COMBINING REPRESENTATION LEARNING AND LOGICAL RULE REASONING FOR KNOWLEDGE GRAPH INFERENCE

Yizhou Sun

Department of Computer Science

University of California, Los Angeles

yzsun@cs.ucla.edu

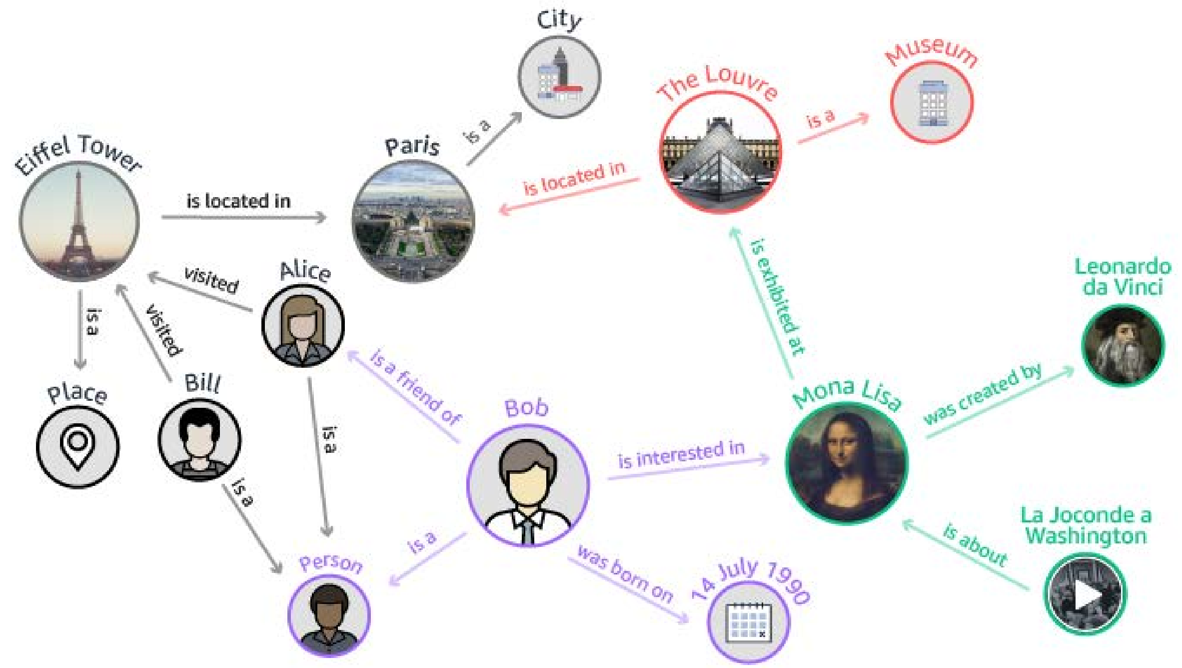
August 15, 2021

Outline

- Introduction
- Bringing First-Order Logic into Uncertain KG Embedding
- UniKER: Integrating Horn Rule Reasoning into KGE
- Summary

Knowledge Graph

- What are knowledge graphs?
  - Multi-relational graph data
    - (heterogeneous information network)
  - Provide structured representation for semantic relationships between real-world entities



A triple (h, r, t) represents a fact, ex:
(Eiffel Tower, is located in, Paris)

