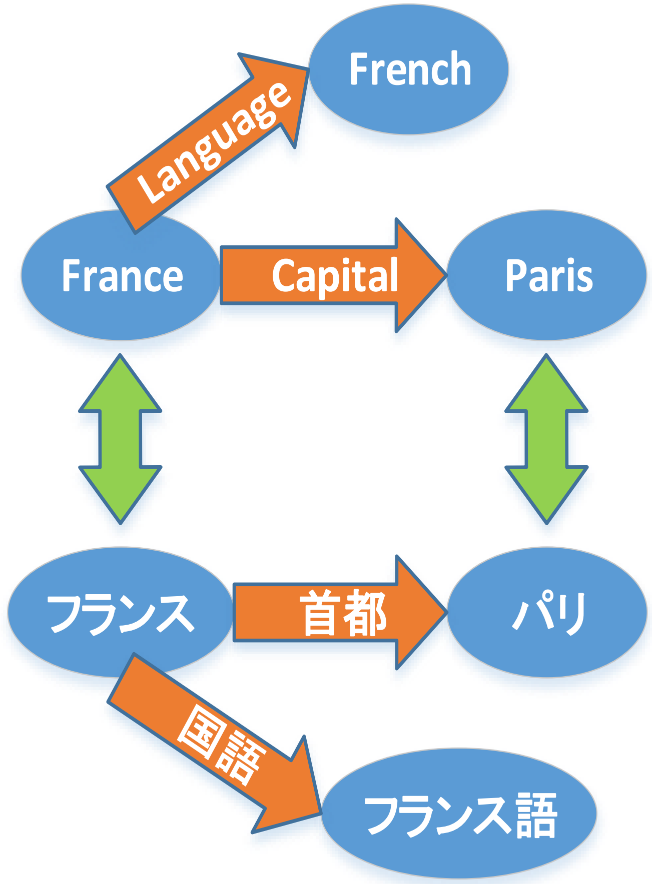
Polypharmacy side effect detection
Disease and phenotype matching
Clinical event prediction

Challenges

- **Limited supervision for knowledge association**
- Auxiliary supervision from alternative information (attributes, descriptions, schemata, etc.)
- **Heterogeneous forms of knowledge association (1-to-1, multi-granular, fuzzy alignment, etc.)**
- Inconsistent structures and different scales of data



Multilingual KGs: An Exemplary Scenario



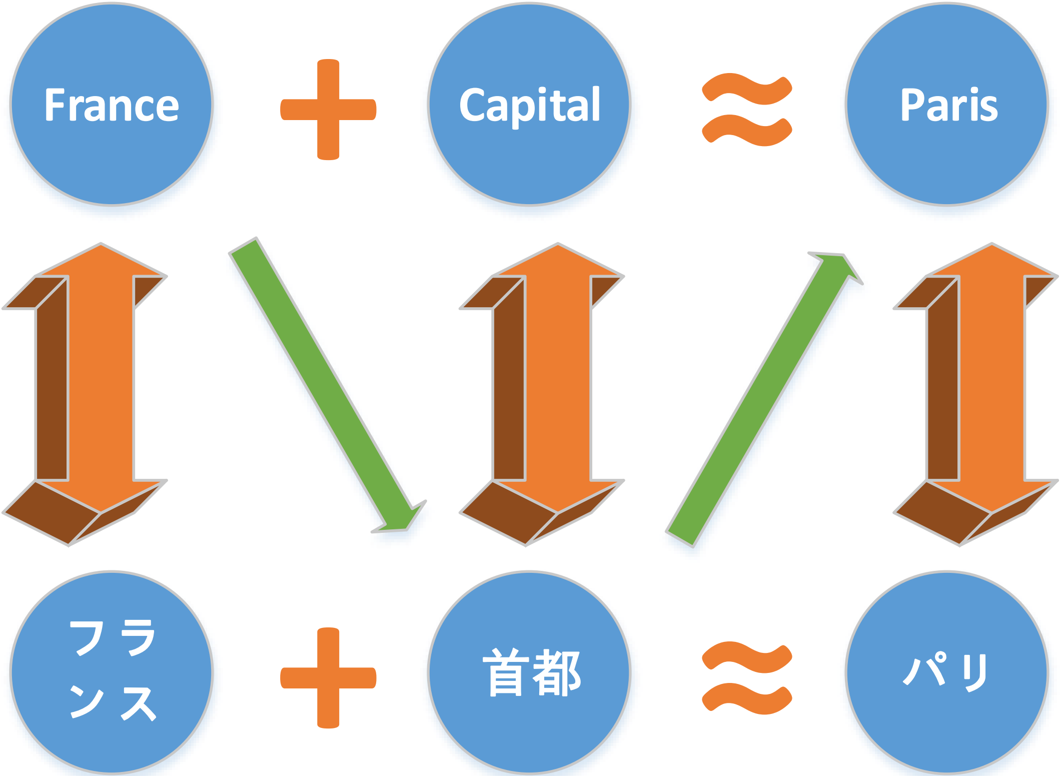
Separately managed language-specific KGs

- DBpedia has **125 languages** ; ConceptNet has **10 core languages**

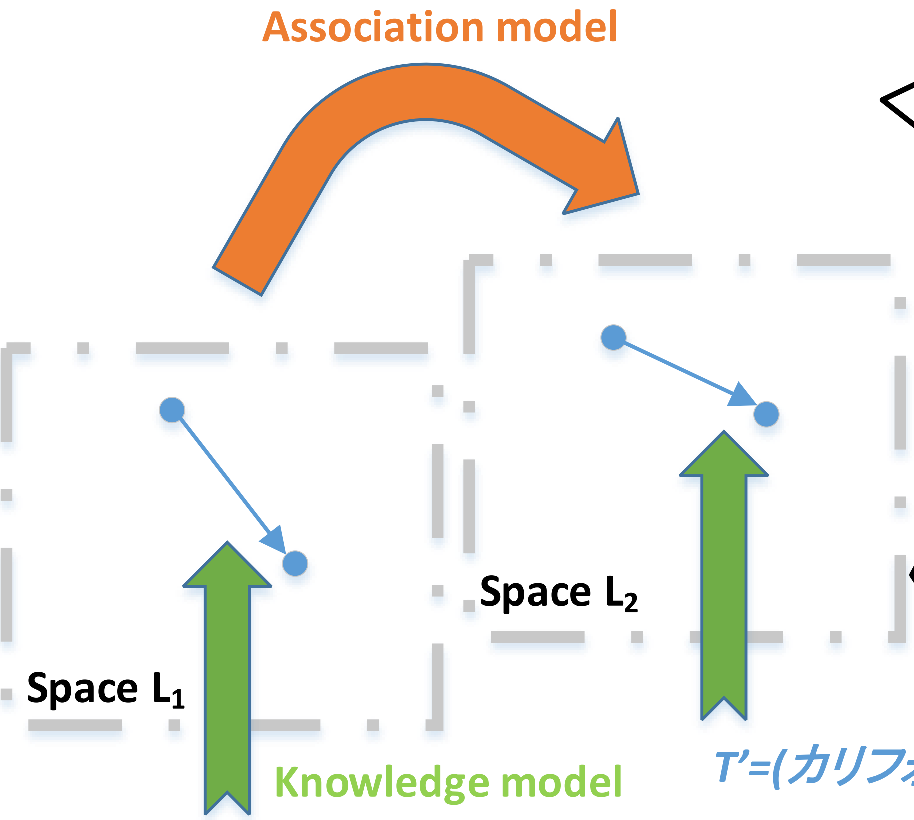
A Pilot Study: Simple Translational Model + Supervised Association Learning (MTransE*)

*[Chen+ IJCAI-17]

- Training data: multiple language-specific KGs + seed entity alignment
- Enabling: cross-lingual semantic transfer + monolingual relational inferences



Joint Learning of MTransE



T'=(カリフォルニア, 州都, サクラメント)

T=(California, capital, Sacramento)

Association model: an embedding transformation learned with seed alignment

$$S_A = \sum_{(e,e')\in\delta(L_i,L_j)} \left\| \mathbf{M}_{ij}\mathbf{e} - \mathbf{e}' \right\|$$

Knowledge model: encoding entities and relations of each language as a **translational embedding**

$$S_K = \sum_{L\in\{L_i,L_j\}} \sum_{(h,r,t)\in G_L \wedge (\hat{h},r,\hat{t})\notin G_L} \left[f_r(h,t) - f_r(\hat{h},\hat{t}) + \gamma \right]_+$$

$$\text{s.t. } f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2$$

- **Joint training loss**

$$S_J = S_K + \alpha S_A$$

*[Chen+,IJCAI-17]

Application: Knowledge Alignment

Table 8: Examples of cross-lingual entity matching.

Entity	Target	Candidates (in ascending order of rank by Euclidean distance)
Barack Obama	French	**Barack Obama**, *George Bush*, *Jimmy Carter*, George Kalkoa
	German	**Barack Obama**, *Bill Clinton*, *George h. w. Bush*, Hamid Karzai
Paris	French	**Paris**, *Amsterdam*, **à Paris**, *Manchester*, De Smet
	German	**Paris**, *Languedoc*, *Constantine*, *Saint-maurice*, *Nancy*
California	French	*San Francisco*, *Los Angeles*, *Santa Monica*, **Californie**
	German	**Kalifornien**, *Los Angeles*, *Palm Springs*, *Santa Monica*

Table 9: Examples of cross-lingual relation matching.

Relation	Target	Candidates (in ascending order of rank by Euclidean distance)
capital	French	**capitale**, *territoire*, pays accrèditant, lieu de veneration
	German	**hauptstadt**, *hauptort*, *gründungsort*, *city*
nationality	French	**nationalié**, **pays de naissance**, *domicile*, *résidence*
	German	**nationalität**, **nation**, letzter start, *sterbeort*
language	French	**langue**, réalisations, lieu deces, *nationalitè*
	German	**sprache**, **originalsprache**, **lang**, *land*

Bold-faced ones are correct answers, *italic* ones are close answers.
Answers do not include those that have pre-existed in training.

This pilot study got ~30% Hits@1 on DBP15k. But we will introduce lots of improvement to it shortly.

Discovering Cross-lingual Relation Facts, e.g.

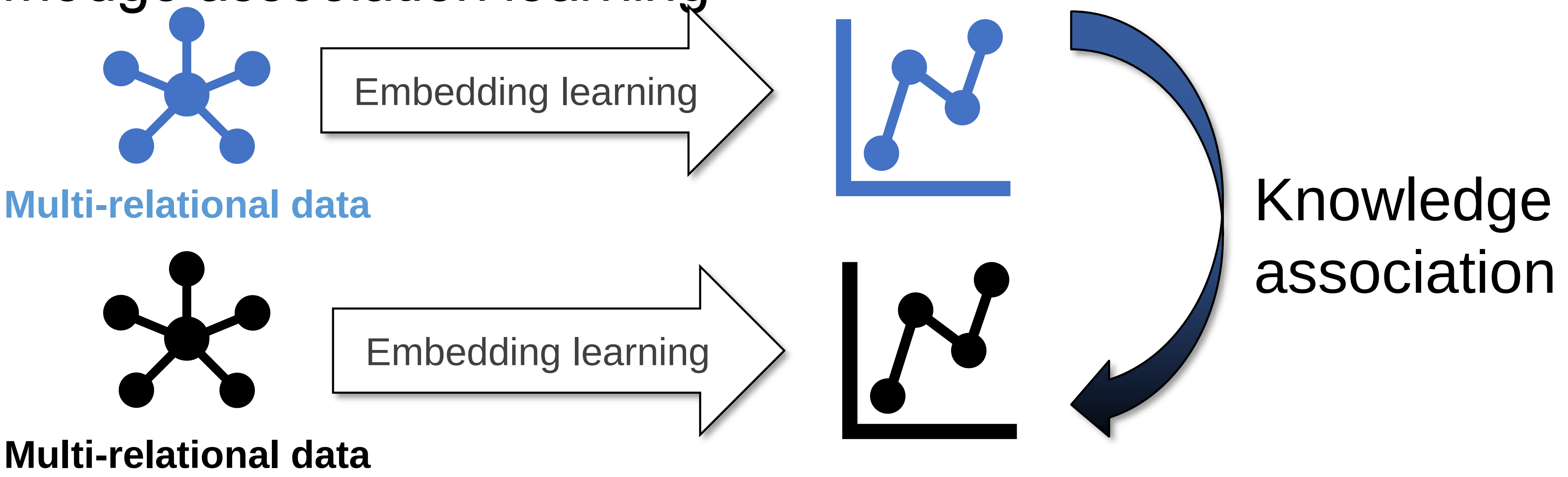
Table 10: Examples of cross-lingual triple completion.

Query	Target	Candidates (in ascending order of rank)
(Adam Lambert, genre, ?t)	French	*musique indèpendante*, **musique alternative**, ode, **glam rock**
	German	**popmusik**, **dance-pop**, no wave, *soul*
(Ronaldinho, position, ?t)	French	**milieu offensif**, **attaquant**, *quarterback*, *latèral gauche*
	German	**stürmer**, *linker flügel*, **angriffsspieler**, *rechter flgel*
(Italy, ?r, Rome)	French	**capitale**, **plus grande ville**, **chef-lieu**, garnison
	German	**hauptstadt**, **hauptort**, verwaltungssitz, stadion
(Barack Obama, ?r, George Bush)	French	*ministre-prèsident*, **prèdècesseur**, *premier ministre*, *prèsident du conseil*
	German	**vorgänger**, **vorgängerin**, besetzung, lied

Bold-faced ones are correct answers, *italic* ones are close answers.
Answers do not include those that have pre-existed in training.

