



School of Engineering

Transferable Representation Learning for Multi-relational Data (or KGs)

Recent Advances in Transferable Representation Learning (Part III)

Muhao Chen
USC

Feb 2020

AAAI Tutorials

Recent Advances in Transferable Representation Learning

<https://cogcomp.seas.upenn.edu/page/tutorial.202002/>

- 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

These terms are used interchangeably

- **Knowledge Graphs (KGs)** \approx Multi-relational Data
- **KG Embedding** \approx Relational Embedding
- Representation learning \approx Embedding
- **Embedding space** \approx Vector space

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)

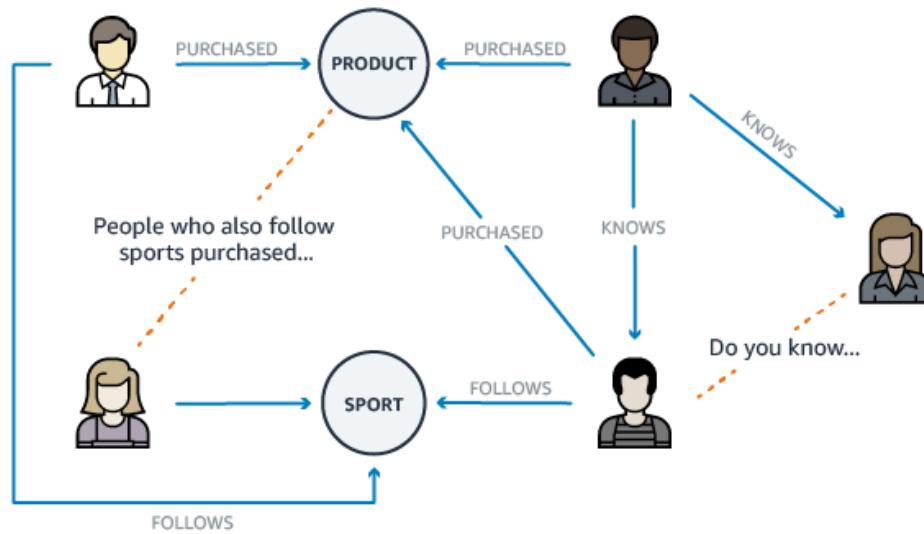
787B

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



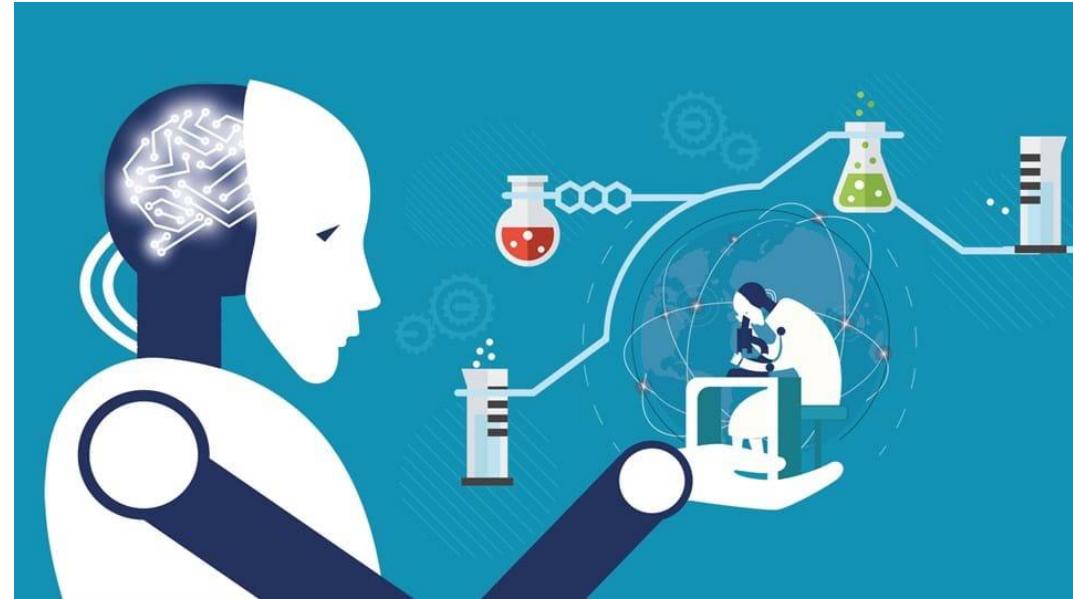
Understanding Relations Is Prominent In Practice

E-Commerce



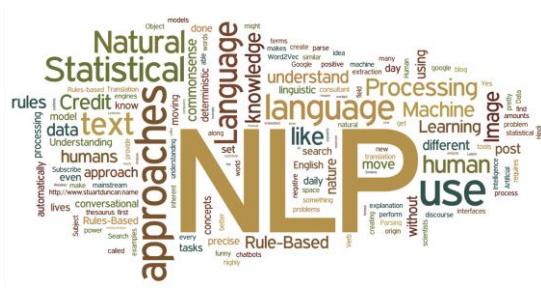
Co-purchase relations of products
Social relations of users

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

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

Multi-relational Data (or KGs)

Open-domain Graphs



Product & E-Commerce Graphs



Bio-med Ontologies /Data Banks



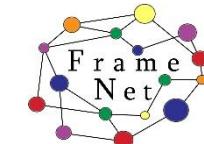
GENEONTOLOGY
Unifying Biology



Lexical and Concept Graphs

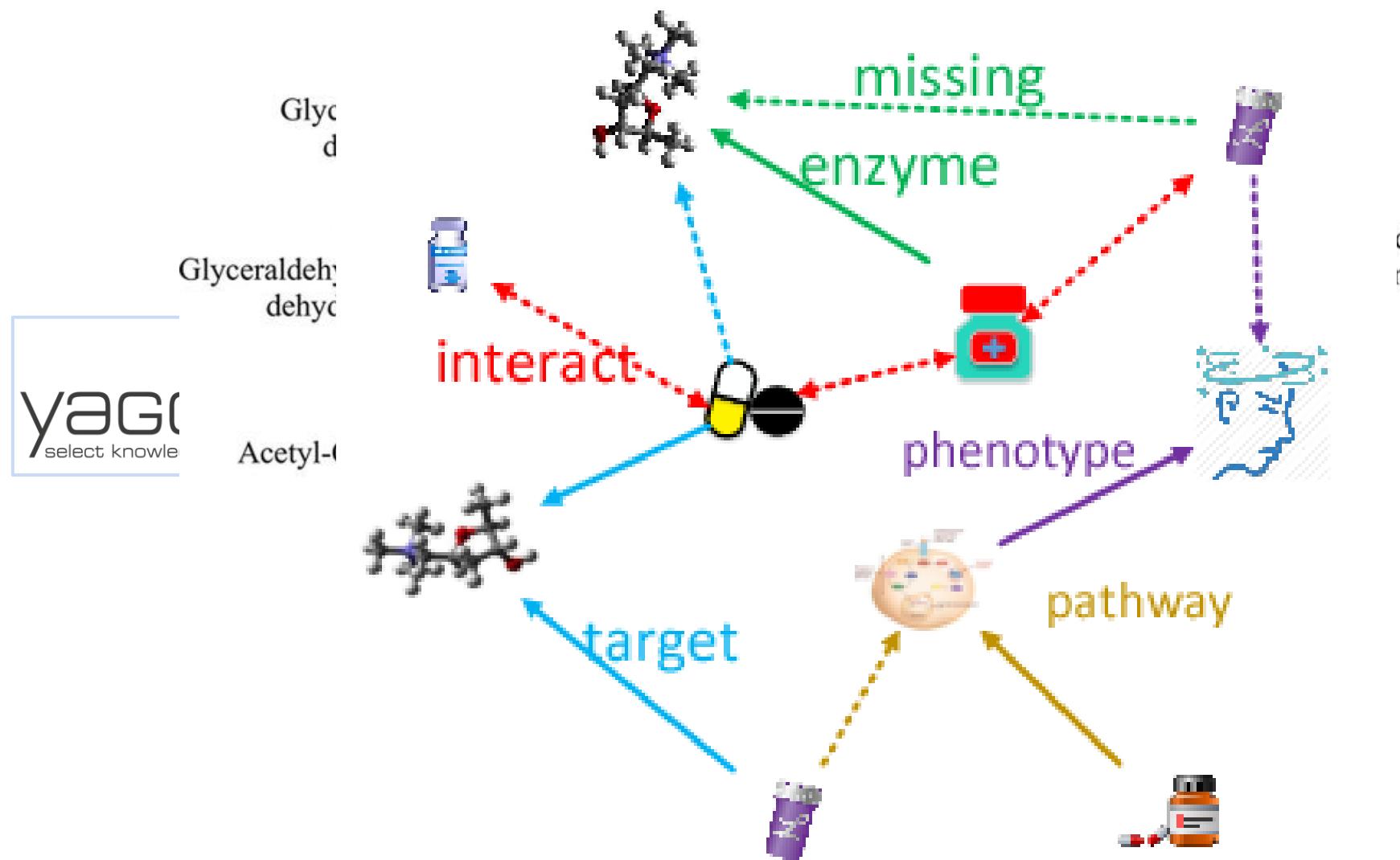


An open, multilingual knowledge graph



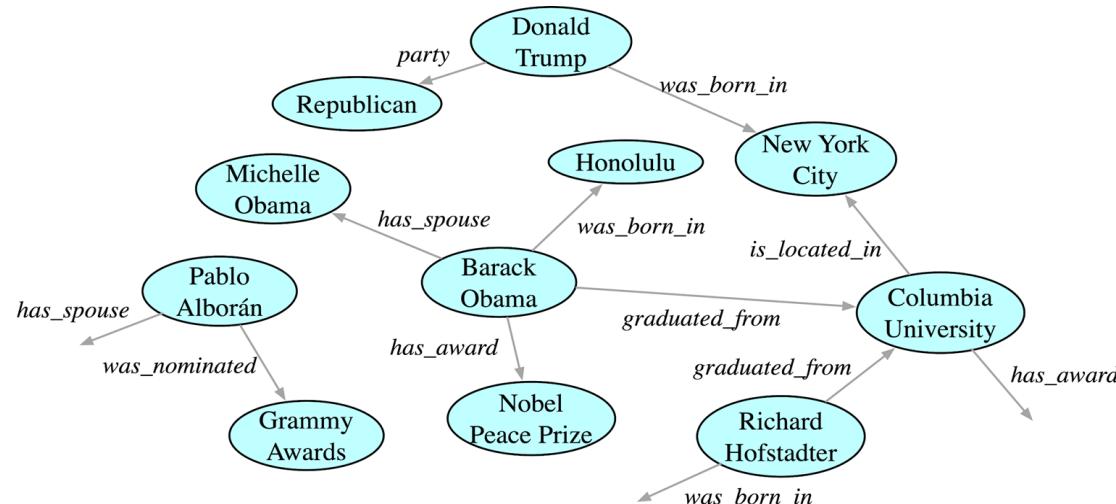
Wiktionary
The free dictionary

Multi-relational Data (or KGs)



A multi-relational dataset is formally defined as an edge-typed graph \mathbf{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

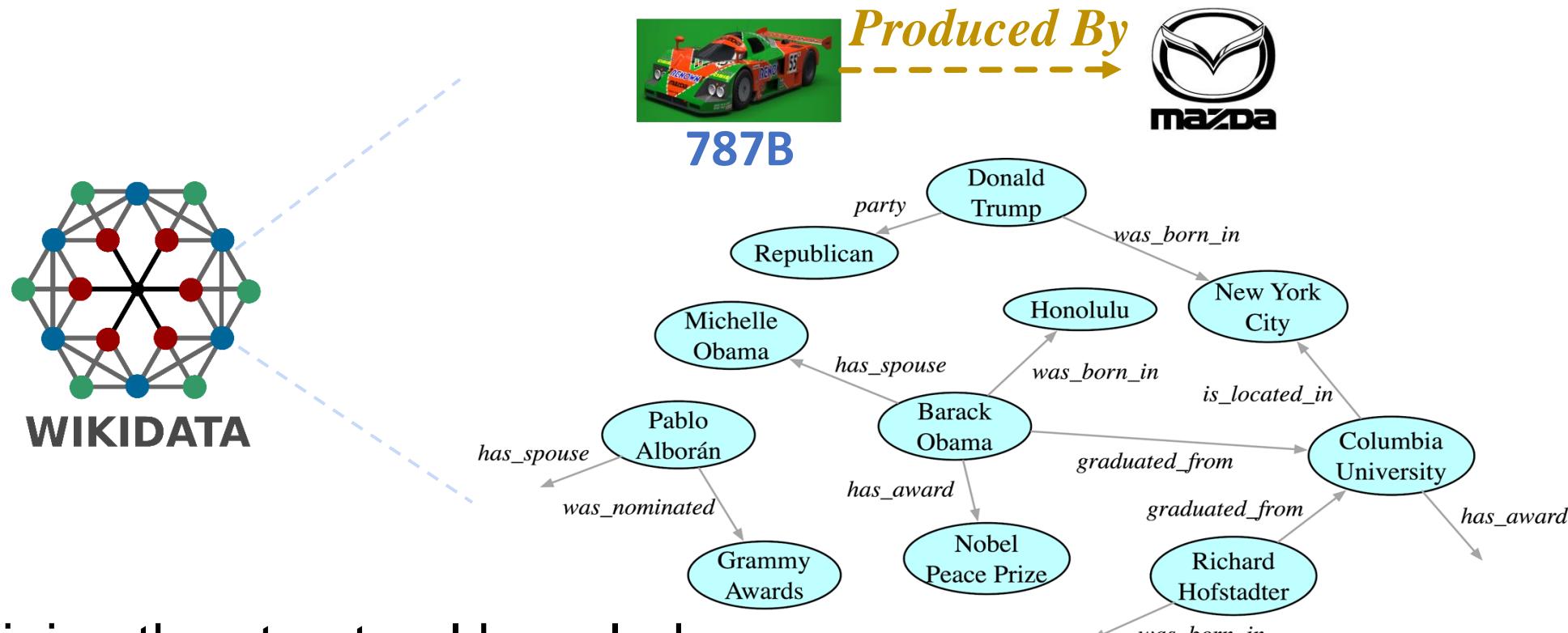


\mathbf{G}



$T=(h, r, t)$

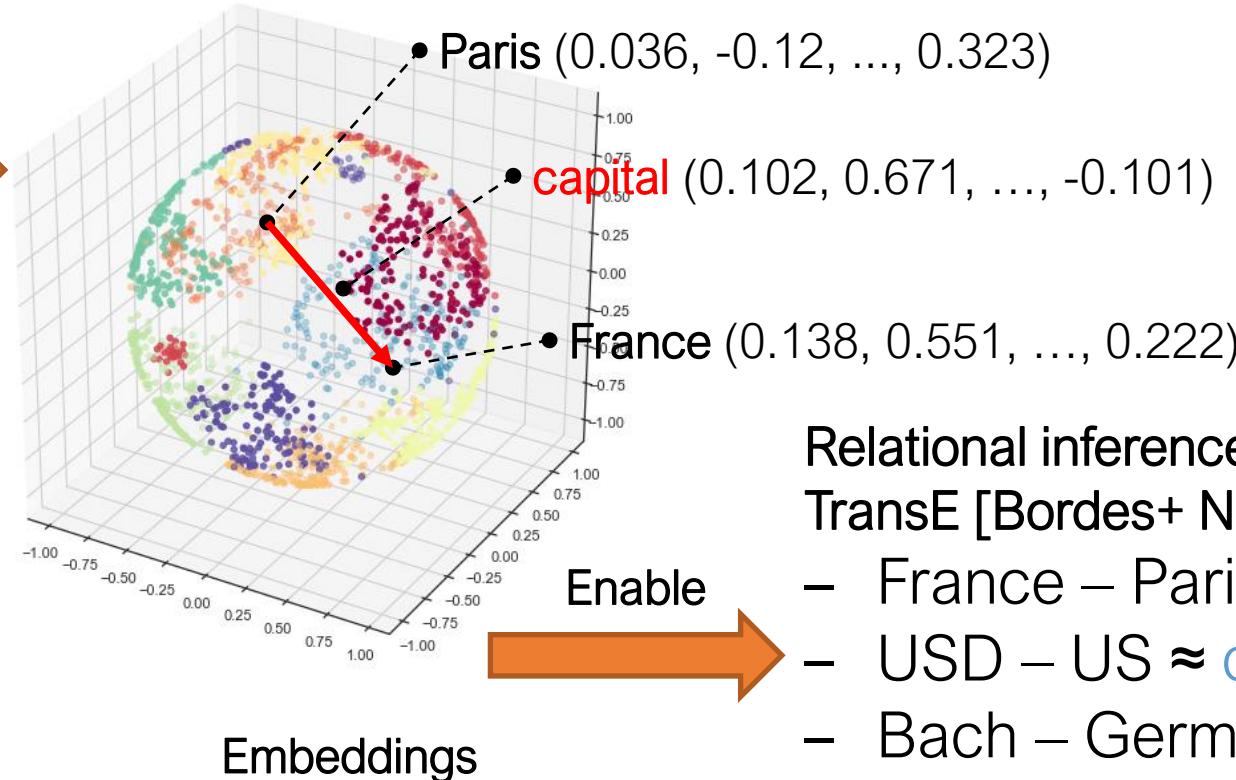
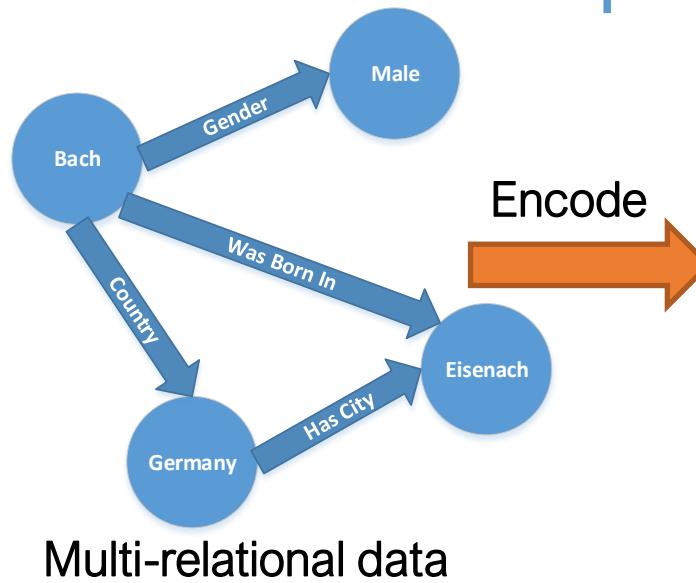
Multi-relational Data: Precise But **Expensive** Knowledge Representation



Obtaining the structural knowledge

- Is expensive (Avg \$5.71 per triple [Paulheim+, ISWC-18] in open domain; higher cost in scientific domains).
- Has relied on massive human efforts.
- Has never been close to complete.