Examples of KG

General-purpose KGs

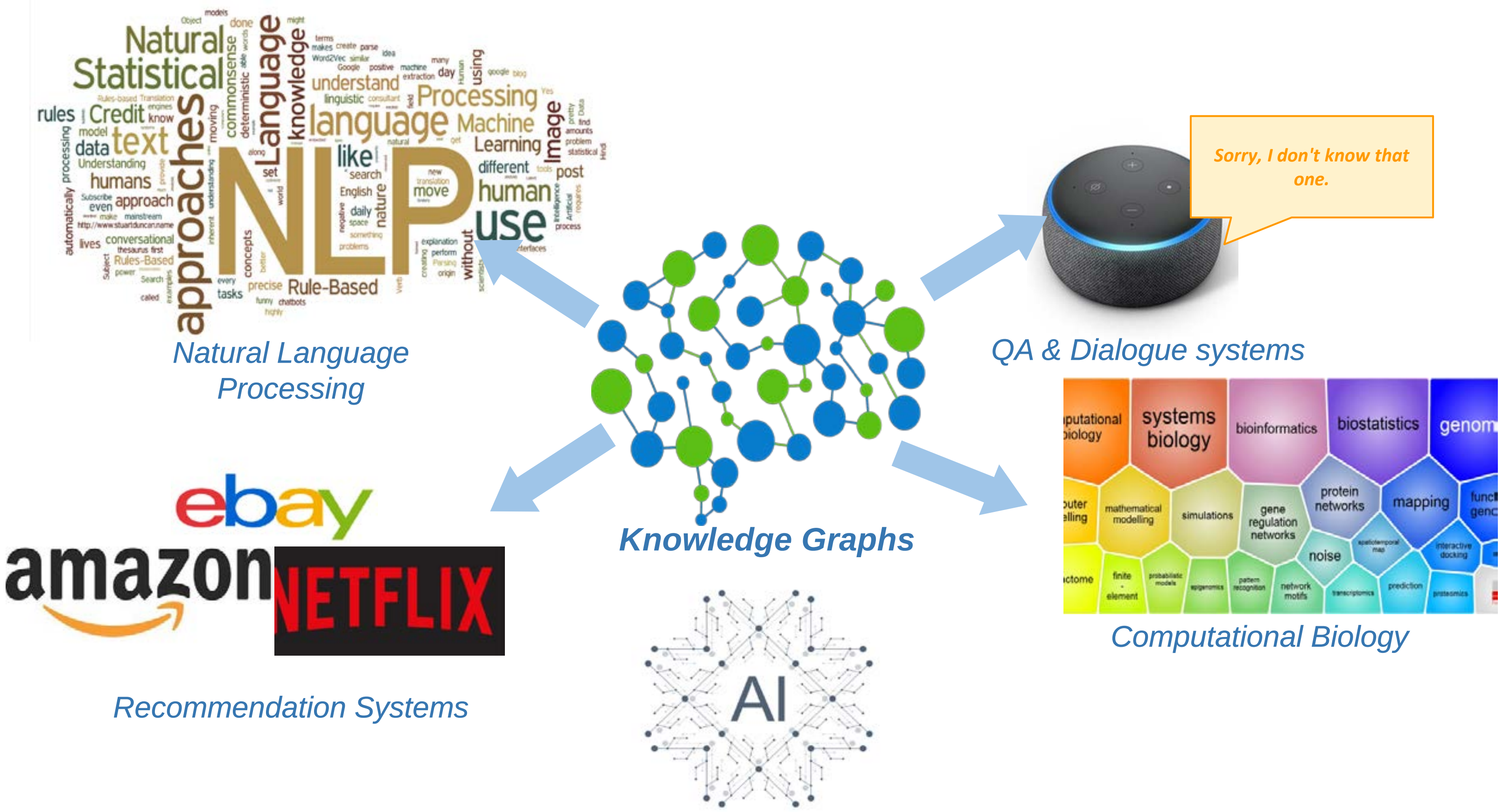
Bio & Medical KGs

Product Graphs & E-commerce

Common-sense KGs & NLP

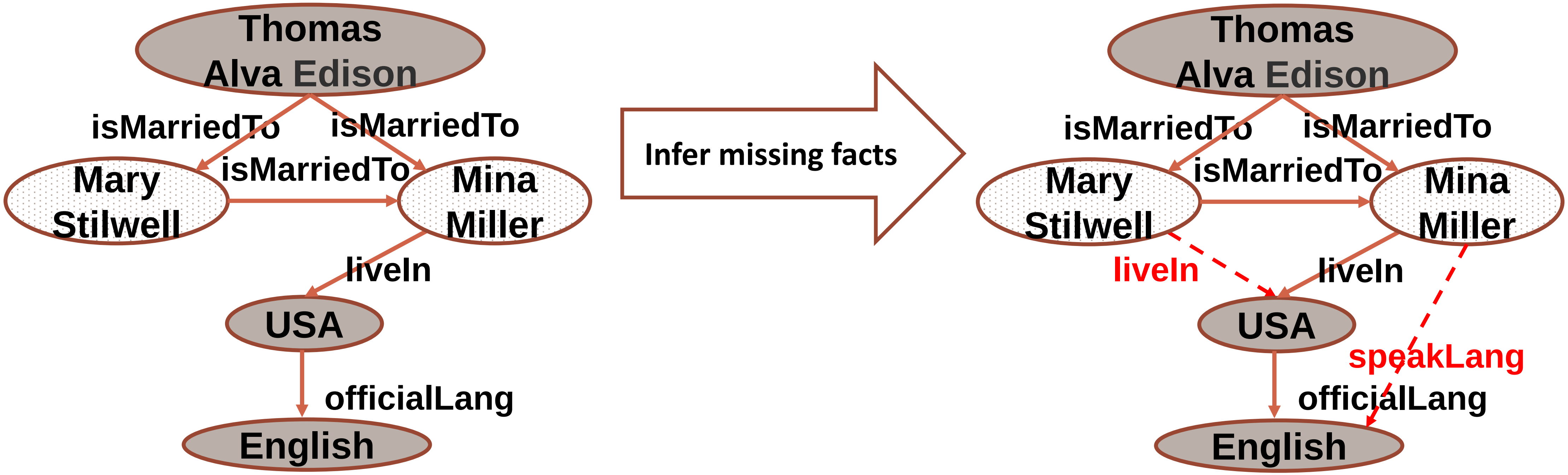
Applications of KGs

- Foundational to knowledge-driven AI systems
- Enable many downstream applications (NLP tasks, QA systems, etc)



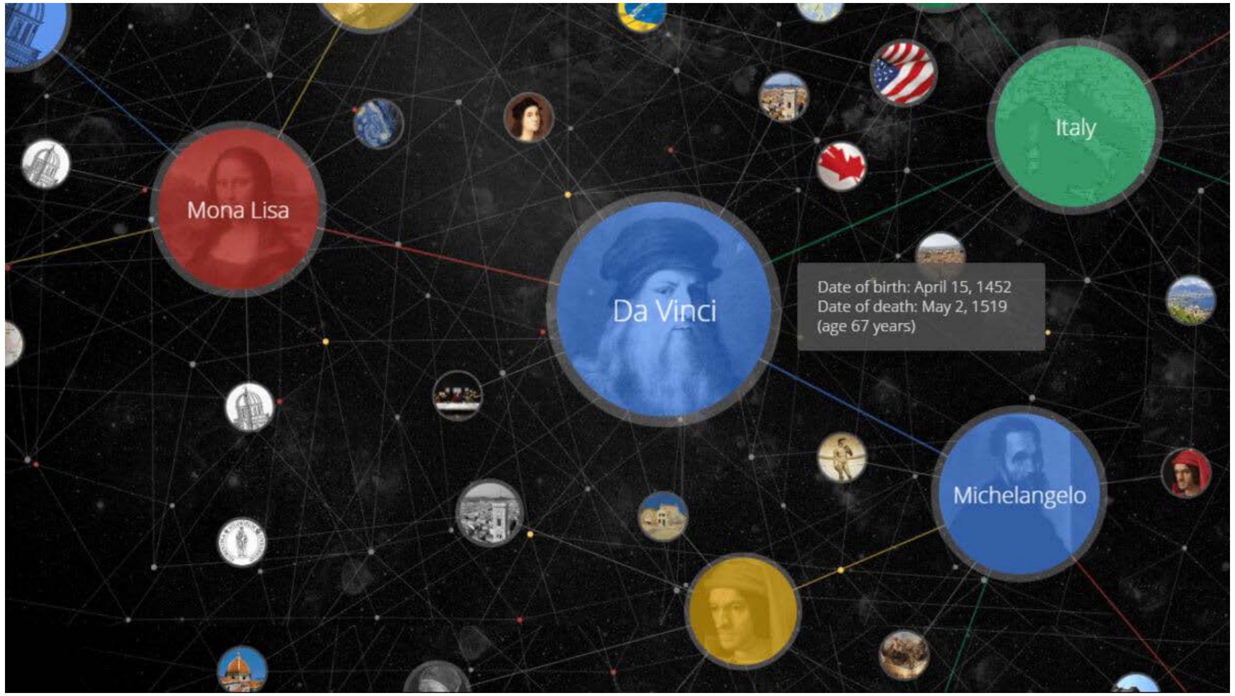
Reasoning over Knowledge Graph

- Knowledge graph reasoning aims at inferring missing knowledge through the existing facts.



Knowledge Graph Embedding

- Entities: low dimensional vectors
- Relations: parametric algebraic operators
- Triples: representation-based score function



Summary of Existing Approaches

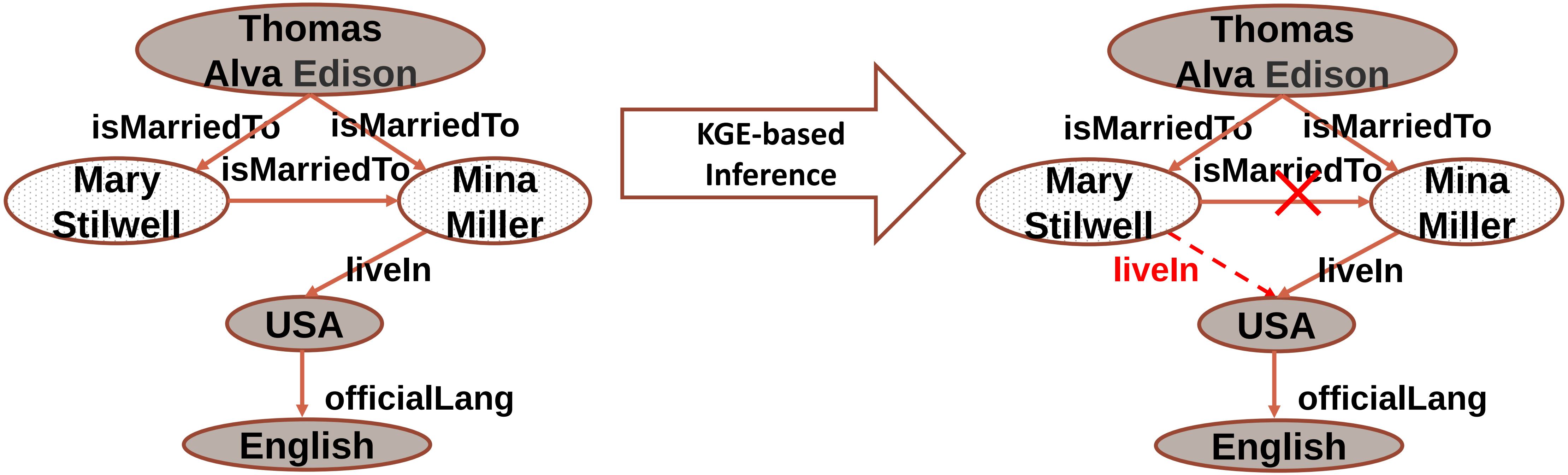
- Define a score function for a triple: $f_r(\boldsymbol{h}, \boldsymbol{t})$
  - According to entity and relation representation
- Define a loss function to guide the training
  - E.g., an observed triple scores higher than a negative one