General Methodology and Further Improvement

Jointly or iteratively conduct two learning processes: **embedding learning** and **knowledge association learning**



Three directions to improvement

1. Better embedding learning techniques for inconsistent structures
2. Knowledge association learning under minimal supervision
3. Auxiliary supervision from entity profile information

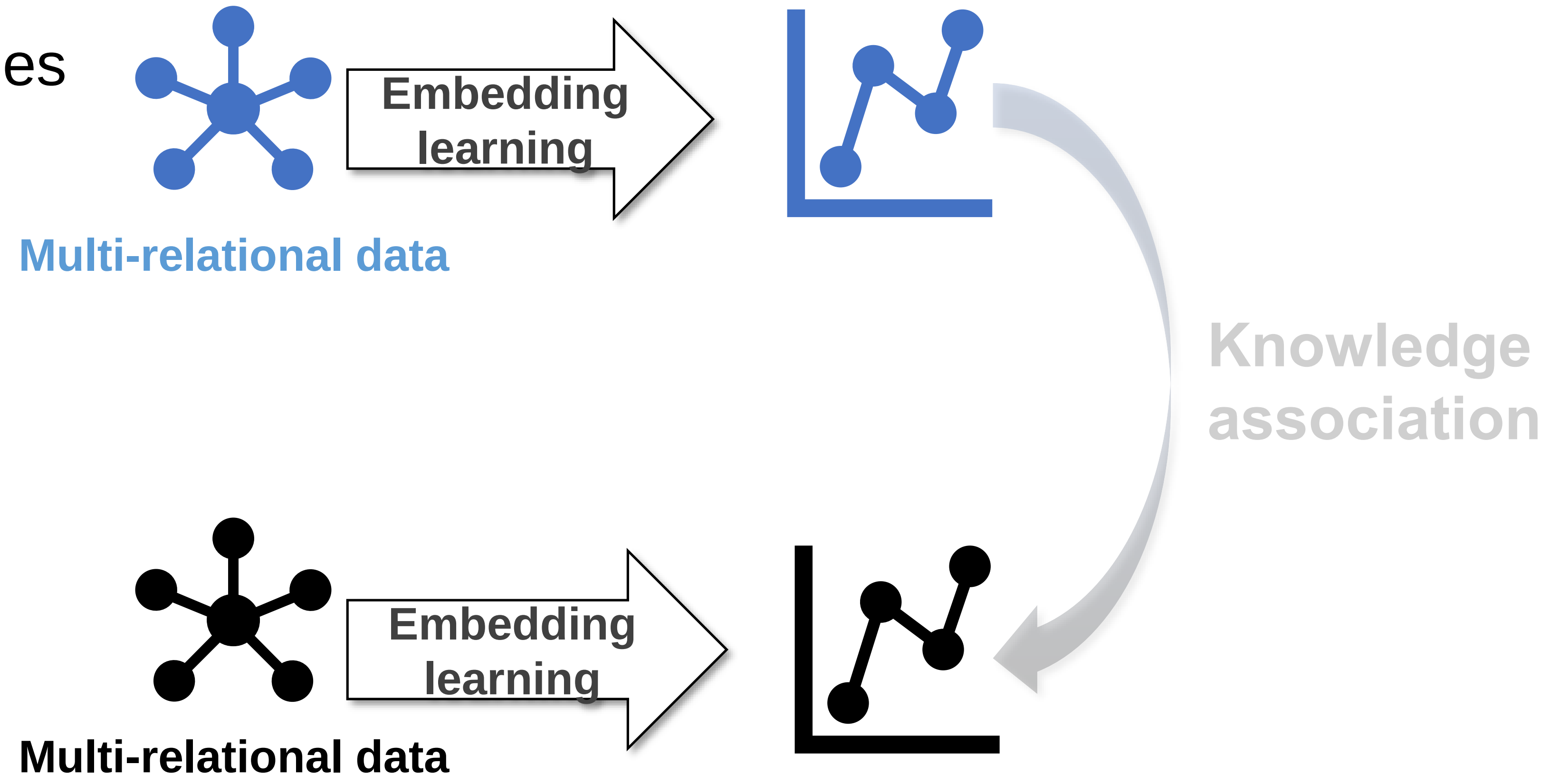
Outline

- Background and Motivation
- **Relational embedding learning methods**
  - First-order and high-order methods
  - Non-Euclidean methods
- Knowledge Association Methods
  - Supervised and semi-supervised methods
  - Auxiliary supervision methods
- Cross-domain and interdisciplinary tasks
  - KBP tasks
  - Computational bio-med tasks

The Embedding Learning Process

- Distributing each domain-specific multi-relational dataset in a separate embedding space.

- Three categories of techniques
  - ☐ First-order methods
  - ☐ High-order methods
  - ☐ Non-Euclidean methods



First-order Methods

- A function $f_r(h, t)$ locally measures the plausibility of each triple *T*=(*h*, *r*, *t*)

Correct Triple UCLA *Located In* Los Angeles → High plausibility score

Training cases $T \in G$

Incorrect Triple UCLA *Located In* Philadelphia

Incorrect Triple Penn *Located In* Los Angeles

→ Low plausibility scores

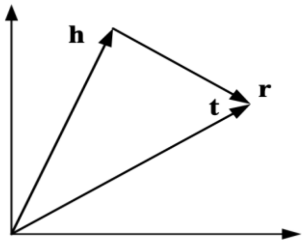
Negative samples $T' \notin G$

Plausibility Scoring Functions

- Translational technique [Bordes+, NIPS-13]

  $$f_r(h,t) = -\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_2$$

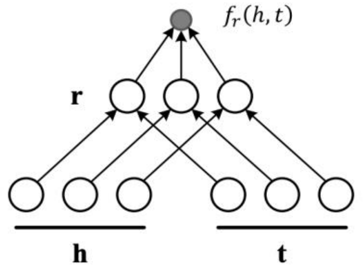
  - **France** + ***Capital*** ≈ **Paris**



- Element-wise product [Yang+, ICLR-15]

  $$f_r(h,t) = (\mathbf{h} \circ \mathbf{t}) \cdot \mathbf{r}$$

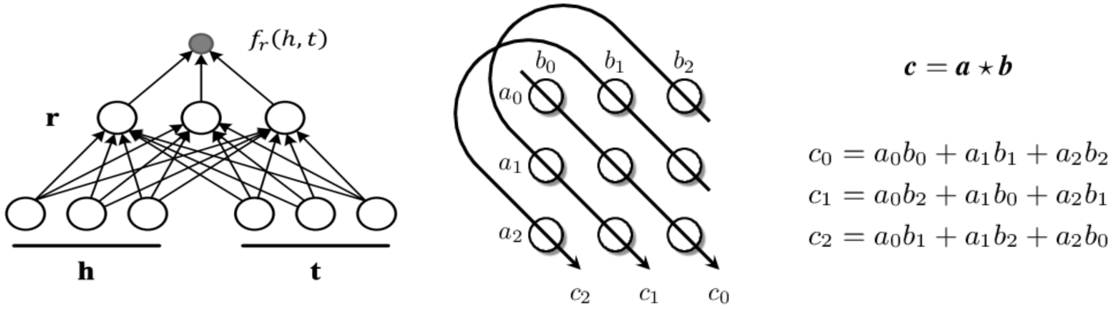
  - $\circ$ denotes element-wise product



- Circular correlation [Nickel+, AAAI-16]

  $$f_r(h,t) = (\mathbf{h} \star \mathbf{t}) \cdot \mathbf{r}$$

  $$[\mathbf{h} \star \mathbf{t}]_d = \sum_{i=0}^{k} \mathbf{h}_i\, \mathbf{t}_{(\mathrm{d+i})\ \mathrm{mod}\ k}$$



First-order Methods

- Learning objective
  - Marginal ranking loss

Expecting a negative sample to be scored less than a positive sample by at least γ.

$$L_K = \sum_{T \in G \wedge T' \notin G} \boxed{\max(0, \gamma + f_r(h', t') - f_r(h, t))}$$

- $f_r(h, t)$: plausibility function, the higher indicates a more plausible triple
- γ: a positive margin
- *T'*=(*h'*,r,*t'*): a negative sample created by corrupting either *h* or *t* in a positive case *T=(h,r,t)*

  - Limit-based loss ($\gamma_1 > \gamma_2$)

$$L_K = \sum_{T \in G} \max(0, f_r(h, t) - \gamma_1) + \sum_{T' \notin G} \max\left(0, \gamma_2 - f_r(h', t')\right)$$

  - Log softmax loss

$$L_K = \sum_{T \in G} \log \frac{\exp(f_r(h, t))}{\sum_{T' \notin G} \exp(f_r(h', t'))}$$

Pros and Cons of First-order Methods

- Pros
  - Low parameter complexity
  - Facilitates inference of relations
  - Robust against data sparsity

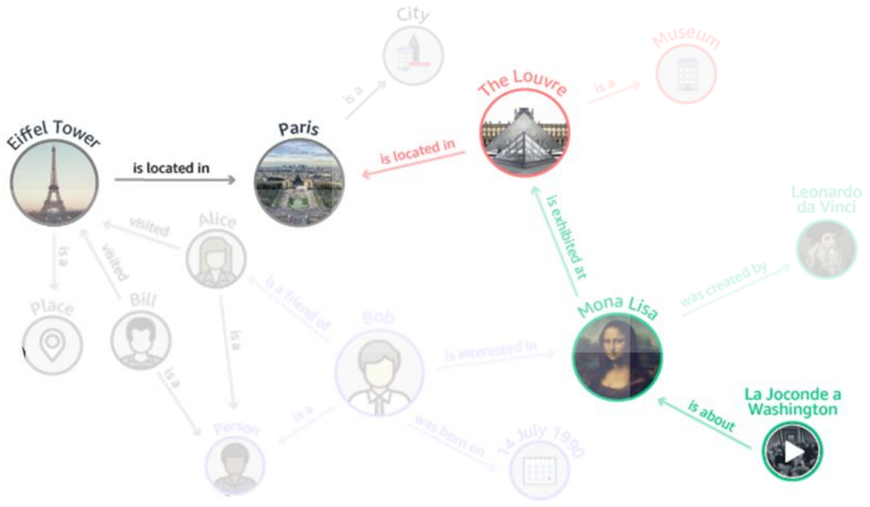
- Cons
  - Less precise modeling of node proximity (may hinder knowledge association)
  - Less robust against structural heterogeneity

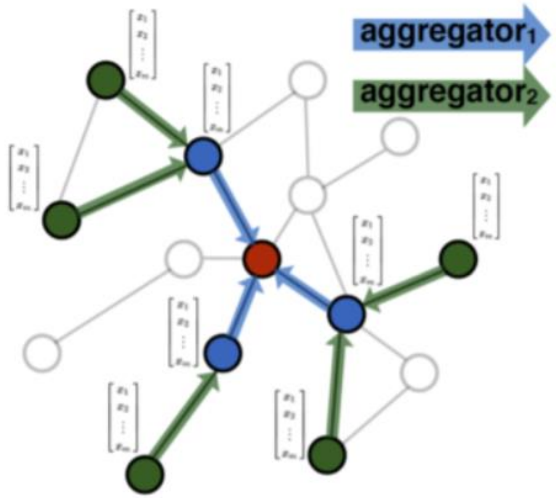
- Transferable representation learning models with first-order methods
  - **For KG alignment/entity resolution**: MTransE [Chen+, IJCAI-17], JAPE [Sun+, ISWC-17], LIN [Otani+ COLING-18], BootEA [Sun+, IJCAI-18], KDCoE [Chen+, IJCAI-18], AttrE [Trsedya, AAAI-19], MultiKE [Zhang+, IJCAI-19], OTEA [Pei+, IJCAI-19], SEA [Pei+, WWW-19]
  - **For entity typing**: JOIE [Hao+, KDD-19]

High-order Methods

- Modeling nodes (objects) based on contexts of the graph

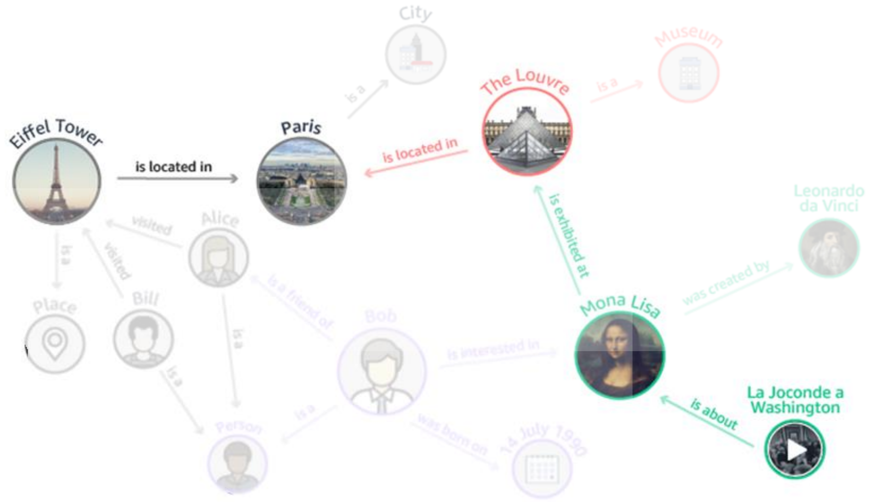
- Two types of context modeling techniques
  - Relation path based techniques
  - Neighborhood aggregation techniques (GNNs)





Relation Path Based Techniques

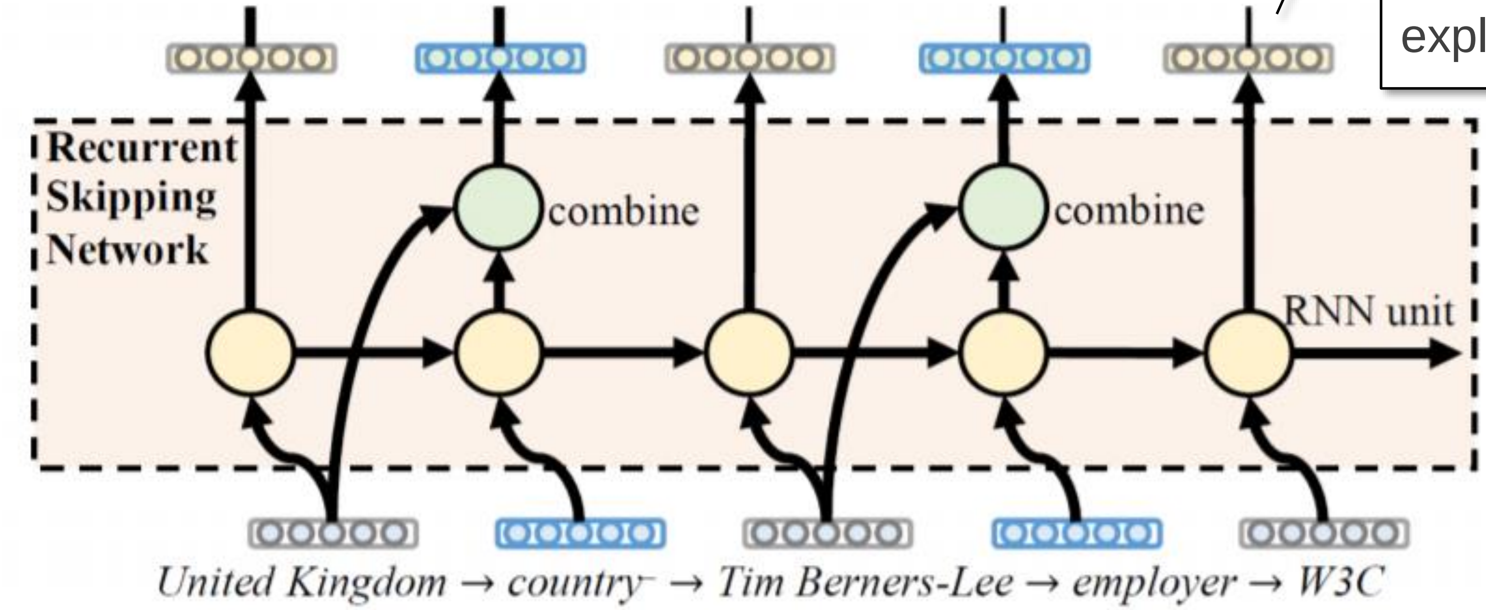
- A relation path is an entity-relation chain, where entities and relations appear alternately
  - United Kingdom → *country* → Tim Berners-Lee → *employer* → W3C
- PTransE [Lin+, EMNLP-15], Bi-Diag [Guu+, EMNLP-15]
  - Given l-length relation paths $\mathbf{p} = (e_0, r_1, e_1, \dots, r_l, e_l)$
    - Minimize $||\mathbf{e}_0 - \mathbf{p} + \mathbf{e}_l||_2$
  - Multiple representations of p
    - Addition (PTransE): $\mathbf{p} = \sum_{i=1}^{l} \mathbf{r}_i$
    - Multiplication (Bi-Diag): $\mathbf{p} = \prod_{i=1}^{l} \mathbf{r}_i$
    - RNN-aggregation (PTransE)
- Path selection
  - Random walk (Bi-Diag)
  - All 3-hop paths (PTransE)