Representation Learning: Cheap Knowledge Acquisition from The Embedding Space



Relational inference as inductive bias (e.g. TransE [Bordes+ NIPS-13])

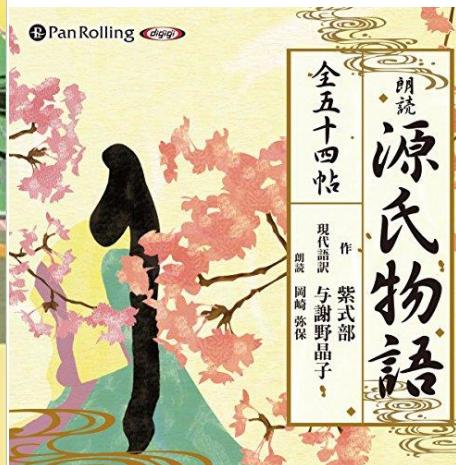
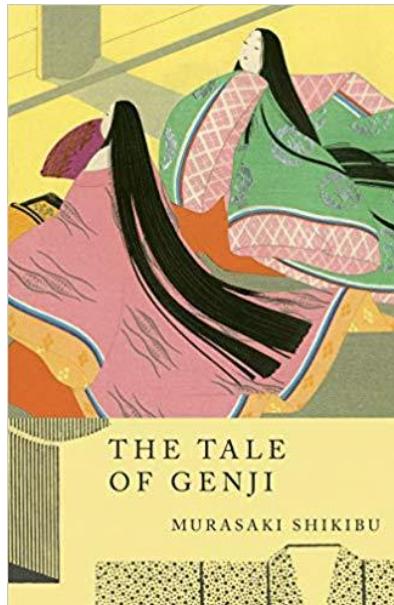
- France – Paris ≈ capital
- USD – US ≈ currency
- Bach – German ≈ nationality
- ...

Automatically predicting knowledge: 787B + ProducedBy ≈ Mazda

- A much less expensive way for knowledge acquisition
- Yet can still suffer from sparsity and noise of known knowledge [Pujara+, EMNLP-17]

Knowledge Is Not Isolated

Different sources of data can possess **complementary** knowledge



(The Tale of Genji, Genre, ?e)



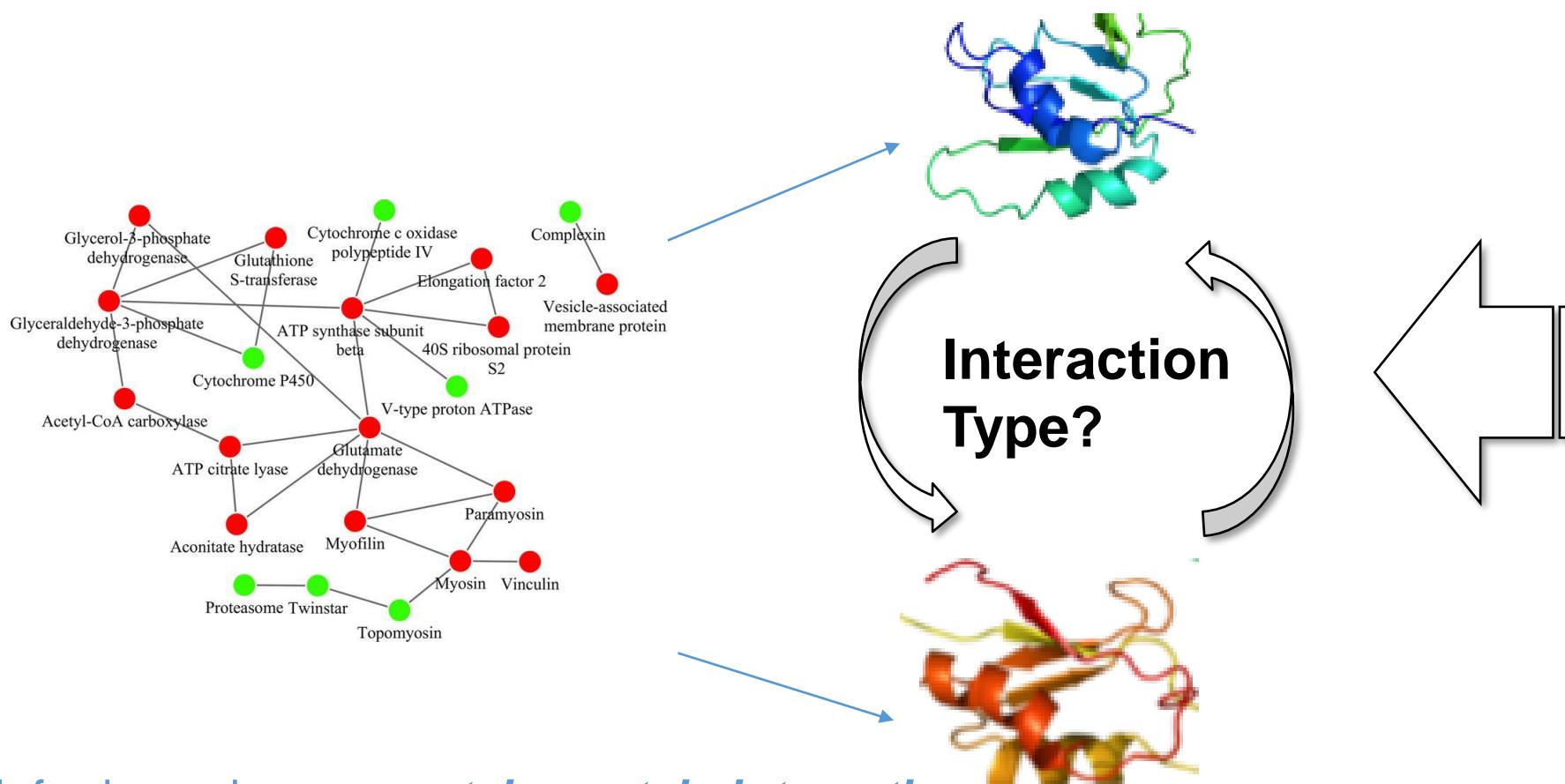
Novel



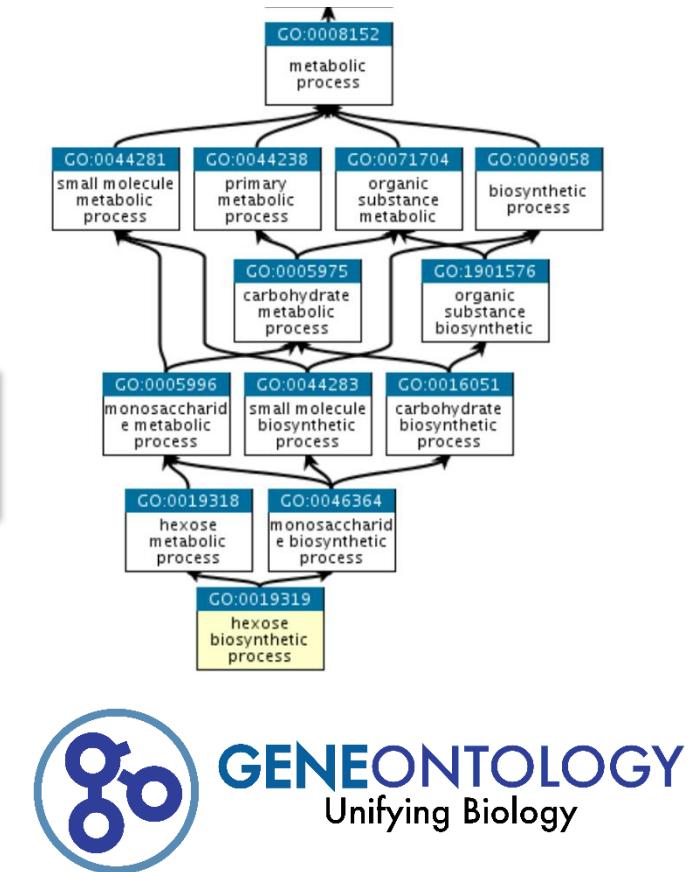
Monogatari (story)
Love story
Royal family story
Realistic novel
Ancient literature

Knowledge Is Not Isolated

Different sources of data can possess **complementary** knowledge



Inferring unknown **protein-protein interaction** information in a proteomic knowledge base

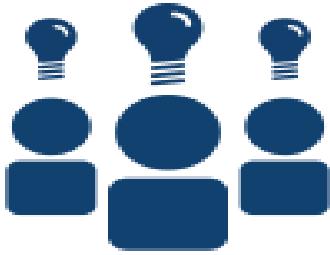


Interchangeable knowledge in various scenarios

- 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
- ...

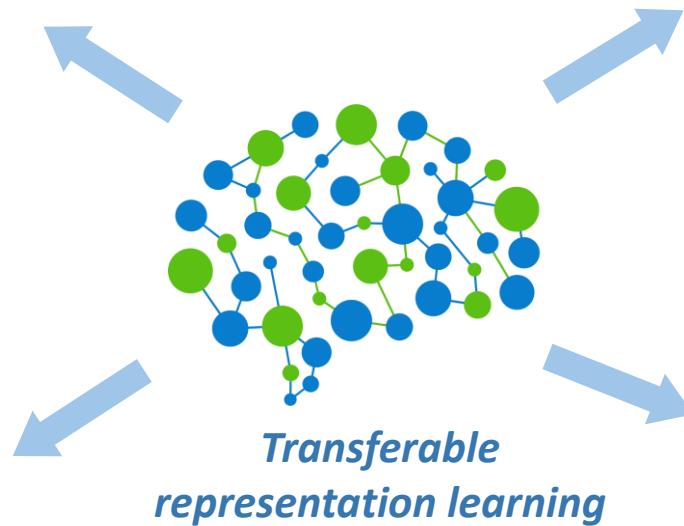
1. How to support the inference of relations in the embedding space?
2. How to capture the association of knowledge with minimal supervision?
3. How to leverage knowledge transfer to populate missing knowledge?

A General Methodology to Benefit A Wide Range of Tasks



Knowledge Base

Knowledge alignment
KG completion
Ontology population



Semantic search
Entity typing
Dialogue state tracking
Paraphrase identification

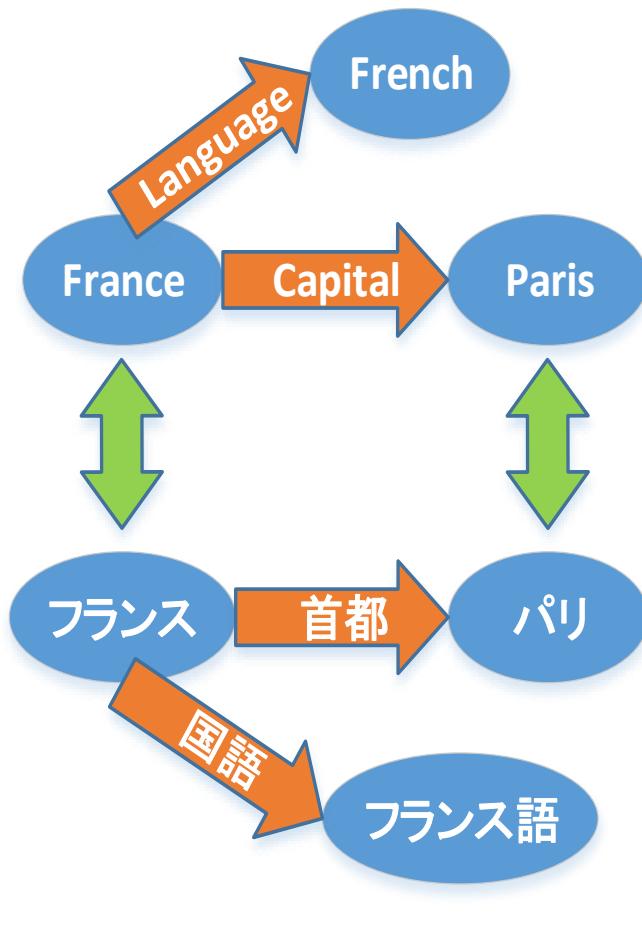


Protein-protein interaction prediction
Protein binding affinity estimation
Single cell RNA-sequence imputation
Gene Ontology term assignment



Polypharmacy side effect detection
Disease and phenotype matching
Clinical event prediction

Scenario 1: KGs in Different Languages



Separately created language-specific KGs

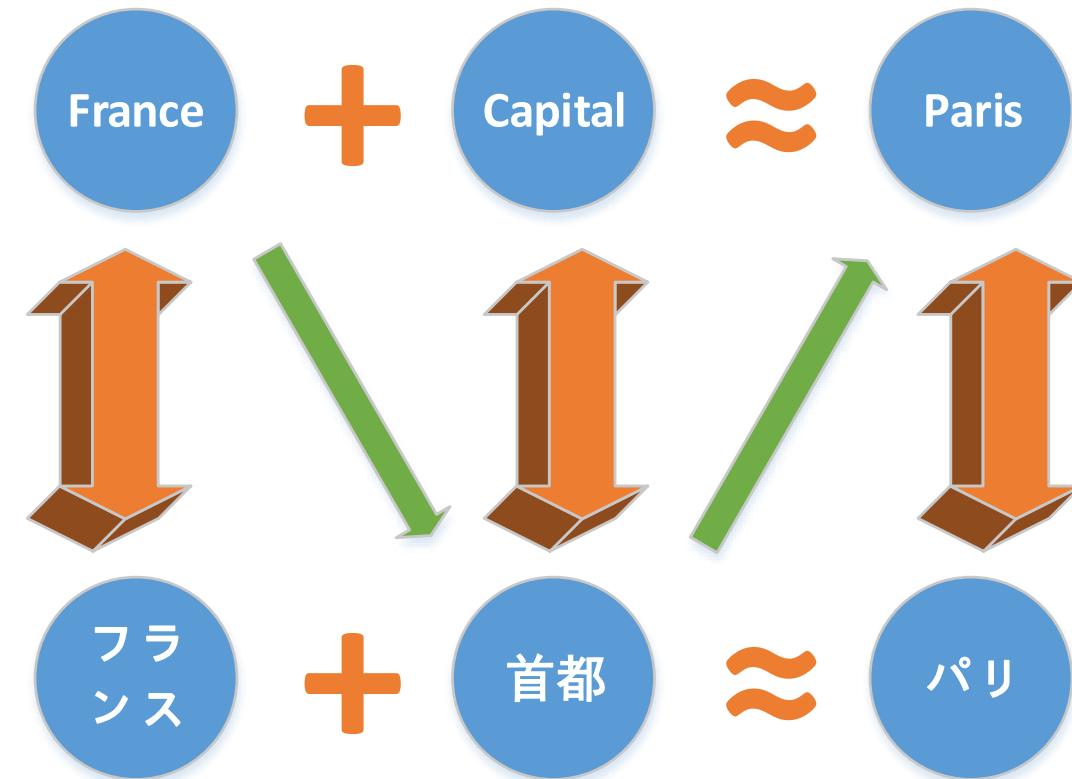
- DBpedia has 125 language-specific versions;
Wikidata has 410 of those.



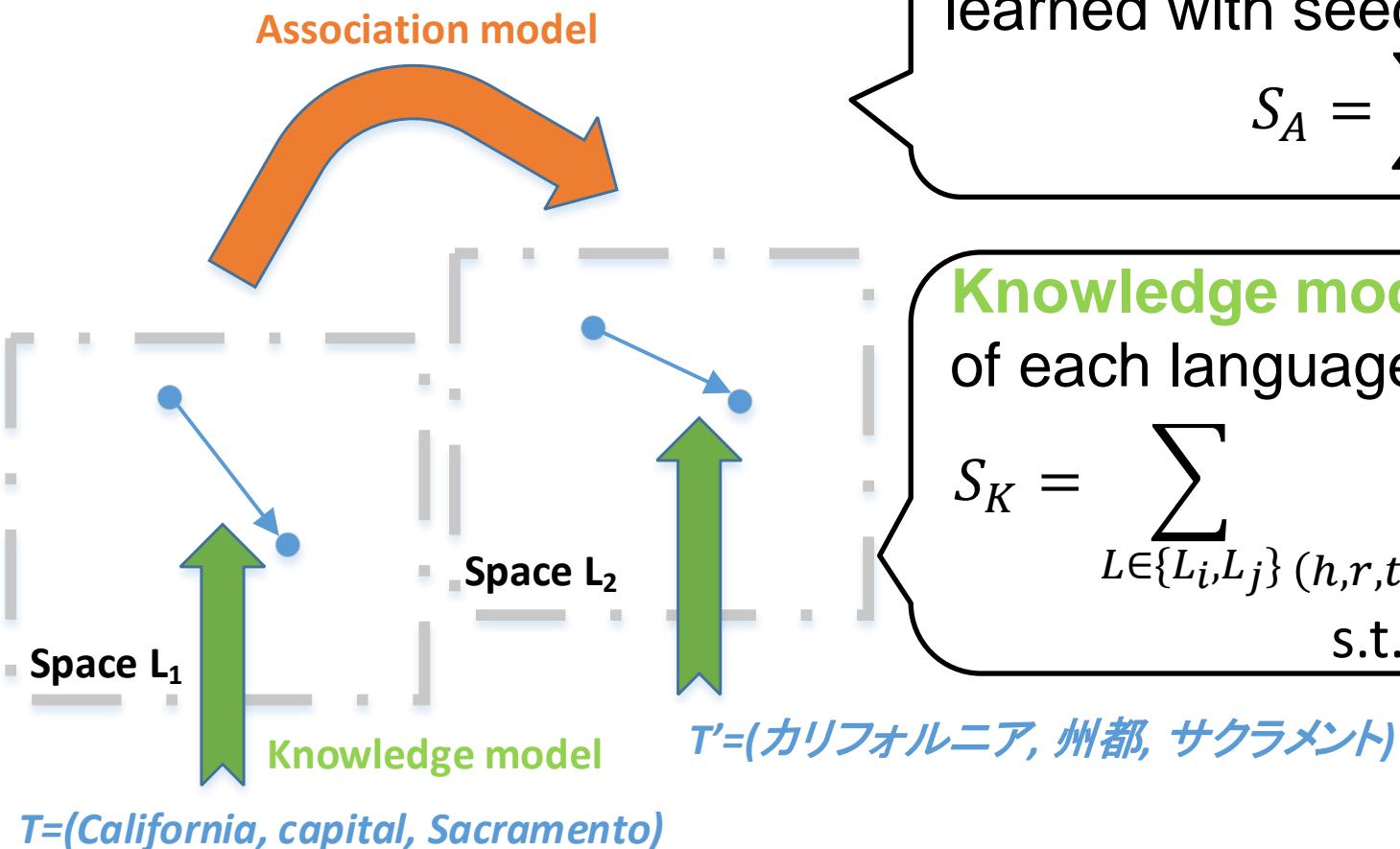
A Pilot Study: Simple Translational Model + Supervised Association Learning (MTransE*)

*[Chen+, IJCAI-17]

- **Training data:** a pair of weakly aligned language-specific KGs
- **Enabling:** cross-lingual semantic transfer + relational inference



Joint Learning of MTransE



Association model: an embedding transformation learned with seed alignment

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

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

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

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 , <i>George Bush</i> , <i>Jimmy Carter</i> , George Kalkoa
	German	Barack Obama , <i>Bill Clinton</i> , <i>George h. w. Bush</i> , Hamid Karzai
Paris	French	Paris , <i>Amsterdam</i> , <i>à Paris</i> , <i>Manchester</i> , De Smet
	German	Paris , <i>Languedoc</i> , <i>Constantine</i> , <i>Saint-maurice</i> , <i>Nancy</i>
California	French	<i>San Francisco</i> , <i>Los Angeles</i> , <i>Santa Monica</i> , Californie
	German	Kalifornien , <i>Los Angeles</i> , <i>Palm Springs</i> , <i>Santa Monica</i>

Table 9: Examples of cross-lingual relation matching.