Model	Score Function	
SE (Bordes et al., 2011)	$-\lVert \boldsymbol{W}_{r,1}\mathbf{h} - \boldsymbol{W}_{r,2}\mathbf{t} \rVert$	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^k, \boldsymbol{W}_{r,\cdot} \in \mathbb{R}^{k \times k}$
TransE (Bordes et al., 2013)	$-\lVert \mathbf{h} + \mathbf{r} - \mathbf{t} \rVert$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
TransX	$-\lVert g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t}) \rVert$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
DistMult (Yang et al., 2014)	$\langle \mathbf{r}, \mathbf{h}, \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ComplEx (Trouillon et al., 2016)	$\mathrm{Re}(\langle \mathbf{r}, \mathbf{h}, \overline{\mathbf{t}} \rangle)$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$
HolE (Nickel et al., 2016)	$\langle \mathbf{r}, \mathbf{h} \otimes \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ConvE (Dettmers et al., 2017)	$\langle \sigma(\mathrm{vec}(\sigma([\overline{\mathbf{r}}, \overline{\mathbf{h}}] * \boldsymbol{\Omega}))\boldsymbol{W}), \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
RotatE	$-\lVert \mathbf{h} \circ \mathbf{r} - \mathbf{t} \rVert^{\boxed{2}}$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k, \lvert r_i \rvert = 1$

Source: Sun et al., RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space (ICLR'19)

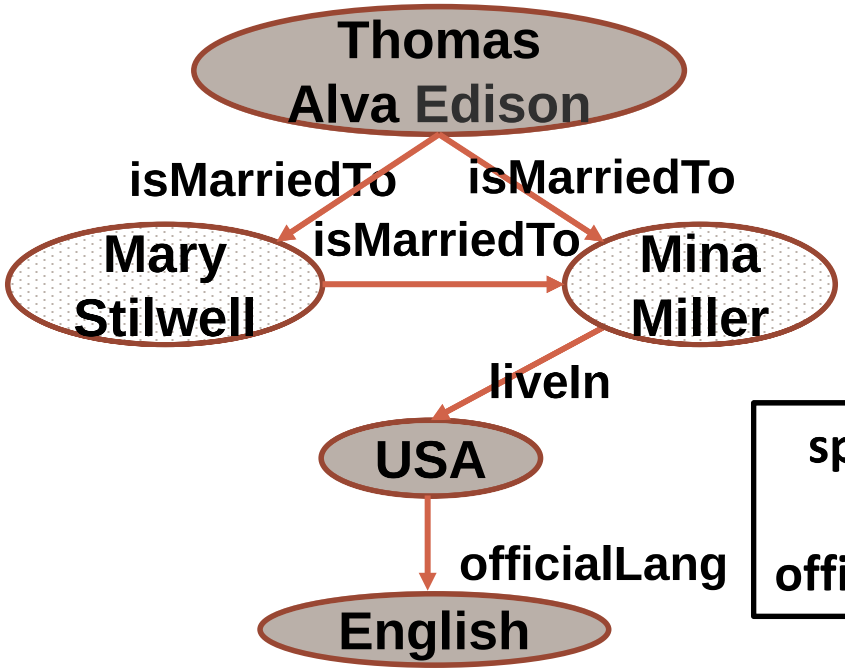
Pros and Cons of KGE

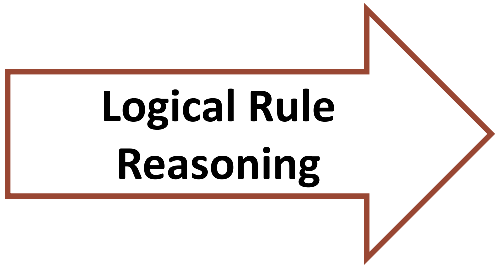
- Knowledge Graph Embedding
  - Shows good scalability as well as robustness.
  - Fails to capture high-order dependency between entities and relations.
  - Can’t handle cold-start entities



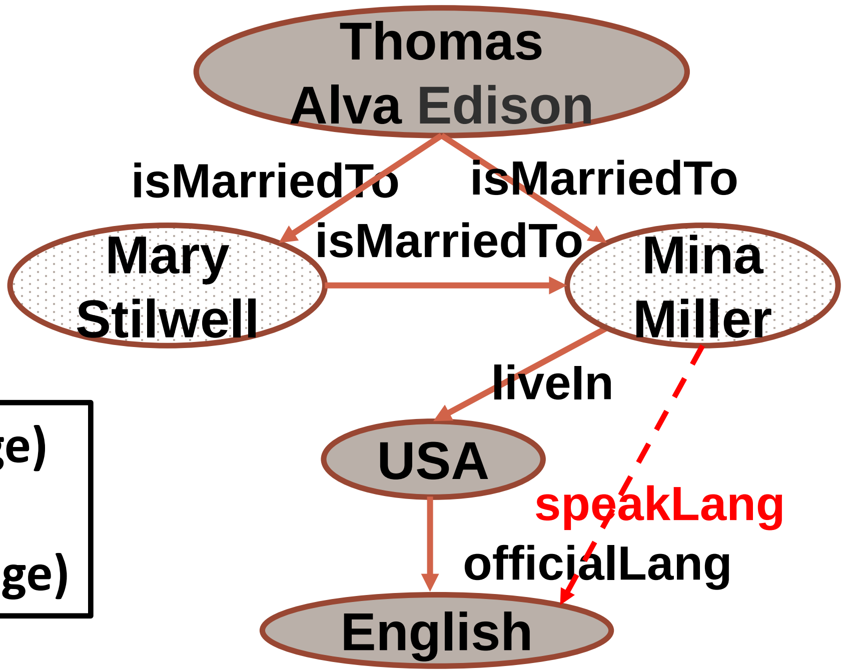
Logical Rule-based KG reasoning

- Find the truth value of each triple to maximize the satisfaction of rules



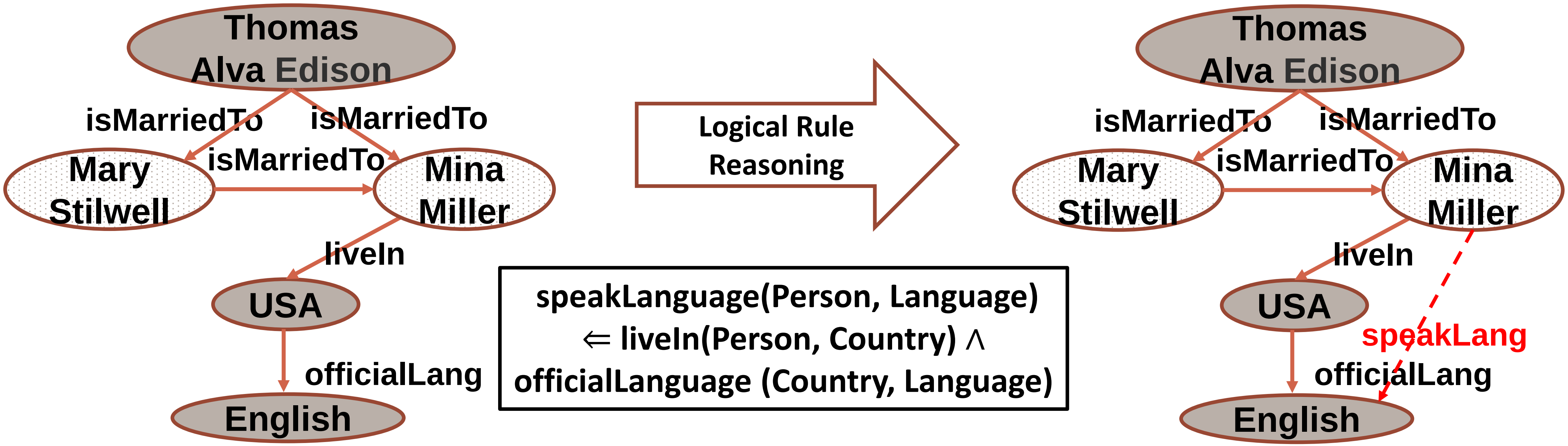


speakLanguage(Person, Language)
⇐ liveIn(Person, Country) ∧
officialLanguage (Country, Language)