Relation Path Based Techniques

- Recurrent skipping network (RSN, Guo+, ICML-19)
  - RNNs perform well on sequential data, but overlooks the basic structure units of triples in a relation path
  - Tri-gram residual mechanism: **shortcut** a **subject entity** to let it **directly** participate in predicting its **object entity**

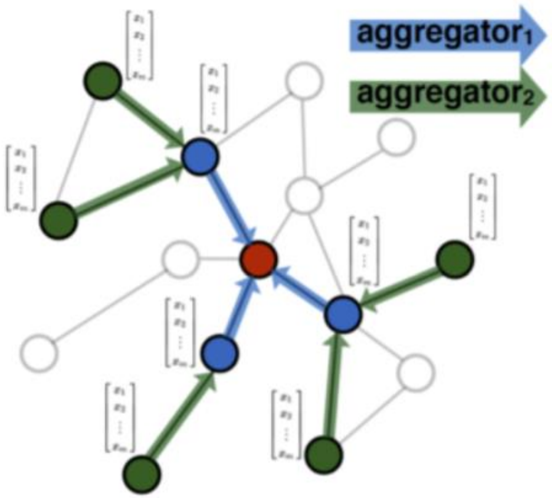
$$\mathbf{h}'_t = \begin{cases} \mathbf{h}_t & x_t \in \mathcal{E} \\ \mathbf{S}_1\mathbf{h}_t + \mathbf{S}_2\mathbf{x}_{t-1} & x_t \in \mathcal{R} \end{cases},$$

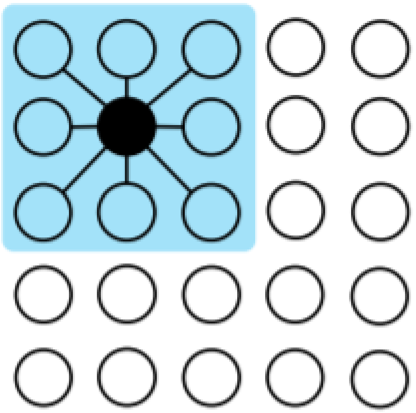


Learning objective: knowledge association (to be explained in the next section)

Neighborhood Aggregation Techniques

- Characterizing an entity based on its neighborhood
- Graph convolutional networks (GCN)
  - Aggregate neighbor information and pass into a neural network.
  - Can be viewed as a center-surround convolution kernel in a CNN





Neighborhood Aggregation Techniques

- GCN representation for an entity *e*

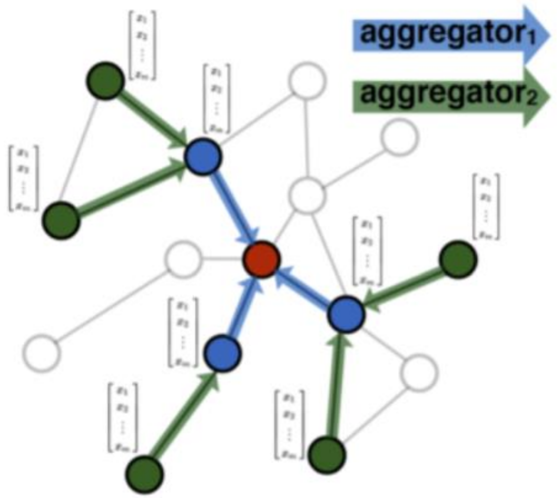
$$\boxed{\mathbf{h}_e^l} = \phi(\boxed{\mathbf{M}^l} \sum_{\boxed{e' \in N(e) \cup e}} \frac{\boxed{\mathbf{h}_{e'}^{l-1}}}{\sqrt{|N(e)||N(e')|}})$$

Trainable convolution kernel

The (l-1)-th layer representation of each neighborhood entity *e'*

The l-th layer representation stacked on *e*

Direct neighborhood of *e*



- GCN produces a trainable representation for the neighborhood of each *e*

Learning objective: knowledge association (to be explained in the next section)

Neighborhood Aggregation Techniques

- How do we consider relations in GCN?
  - R-GCN [Schlichtkrull+, ESWC-18; Wu+, IJCAI-19]: relation-specific convolution kernels

$$\mathbf{h}_e^l = \phi\Big(\sum_{e' \in N(e)} \boxed{\mathbf{M}_r^l} \frac{\mathbf{h}_{e'}^{l-1}}{\sqrt{|N(e)||N(e')|}} + \mathbf{M}^l \frac{\mathbf{h}_e^{l-1}}{\sqrt{|N(e)||N(e')|}}\Big)$$

Relation-specific kernel for each neighboring entity

- Other variants of GNN
  - Graph attention network (GAT, Zhu+, IJCAI-19)
  - Multi-channel GNN [Cao+, ACL-19]
  - Gated Multi-hop GNN [Sun+, AAAI-20]

Pros and Cons of high-order methods

- Pros
  - Better capturing entity proximity (benefitting knowledge association)
  - Robust against structural heterogeneity

- Cons
  - Much higher parameter complexity
  - May not directly support inference of relations
  - Less robust against data sparsity

- Transferable representation learning methods with high-order methods
  - Relation path based: IPTransE [Zhu+, IJCAI-17], RSN [Guo+, ICML-19]
  - GNN-based: GCN-Align [Wang+, EMNLP-18], MuGCN [Cao+, ACL-19], NAEA [Zhu+, IJCAI-19], KECG [Li+, EMNLP-19], HMAN [Yang+, EMNLP-19], MMR [Shi+, EMNLP-19], HGCN [Wu+, EMNLP-19]

Non-Euclidean Methods

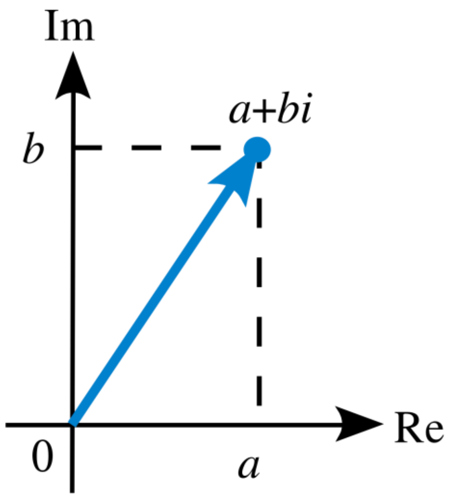
- Complex space relational embeddings
  - ComplEx [Trouillon+, ICML-16] ($Re(.)$ denotes the real part)

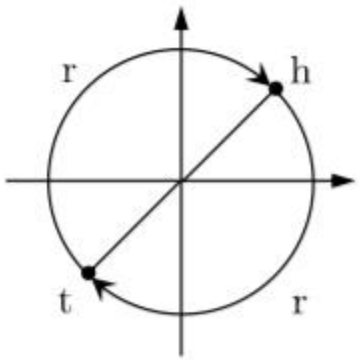
$$f_r(h,t) = Re((\mathbf{h} \circ \mathbf{t}) \cdot \mathbf{r})) \text{ s.t. } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$$

    - Suitable for capturing symmetric and antisymmetric relations
  - RotatE [Sun+, ICLR-19]

$$f_r(h,t) = -\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \text{ s.t. } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$$

    - Suitable for more **relation patterns**: symmetry/anti-symmetry, inversion, and composition





(c) RotatE: an example of modeling symmetric relations $\mathbf{r}$ with $r_i = -1$

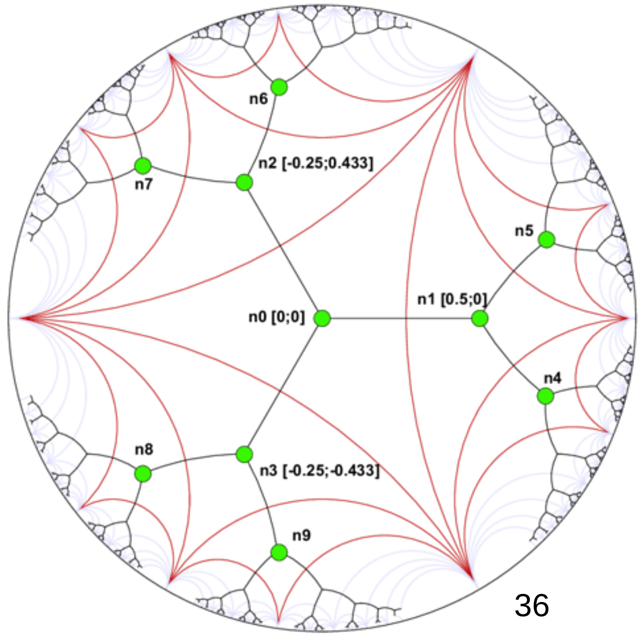
Non-Euclidean Methods

- The **hyperbolic space**: the amount of space has an **exponential growth** w.r.t. the radius [Nickel+ NIPS-17, Ganea+ NeurIPS-18, Liu+ NeurIPS-19]

$$d_{\mathbb{D}}(\mathbf{u},\mathbf{v}) = \operatorname{arccosh}\left(1 + 2\frac{\|\mathbf{u}-\mathbf{v}\|^2}{(1-\|\mathbf{u}\|^2)(1-\|\mathbf{v}\|^2)}\right).$$

- Many data form hierarchies
  - Ontologies, taxonomies syntax trees, org charts, claim provenance in social media, etc.

- Hyperbolic representation learning models:
  - Graph embeddings [Nickel+ NIPS-17, Graph NN [Liu+ NeurIPS-19]

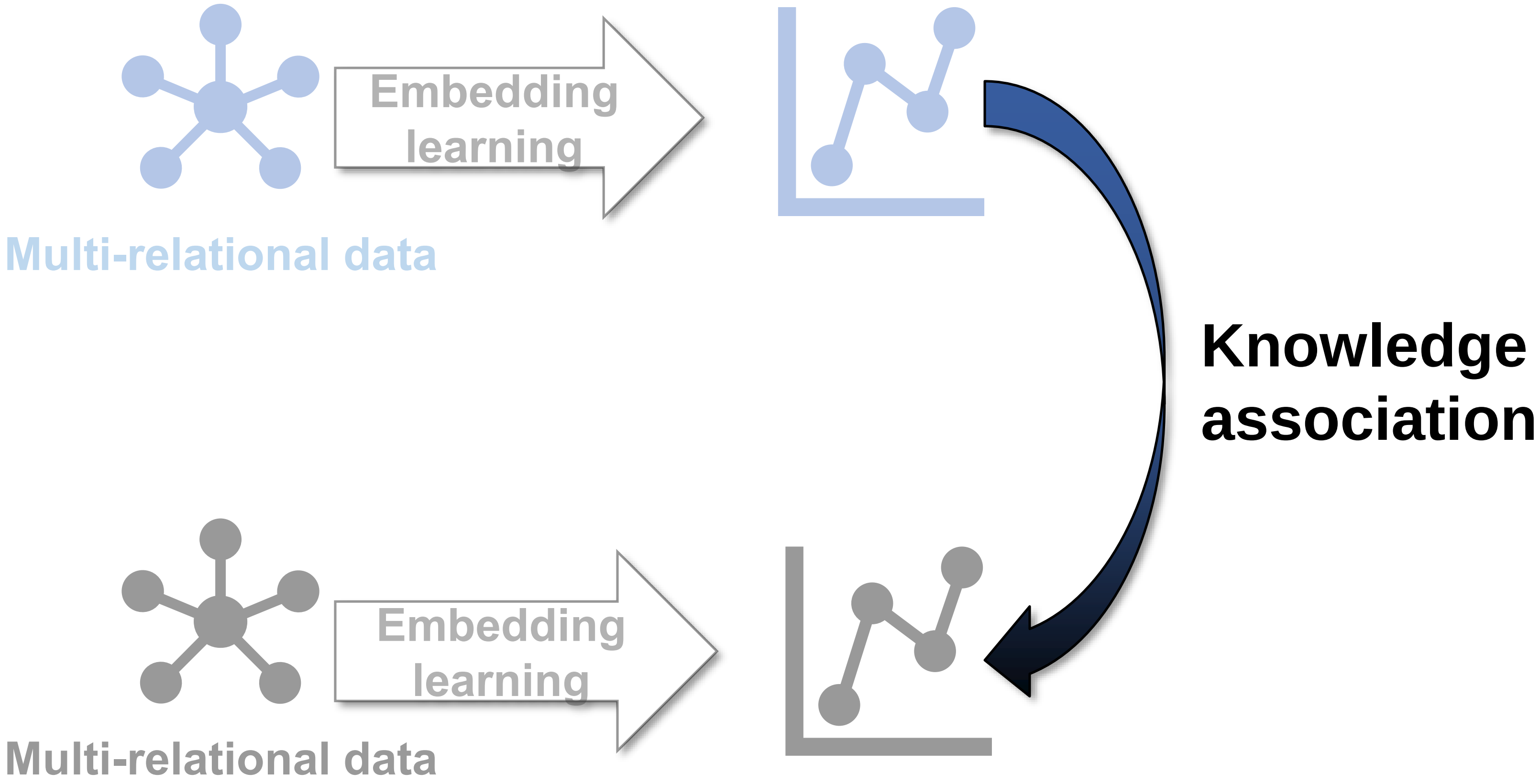


Outline

- Background and Motivation
- Embedding learning methods
  - ☐ First-order and high-order methods
  - ☐ Non-Euclidean methods
- **Knowledge Association Methods**
  - ☐ Supervised and semi-supervised methods
  - ☐ Auxiliary supervision methods
- Cross-domain and interdisciplinary tasks
  - ☐ KBP tasks
  - ☐ Computational bio-med tasks