Relation	Target	Candidates (in ascending order of rank by Euclidean distance)
capital	French	capitale , <i>territoire</i> , pays accréditant, lieu de vénération
	German	hauptstadt , <i>hauptort</i> , <i>gründungsstadt</i> , city
nationality	French	nationalité , pays de naissance , <i>domicile</i> , <i>résidence</i>
	German	nationalität , nation , <i>letzter start</i> , <i>sterbeort</i>
language	French	langue , réalisations, lieu deces, <i>nationalité</i>
	German	sprache , originalsprache , <i>lang</i> , <i>land</i>

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.

Cross-lingual Fact Prediction, e.g.

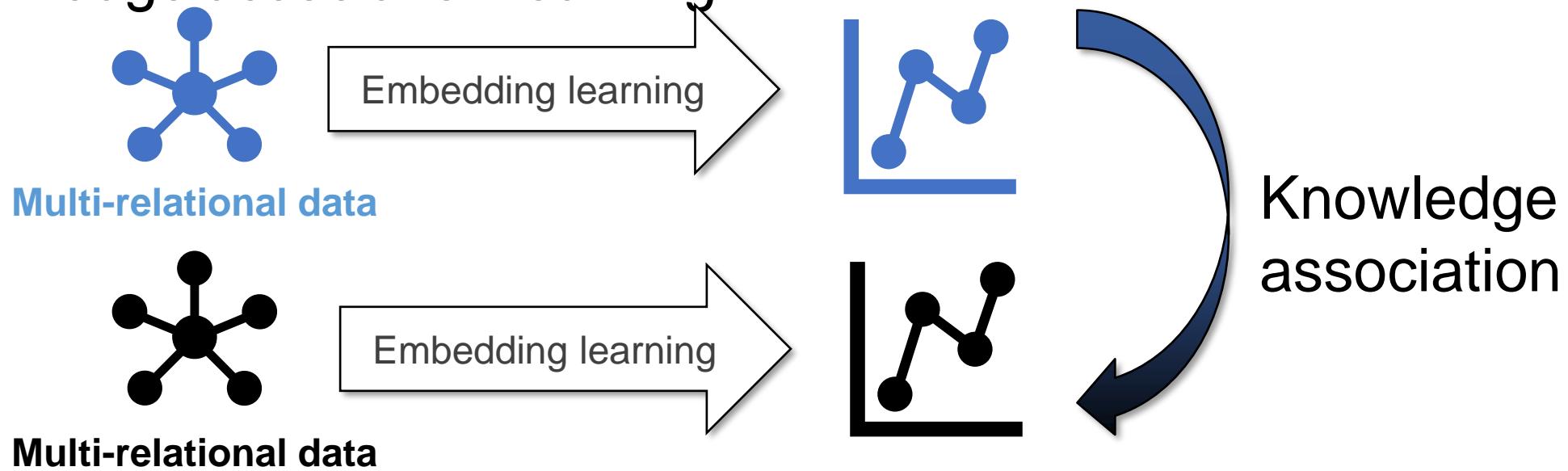
Table 10: Examples of cross-lingual triple completion.

Query	Target	Candidates (in ascending order of rank)
(Adam Lambert, genre, ?t)	French	<i>musique indépendante</i> , musique alternative , <i>ode</i> , glam rock
	German	popmusik , dance-pop , <i>no wave</i> , <i>soul</i>
(Ronaldinho, position, ?t)	French	milieu offensif , attaquant , <i>quarterback</i> , <i>latéral gauche</i>
	German	stürmer , <i>linker flügel</i> , angriffsspieler , <i>rechter flgel</i>
(Italy, ?r, Rome)	French	capitale , plus grande ville , chef-lieu , garnison
	German	hauptstadt , hauptort , verwaltungssitz, stadion
(Barack Obama, ?r, George Bush)	French	<i>ministre-président</i> , prédécesseur , <i>premier ministre</i> , <i>président du conseil</i>
	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

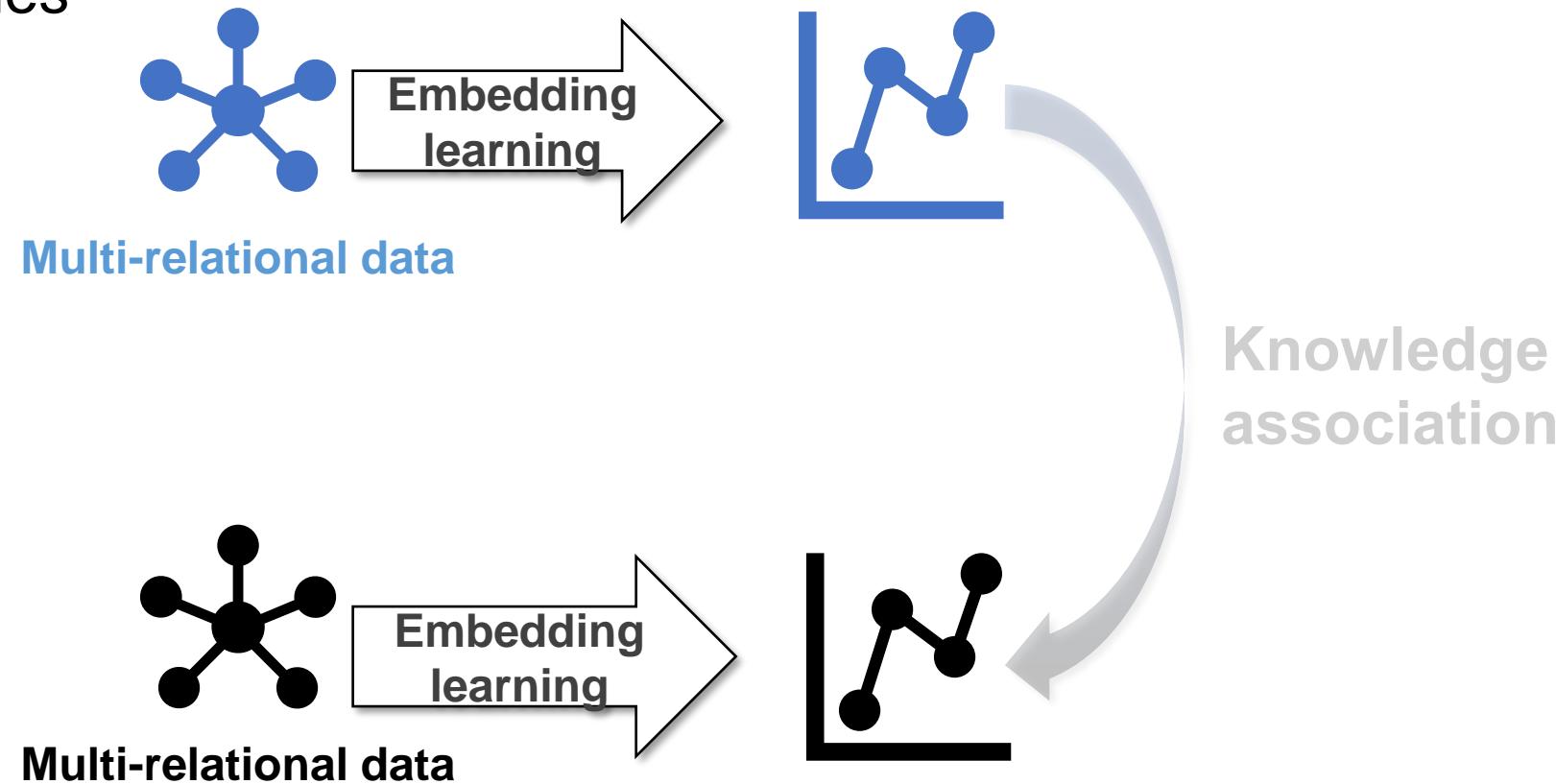
- Background and Motivation
- Relational embedding learning methods
 - First-order and high-order methods
 - Non-Euclidean methods
- Knowledge Association Methods
 - Supervised and semi-supervised methods
 - Auxiliary supervision methods
- Cross-domain and interdisciplinary tasks
 - KBP tasks
 - Computational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

The Embedding Learning Process

- Distributing each KG 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 $T=(h, r, t)$



High plausibility score

Training cases $T \in G$



Low plausibility scores



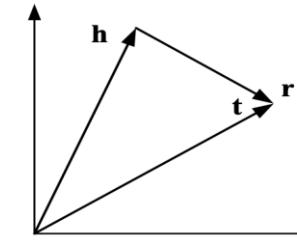
Negative samples $T' \notin G$

Plausibility Scoring Functions

- Translational technique (TransE [Bordes+, NIPS-13])

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

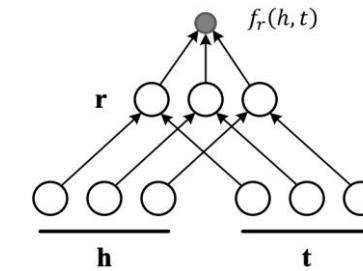
□ France + *Capital* ≈ Paris



- Element-wise product (DistMult [Yang+, ICLR-15])

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

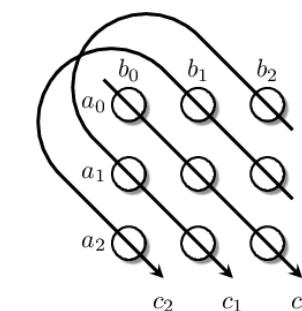
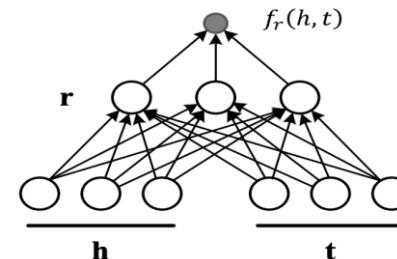
□ \circ denotes element-wise product



- Circular correlation (HoLE [Nickel+, AAAI-16])

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

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



$$c = a \star b$$

$$\begin{aligned} c_0 &= a_0 b_0 + a_1 b_1 + a_2 b_2 \\ c_1 &= a_0 b_2 + a_1 b_0 + a_2 b_1 \\ c_2 &= a_0 b_1 + a_1 b_2 + a_2 b_0 \end{aligned}$$

- Learning objective
 - Margin ranking loss

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

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

- $f_r(h, t)$: 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 ($\gamma_1 > \gamma_2$)

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

- Log softmax loss

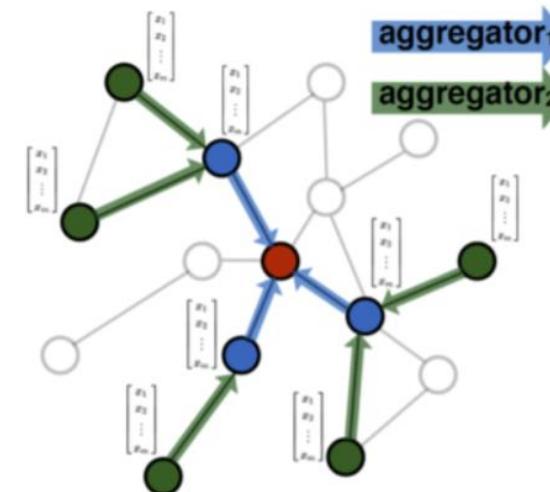
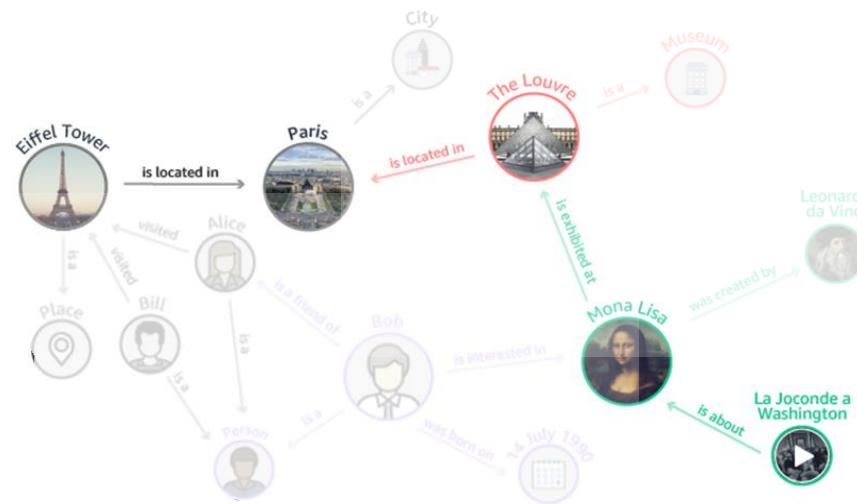
$$L_K = \sum_{T \in G} \log \frac{\exp(f_r(h, t))}{\sum_{T' \notin G} \exp(f_r(h', t'))}$$

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 of KGs to align
- 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](#) [Trsedy, AAAI-19], [MultiKE](#) [Zhang+, IJCAI-19], [OTEA](#) [Pei+, IJCAI-19], [SEA](#) [Pei+, WWW-19]
 - For entity typing: [JOIE](#) [Hao+, KDD-19]

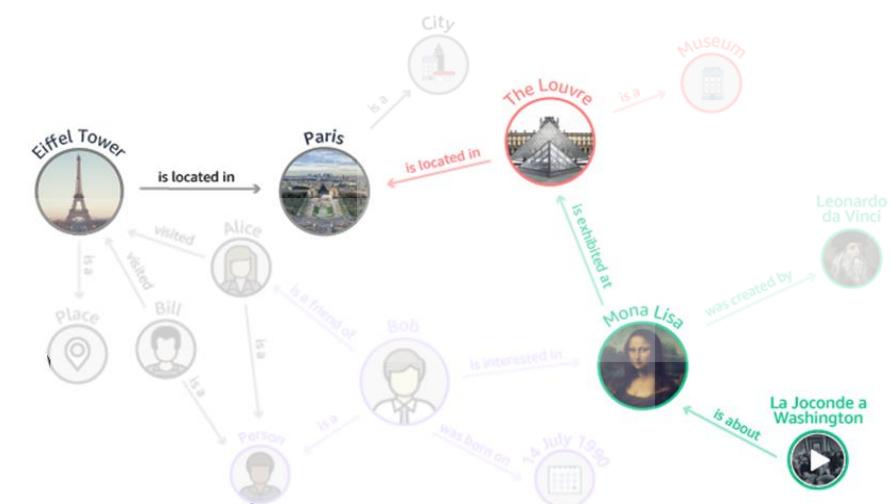
High-order Methods

- Modeling nodes (objects) based on contexts of the graph
- Two types of context modeling techniques
 - Relation path based techniques
 - Neighborhood aggregation techniques (e.g. Graph Neural Networks (GNNs))



Relation Path Based Techniques

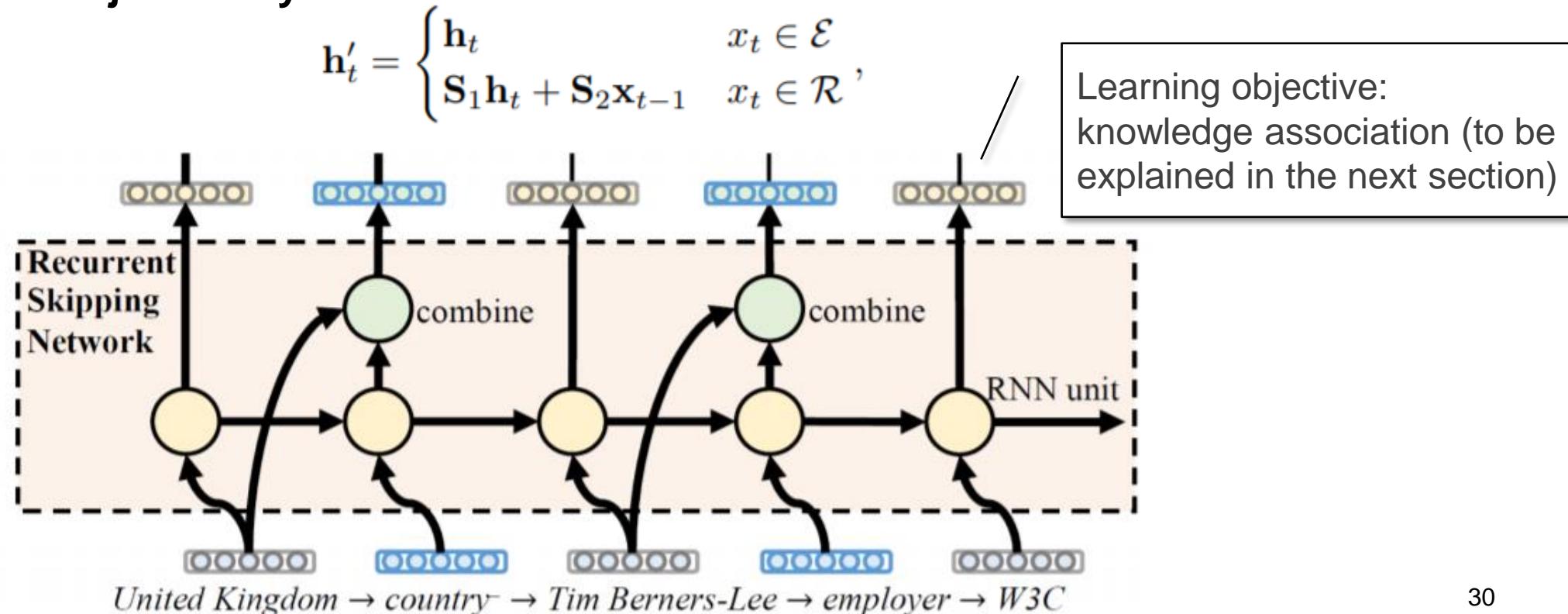
- A relation path is an entity-relation chain, where entities and relations appear alternately
 - United Kingdom → *country* → Tim Berners-Lee → *employer* → W3C
- PTransE [Lin+, EMNLP-15], Bi-Diag [Guu+, EMNLP-15]
 - Given l -length relation paths $p = (e_0, r_1, e_1, \dots, r_l, e_l)$
 - Minimize $\|\mathbf{e}_0 - \mathbf{p} + \mathbf{e}_1\|_2$
 - 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

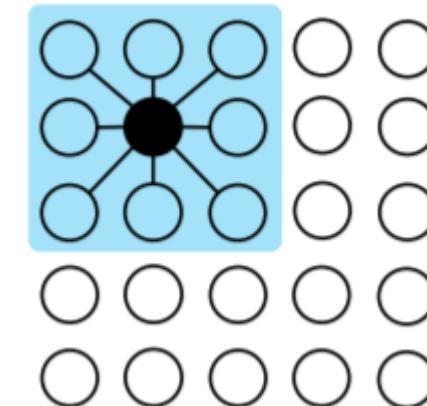
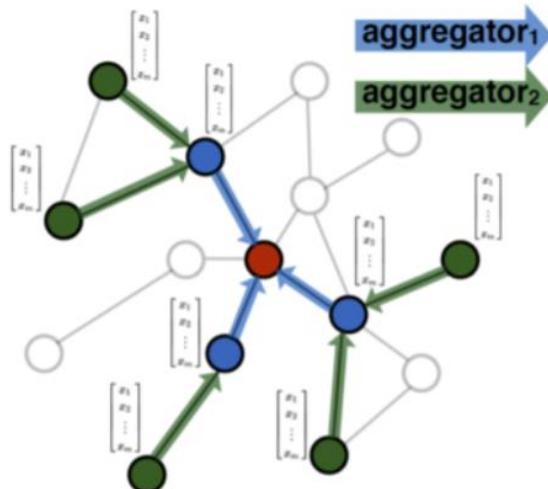
■ Recurrent skipping network (RSN, Guo+, ICML-19)