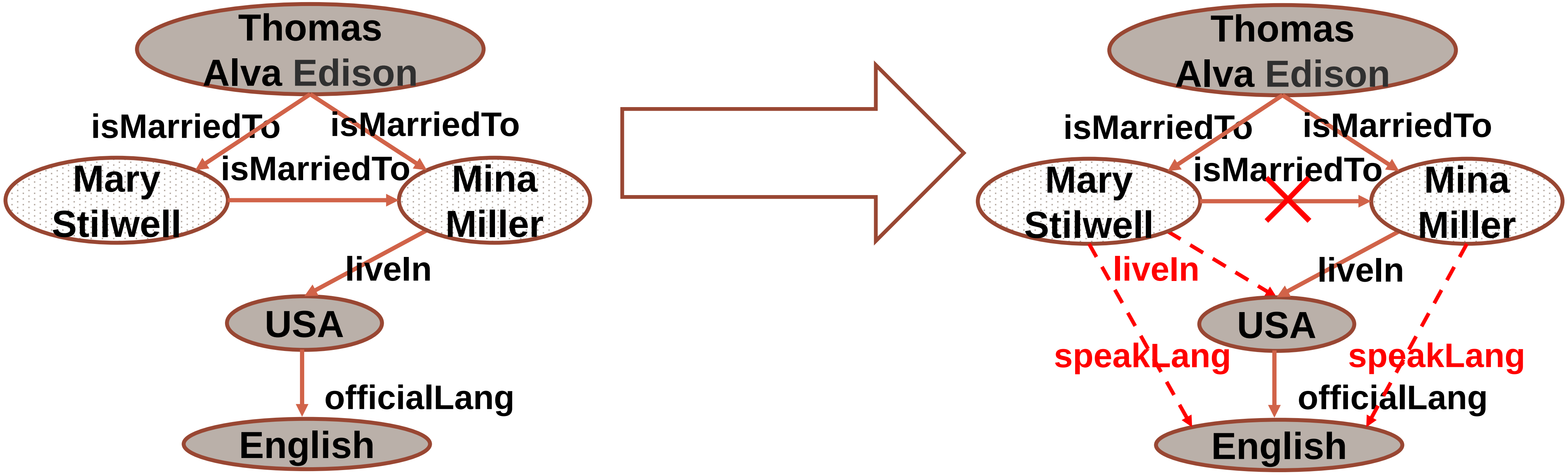


Pros and Cons of Logical Rule-based Reasoning

- Logical Rule-based Reasoning
  - Good at capturing high-order dependency and good interpretability.
  - Unable to handle noisy data as well as suffer from high computation complexity.
  - Coverage is low



Combine both Worlds:1+1>2!



Outline

- Introduction
- Bringing First-Order Logic into Uncertain KG Embedding
- UniKER: Integrating Horn Rule Reasoning into KGE
- Summary

The First Attempt

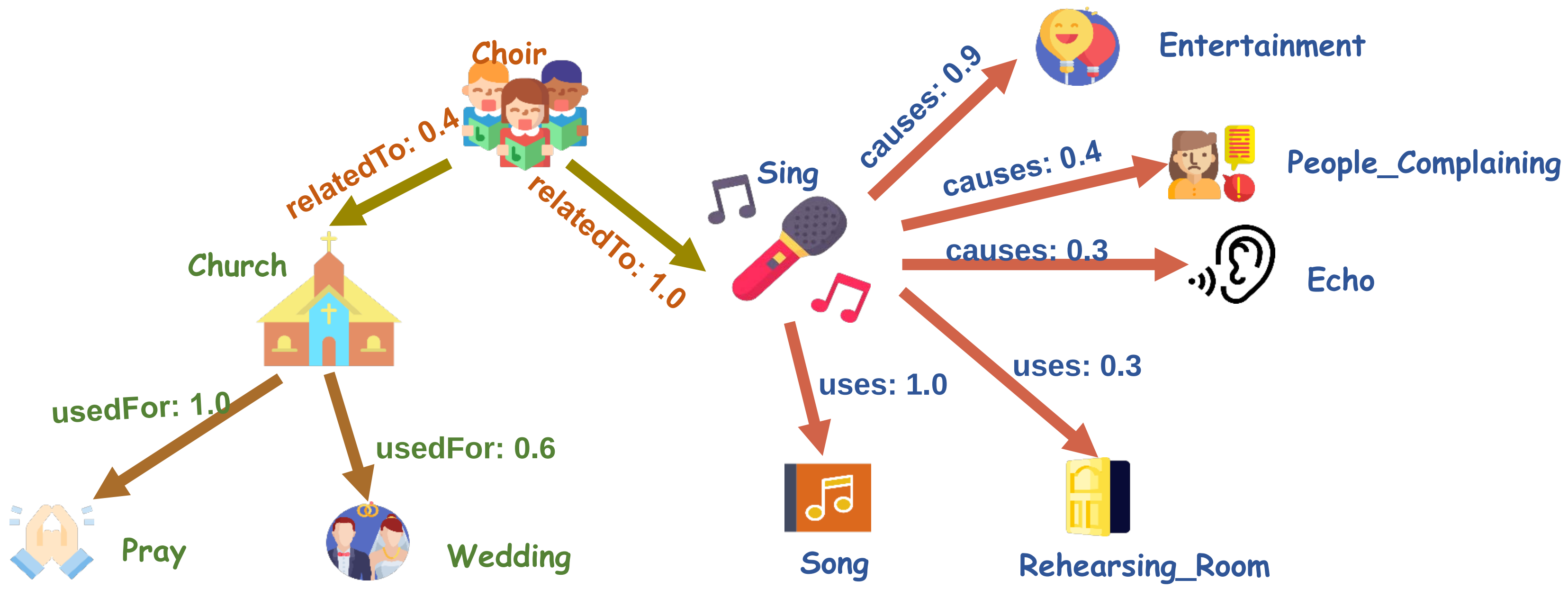
- Chen et al., "*Embedding Uncertain Knowledge Graphs*," AAAI'19

Two Types of Errors in KG

- False positive
  - An observed triple is wrong,
    - e.g., (Obama, is_born_in, Kenya)
- False negative
  - A true fact is missing
    - e.g., (Eiffel Tower, is located in, France)