The Knowledge Association Process

- Capturing the correspondence of objects between the embedding representations of two multi-relational datasets (say G_1 and G_2)



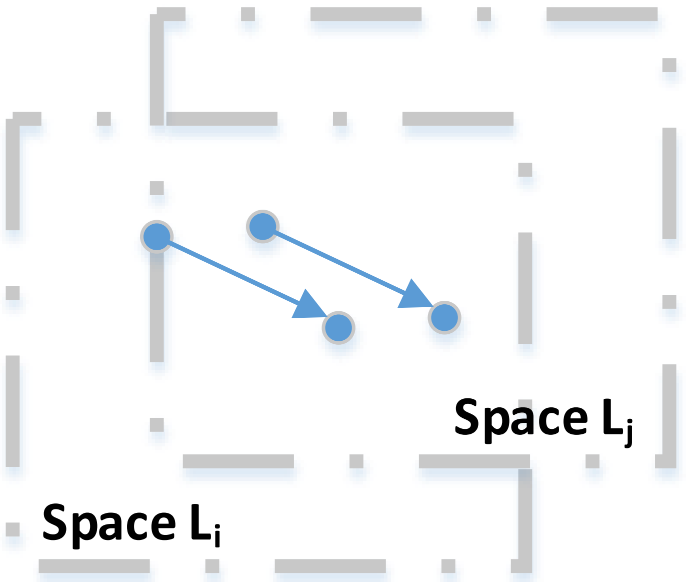
Several Key Questions

- What geometric representations should be used to capture the knowledge association?
  - Connecting same or different sizes of embeddings
  - Types of associations (1-to-1, multi-granular, fuzzy)

- What learning strategies should be used under scenarios with limited supervision?
  - Semi-supervised learning?
  - Auxiliary supervision?

Geometric Forms of Embedding Association

- Distance-based association (a.k.a. axis calibration)



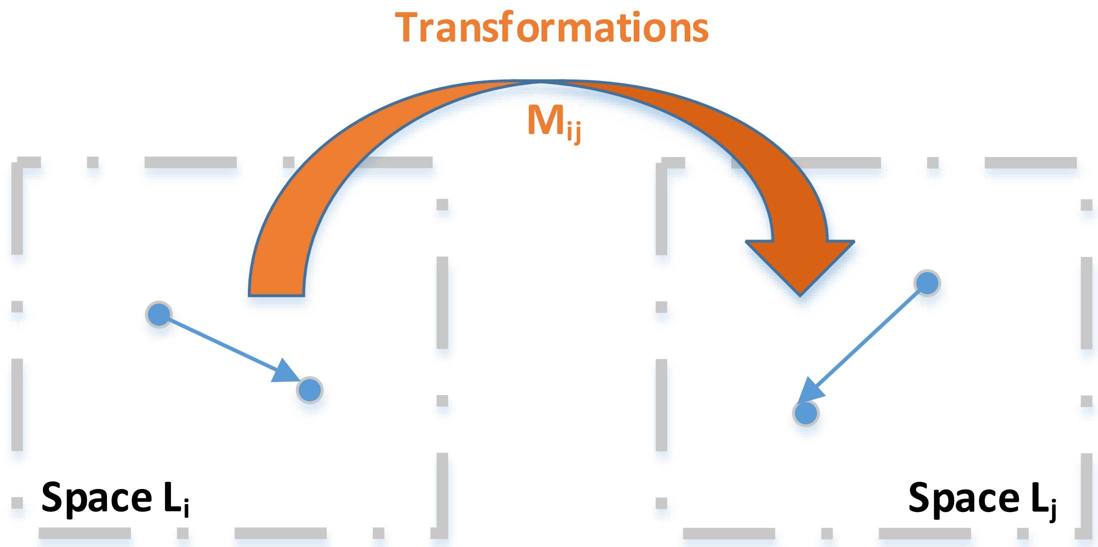
Minimize the distance of embedding representations for associated objects

- Suitable for associations between data with similar structures and sizes (e.g. 1-to-1 entity alignment between well-populated multilingual KGs).

Geometric Forms of Embedding Association

- Transformation-based association
  - Suitable for data of considerably different structures and sizes (allows embedding spaces of different dimensions)

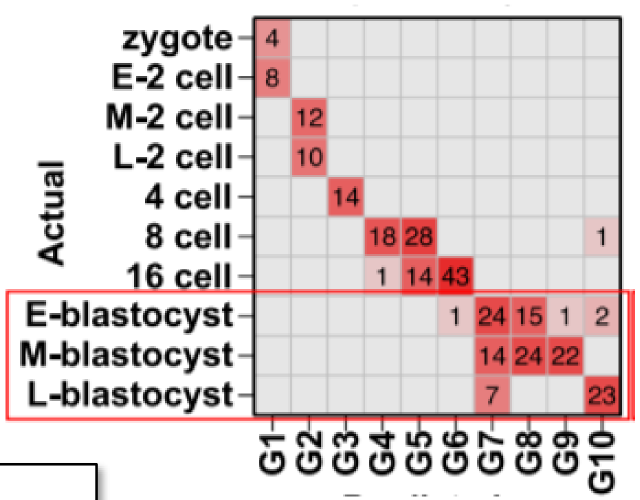
Transformations across embedding spaces of different domains/data



Geometric Forms of Embedding Association

- Matrix factorization based association
  - For knowledge alignment with uncertainty (e.g. RNA sequencing transcripts between domains of *cells* and *genes*)

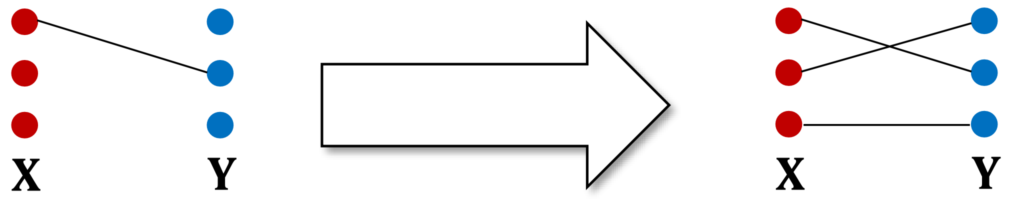
- Techniques
  - **Matrix factorization** (Given S as the alignment data matrix
    - Minimize $\left\|S - \mathbf{E}_1\mathbf{E}_2^{\mathrm{T}}\right\|$
    - $\mathbf{E}_1$ and $\mathbf{E}_2$ are both of dim k
  - **Matrix tri-factorization**
    - Minimize $\left\|S - \mathbf{E}_1\mathbf{U}\mathbf{E}_2^{\mathrm{T}}\right\|$
    - $\mathbf{E}_1$ and $\mathbf{E}_2$ are of dim k_1 and k_2
    - $\mathbf{U}$ is a $k_1 \times k_2$ matrix



Allows two embedding spaces to be of different dims

Semi-supervised learning strategies

- Bootstrapping (BootEA [Sun+, IJCAI-18])
  - Iteratively suggesting new alignment labels for unaligned entities in training
  - Distance-based association
  - A new label (*e*, *e'*) is added to training if
    - The embedding distance of **e**, **e**' are within a threshold
    - They are mutually nearest neighbor of each other



Semi-supervised learning strategies

Co-training (KDCoE [Chen+ IJCAI-18])

- Alternately proposing new labels based on different sets of features (graph structures and entity descriptions)

Inter-lingual Link (ILL): *(**astronomer**@EN, **astronome**@FR)*

EN triple: *(**Ulugh Beg**, occupation, **astronomer**)*

FR triple: *(**Ulugh Beg**, activité, **astronome**)*

An astronomer is a scientist in the field of astronomy who concentrates their studies on a specific question or field outside of the scope of Earth...

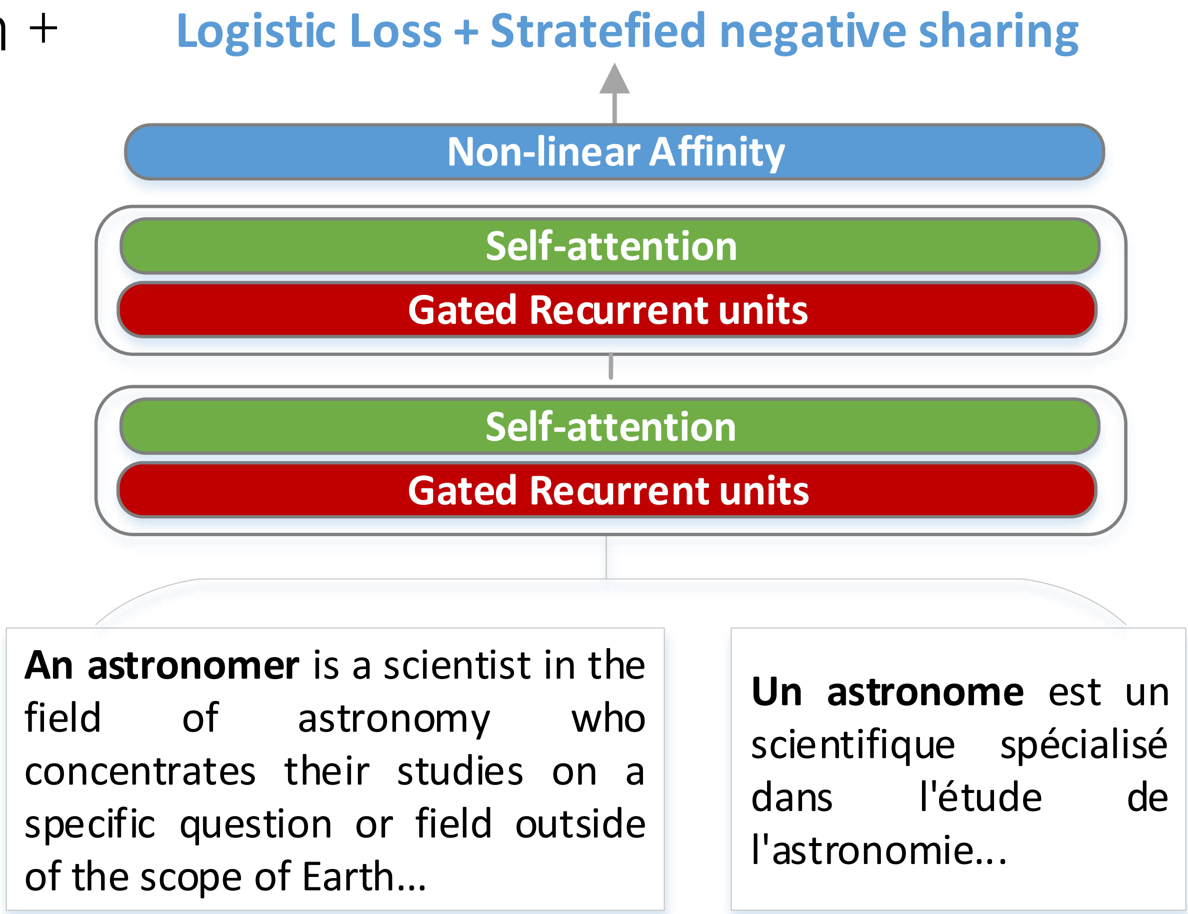
Un astronome est un scientifique spécialisé dans l'étude de l'astronomie...

Semi-supervised learning strategies

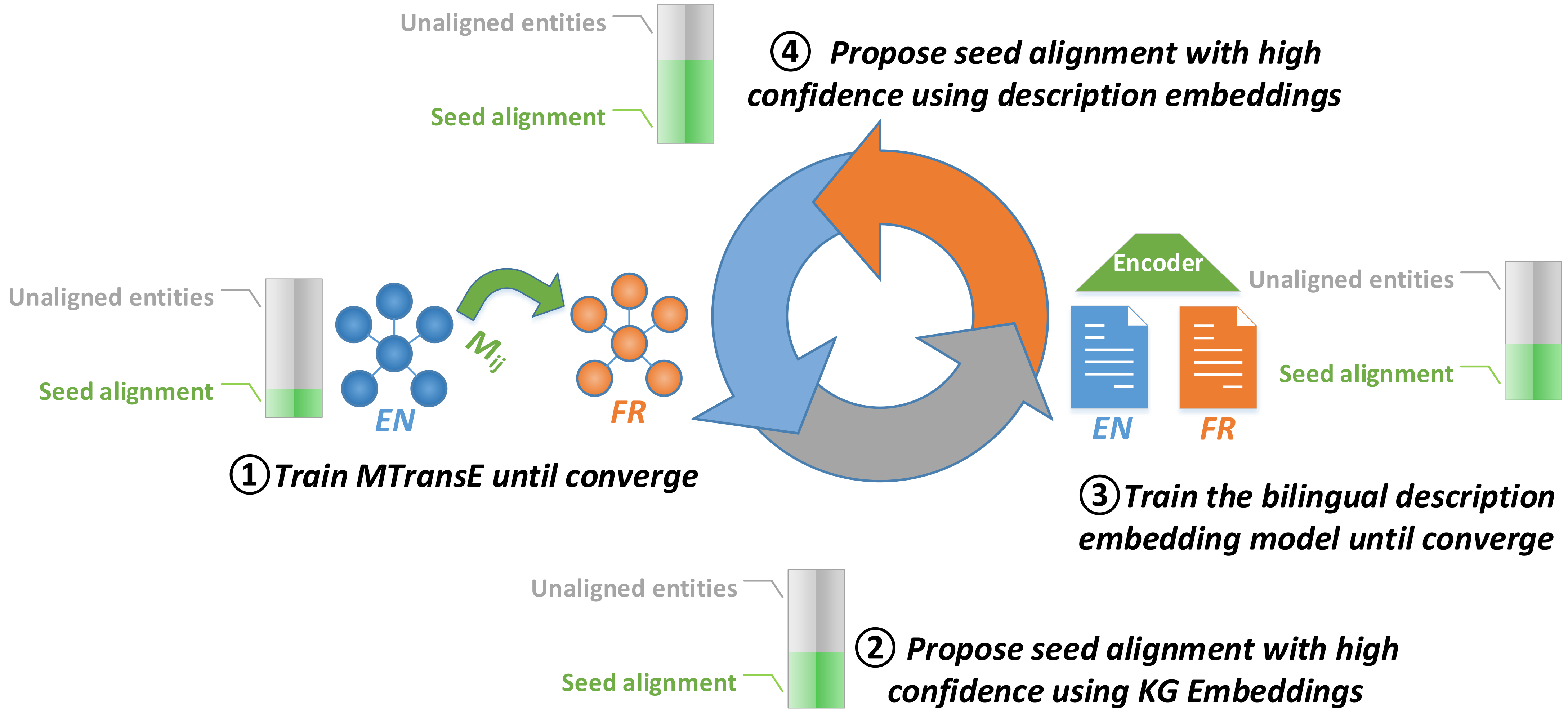
Co-training (KDCoE [Chen+ IJCAI-18])

Siamese document encoder with Self-attention + Pre-trained bilingual word embeddings

To decide whether two multilingual descriptions are describing the same entity.