- 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**



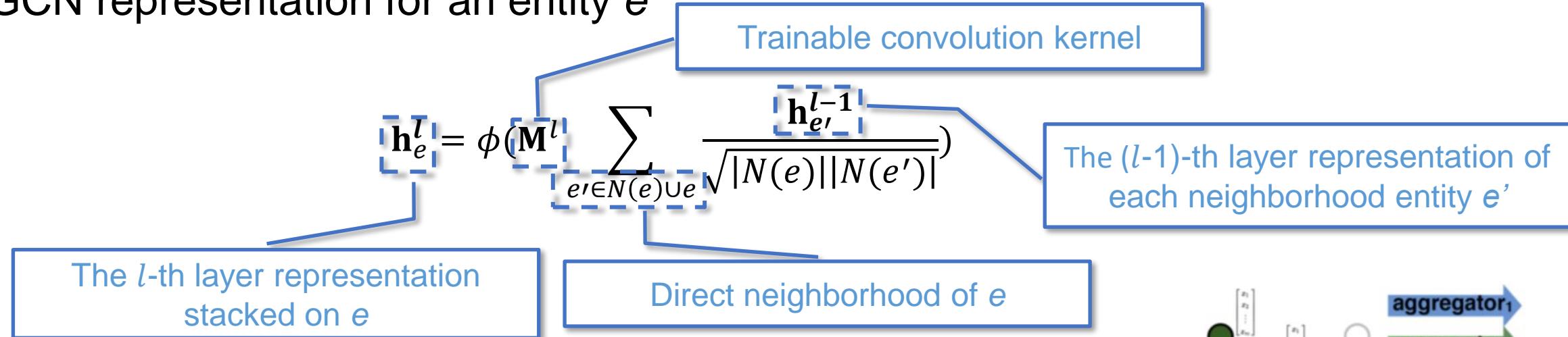
Neighborhood Aggregation Techniques

- 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



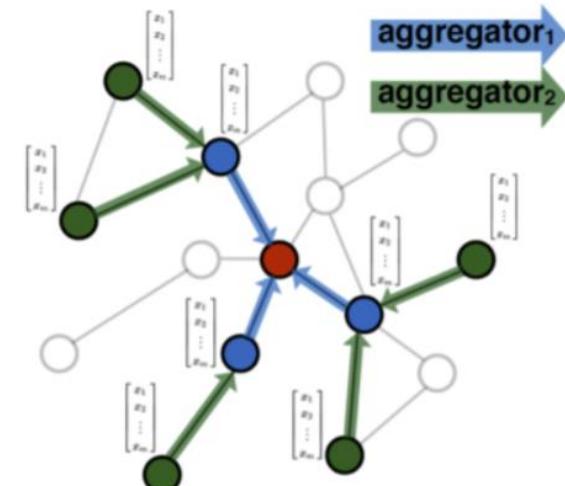
Neighborhood Aggregation Techniques

- GCN representation for an entity e



- 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]: relation-specific convolution kernels

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

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
 - Better capturing entity proximity (benefiting 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]

- Complex space relational embeddings

- CompIE [Trouillon+, ICML-16] ($Re(\cdot)$ denotes the real part)

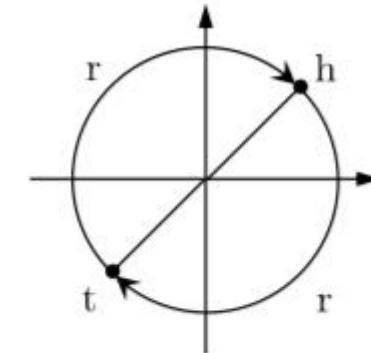
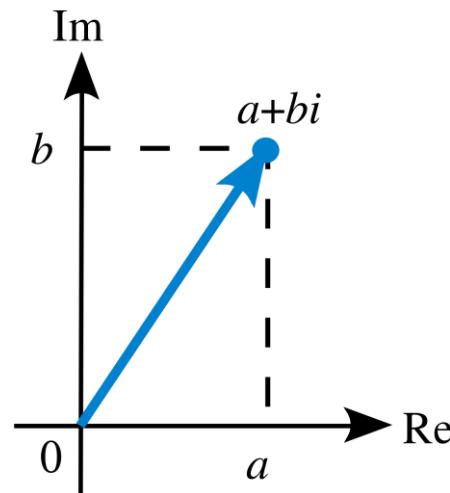
$$f_r(h, t) = Re((\mathbf{h} \circ \mathbf{t}) \cdot \mathbf{r}) \text{ s.t. } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$$

- 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

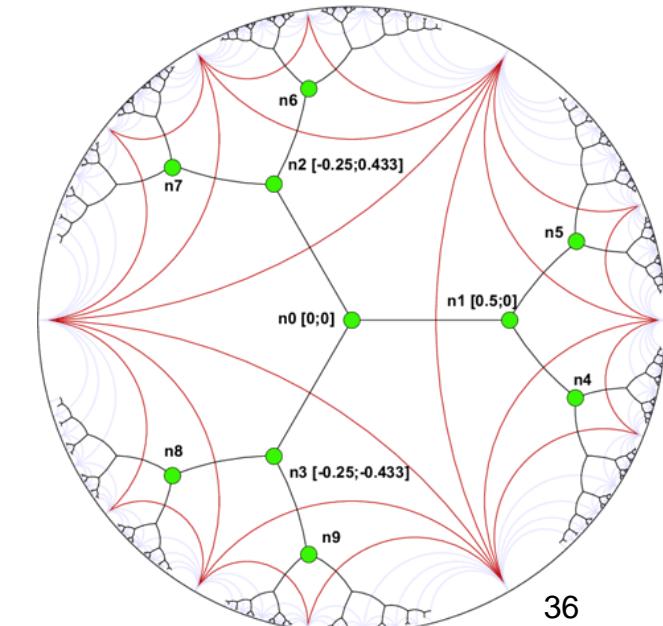


(c) RotatE: an example of modeling symmetric relations \mathbf{r} with $r_i = -1$

- The **hyperbolic space**: where the circumference grows **exponentially** w.r.t. the radius [Nickel+ NIPS-17, Ganea+ NeurIPS-18, Liu+ NeurIPS-19]

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

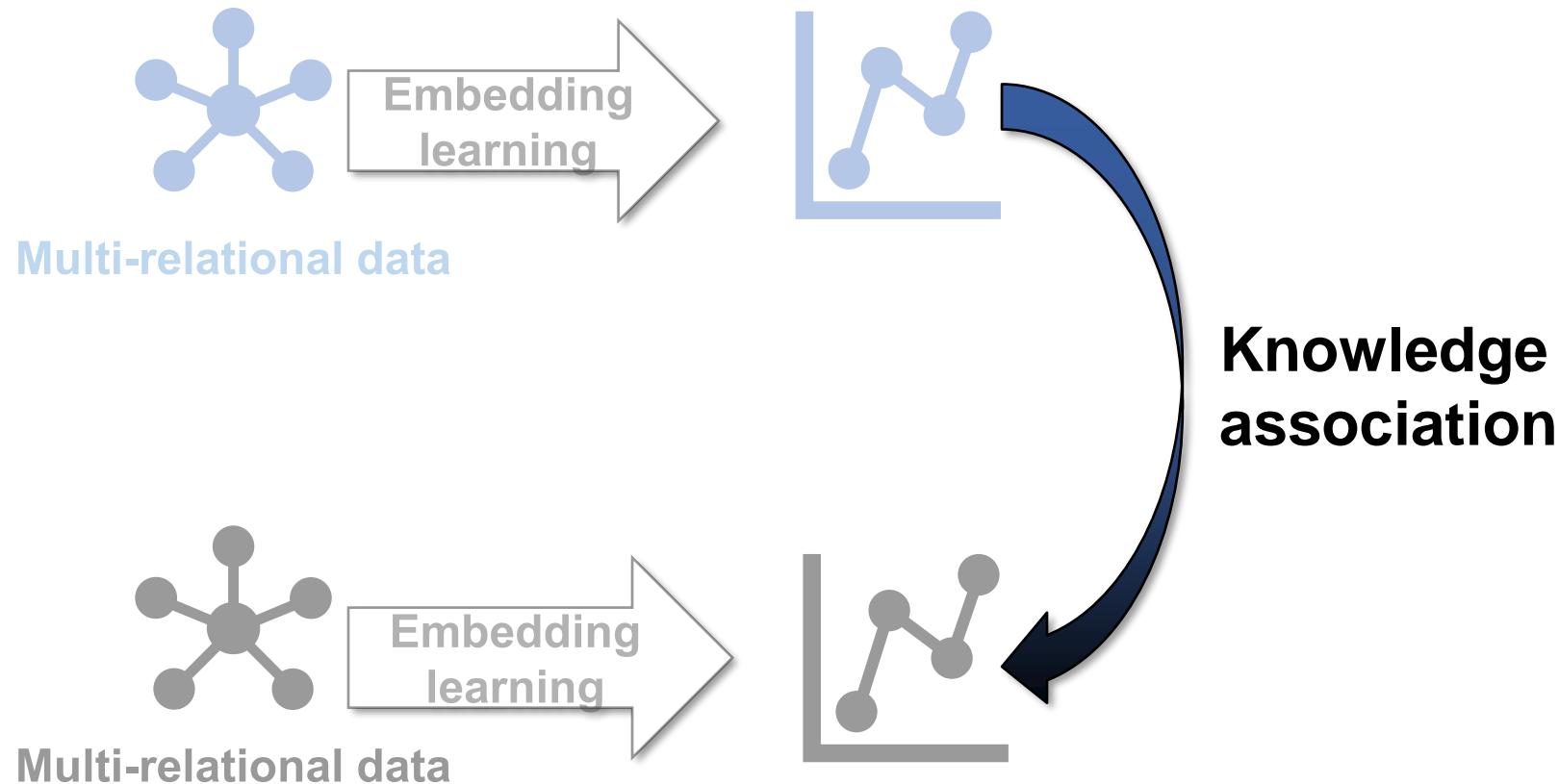
- Many data form hierarchies
 - Ontologies, taxonomies syntax trees, org charts, claim provenance in social media, etc.
- Hyperbolic representation learning models:
 - Hyperbolic Graph Embeddings [Nickel+ NIPS-17]
 - Hyperbolic Graph Neural Networks [Liu+ NeurIPS-19]



- 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)



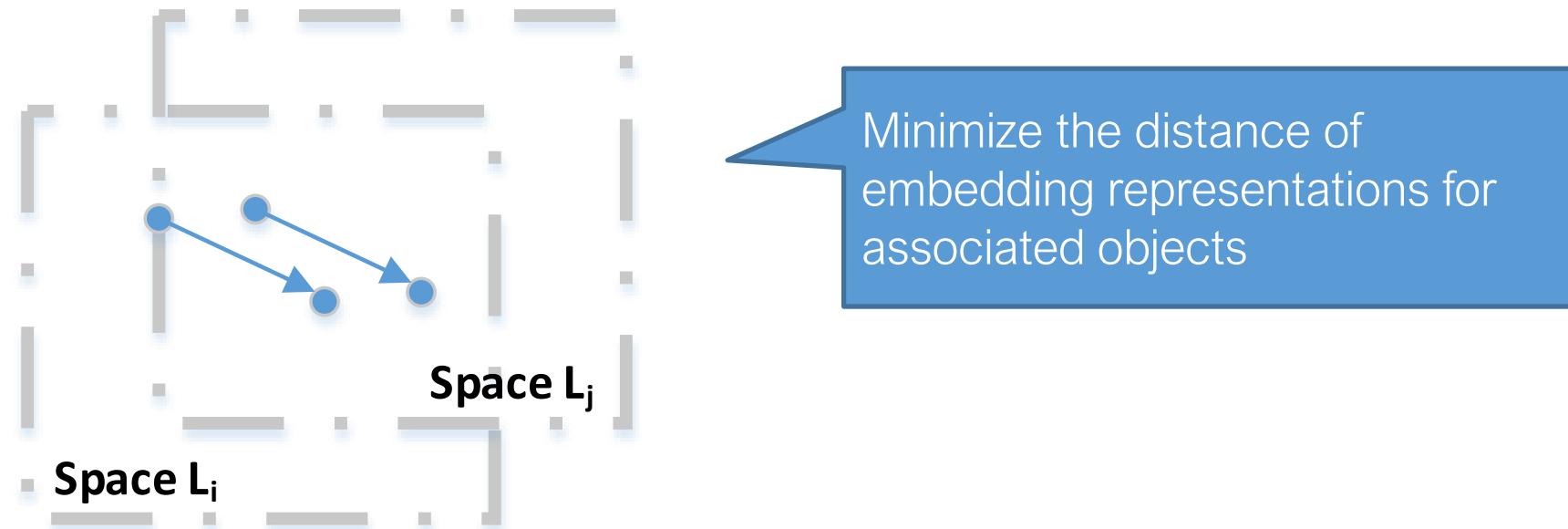
Several Key Questions

- 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)

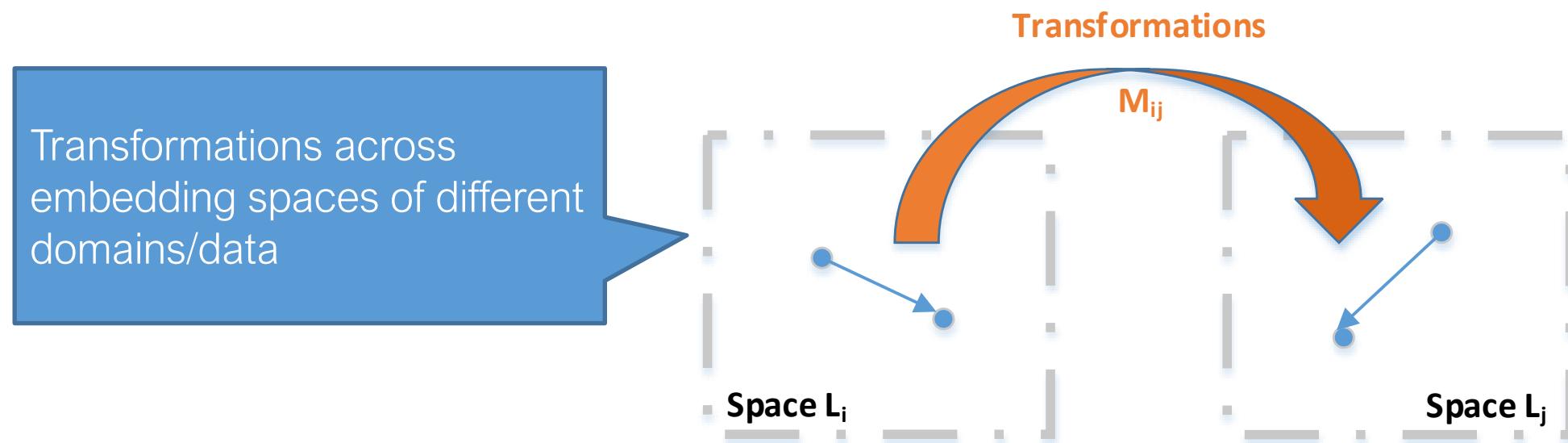


- Suitable for associations between data with similar structures and sizes (e.g. 1-to-1 entity alignment between well-populated multilingual KGs).

Geometric Forms of Embedding Association

- Transformation-based association

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



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

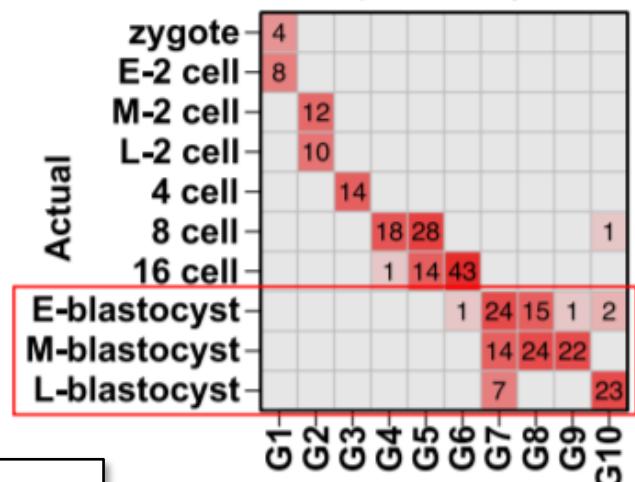
- **Matrix factorization** (Given S as the alignment data matrix)

- Minimize $\|S - \mathbf{E}_1 \mathbf{E}_2^T\|$
- \mathbf{E}_1 and \mathbf{E}_2 are both of dim k

- **Matrix tri-factorization**

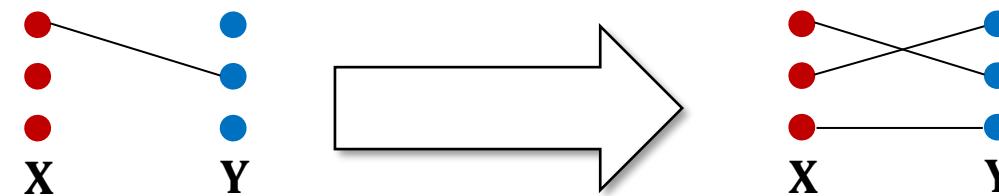
- Minimize $\|S - \mathbf{E}_1 \mathbf{U} \mathbf{E}_2^T\|$
- \mathbf{E}_1 and \mathbf{E}_2 are of dim k_1 and k_2
- \mathbf{U} is a $k_1 \times k_2$ matrix

Allows two embedding spaces to be of different dims



Semi-supervised learning strategies

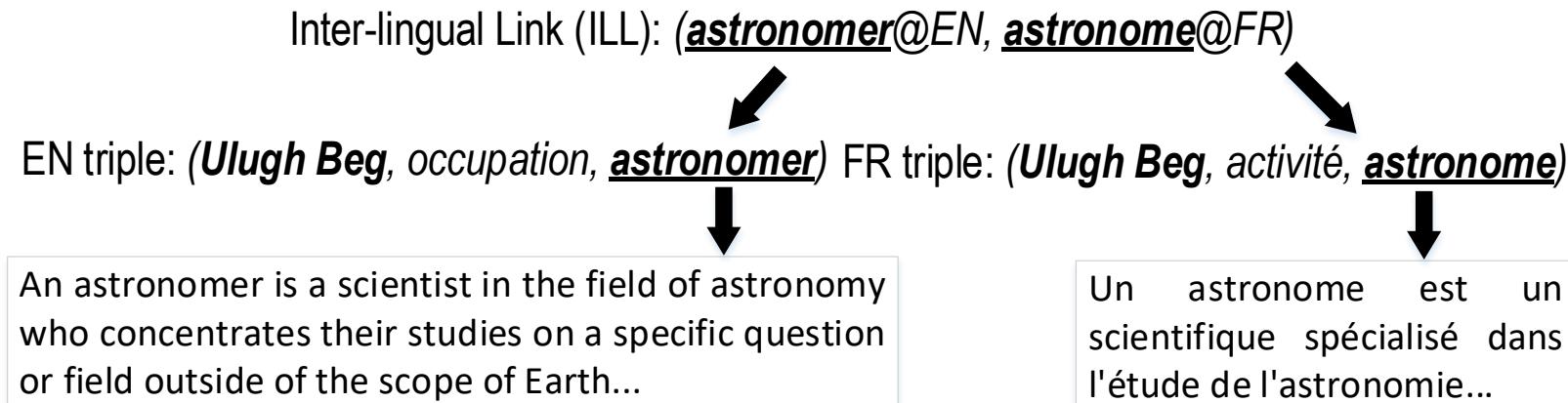
- Bootstrapping (BootEA [Sun+, IJCAI-18])
 - Iteratively suggesting new alignment labels for unaligned entities in training
 - Distance-based association
 - A new label (e, e') is added to training if
 - The embedding distance of e, e' are within a threshold
 - They are mutually nearest neighbor of each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



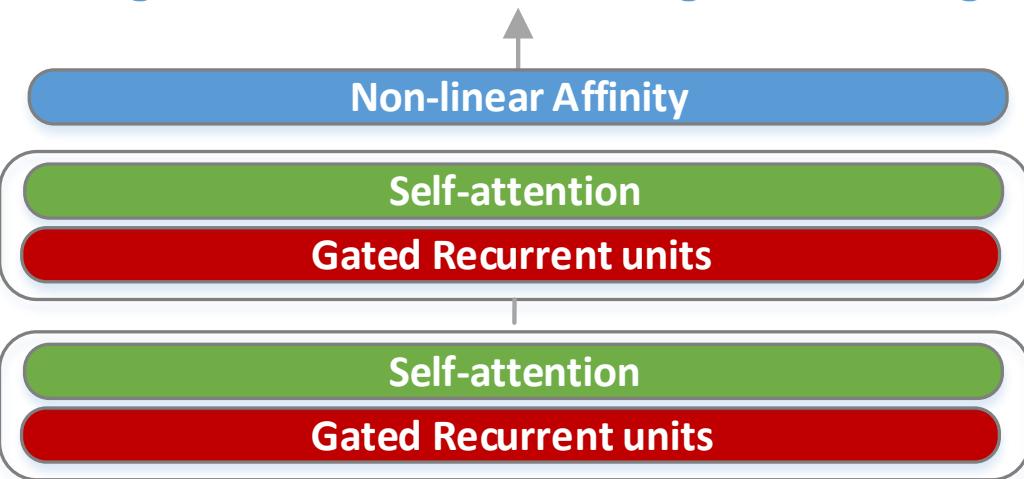
Semi-supervised learning strategies

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.