Handling Uncertainty in Triples

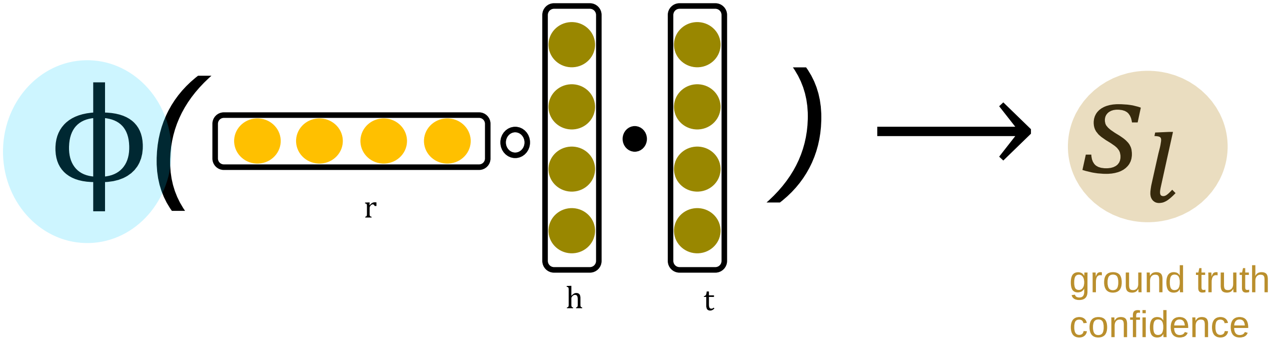
- False positive errors can be alleviated by introducing uncertainty
  - E.g., (Obama, is_born_in, Kenya): 0.01



  - Fit $f_r(\boldsymbol{h}, \boldsymbol{t})$ to uncertainty scores

From score function to uncertainty score

- Given a triple $l = (h, r, t)$ with uncertainty score s_l
  - Transform $f_r(\boldsymbol{h}, \boldsymbol{t})$ into a score in the range [0,1]
    - E.g., for DisMult score function

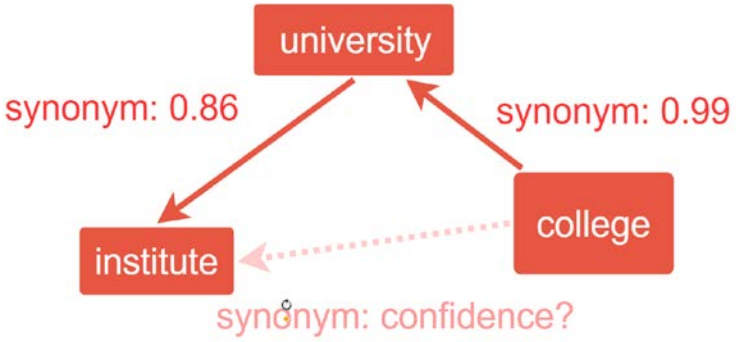


    - Where $\phi(\cdot)$ can be defined as

      - **Logistic function** $\phi(x) = \dfrac{1}{1 + e^{-(\mathbf{w}x + \mathbf{b})}}$ **UKGE(logi)**
      - **Bounded Rectifier** $\phi(x) = \min(\max(\mathbf{w}x + \mathbf{b}, 0), 1)$ **UKGE(rect)**

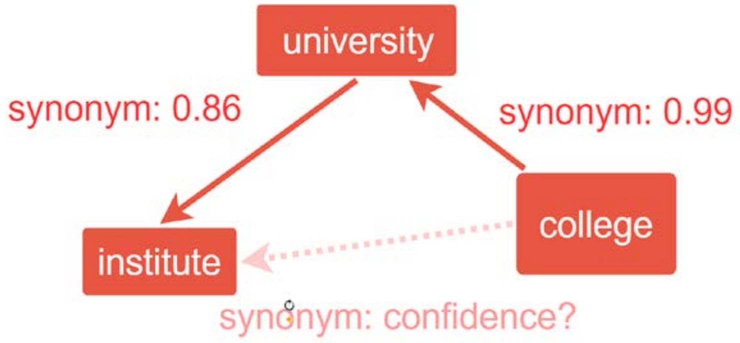
Handling Missing Facts

- Are unseen triples still needed?
  - Yes, negative triples are still data points!
- Can we treat them as false, i.e., $s_l = 0$, if triple l is unseen?
  - No, we are going to make too many mistakes!
    - The potential probability of an unseen triple could be higher than an observed triple with low confidence



Bringing Logic Rules

- What are logic rules?
  - Logic rule (Template)
    - (A , synonym, B) ∧ (B , synonym, C) → (A, synonym, C)
  - Ground rule (Instance)
    - (college, synonym, university) ∧ (university , synonym, institute) → (college, synonym, institute)
- Why are they helpful?
  - Help us to infer the score for unseen triples



Probabilistic Soft Logic

- Quantify a ground rule using PSL
  - Lukasiewicz t-norm, from Boolean logic to soft logic

$$l_1 \land l_2 = \max\{0, I(l_1) + I(l_2) - 1\}$$
$$l_1 \lor l_2 = \min\{1, I(l_1) + I(l_2)\}$$
$$\neg l_1 = 1 - I(l_1)$$

  - Probability of a ground rule $\gamma \equiv \gamma_{body} \rightarrow \gamma_{head}$
    - $p_\gamma = I(\neg\gamma_{body} \lor \gamma_{head}) = \min\{1, 1 - I(\gamma_{body}) + I(\gamma_{head}))$
  - Distance to satisfaction $d_\gamma = 1 - p_\gamma = \max\{0, I(\gamma_{body}) - I(\gamma_{head})\}$
    -

More publications on PSL: https://psl.linqs.org/

The Goal: Minimize Distance to Satisfaction

- Example: Consider the following ground rule

l_1 confidence: 0.99 l_2 confidence: 0.86

- (college, synonym, university) ∧ (university , synonym, institute) → (college, synonym, institute)

l_3 confidence: ?

- Recall, l_3 confidence: ?

$$d_\gamma = \max\{0, I(l_1 \wedge l_2) - I(l_3)\}$$

0.99 0.86

$$= \max\{0, s_{l_1} + s_{l_2} - 1 - f(l_3)\}$$

$$= \max\{0, 0.85 - f(l_3)\}$$

- (college, synonym, university) ∧ (university , synonym, institute) → (college, synonym, institute)
- Recall,

Say, our embedding model predicts it as 0.65.
How good is this prediction?

The New Embedding Model

- For observed triples, force its score close to ground truth score
- For unseen triples, minimize the distance to satisfaction in ground rules they are involved

$$\mathcal{J} = \sum_{l \in \mathcal{L}^+} |\boxed{f(l)} - s_l|^2 + \sum_{l \in \mathcal{L}^-} \sum_{\gamma \in \Gamma_l} |\boxed{\psi_\gamma(f(l))}|^2$$

Embedding-based confidence function

Distance to satisfaction for a ground rule γ, where $triple\ l$ is involved in

Experiments

- Datasets

Dataset	#Ent.	#Rel.	#Rel. Facts	Avg(s)	Std(s)
CN15k	15,000	36	241,158	0.629	0.232
NL27k	27,221	404	175,412	0.797	0.242
PPI5k	5,000	7	271,666	0.415	0.213

- Logic Rules

(A, relatedto, B)∧(B,relatedto,C)→(A,relatedto,C)
(A, causes, B)∧(B,causes,C)→(A,causes,C)

(A, competeswith, B)∧(B,competeswith,C)→(A,competeswith,C)
(A, atheletePlaysForTeam, B)∧(B, teamPlaysSports, C)→(A, atheletePlaysSports, C)

(A, binding, B)∧(B,binding,C)→(A,binding,C)

Baselines

- Deterministic KG embedding models, which does not model confidence scores explicitly
  - TransE [Bordes et al. 2013)]
  - DistMult [Yang et al. 2015]
  - ComplEx [Trouillon et al. 2016]
- Uncertain Graph Embedding, which only provides node embeddings
  - URGE [Hu et al. 2017]
- Two simplified version of our models
  - Without Negative Sampling (UKGE_n-)
    - Can we just ignore the negative links during training?
  - Without PSL (UKGE_p-)
    - Will simply treating unseen relations as 0 a good strategy?