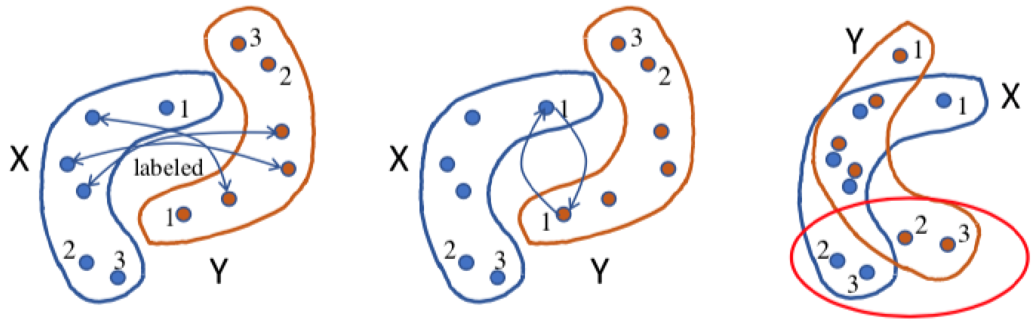


Semi-supervised learning strategies



Semi-supervised learning strategies

- **Optimal transport** (OTEA [Pei+, IJCAI-19]): matching the distribution of embeddings



(a) (b) (c)

Transferred embedding

$$\mathcal{L}_{ot}(\mathbf{p}^{\boxed{\mathbf{M^1}\theta_e^i}}, \mathbf{q}^{\theta_e^j}) = \frac{1}{K} \sup_{\|f\|_L \le K} \mathbb{E}_{\boxed{y} \sim \mathbf{q}^{\theta_e^j}}[f(y)] - \mathbb{E}_{y \sim \mathbf{p}^{\mathbf{M^1}\theta_e^i}}[f(y)]$$

entity embedding

WGAN

$$\mathcal{L}_{g_1} = \min_{\mathbf{M^1}} \max_{D_1} \mathbb{E}_{y \sim \mathbf{q}^{\theta_e^j}}[f_{D_1}(y)] - \mathbb{E}_{x \sim \mathbf{p}^{\theta_e^i}}[f_{D_1}(\mathbf{M^1}x)]$$

$$\mathcal{L}_{g_2} = \min_{\mathbf{M^2}} \max_{D_2} \mathbb{E}_{y \sim \mathbf{q}^{\theta_e^i}}[f_{D_2}(y)] - \mathbb{E}_{x \sim \mathbf{p}^{\theta_e^j}}[f_{D_2}(\mathbf{M^2}x)]$$

Optimization for embedding transformations of two directions

Learning with Auxiliary Information

- Attribute-based embedding association
- JAPE [Sun+, ISWC-17]:
  - Using a weighted Skip-gram language model [Mikolov+, NIPS-13] that predicts entities based on attributes

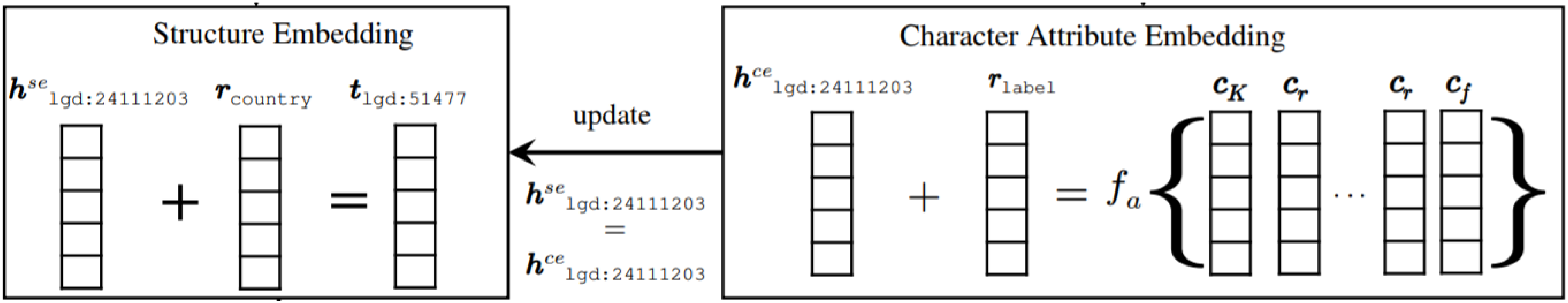
longitude
latitude
place name
…

birthday
birthplace
gender
…

Entities with correlated attributes will have similar embedding vectors.

Learning with Auxiliary Supervision

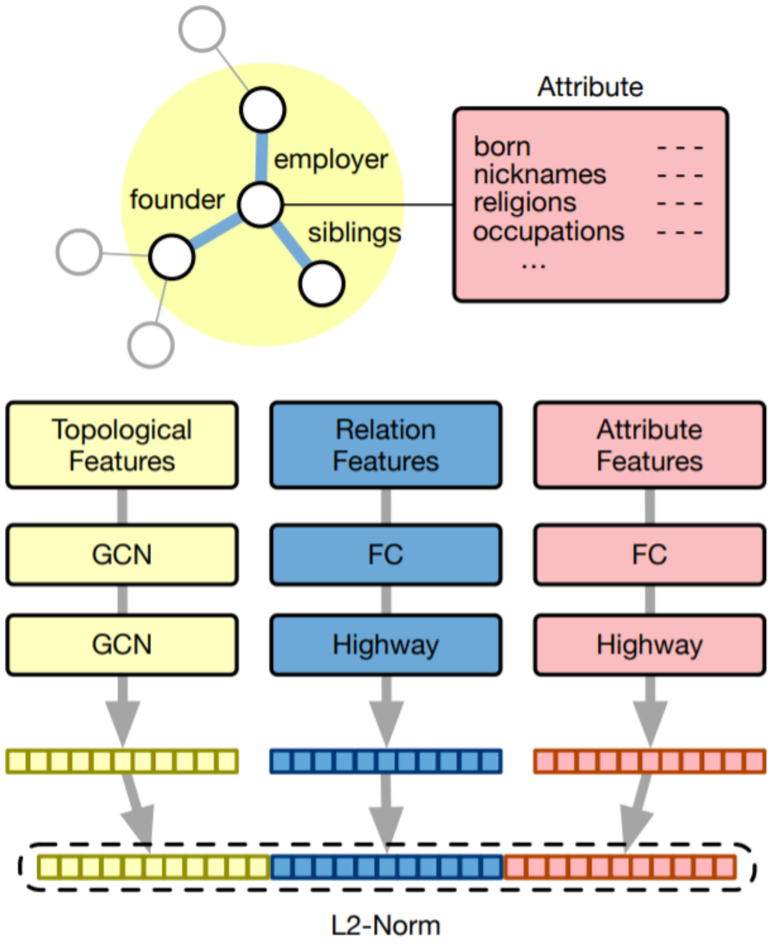
- Attribute-based embedding association
- AttrE [Trisedya+, AAAI-19]
  - Using a Char-LSTM to encode attributes of each entity
  - The translational embedding model is jointly trained with the attribute Char-LSTM



Learning with Auxiliary Supervision

- Multi-view learning: using different views of entities to bridge between two domains
  - MultiKE [Zhang+, IJCAI-19], HMAN [Yang+, EMNLP-19]
  - Combining all modalities
    - Structures: translational (MultiKE) and GCN (HMAN) encoders
    - Attributes: CNN (MultiKE) and FFNN (HMAN)
    - Literals/descriptions: BiLSTM (MultiKE), BERT (MultiKE)
  - Embedding combination
    - Concatenation (HMAN)
    - Weighted avg or FFNN (MultiKE)

- What about multi-media?



HMAN [Yang+, EMNLP-19]

Cross-domain and Interdisciplinary Tasks

- KBP tasks
  - Knowledge alignment
  - Knowledge synchronization
  - Entity typing
  - Ontology population
- Computational Bio-med tasks
  - Protein-protein interaction prediction
  - Single-cell RNA sequence imputation
  - Polypharmacy side effect detection

Scenario 1: Knowledge alignment / Entity resolution

- Task: to identify the match of entities in different KGs
- Cross-lingual entity alignment
  - DBP15k dataset
  - 15k aligned entities between each two KGs
  - <30% training set
- Monolingual entity alignment
  - DWY100K dataset
  - 100k aligned entities between each two KGs
  - <30% training set

Datasets		DBP15K				
		Entities	Rel.	Attr.	Rel.triples	Attr.triples
ZH-EN	Chinese	66,469	2,830	8,113	153,929	379,684
	English	98,125	2,317	7,173	237,674	567,755
JA-EN	Japanese	65,744	2,043	5,882	164,373	354,619
	English	95,680	2,096	6,066	233,319	497,230
FR-EN	French	66,858	1,379	4,547	192,191	528,665
	English	105,889	2,209	6,422	278,590	576,543

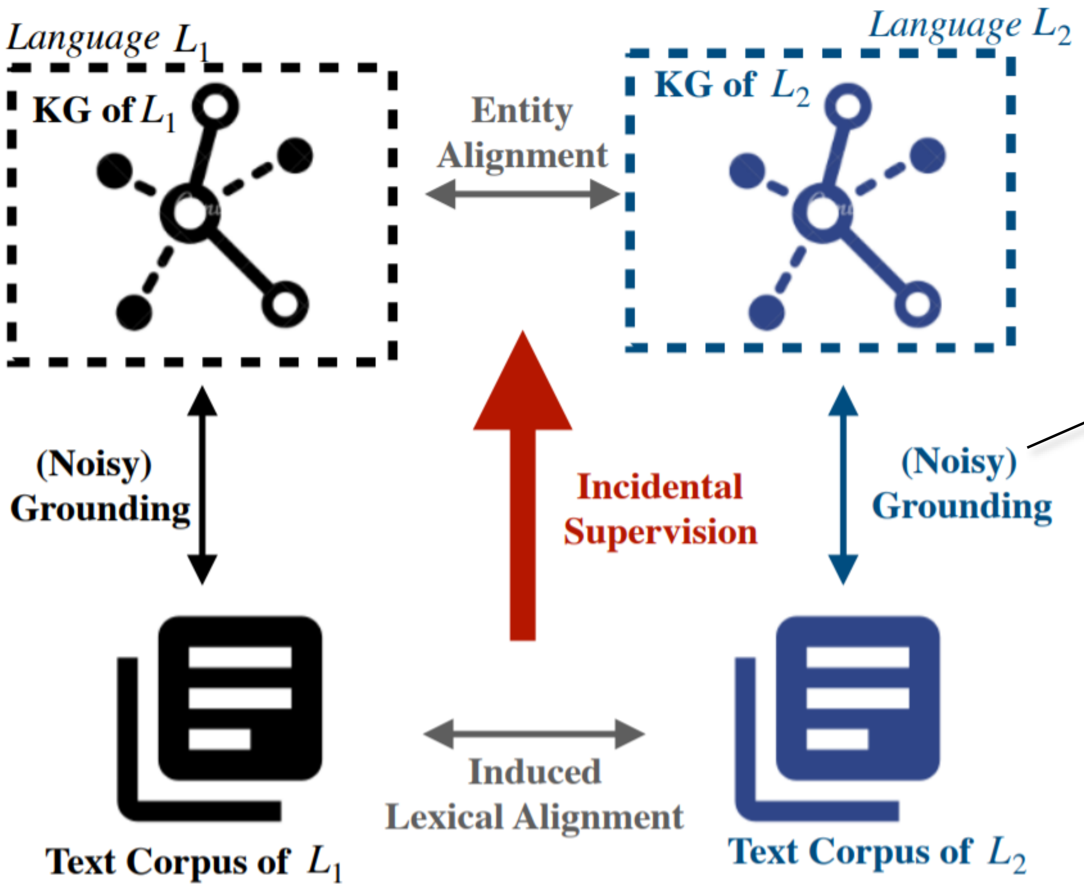
https://github.com/nju-websoft/JAPE

Datasets		# Ent.	# Rel.	# Attr.	# Rel tr.	# Attr tr.
DBP-WD	DBpedia	100,000	330	351	463,294	381,166
	Wikidata	100,000	220	729	448,774	789,815
DBP-YG	DBpedia	100,000	302	334	428,952	451,646
	YAGO3	100,000	31	23	502,563	118,376

https://github.com/nju-websoft/MultiKE

A paper list for entity alignment/resolution: https://github.com/THU-KEG/Entity_Alignment_Papers