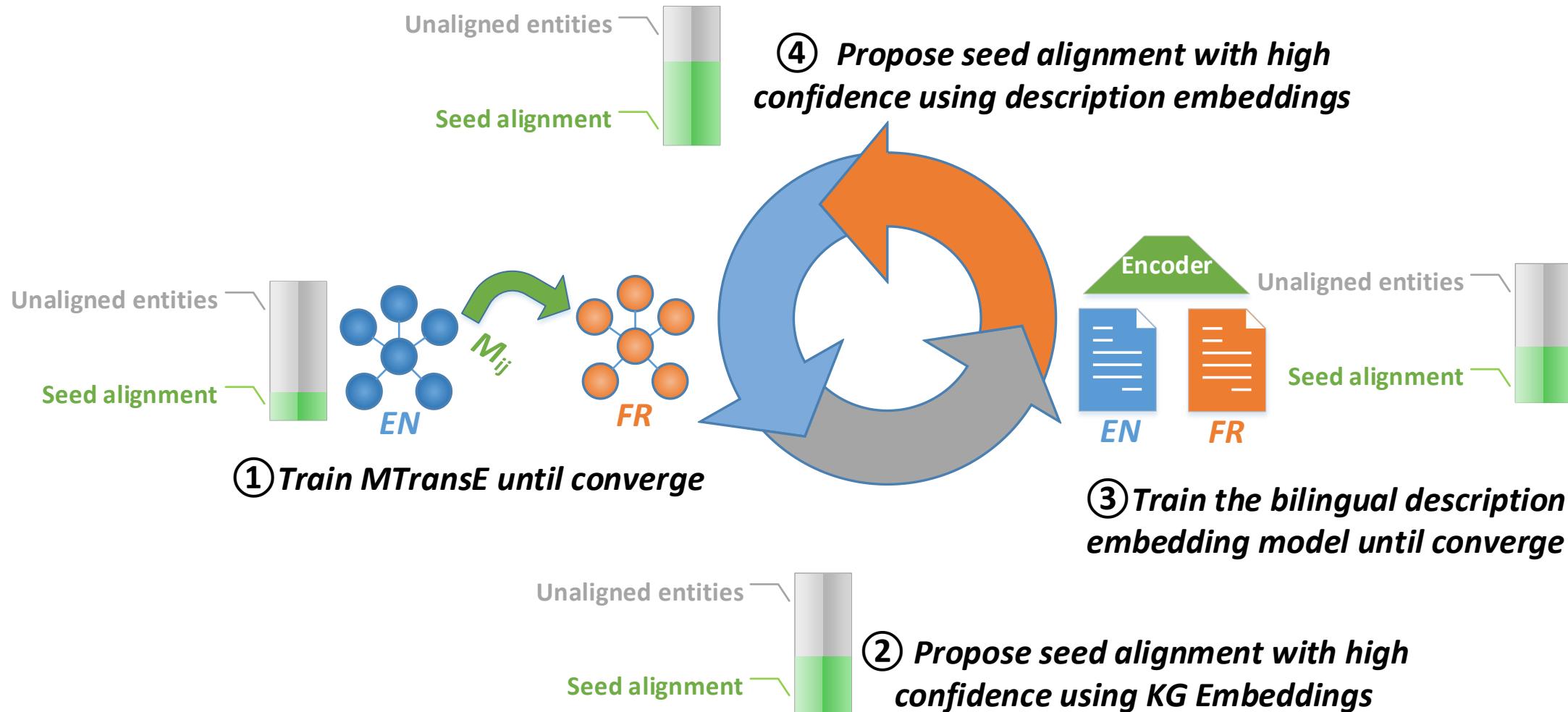
Logistic Loss + Stratified negative sharing



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...

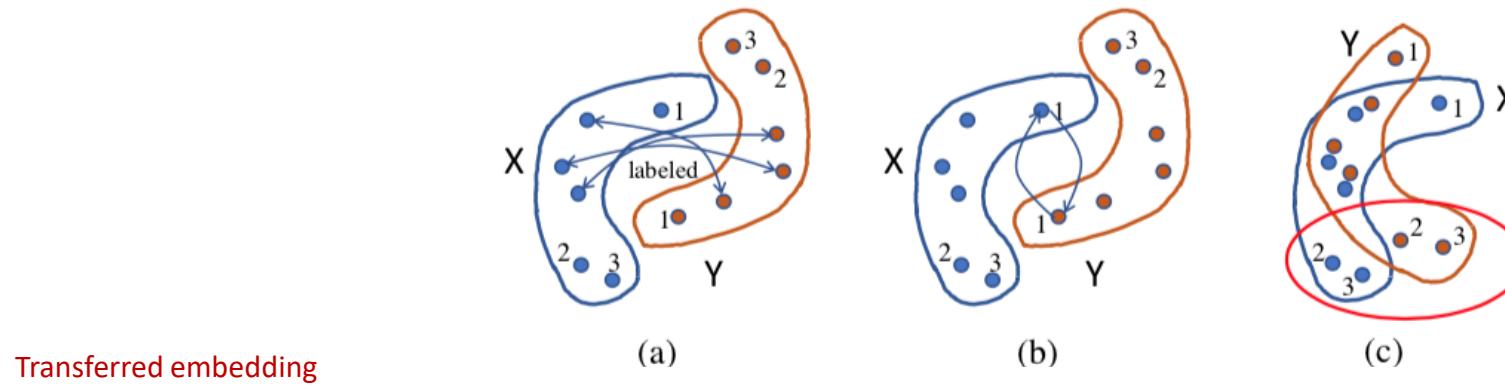
Un **astronome** est un scientifique spécialisé dans l'étude de l'astronomie...

Semi-supervised learning strategies



Semi-supervised learning strategies

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



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

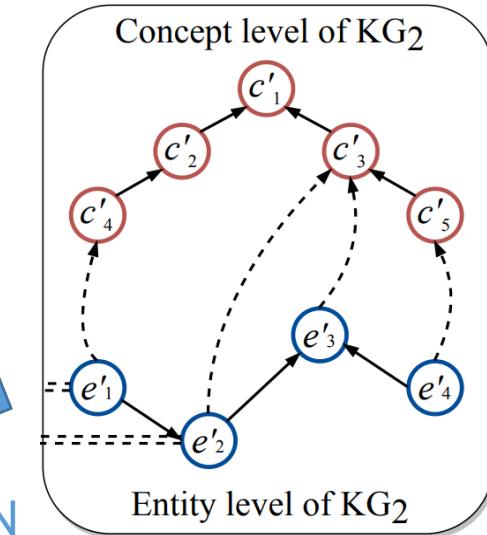
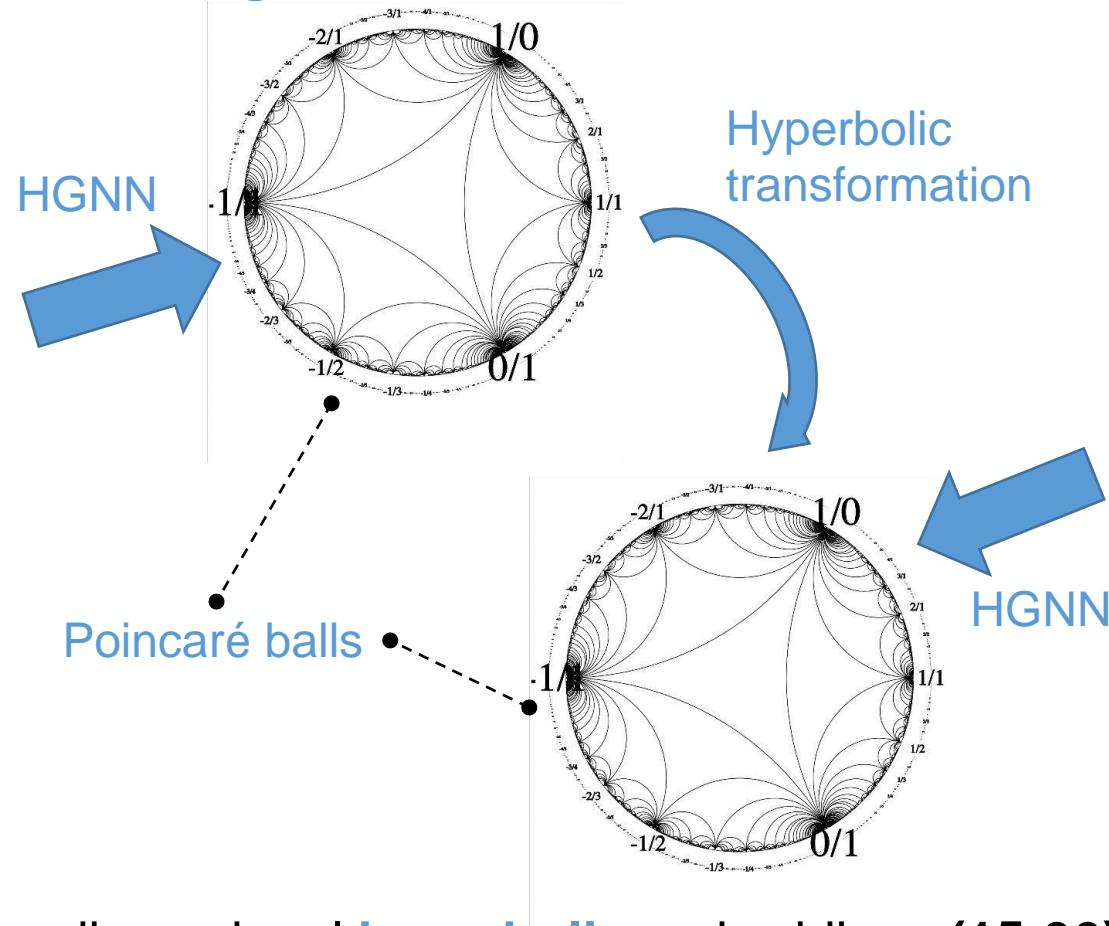
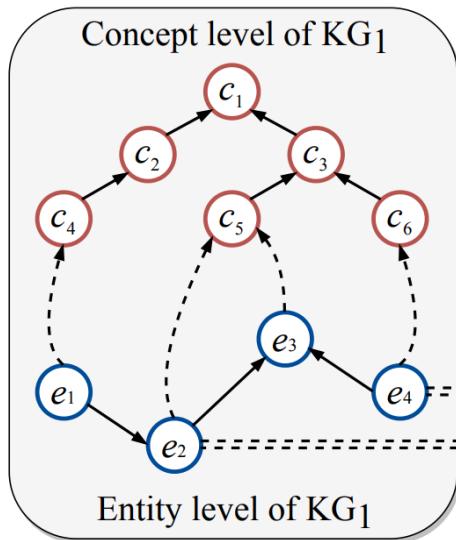
WGAN
→

$$\mathcal{L}_{g_1} = \min_{\mathbf{M}^1} \max_{D_1} \mathbb{E}_{y \sim \mathbf{q}^{\theta_e^j}} [f_{D_1}(y)] - \mathbb{E}_{x \sim \mathbf{p}^{\theta_e^i}} [f_{D_1}(\mathbf{M}^1 x)]$$
$$\mathcal{L}_{g_2} = \min_{\mathbf{M}^2} \max_{D_2} \mathbb{E}_{y \sim \mathbf{q}^{\theta_e^i}} [f_{D_2}(y)] - \mathbb{E}_{x \sim \mathbf{p}^{\theta_e^j}} [f_{D_2}(\mathbf{M}^2 x)]$$

Optimization for
embedding
transformations of
two directions

(3) Hyperbolic Knowledge Association*

*[Sun+, EMNLP-20]



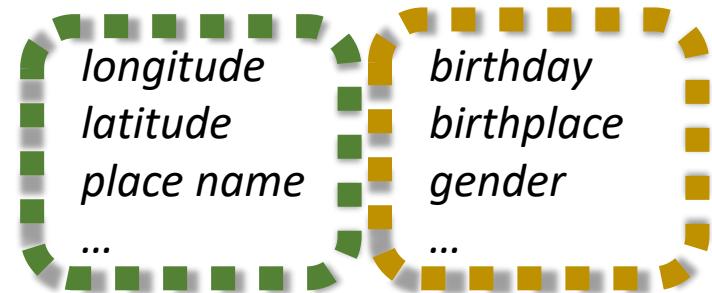
Transferable, ultra low-dimensional **hyperbolic** embeddings (15-30).

- Non-linear distance metric $d_{\mathbb{D}}(\mathbf{u}, \mathbf{v}) = \text{arccosh}\left(1 + 2 \frac{\|\mathbf{u} - \mathbf{v}\|^2}{(1 - \|\mathbf{u}\|^2)(1 - \|\mathbf{v}\|^2)}\right)$
- Suitable for capturing knowledge association between **hierarchical** KGs.
- and KGs with **significantly different scales** (e.g. an instance-graph vs a concept graph).

Also applied to entity type inference.

Learning with Auxiliary Information

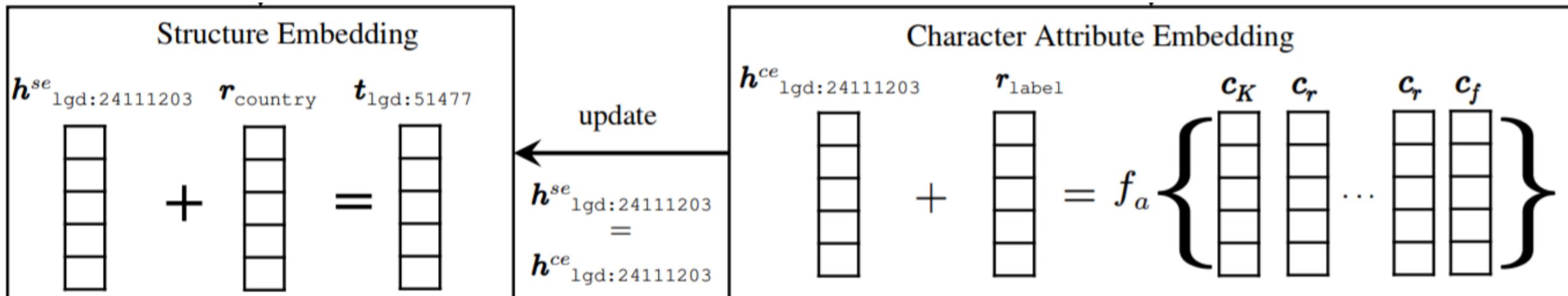
- Attribute-based embedding association
- JAPE [Sun+, ISWC-17]:
 - Using a weighted Skip-gram language model [Mikolov+, NIPS-13] that predicts entities based on attributes



Entities with correlated attributes will have similar embedding vectors.

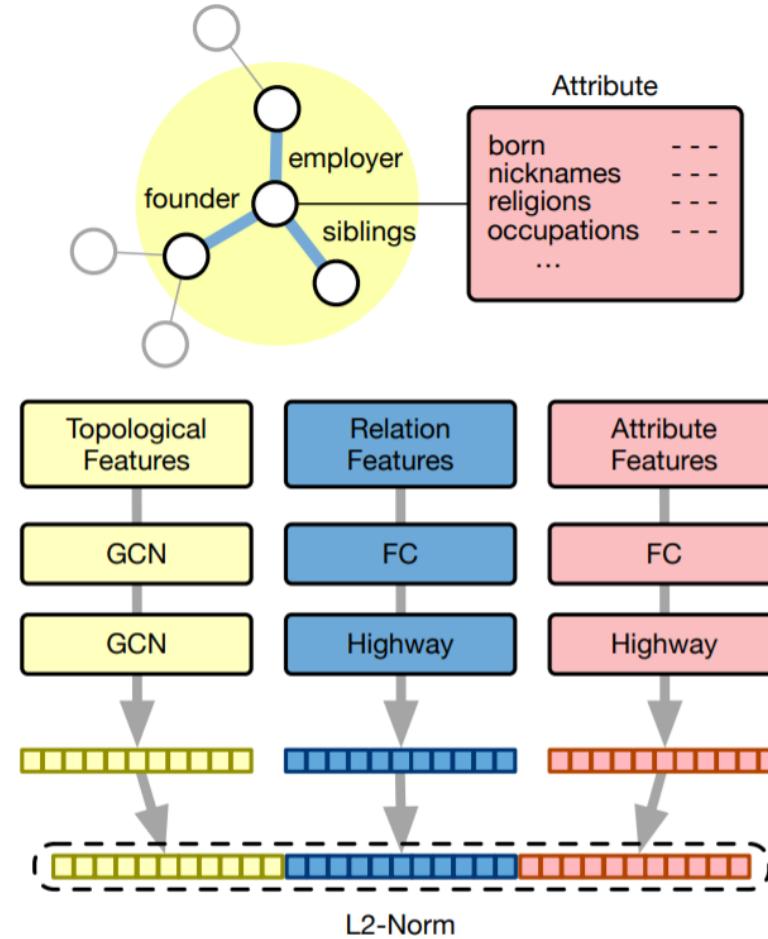
Learning with Auxiliary Supervision

- 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)