Relation Fact Confidence Score Prediction

- Given an unseen triple (h,r,t), predict its confidence
- Metrics: MSE and MAE ($\times 10^{-2}$)

Dataset	CN15k		NL27k		PPI5k	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE
URGE	10.32	22.72	7.48	11.35	1.44	6.00
UKGE_{n-}	23.96	30.38	24.86	36.67	7.46	19.32
UKGE_{p-}	9.02	20.05	2.67	7.03	0.96	4.09
UKGE_{rect}	**8.61**	**19.90**	**2.36**	**6.90**	**0.95**	**3.79**
UKGE_{logi}	9.86	20.74	3.43	7.93	0.96	4.07

Relation Fact Ranking

- Given a query (h, r, ?t), rank all entities in our vocabulary as tail candidates
- Metrics: normalized Discounted Cumulative Gain (nDCG) (linear gain and exp gain)

metrics	CN15K		NL27k		PPI5k	
Dataset	linear	exp.	linear	exp.	linear	exp.
TransE	0.601	0.591	0.730	0.722	0.710	0.700
DistMult	0.689	0.677	0.911	0.897	0.894	0.880
ComplEx	0.723	0.712	0.921	0.913	0.896	0.881
URGE	0.572	0.570	0.593	0.593	0.726	0.723
UKGE_{n-}	0.236	0.232	0.245	0.245	0.514	0.517
UKGE_{p-}	0.769	0.768	0.933	0.929	0.940	0.944
UKGE_{rect}	0.773	0.775	0.939	0.942	0.946	0.946
UKGE_{logi}	**0.789**	**0.788**	**0.955**	**0.956**	**0.970**	**0.969**

Relation Fact Ranking – Case Study

			Ground Truth		Predictions		
			Entity	Score	Entity	Predicted Score	True Score
CN15k	house	usedfor	sleeping	1.0	**relaxing**	**0.86**	**N/A**
			rest	0.98	sleeping	0.85	1.0
			bed away from home	0.71	rest	0.82	0.98
			stay overnight	0.71	**hotel room**	**0.80**	**N/A**
NL27k	Toyota	competeswith	Honda	1.0	Honda	0.94	1.0
			Ford	1.0	Hyundai	0.91	0.72
			BMW	0.96	**Chrysler**	**0.90**	**N/A**
			General Motors	0.90	Nissan	0.89	0.86

Outline

- Introduction
- Bringing First-Order Logic into Uncertain KG Embedding
- UniKER: Integrating Horn Rule Reasoning into KGE
- Summary

The Second Attempt

- Cheng et al., *UniKER: A Unified Framework for Combining Embedding and Horn Rules for Knowledge Graph Inference*, In Submission

Existing Literature on Combining Both Worlds

- Probabilistic logic is widely used to integrate both worlds
  - PSL-based Regularization in Embedding Loss
    - Leverage Probabilistic Soft Logic (PSL) [7] for satisfaction loss calculation
    - Treat logical rules as additional regularization to embedding models, where the satisfaction loss of ground rules is integrated into the original embedding loss.
    - Limitation: only utilize a sample set of rule instances
  - Embedding-based Variational Inference for MLN.
    - Extends Markov Logic Network (MLN) [8]
    - Leverage graph embedding to define variational distribution for all possible hidden triples to conduct variational inference of MLN.
    - Limitation: efficiency issue, sampling is required

Combining Both Worlds

Categories	Methods	Interactive	Exact Logical Inference
PSL-based Regularization	KALE [1]	×	×
	RUGE [2]	√	×
	Rocktaschel et al [3]	×	×
Embedding-based Variational Inference to MLN	pLogicNet [4]	√	×
	ExpressGNN [5]	√	×
	pGAT [6]	√	×

Our Proposed Work: UniKER for Horn Rules

- Idea 1: use forward chaining to conduct exact inference
- Idea 2: combine embedding and logical rules in an iterative manner.
- Idea 3: remove potential incorrect triples during learning to ensure robustness

Traditional Logical Inference: MAX-SAT problem

Knowledge Graph

Entities

Country
USA

Person
Thomas Alva Edison, Mary Stilwell, Mina Miller

Language
English

Predicates

isMarriedTo
liveIn
isParentOf
isSiblingOf
officialLanguage
speakLanguage

Observed Facts

isMarriedTo (Thomas Alva Edison, Mary Stilwell)
isMarriedTo (Thomas Alva Edison, Mina Miller)
isMarriedTo (Mary Stilwell, Mina Miller)
liveIn (Mina Miller, USA)
officialLanguage (USA, English)

All ground predicates

liveIn (Thomas Alva Edison, USA)	**T**
…	
liveIn (Mary Stilwell , USA)	**?**

speakLanguage(Person, Language) ⇐ liveIn(Person, Country) ∧ officialLanguage (Country, Language)

Definite Horn rule

All ground rules

speakLanguage (Thomas Alva Edison, English) ⇐ liveIn (Thomas Alva Edison, USA) ∧ officialLanguage (USA, English)
…
speakLanguage (Mary Stilwell, English) ⇐ liveIn (Mary Stilwell, USA) ∧ officialLanguage (USA, English)

NP-complete **SAT Solver**

New fact speakLanguage (Mina Miller, English)

Forward Chaining for Horn rules: Exact and Fast

Observed Facts
isMarriedTo (Thomas Alva Edison, Mary Stilwell)
isMarriedTo (Thomas Alva Edison, Mina Miller)
isMarriedTo (Mary Stilwell, Mina Miller)
liveIn (Mina Miller, USA)
officialLanguage (USA, English)

involve only a small subset of active ground predicates/rules

Country = USA
Language = English

Person = Mina Miller

speakLanguage(Person, Language) ⇐ liveIn(Person, Country) ∧ officialLanguage (Country, Language)

Definite Horn rule

Forward Chaining

speakLanguage (Mina Miller, English)

