How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View

李韧,[1] 曹亚男,[1] 朱倩男,[2] 毕冠群,[1] 方芳,[1] 柳毅,[3] 李倩 [4]

[1] 中科院信息工程研究所
[2] 中国人民大学高瓴人工智能学院
[3] 国家计算机网络应急技术处理协调中心
[4] 悉尼科技大学

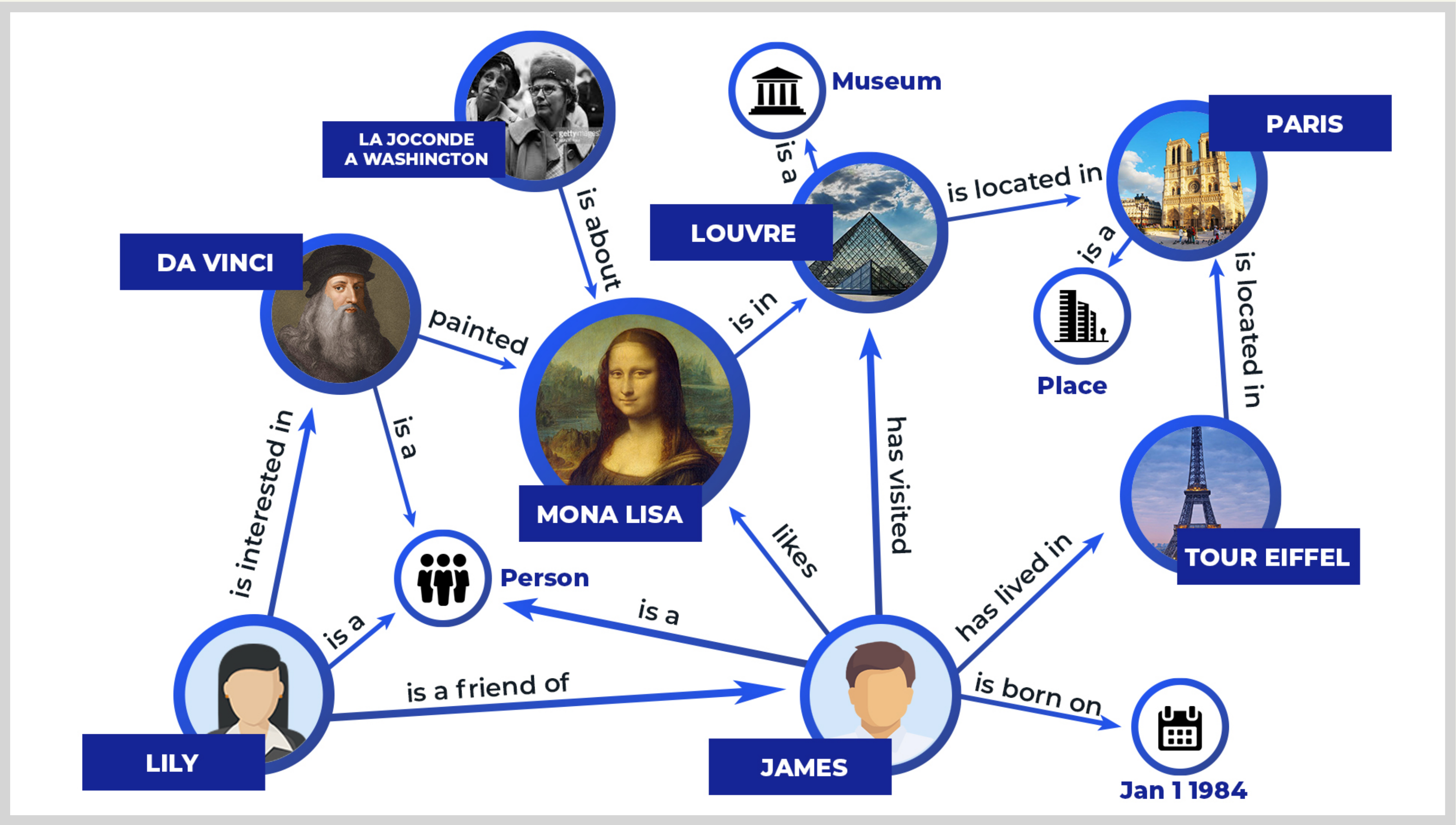
中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