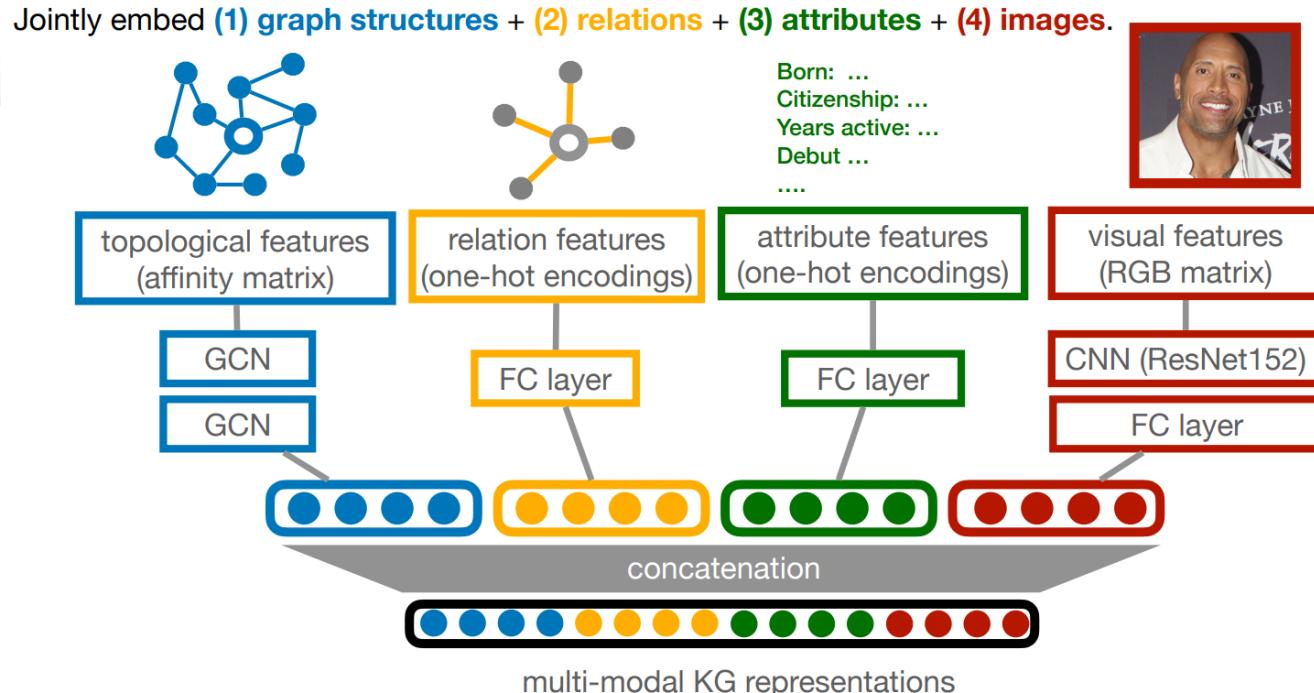


HMAN [Yang+, EMNLP-19]

Learning with Auxiliary Supervision

- Visual pivoting [Liu+, AAAI-21]
 - Find the correspondence of entities based on images
 - The visual modalities helps aligning concrete entities

The screenshot shows a Wikipedia page for "Dwayne Johnson". The top navigation bar includes links for "Page", "Talk", "Read", "Change", "Change source", "View history", and a search bar. Below the header, the title "Dwayne Johnson" is displayed, followed by a brief summary: "Dwayne 'The Rock' Johnson OC (born May 2, 1972 in Hayward, California) is an American actor and semi-retired professional wrestler. He is best known for his work with the WWE." A sidebar on the left contains a "Contents" section with links to various sections of the article, including "Early life", "Professional wrestling", "Acting", "Personal life", "References", and "Other websites". To the right of the summary is a large portrait photo of Dwayne Johnson, with the caption "Johnson in March 2013".



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>

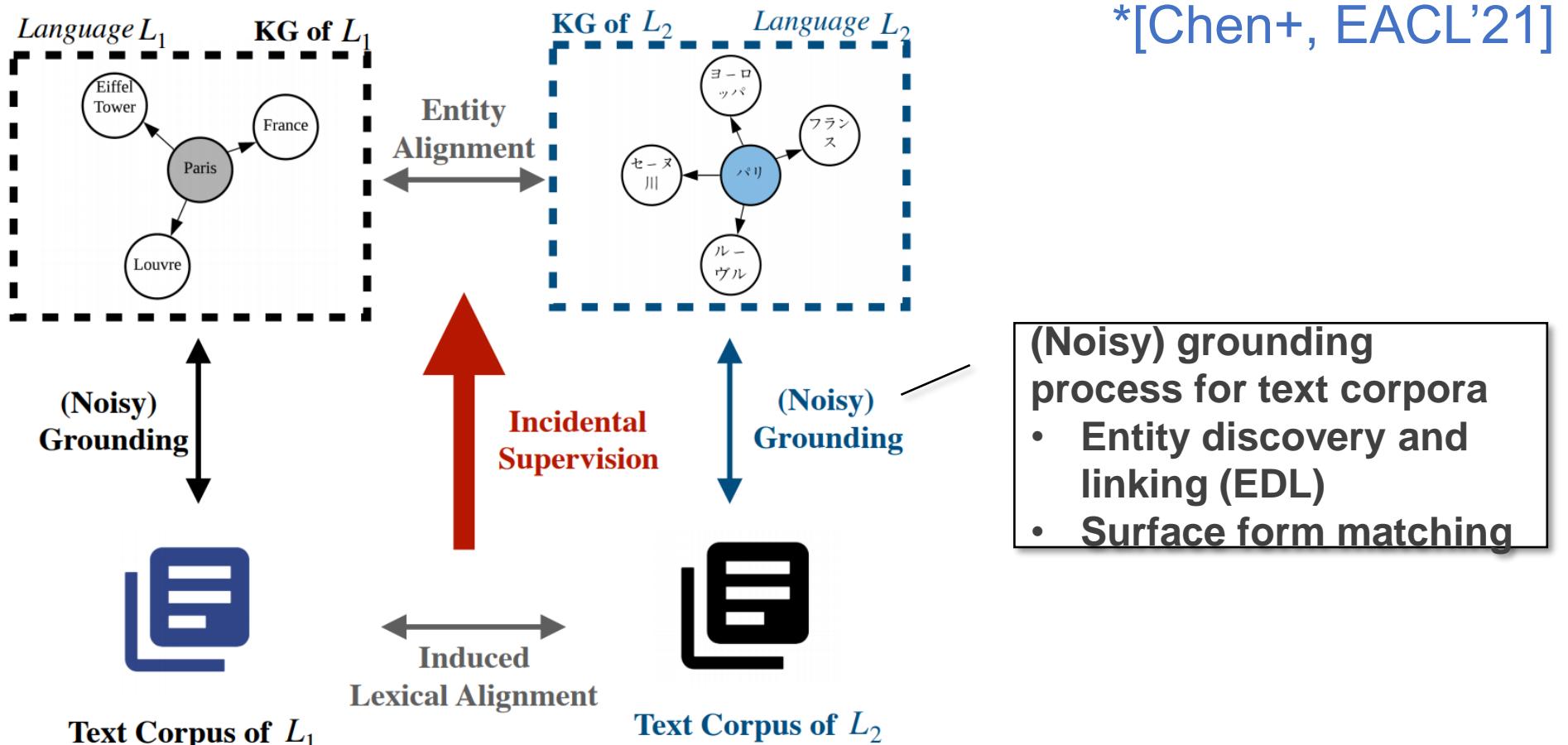
Datasets	# Ent.	# Rel.	# Attr.	# Rel tr.	# Attr tr.
DBP-WD	DBpedia	100,000	330	351	463,294
	Wikidata	100,000	220	729	448,774
DBP-YG	DBpedia	100,000	302	334	428,952
	YAGO3	100,000	31	23	502,563

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

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

Survey: Sun+. A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs. **PVLDB**, Vol. 13. 2020

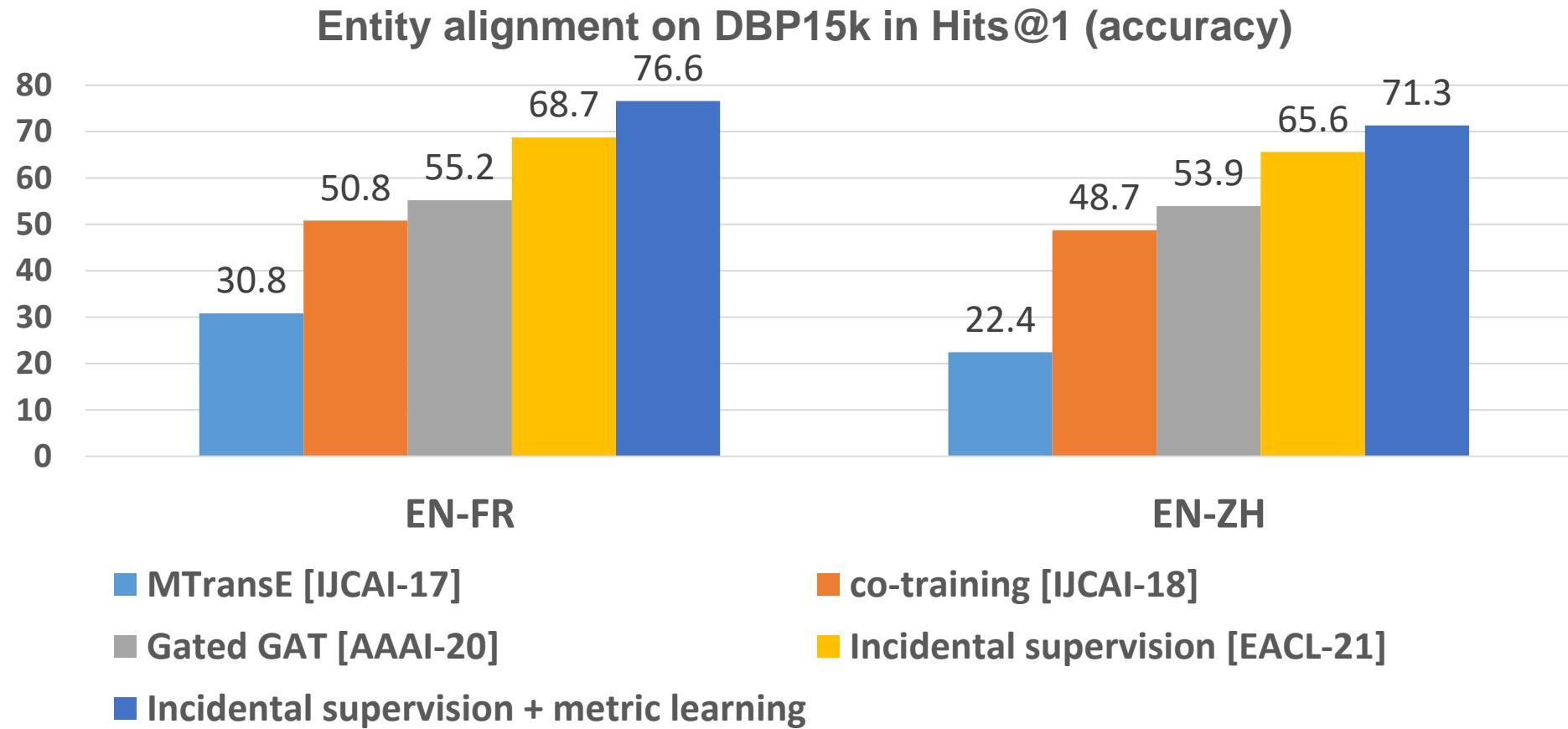
Entity Alignment Using Incidental Supervision From Free Text*



Three steps

1. **(Noisy) grounding:** connecting KGs and text corpora
2. **Embedding learning:** GNN + a neural language model
3. **Alignment learning:** self-learning for both entity and lexical alignment

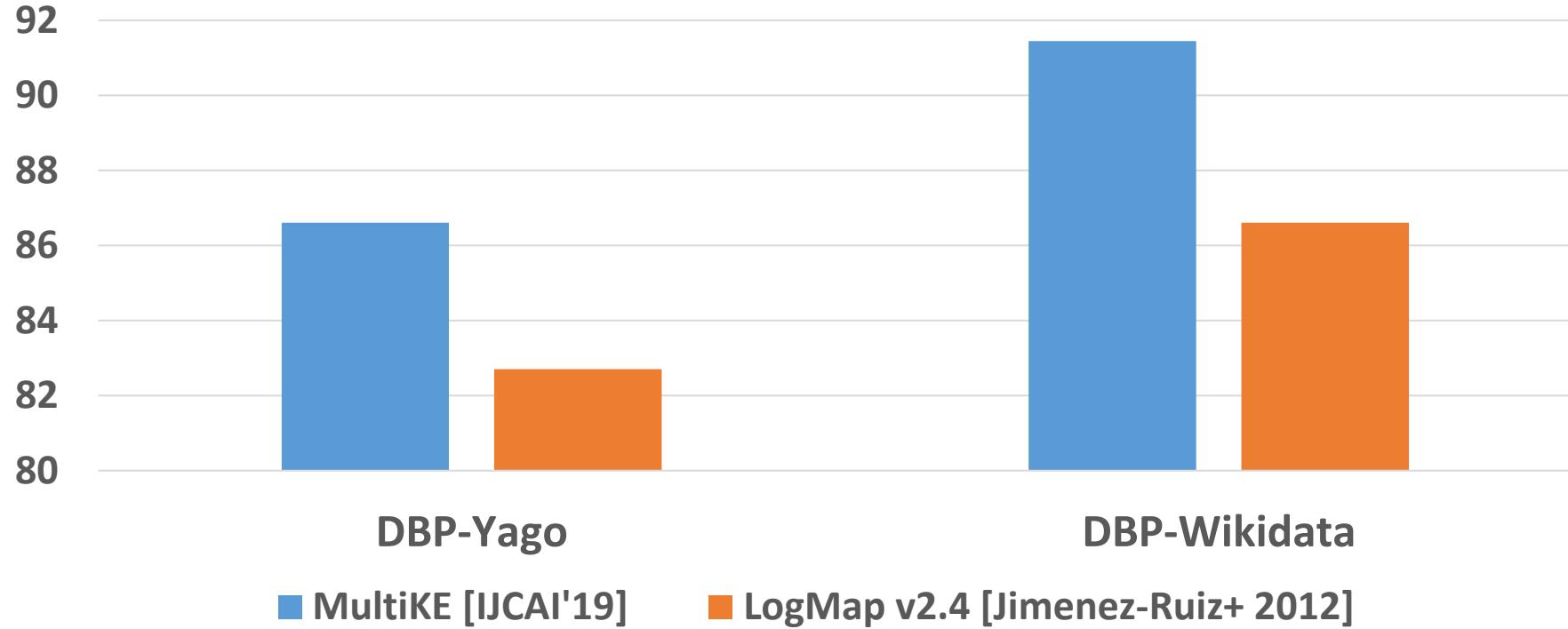
Cross-lingual Entity Alignment Results



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

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

Multi-KE vs. LogMap2.4 on Aligning Subsets of DBPedia to Yago and Wikidata

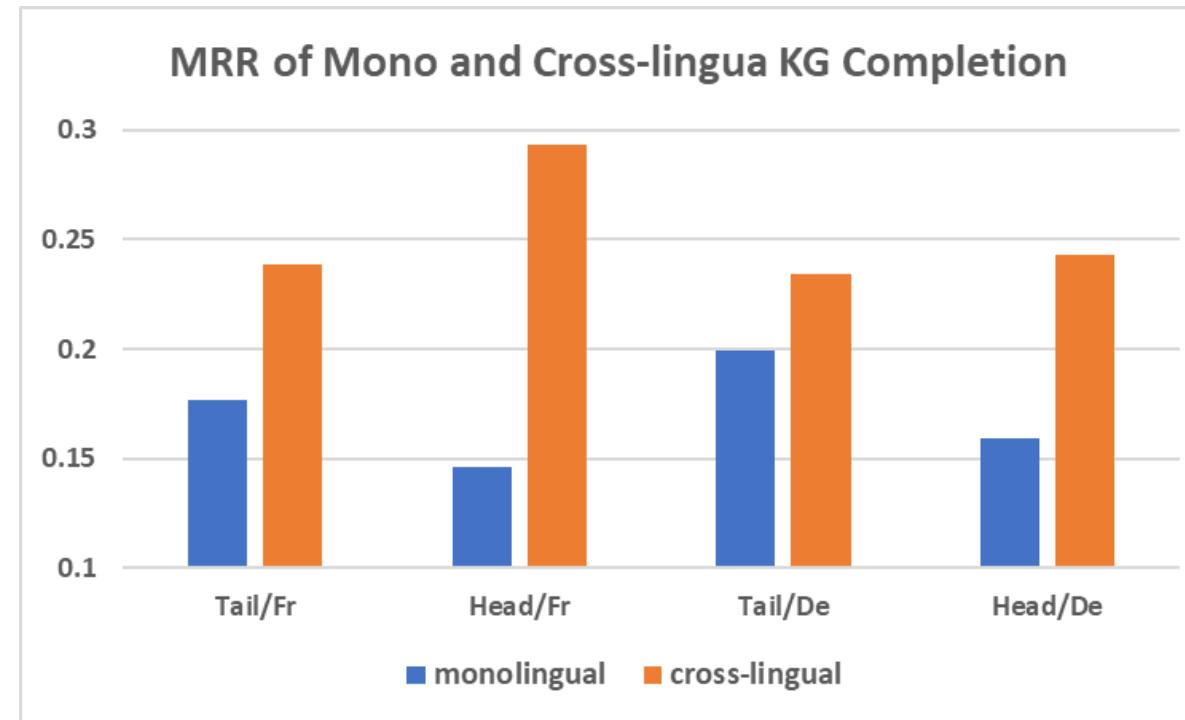


***MultiKE [Sun+ IJCAI'19]** is a monolingual ontology matching system in which multi-view embeddings of structures, literals, descriptions and attributes are combined.

Cross-lingual Knowledge Transfer

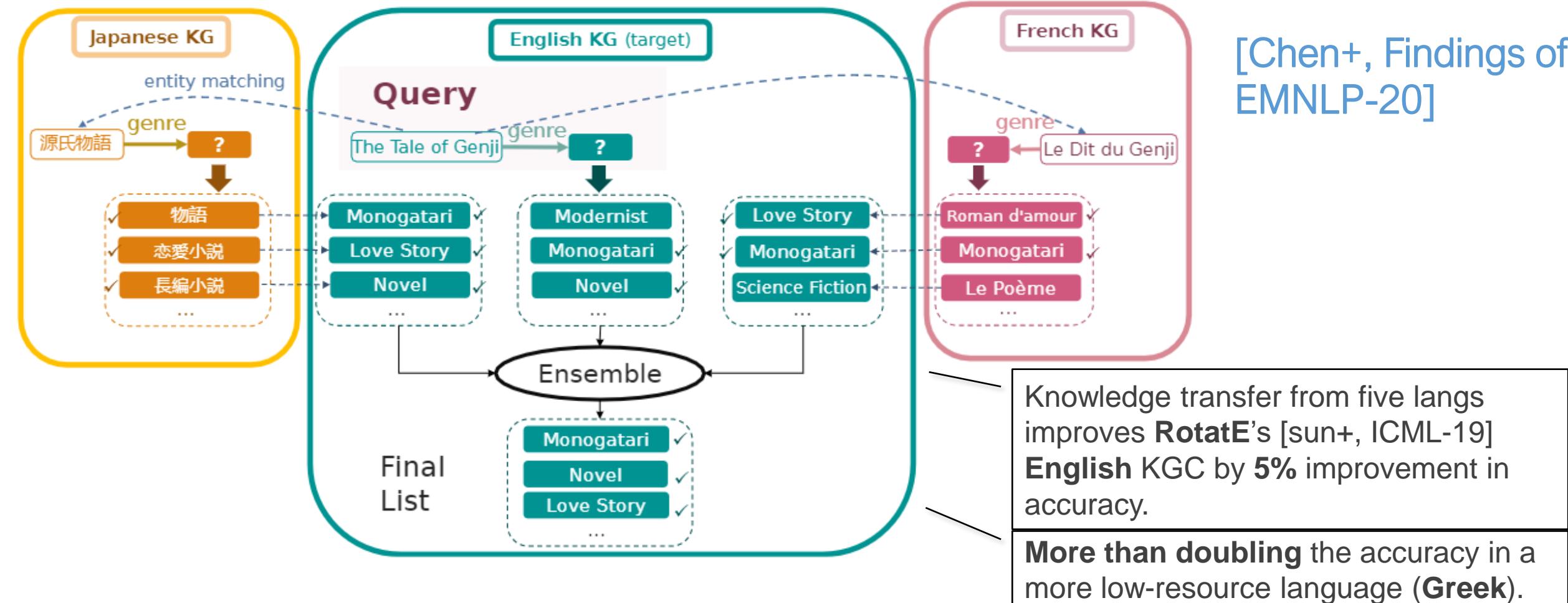
Knowledge transfer to a sparser KG (e.g. French)

- 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.