Entity Alignment with Incidental Supervision From Free Text*



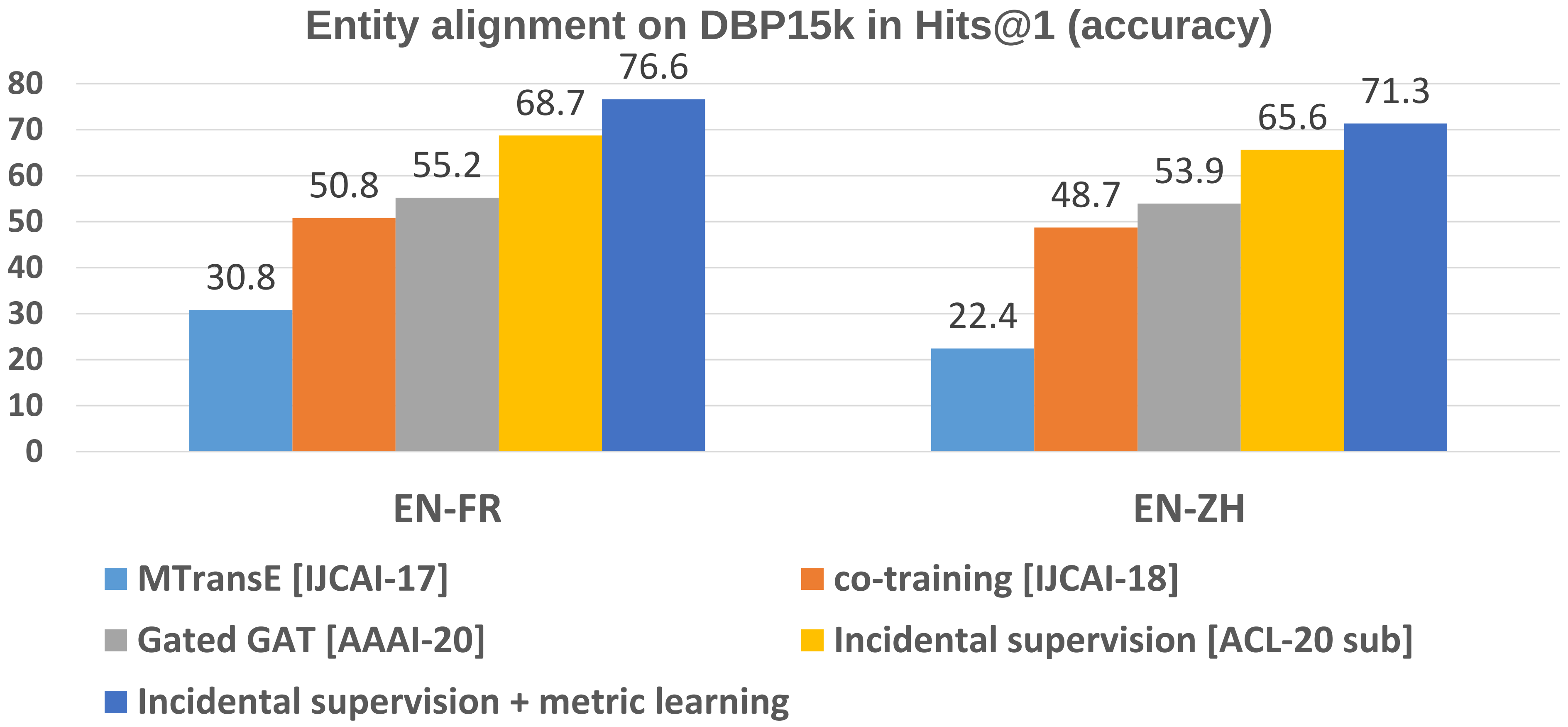
(Noisy) grounding process for text corpora

- EDL
- Surface form matching

Three steps

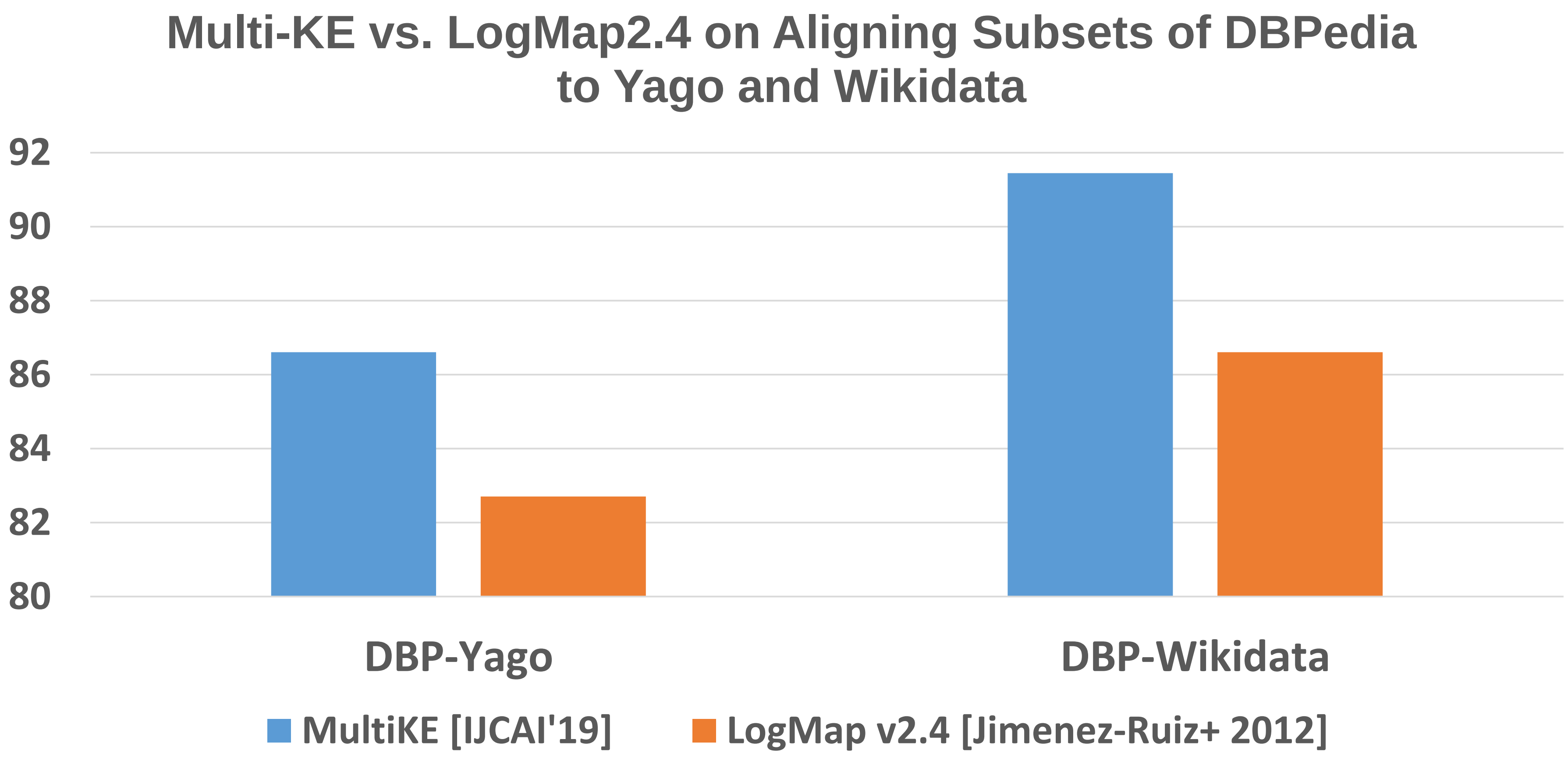
1. ***(Noisy) grounding***: connecting KGs and text corpora
2. ***Embedding learning***: Translational GCN + a neural language model
3. ***Alignment learning***: self-learning + optimal transport

Cross-lingual Entity Alignment Results



***Candidate space is 63k~98k entities in each language**

Representation Learning Method vs. SotA Ontology Matching System (LogMap v2.4)

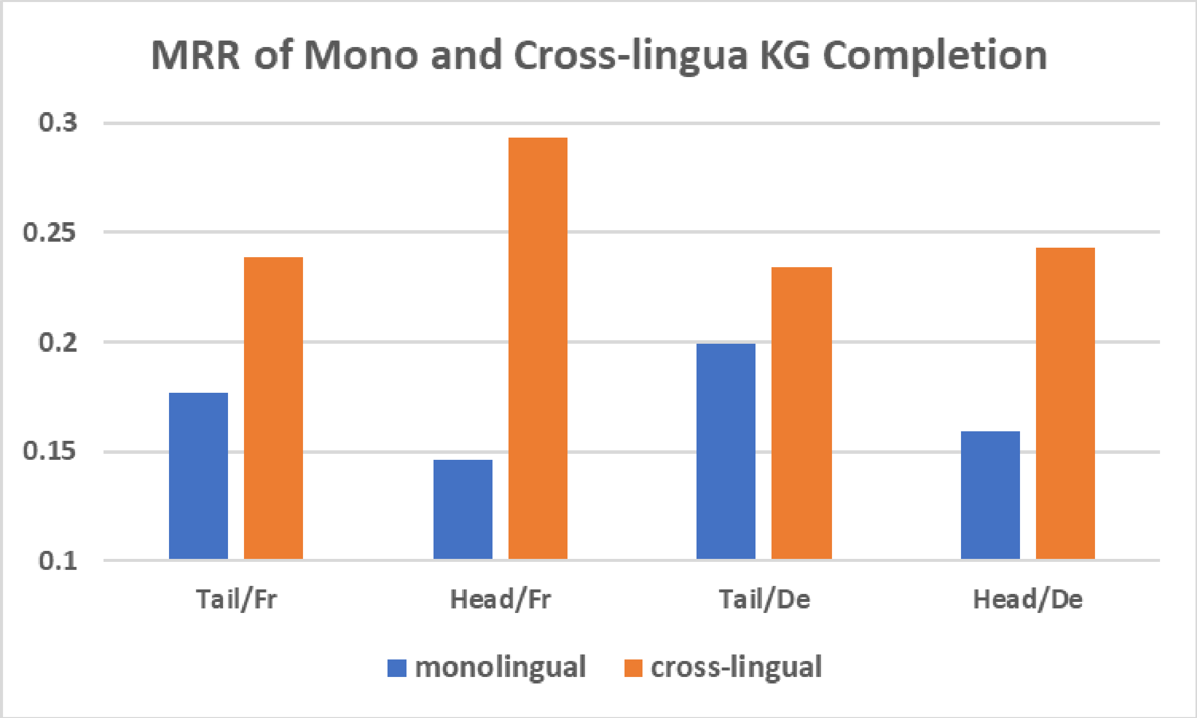


***MultiKE [Sun+ IJCAI’19]** is a monolingual ontology matching system in which multi-view embeddings of structures, literals, descriptions and attributes are combined.

Cross-lingual Knowledge Projection

Knowledge transfer to a sparser KG (e.g. French)

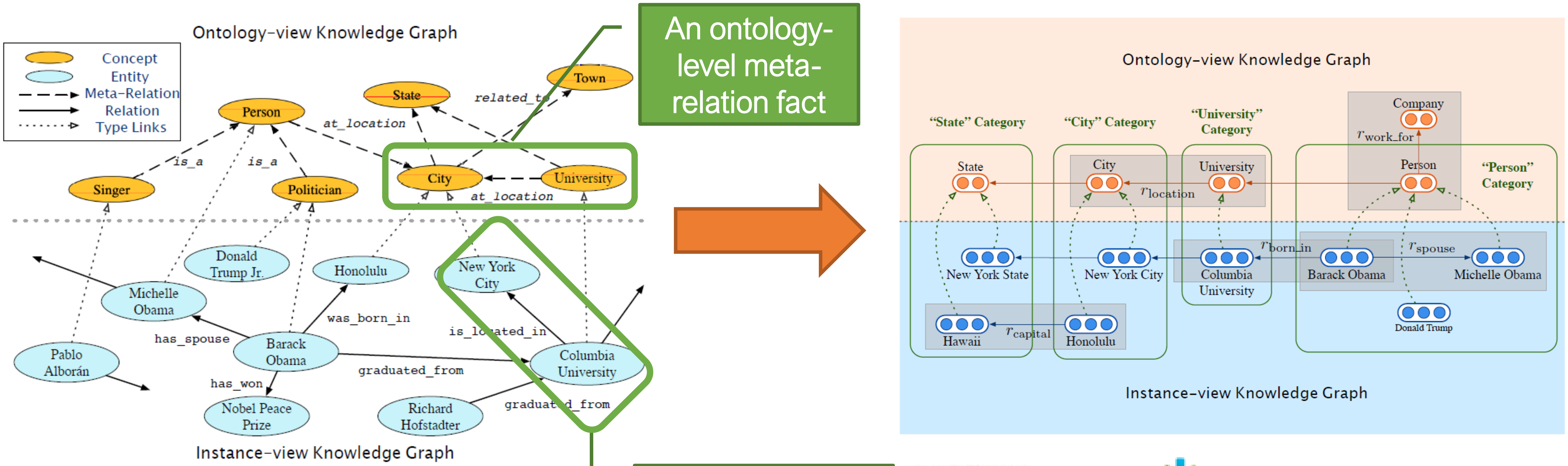
- Obtain the answer of queries *(h, r, ?t)* in the embedding space of a well-populated version (e.g. English) of KG



Cross-lingual knowledge transfer improves sparse KG completion.

Scenario 2: Instance Knowledge and Ontological Concepts

Ontology view: meta-relations of commonsense concepts
Instance view: relations of entities instantiated from concepts



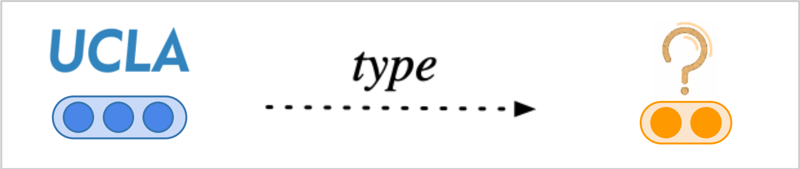
An instance-level relation fact

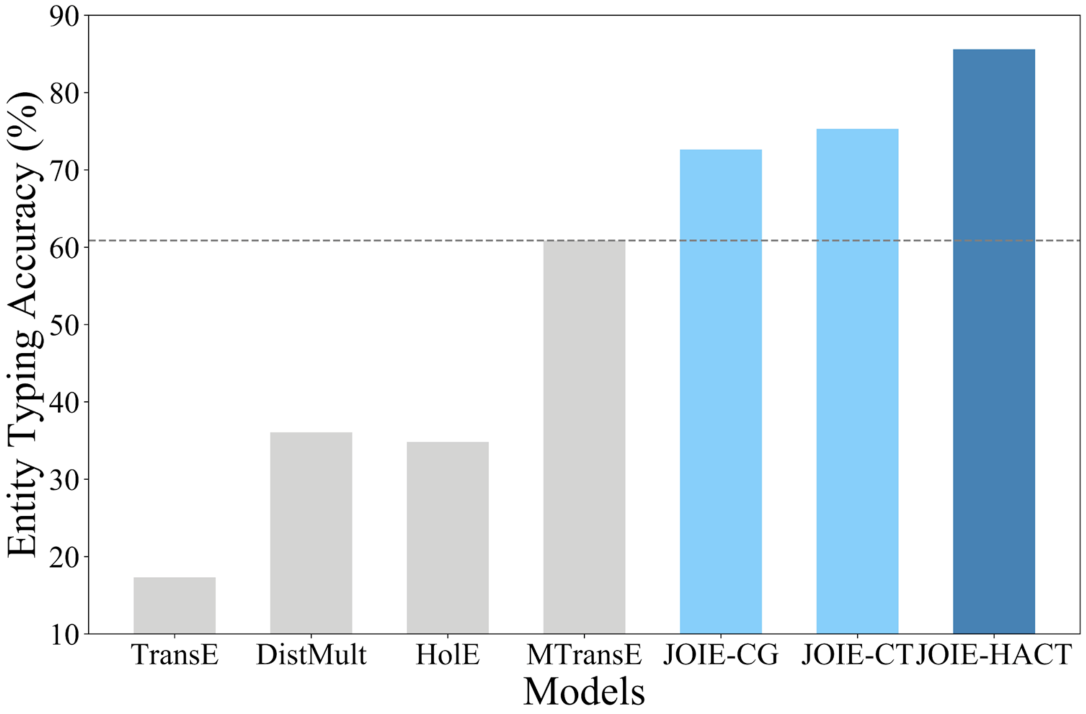
JOIE [Hao+, KDD-19]

Entity Typing

- Given an entity without a known type, what is the most likely type (concept) that it associates with?