New fact

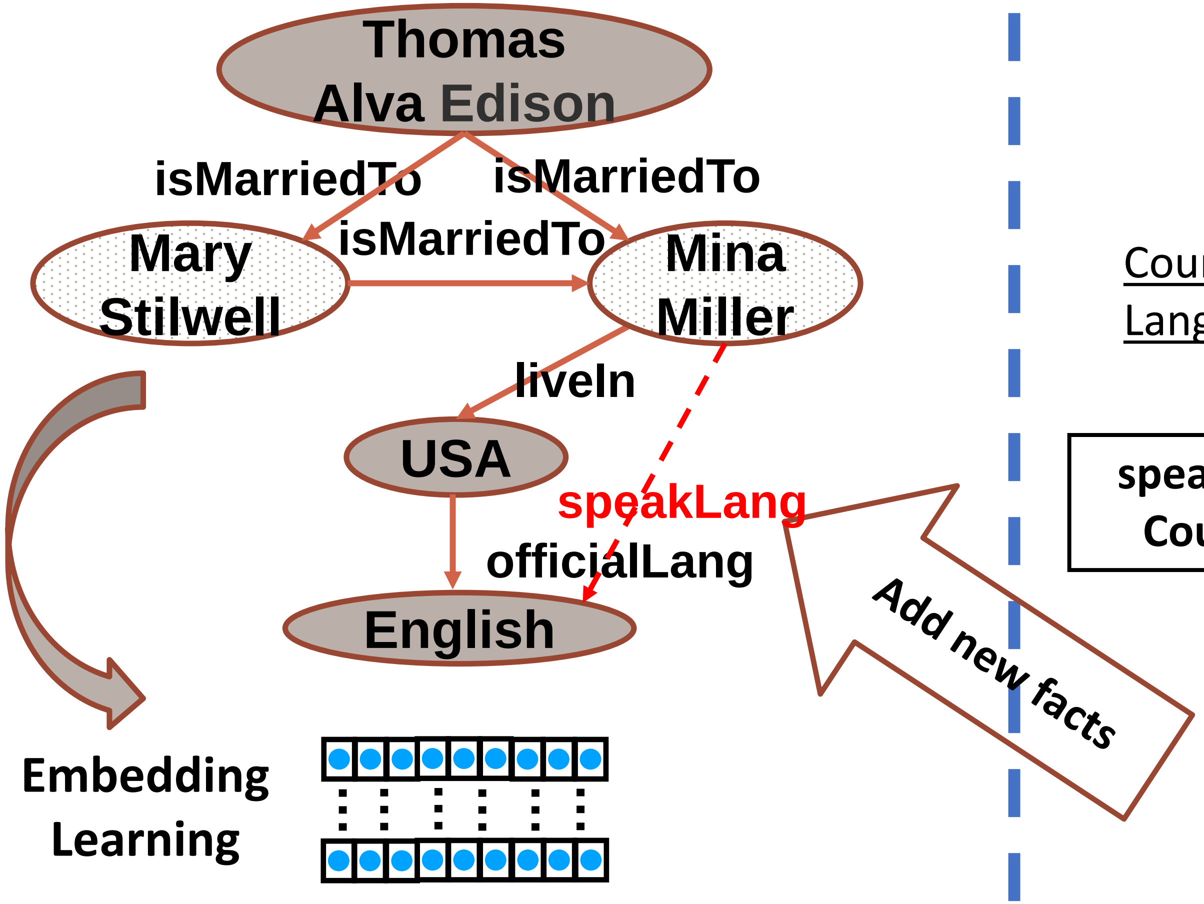
Iterative Mutual Enhancement

- Enhance KGE via logical inference
  - Update KG via forward chaining-based logical reasoning

- Enhance logical inference via KGE
  - Excluding potential incorrect triples
  - Including potential useful hidden triples

Update KG via Forward Chaining-based Logical Reasoning

Knowledge Graph Embedding



Logical Rule-based Reasoning

liveIn (Mina Miller, USA)
officialLanguage (USA, English)

Observed Facts

Country = USA
Language = English

speakLanguage(Person, Language) ⇐ liveIn(Person, Country) ∧ officialLanguage (Country, Language)

Definite Horn rule

Logical Inference

speakLanguage (Mina Miller, English)

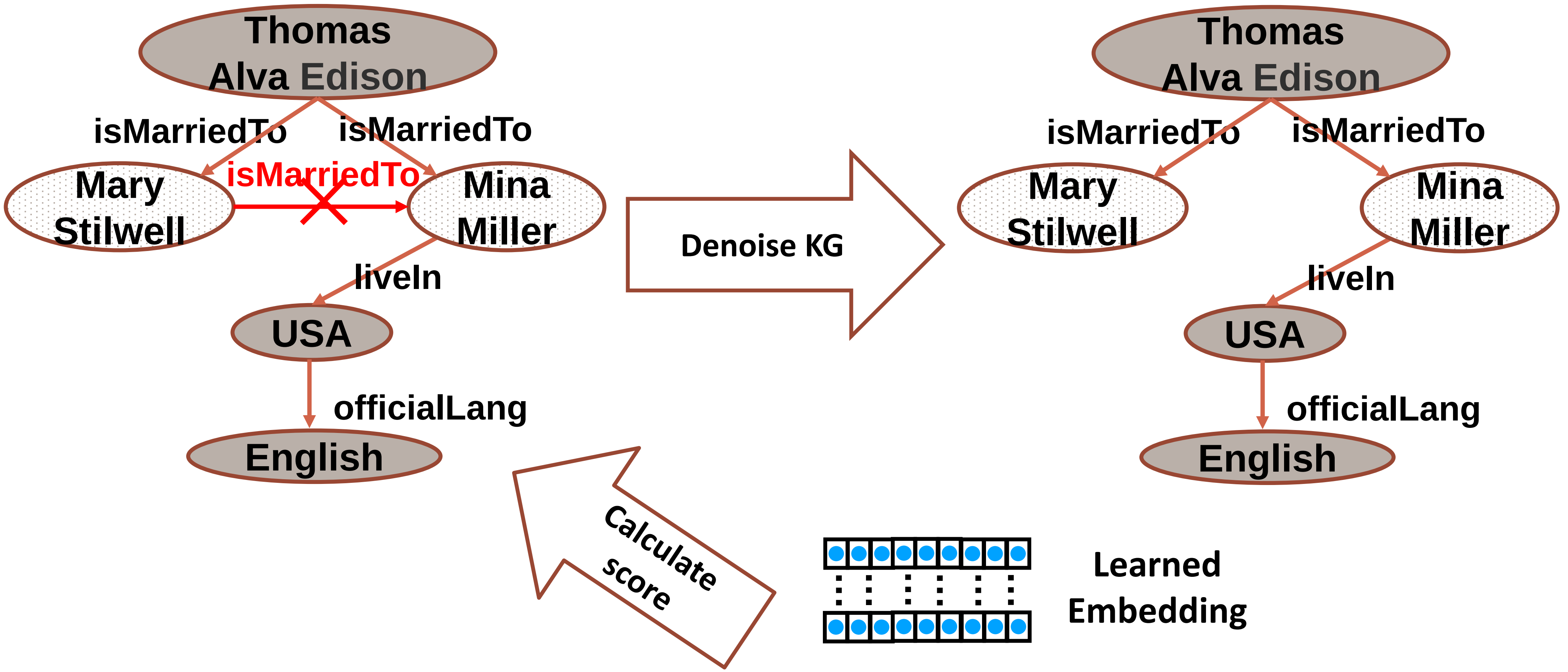
New fact

Iterative Mutual Enhancement

- Enhance KGE via logical inference
  - Update KG via forward chaining-based logical reasoning

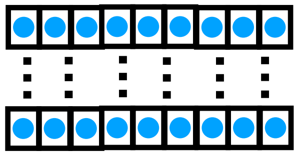
- Enhance logical inference via KGE
  - Excluding potential incorrect triples
  - Including potential useful hidden triples

Excluding potential incorrect triples



Including potential useful hidden triples

Learned Embedding



√ triples in KGs
? triples not in KGs

Observed Facts

liveIn (Mary Stilwell, USA) **?**
officialLanguage (USA, English) **√**

Exist?

Country = USA
Language = English

Person = Mary Stilwell

speakLanguage(Person, Language) ⇐ liveIn(Person, Country) ∧ officialLanguage (Country, Language)

Definite Horn rule

Forward Chaining

Including potential useful hidden triples

√ triples in KGs

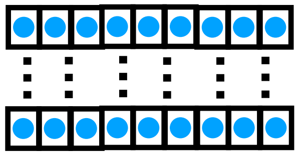
? triples not in KGs

Observed Facts

liveIn (Mary Stilwell, USA) **√**
officialLanguage (USA, English) **√**

Exist?

Learned Embedding



Add!