Ensemble Knowledge Transfer

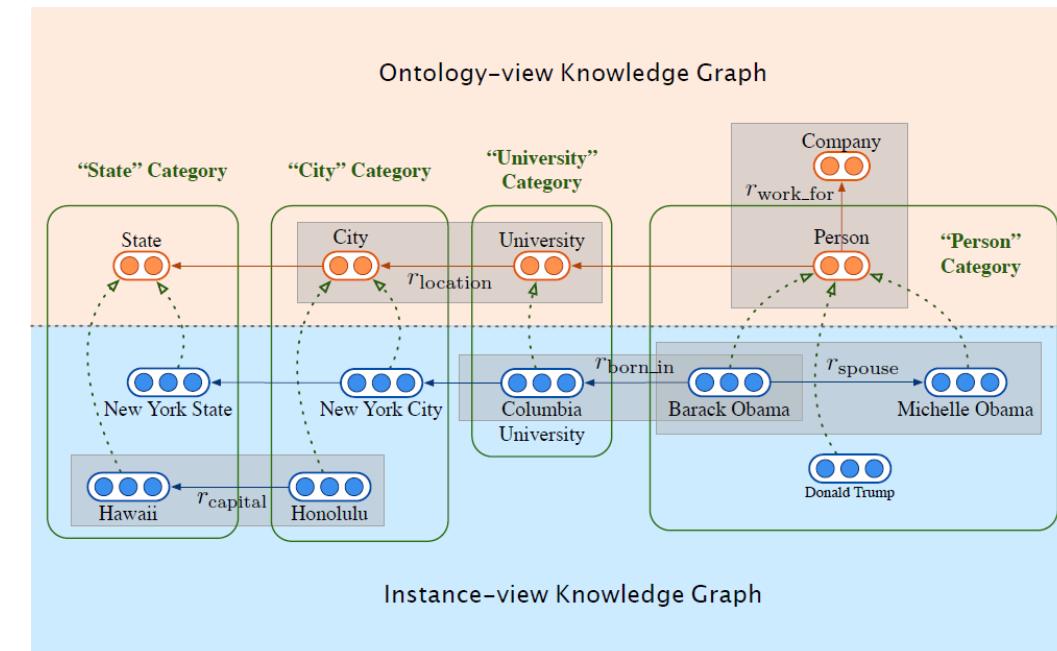
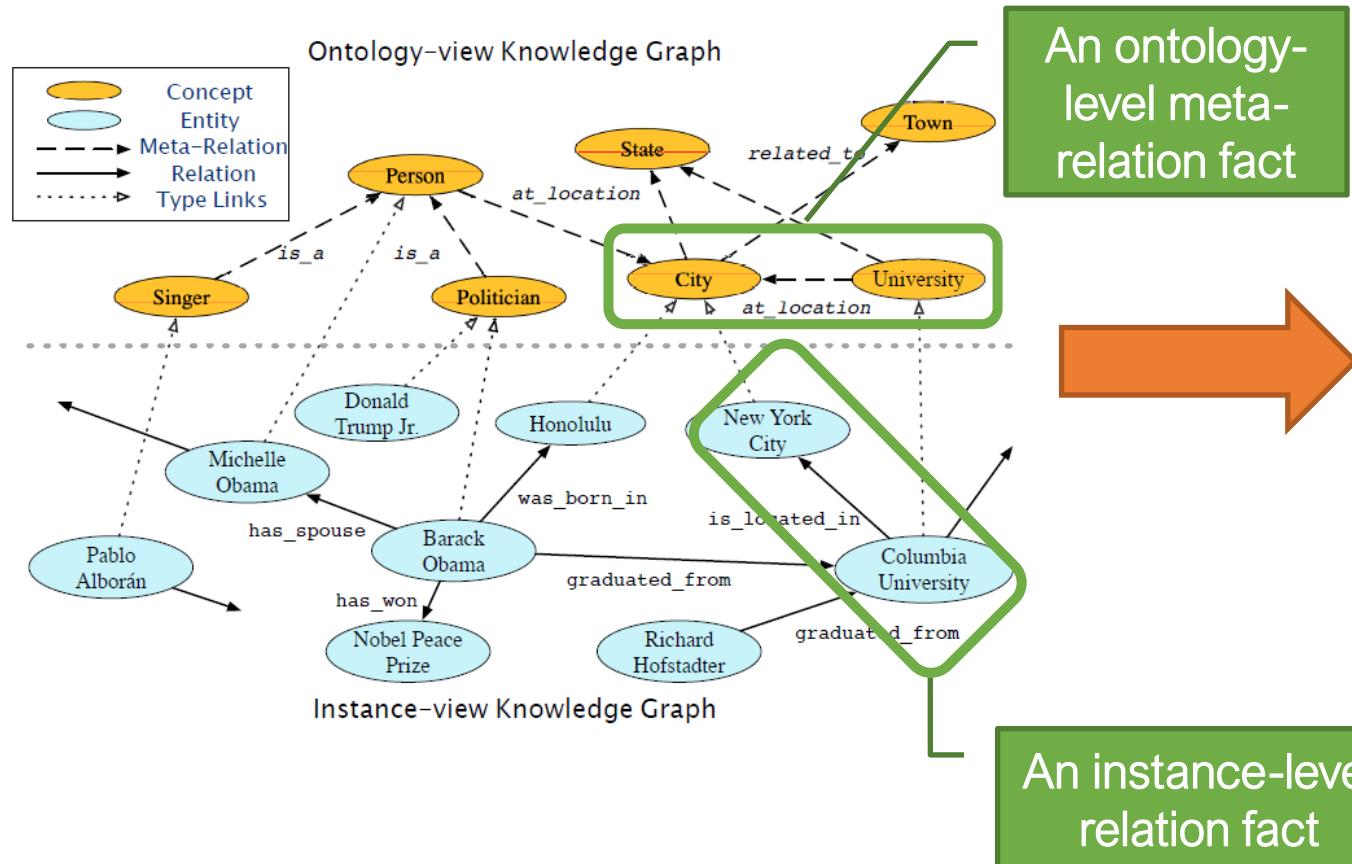


Transferable Embeddings + **Meta-learning** w/ RankBoost-based Model Weights

Scenario 2: Instance Knowledge and Ontological Concepts

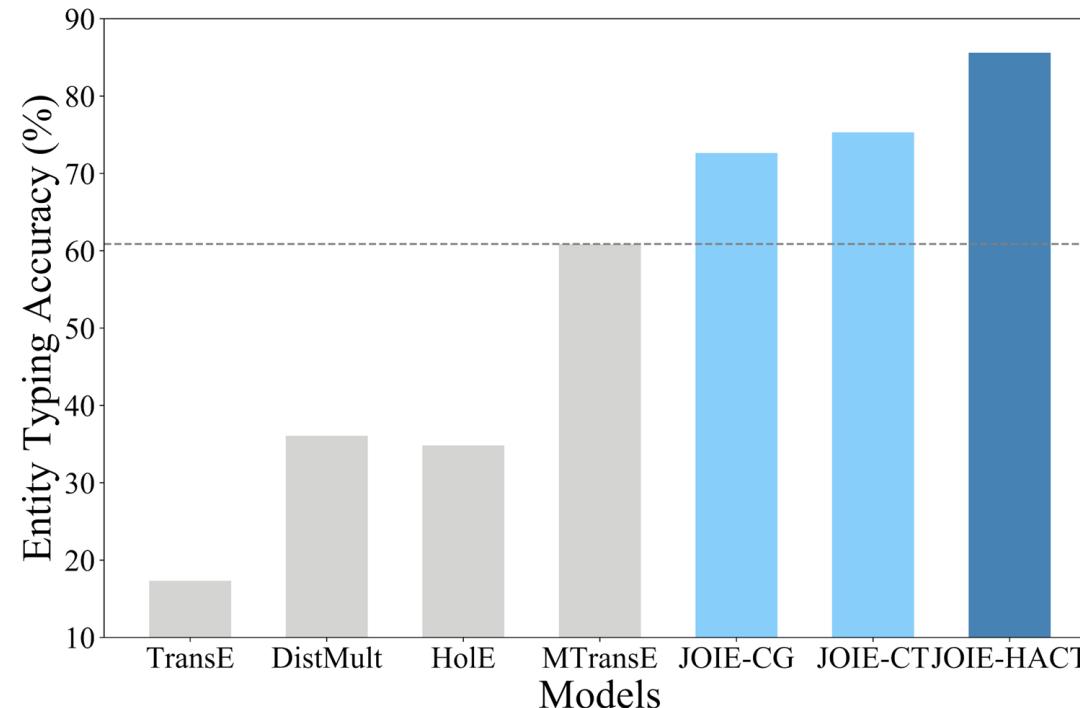
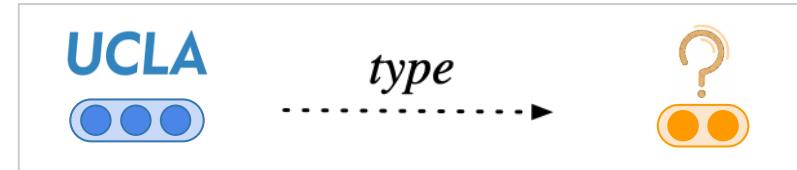
Ontology view: meta-relations of commonsense concepts

Instance view: relations of entities instantiated from concepts



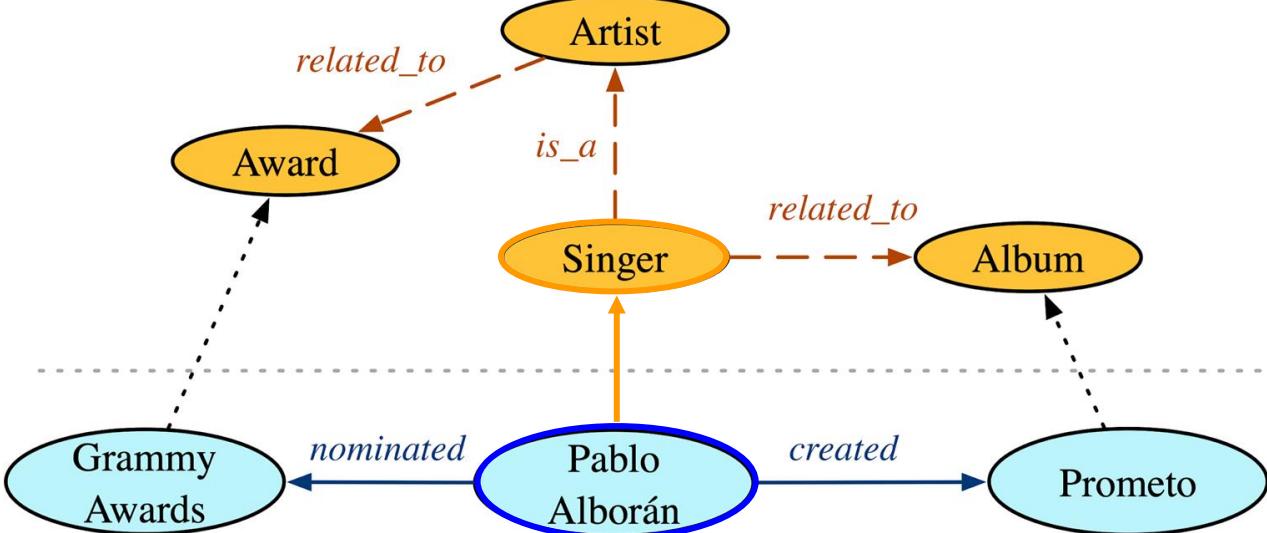
- Given an entity without a known type, what is the most likely type (concept) that it associates with?

JOIE [Hao+, KDD-19]



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

Long-tail Entity Typing (Least Frequent 15%)



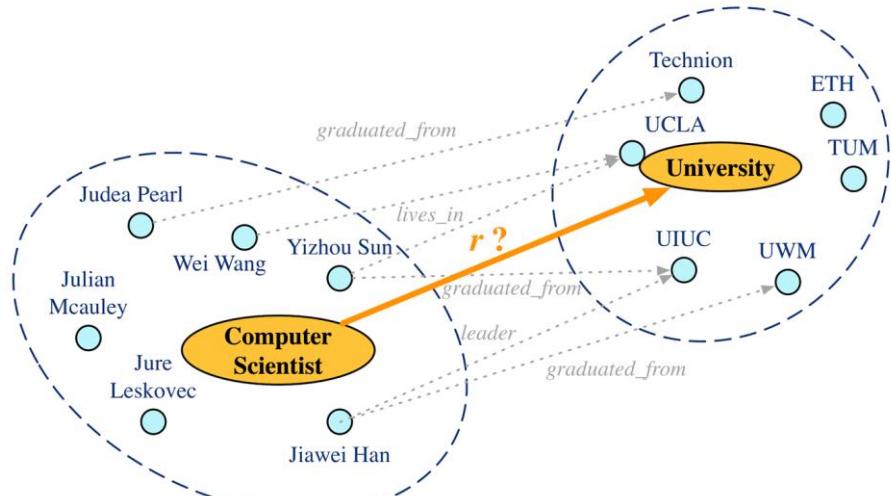
Example of long-tail entity typing

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

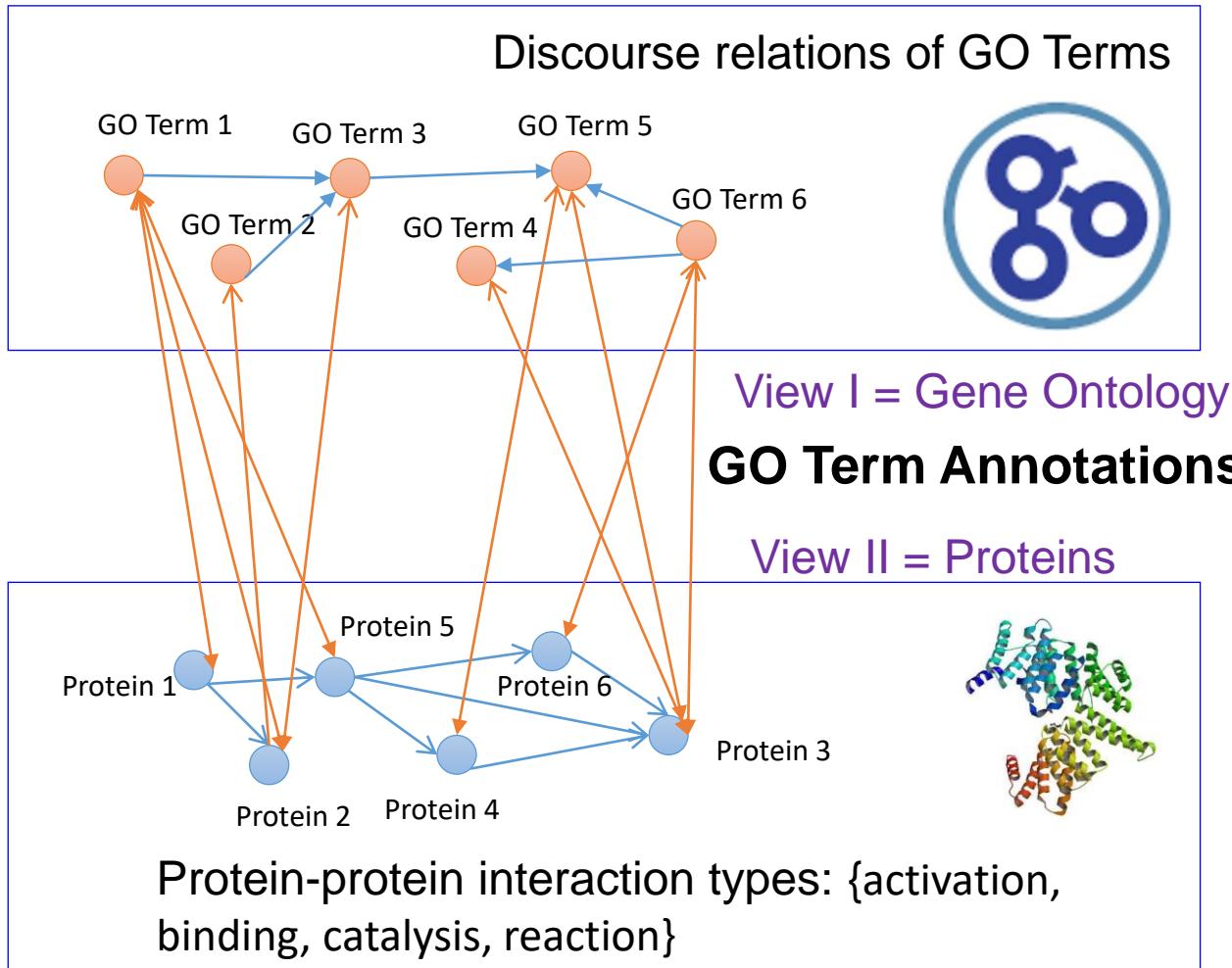


Populating unseen ontological facts by transferring knowledge from instance-level facts.