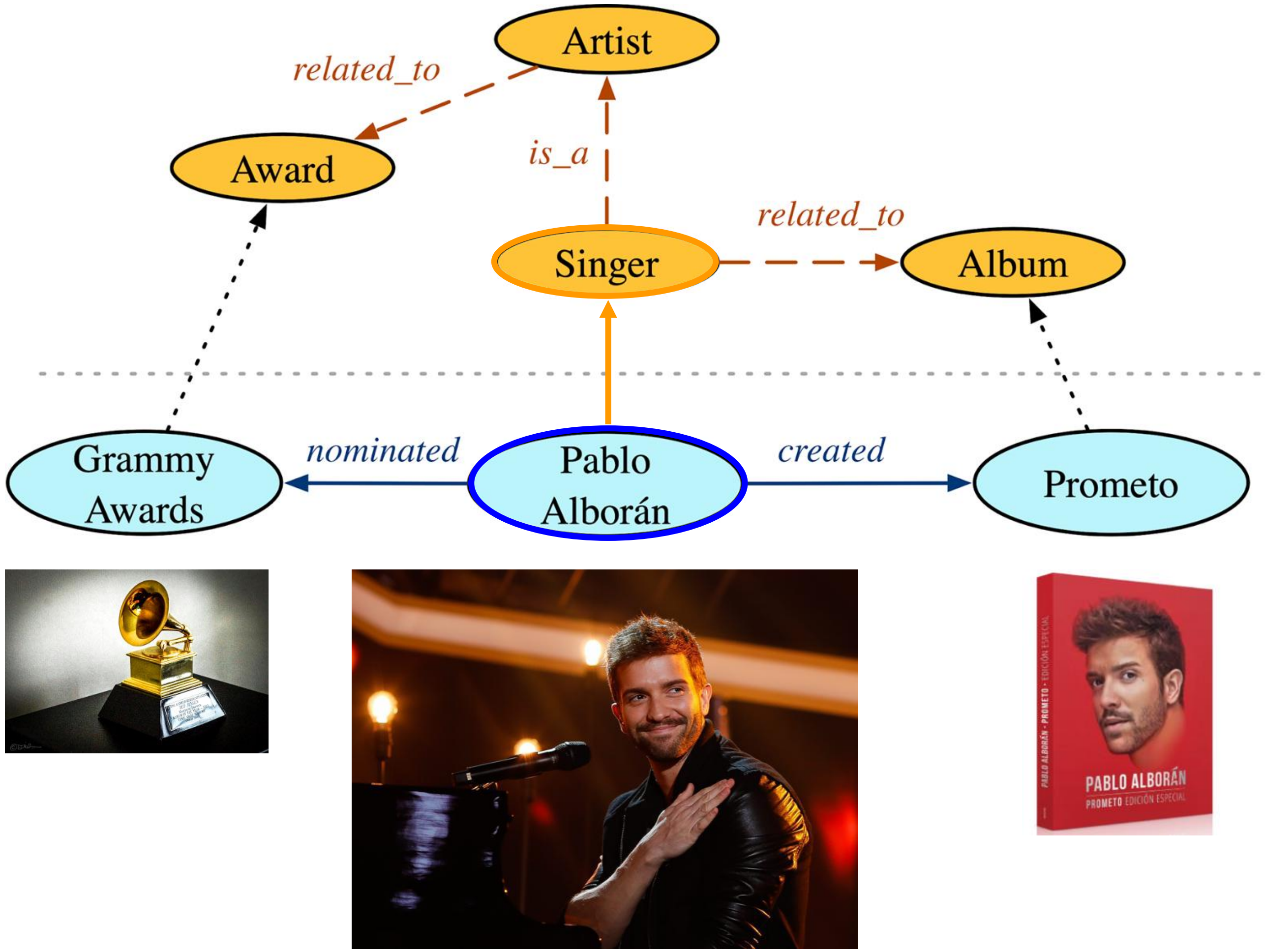
JOIE [Hao+, KDD-19]





Type inference (906 labels) on 40% of >111k entities in YAGO.

Long-tail Entity Typing (Least Frequent 15%)



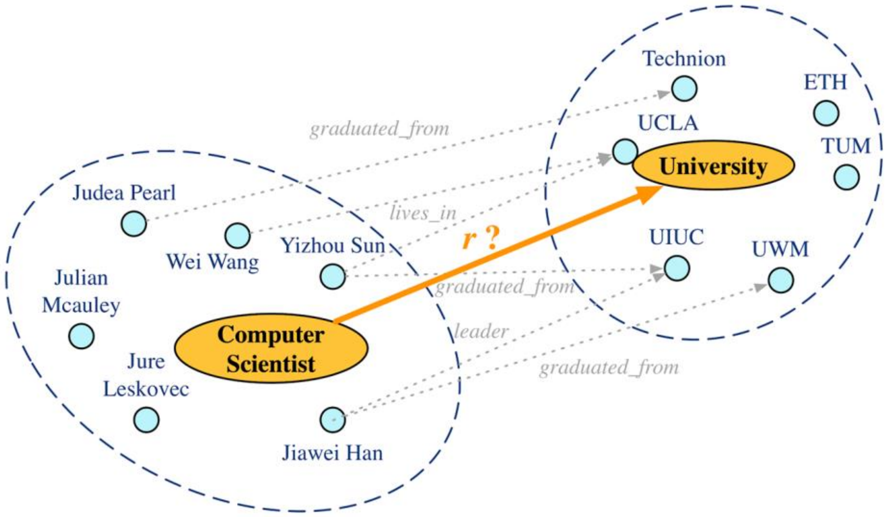
Example of long-tail entity typing

Entity	Model	Top 3 Concept Prediction
Laurence Fishburne	DistMult	football team, club, team
	MTransE	writer, **person**, artist
	JOIE	**person**, artist, philosopher
Warangal City	DistMult	country, village,**city**
	MTransE	administrative region, **city**, settlement
	JOIE	**city**, town, country
Royal Victor -ian Order	DistMult	person, writer, administrative region
	MTransE	election, award, **order**
	JOIE	award, **order**, election

Entity typing accuracy on long-tail entities

Datasets	YAGO26K-906		
Metrics	MRR	Acc.	Hit@3
DistMult	0.156	10.89	25.33
MTransE	0.526	46.45	67.25
JOIE-TransE-CG	0.708	59.97	79.80
JOIE-TransE-CT	0.737	62.05	82.60
JOIE-HATransE-CT	**0.802**	**69.66**	**87.75**

Transfer Instance-level Knowledge for **Ontology Population**

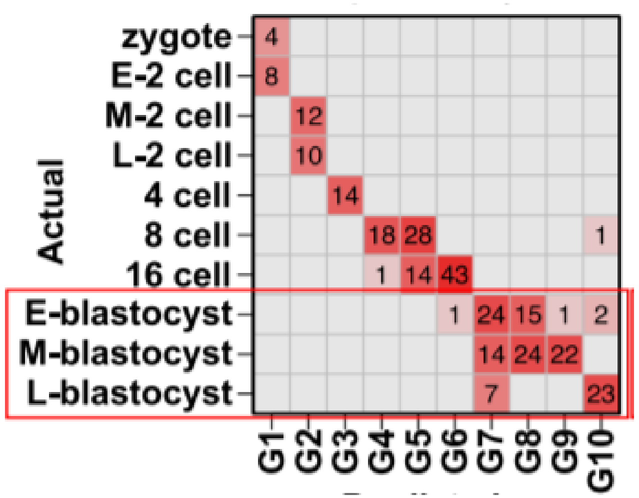


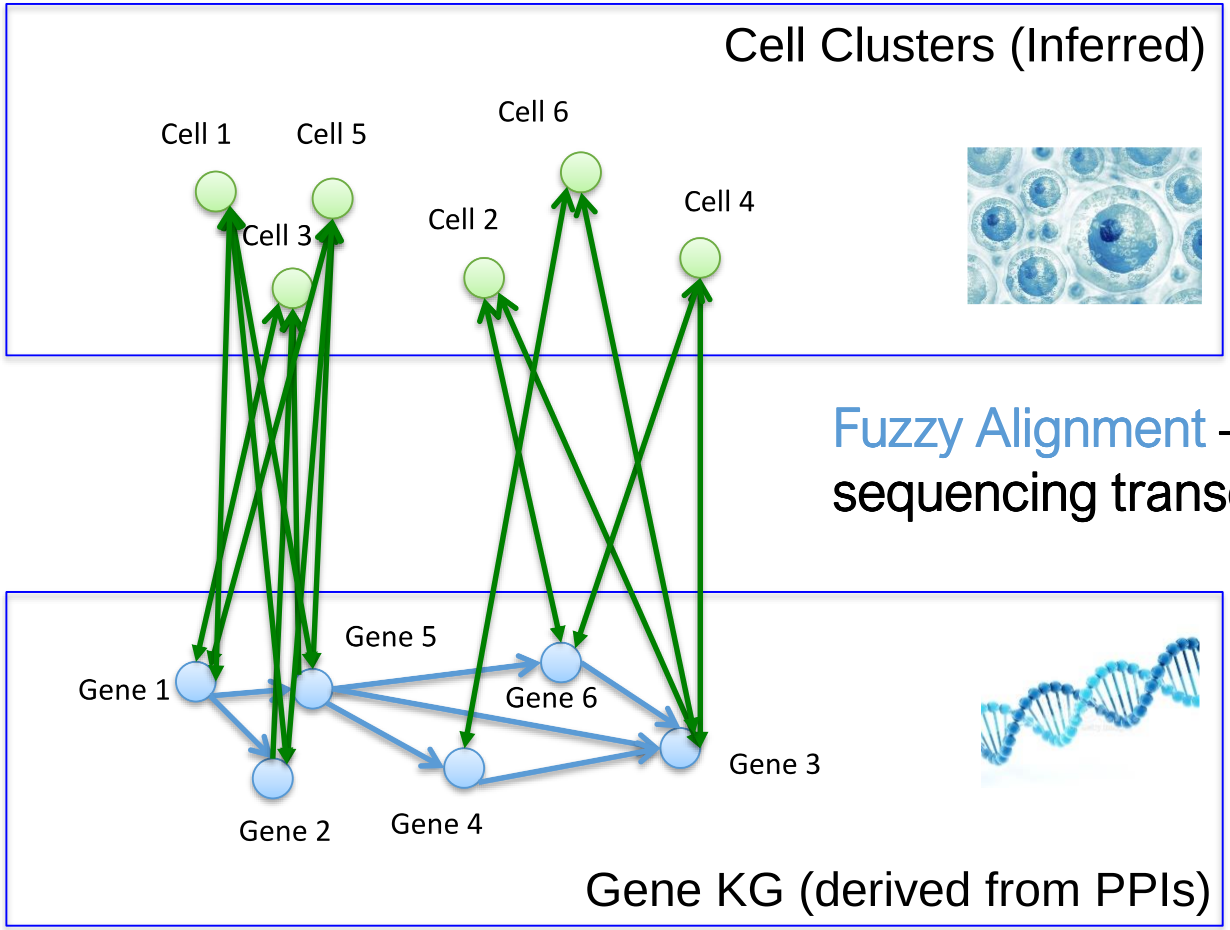
Populating unseen ontological relation facts by transferring instance-view relations.

Examples of ontology population

Query	Top 3 Populated Triples with distances
(scientist,?r, university)	scientist, ***graduated from***, university (0.499) scientist, ***isLeaderOf***, university (1.082) scientist, *isKnownFor*, university (1.098)
(boxer, ?r, club)	boxer, ***playsFor***, club (1.467) boxer, *isAffiliatedTo*, club (1.474) boxer, ***worksAt***, club (1.479)
(scientist, ?r, scientist)	scientist, ***doctoralAdvisor***, scientist (0.204) scientist, ***doctoralStudent***, scientist (0.221) scientist, ***relative***, scientist (0.228)

Scenario 3.a: Single-cell Gene Expressions





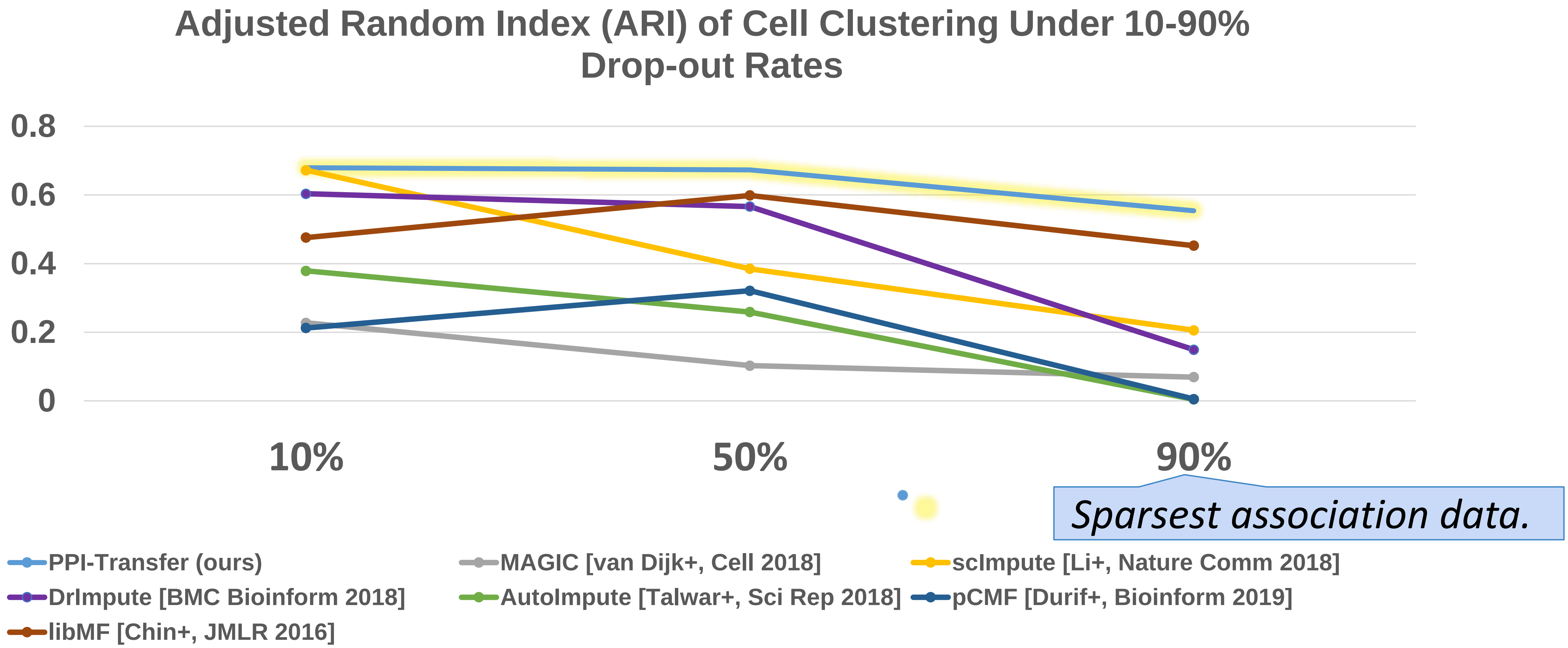
View I = Cells

Fuzzy Alignment – Single-cell RNA sequencing transcripts

View II = Genes

Relations = {binding, activation, reaction, catalyst, expression, inhibition, ptmod}