Country = USA
Language = English

Person = Mary Stilwell

speakLanguage(Person, Language) ⇐ liveIn(Person, Country) ∧ officialLanguage (Country, Language)

Definite Horn rule

Forward Chaining

speakLanguage (Mina Miller, English)

New fact

Experimental Results

- KG completion task

Model	Kinship Hit@1	Kinship Hit@10	Kinship MRR	FB15k-237 Hit@1	FB15k-237 Hit@10	FB15k-237 MRR	WN18RR Hit@1	WN18RR Hit@10	WN18RR MRR
RESCAL	0.489	0.894	0.639	0.108	0.322	0.179	0.123	0.239	0.162
SimplE	0.335	0.888	0.528	0.150	0.443	0.249	0.290	0.351	0.311
HypER†	0.364	0.903	0.551	0.252	0.520	0.341	0.436	0.522	0.465
TuckER†	0.373	0.898	0.567	0.266	0.544	0.358	**0.443**	0.526	0.470
BLP†	-	-	-	0.062	0.150	0.092	0.187	0.358	0.254
MLN	0.655	0.732	0.694	0.067	0.160	0.098	0.191	0.361	0.259
KALE	0.433	0.869	0.598	0.131	0.424	0.230	0.032	0.353	0.172
RUGE	0.495	0.962	0.677	0.098	0.376	0.191	0.251	0.327	0.280
ExpressGNN	0.105	0.282	0.164	0.150	0.317	0.207	0.036	0.093	0.054
pLogicNet	0.683	0.874	0.768	0.261	0.567	0.364	0.301	0.410	0.340
pGAT†	-	-	-	0.377	0.609	0.457	0.395	0.578	0.459
BoxE†	-	-	-	-	0.538	0.337	-	0.541	0.451
TransE	0.221	0.874	0.453	0.231	0.527	0.330	0.007	0.406	0.165
UniKER-TransE	0.873	**0.971**	0.916	0.463	**0.630**	0.522	0.040	0.561	0.307
DistMult	0.360	0.885	0.543	0.220	0.486	0.308	0.304	0.409	0.338
UniKER-DistMult	0.770	0.945	0.823	**0.507**	0.587	0.533	0.432	0.538	0.485
RotatE	0.787	0.933	0.862	0.237	0.526	0.334	0.421	0.563	0.469
UniKER-RotatE	**0.886**	**0.971**	**0.924**	0.495	0.612	**0.539**	0.437	**0.580**	**0.492**

Experimental Results

- A few iterations is good enough

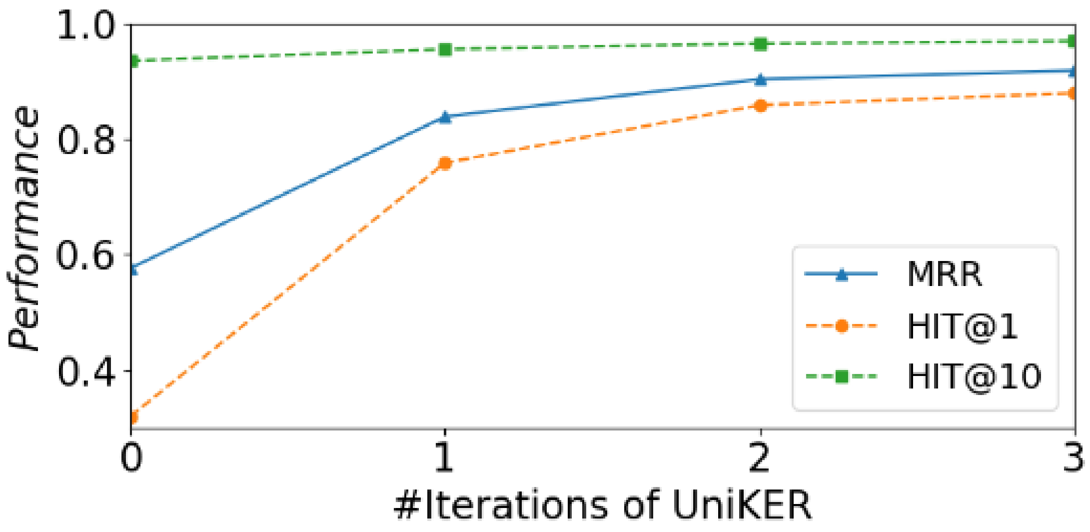


Figure 3: Impact of #iterations on UniKER (KG completion task on Kinship dataset).

Robust to Noise

- construct a noisy dataset with noisy triples to be 40% of original data.

Model	θ	Hit@1	Hit@10	MRR
TransE	-	0.026	0.800	0.319
UniKER-TransE	10	0.286	0.776	0.466
	20	0.311	0.816	0.503
	30	0.322	**0.833**	0.520
	40	**0.352**	0.812	**0.523**
	50	0.292	0.791	0.486

Table 3: Ablation study on noise threshold $\theta\%$ on Kinship dataset (whose train set is injected with noise)

Effcient

- Evaluate the scalability of forward chaining against a number of SOTA inference algorithms for MLN

Model	sub-YAGO3-10	sub-Kinship	RC1000	Kinship	FB15k-237	WN18RR
MCMC	76433s	-	-	-	-	-
MCSAT	1292s	25912s	-	-	-	-
BP	10s	16343s	-	-	-	-
liftedBP	15s	16075s	-	-	-	-
Tuffy	0.849s	1.398s	4.899s	-	-	-
Forward Chaining	0.003s	0.034s	0.007s	0.593s	186s	30s

Table 7: Comparison of Inference Time for Forward Chaining vs. MLN.

Outline

- Introduction
- Bringing First-Order Logic into Uncertain KG Embedding
- UniKER: Integrating Horn Rule Reasoning into KGE
- Summary

Take Away

- Two methodologies in KG inference
  - Embedding-based approach
  - Logical rule-based reasoning
- Combination of the two worlds is the promising direction
  - Embedding can handle noise and uncertainty of KG
  - Logical rules provide higher-order dependency constraints among entities and relations
- Different ways of combination
  - UniKER is the best solution if the logical rules are confined to Horn rules

References

- [1] Guo S, Wang Q, Wang L, et al. Jointly embedding knowledge graphs and logical rules[C]//Proceedings of the 2016 conference on empirical methods in natural language processing. 2016: 192-202.
- [2] Guo S, Wang Q, Wang L, et al. Knowledge graph embedding with iterative guidance from soft rules[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).
- [3] Rocktäschel T, Singh S, Riedel S. Injecting logical background knowledge into embeddings for relation extraction[C]//Proceedings of the 2015 conference of the north American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015: 1119-1129.
- [4] Qu M, Tang J. Probabilistic logic neural networks for reasoning[J]. arXiv preprint arXiv:1906.08495, 2019.
- [5] Zhang Y, Chen X, Yang Y, et al. Can Graph Neural Networks Help Logic Reasoning?[J]. arXiv preprint arXiv:1906.02111, 2019.
- [6] Harsha Vardhan L V, Jia G, Kok S. Probabilistic logic graph attention networks for reasoning[C]//Companion Proceedings of the Web Conference 2020. 2020: 669-673.
- [7] Kimmig A, Bach S, Broecheler M, et al. A short introduction to probabilistic soft logic[C]//Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications. 2012: 1-4.
- [8] Richardson M, Domingos P. Markov logic networks[J]. Machine learning, 2006, 62(1-2): 107-136.
- [9] Ren, H., Hu, W., and Leskovec, J. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. ICLR'2020.
- [10] Ren, H. and Leskovec, J. Beta embeddings for multi-hop logical reasoning in knowledge graphs. NeurIPS'2020.
- [11] Hamilton, W. L., Bajaj, P., Zitnik, M., Jurafsky, D., and Leskovec, J. Embedding logical queries on knowledge graphs. NeurIPS'2018.

Q & A

- Thanks to my collaborators:
  - Shirley Chen, Vivian Cheng, Junheng Hao, Muhao Chen, Wei Wang, Carlo Zaniolo