Examples of ontology population

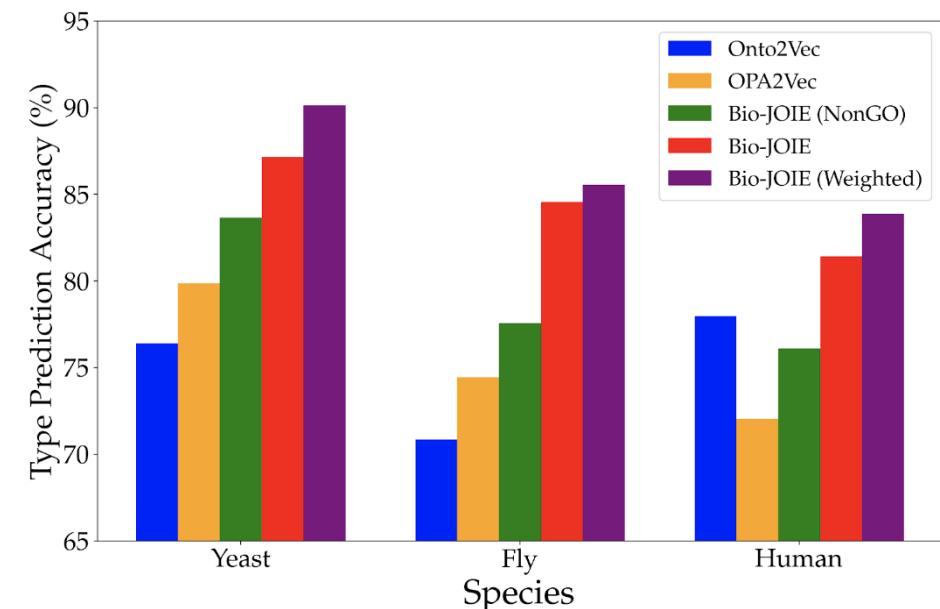
Query	Top 3 Populated Triples with distances
(scientist, ?r, university)	scientist, <i>graduated_from</i> , university (0.499) scientist, <i>isLeaderOf</i> , university (1.082) scientist, <i>isKnownFor</i> , university (1.098)
(boxer, ?r, club)	boxer, <i>playsFor</i> , club (1.467) boxer, <i>isAffiliatedTo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

Scenario 3.a: Proteomics and Gene Ontologies *[Hao+ ACM BCB-20]



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

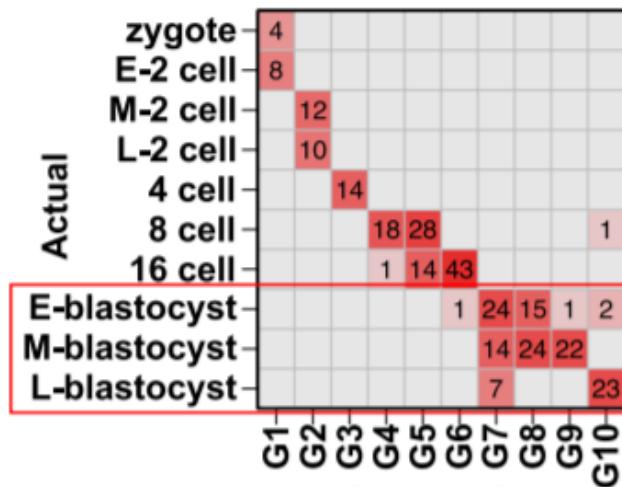
~10% of ACC improvement over SOTA
(Opa2Vec, *Bioinformatics* [J] 2019).



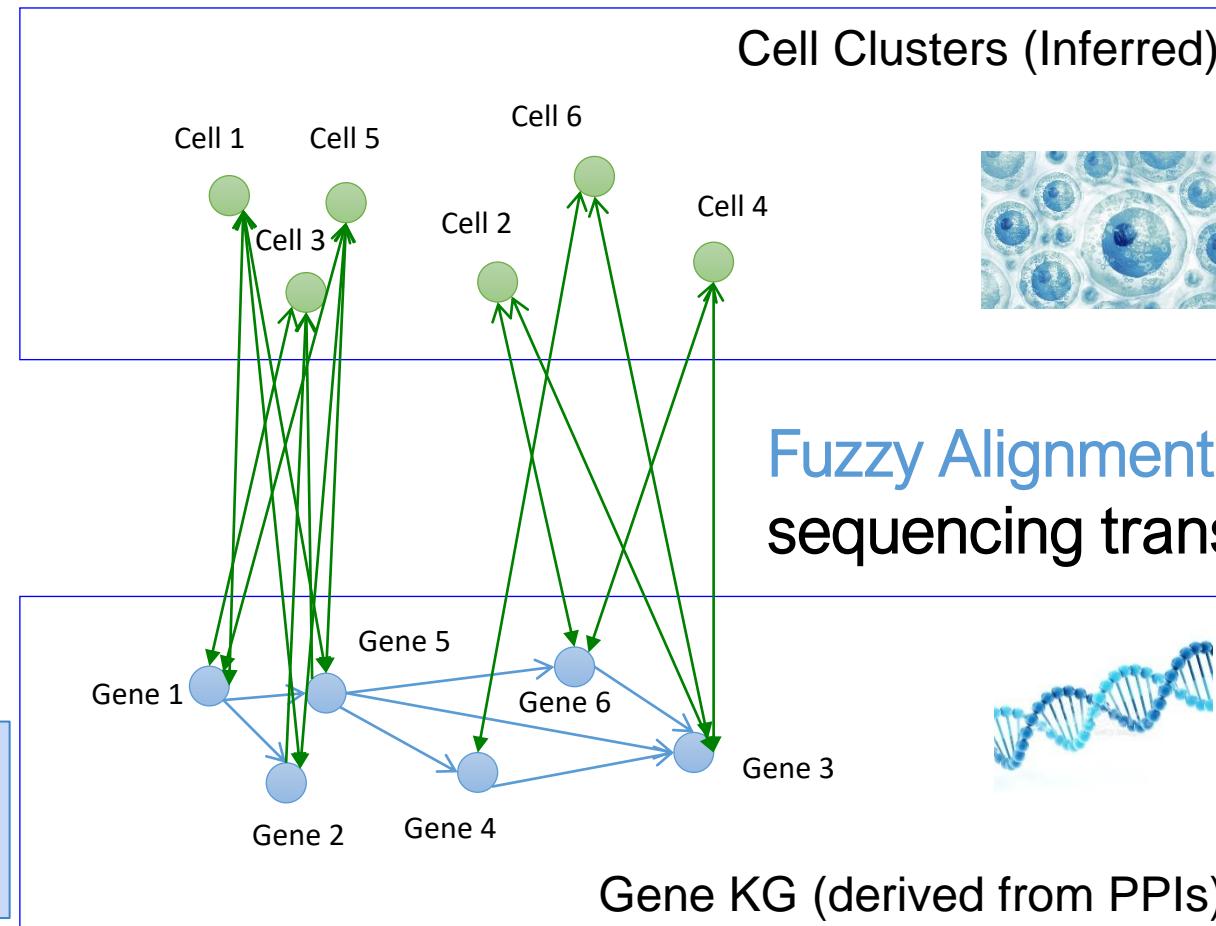
and helps disease target prediction for COVID-19 related viral proteins.

Scenario 3.b: Cell Clustering

At least 10-15% of ARI improvement over pCMF [Durif+, *Bioinform.* 2019] and others.

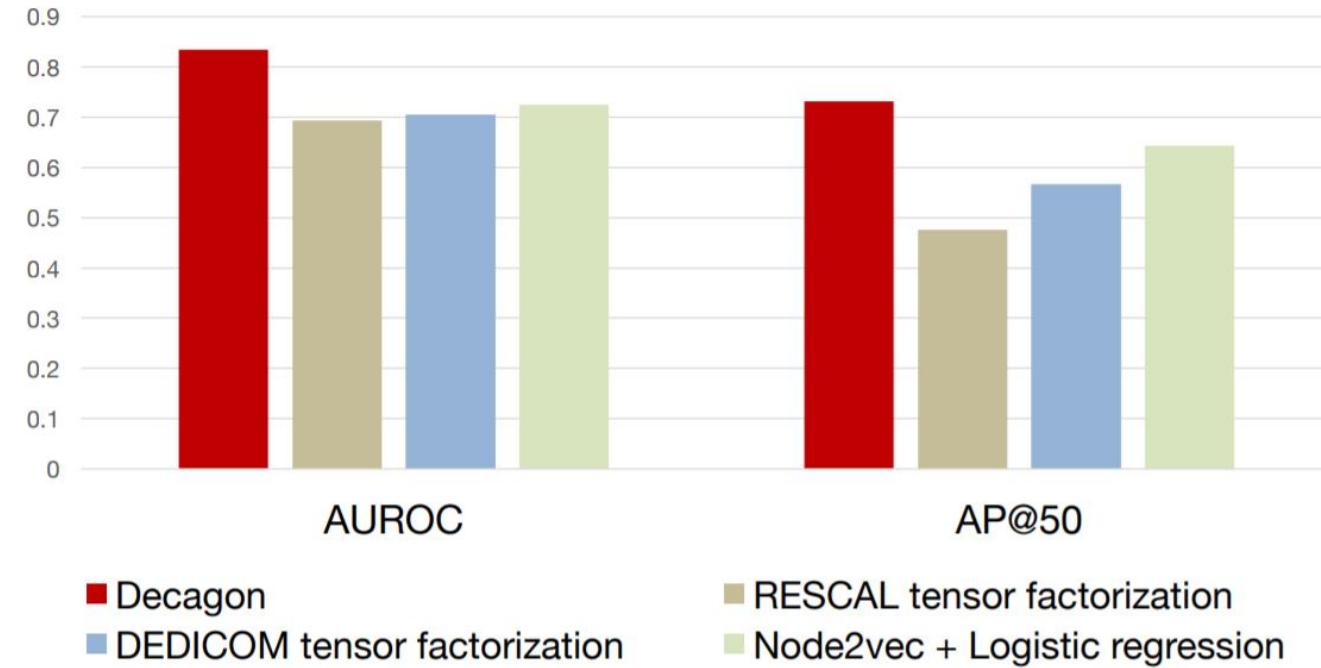
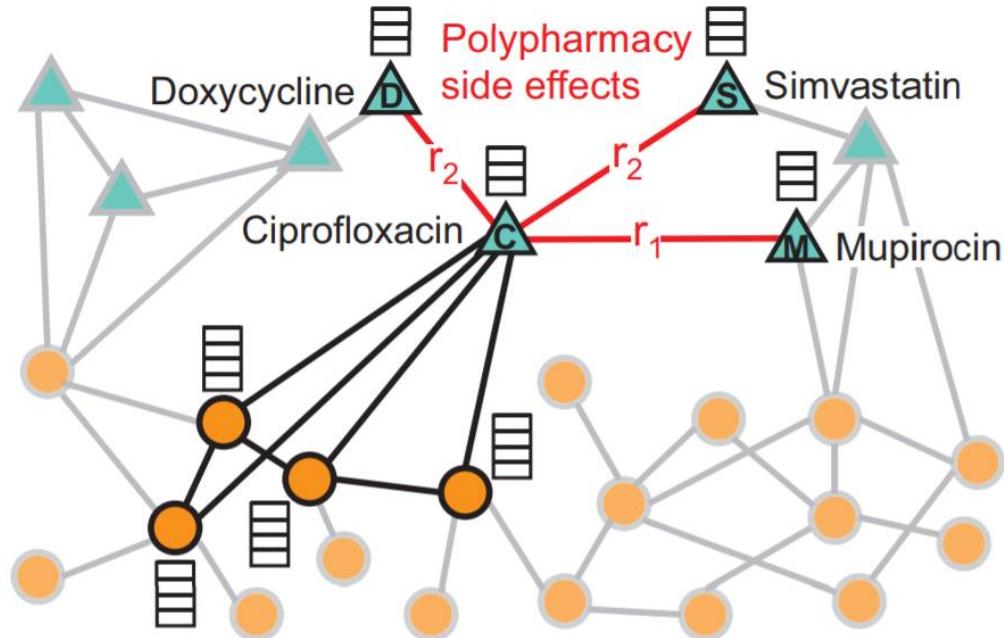


Non-negative Tri-Factorization
 $\text{argmin}_{\theta} \|S - E_1 U E_2^T\|$



Experiment done on the Zeisel dataset [Zeisel+, *Science* 2015]

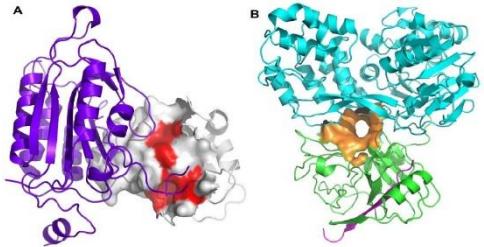
Scenario 4: Polypharmacy side effect detection



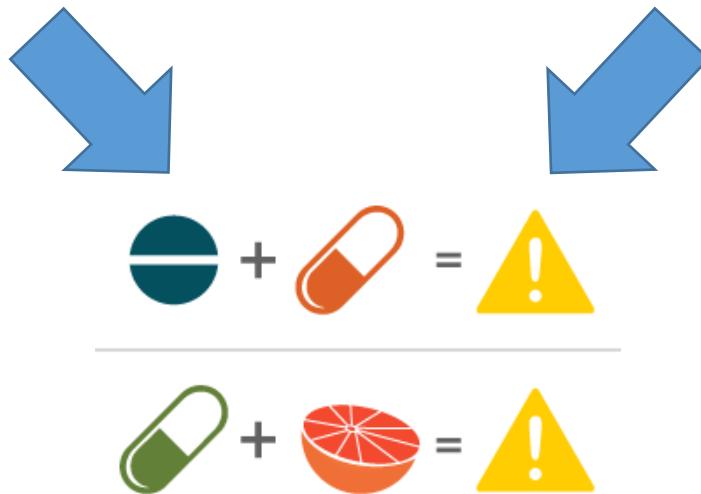
Decagon [Zitnik+, ISMB-18]

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

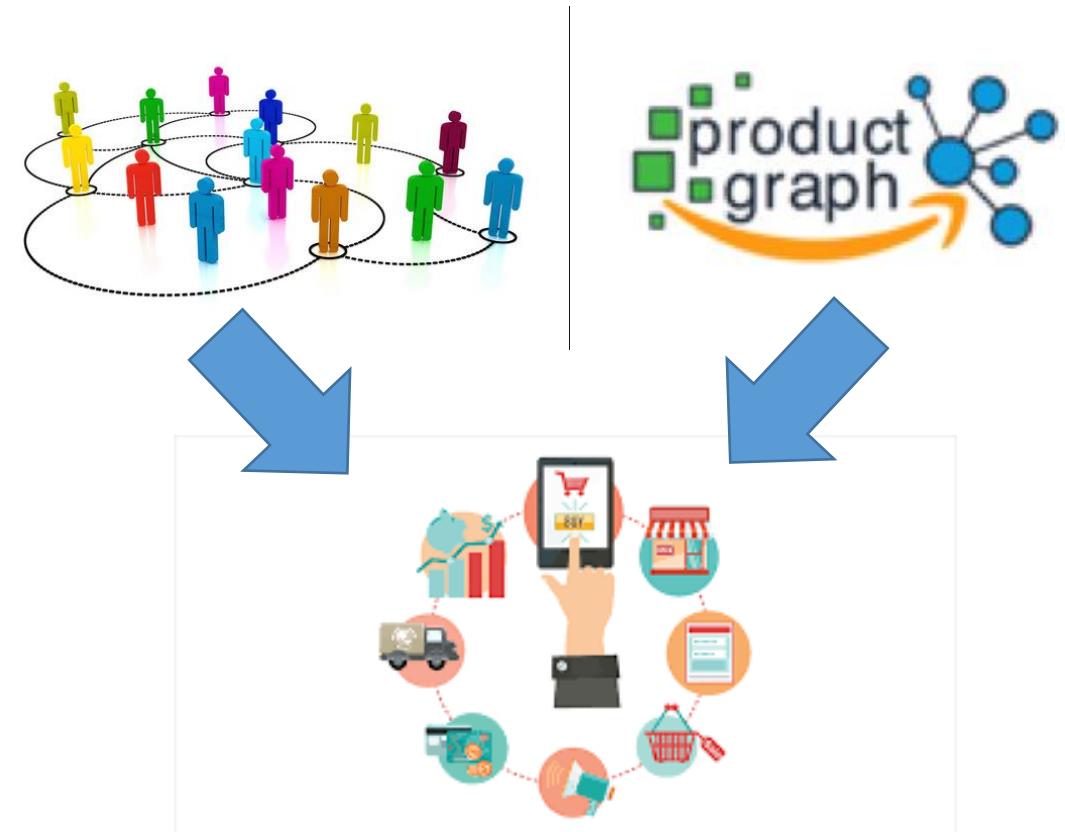
More Applications To Be Explored



UMLS
DISEASE
ONTOLOGY



Polypharmacy (drug-drug) interaction
or drug-target prediction



Product recommendation

*[Chen+ /ISMB'19, *Bioinformatics* 2019]

Protein profile 1

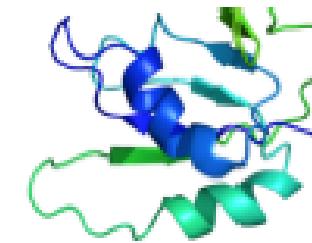


MQSPYPMTQVSNVDDGSLLK...

Interact or not?

Interaction type? (Binding, catalysis, inhibition, ...)

Protein profile 2

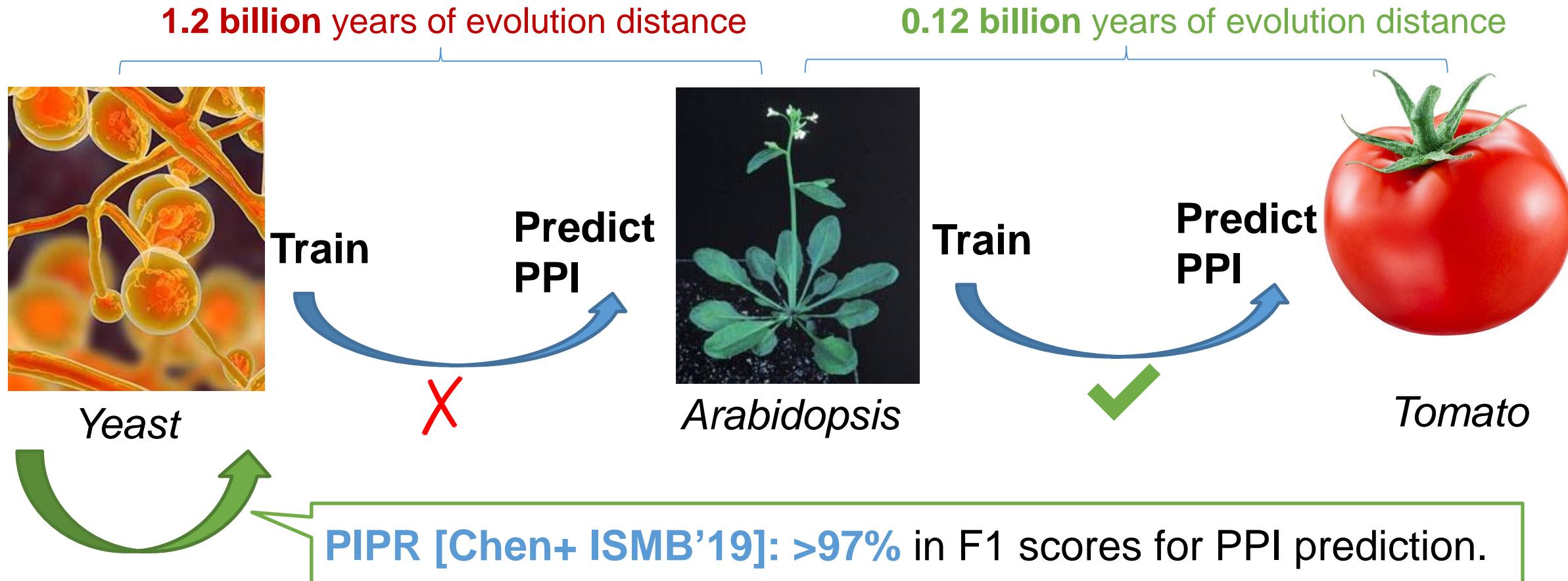


MLERIQQLVNAVNDPRSDVAT...

Binding affinity? (ΔG)

An example task: Protein-protein interaction prediction.

Cross-species Transferability: Why Important



Emerging topic: transferability across species

Cross-species Transferability: What Are Needed?

- <3.5k “high-resource” species vs 1.5M “low-resource” species
 - Complex organisms without full genomes
 - Newly discovered ones
- Transfer learning is important for *de novo* prediction on 1.5M “low-resource” species
 - Reliable *de novo* prediction can be used to guide wet lab experiments
- New technologies for the community
 - Adversarial learning for “species-invariant” sequence representations
 - Massively pre-trained language models for amino acid sequences

Thank You

Feb 2020

AAAI Tutorials

Recent Advances in Transferable Representation Learning

<https://cogcomp.seas.upenn.edu/page/tutorial.202002/>