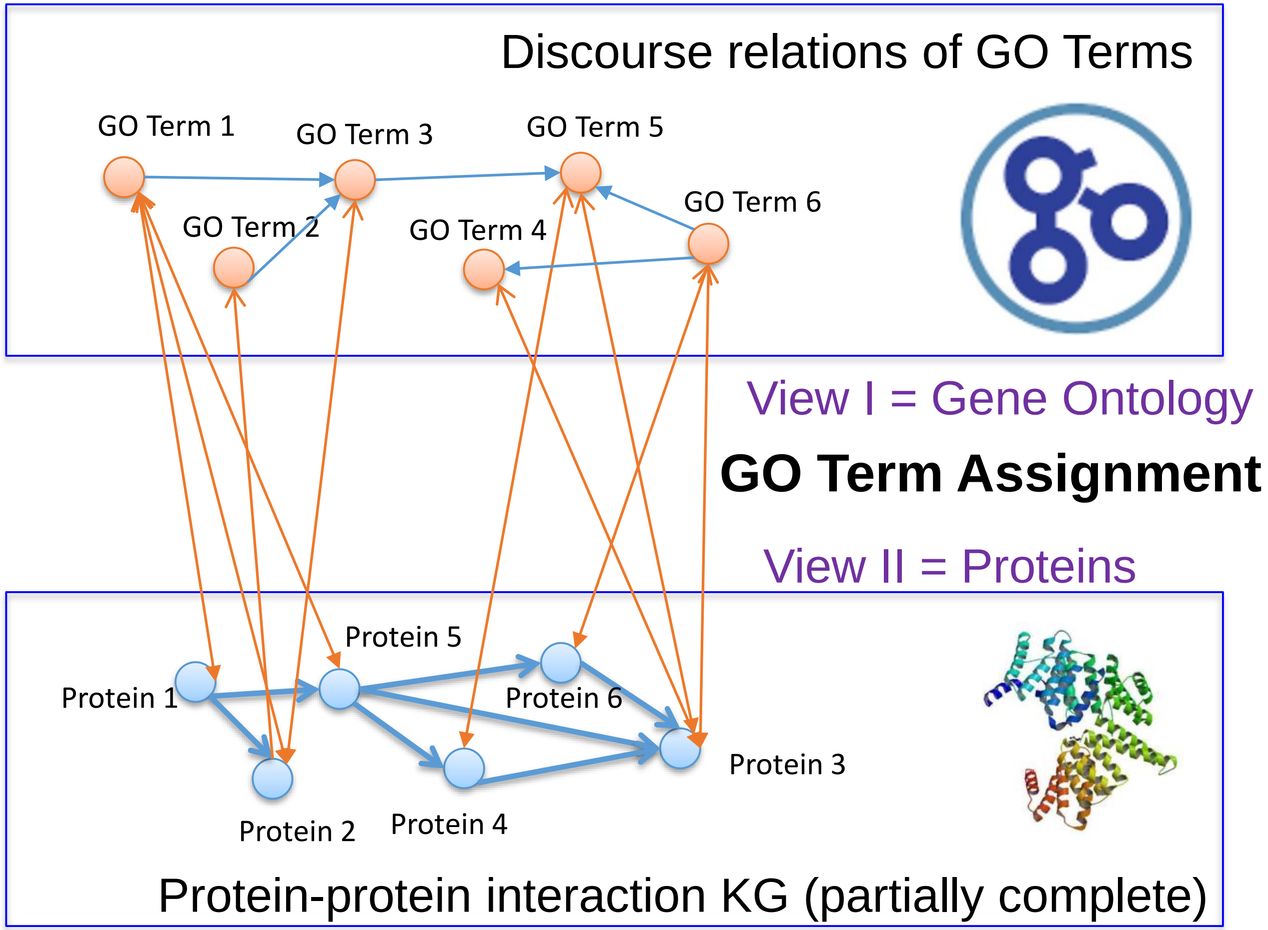
Scenario 3.a: Single-cell Gene Expressions



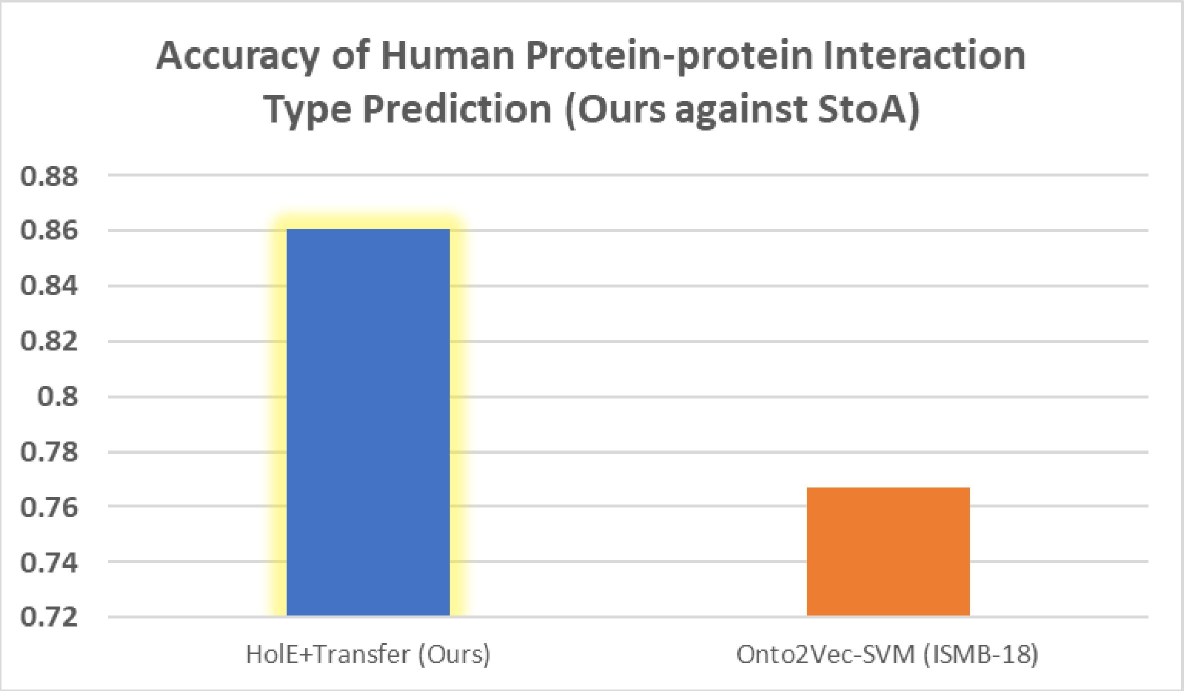
***Based on Mouse cortex and hippocampus data [Ziesel+, *Science 2015*]**

Transferring gene-interaction knowledge improves cell clustering, especially when the gene-cell association data are very sparse.

Scenario 3.b: PPI and Gene Ontologies

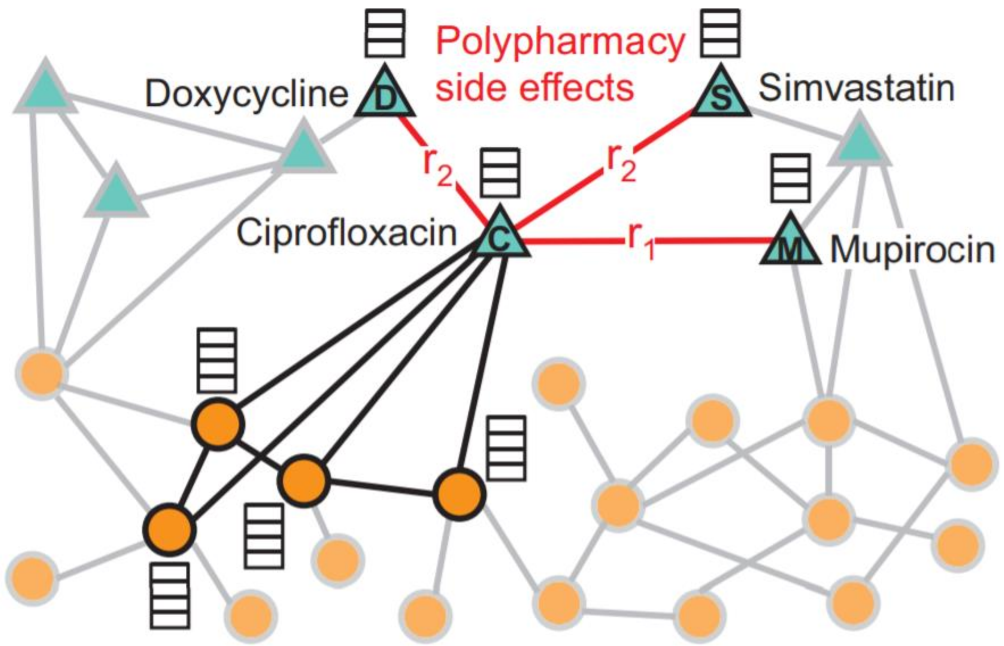


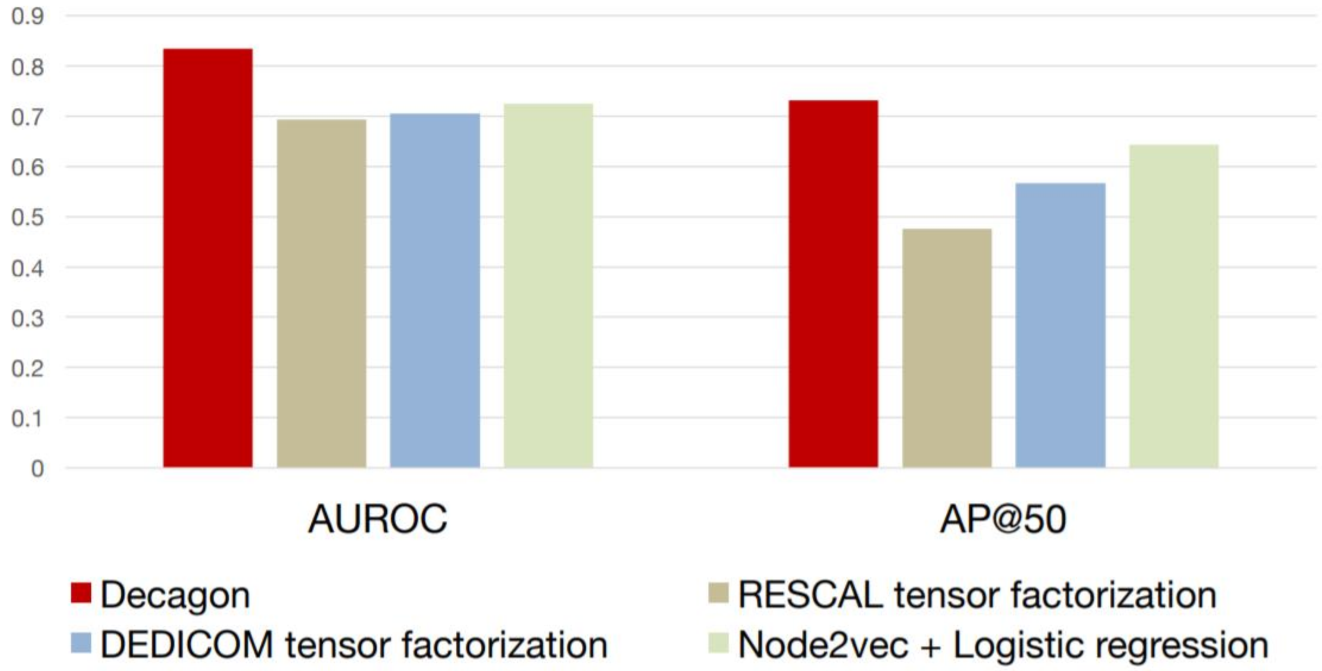
Task: Protein-protein interaction prediction for *Homo Sapiens*



Transferring knowledge from the gene ontology improves protein-protein interaction type prediction.

Scenario 4: Polypharmacy side effect detection

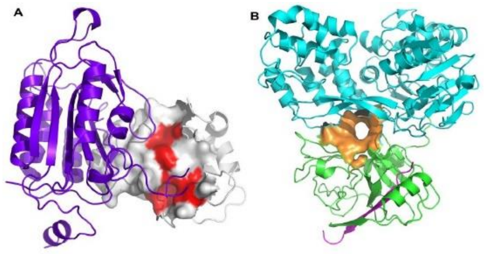




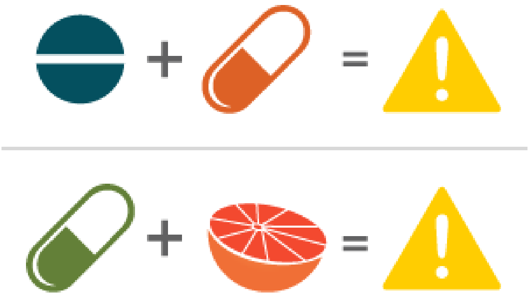
Decagon [Zitnik+, ISMB-18]

- Aggregating protein-protein interaction knowledge for predicting the interaction of drugs.

More Application Scenarios



DISEASE ONTOLOGY



Drug Repurposing

product graph

Product Recommendation