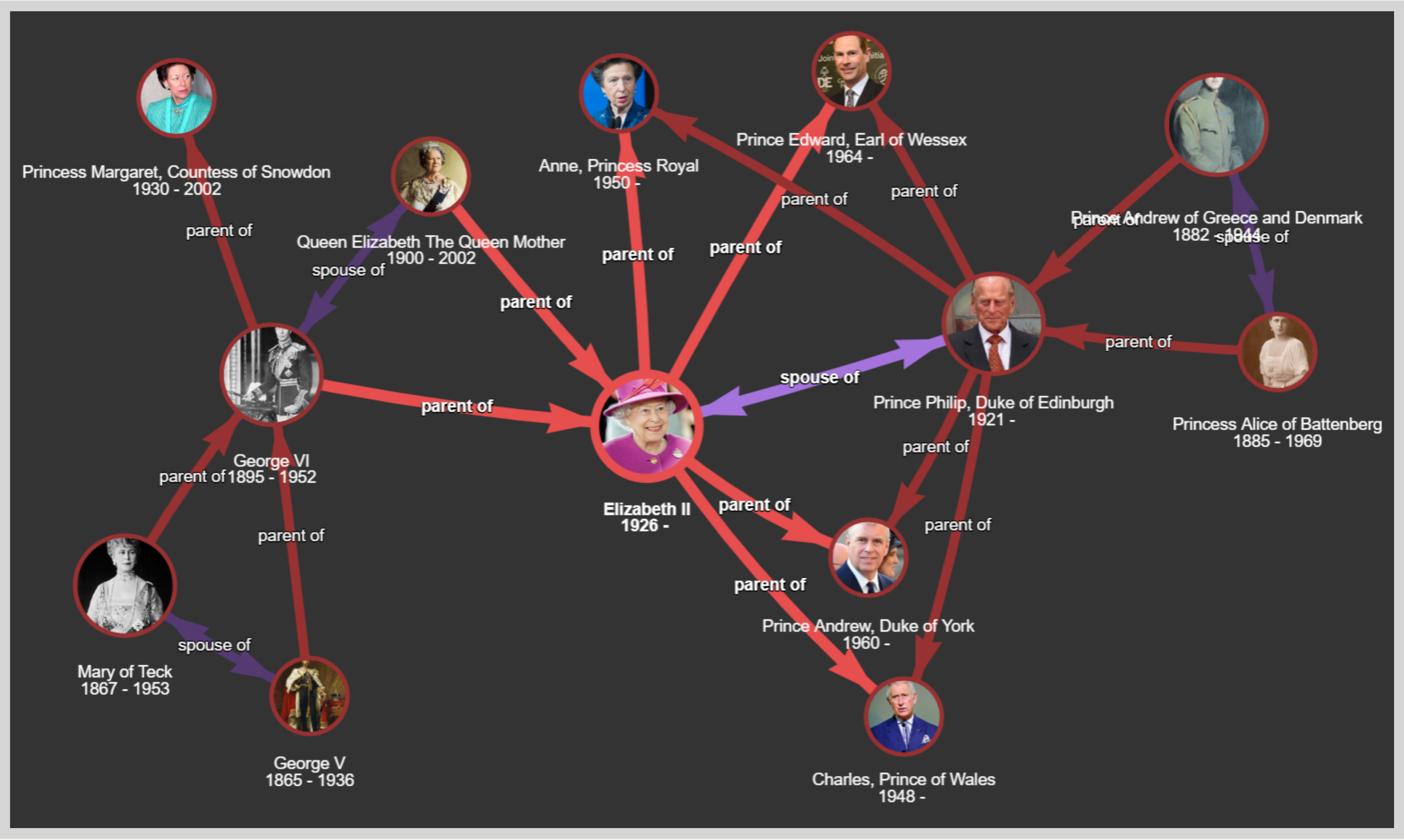
报告人: 李韧

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Background

- Knowledge Graphs (KGs): graph structured knowledge representation in triple form (h, r, t), denoting "*head entity* h *and tail entity* t *satisfy relation* r" .



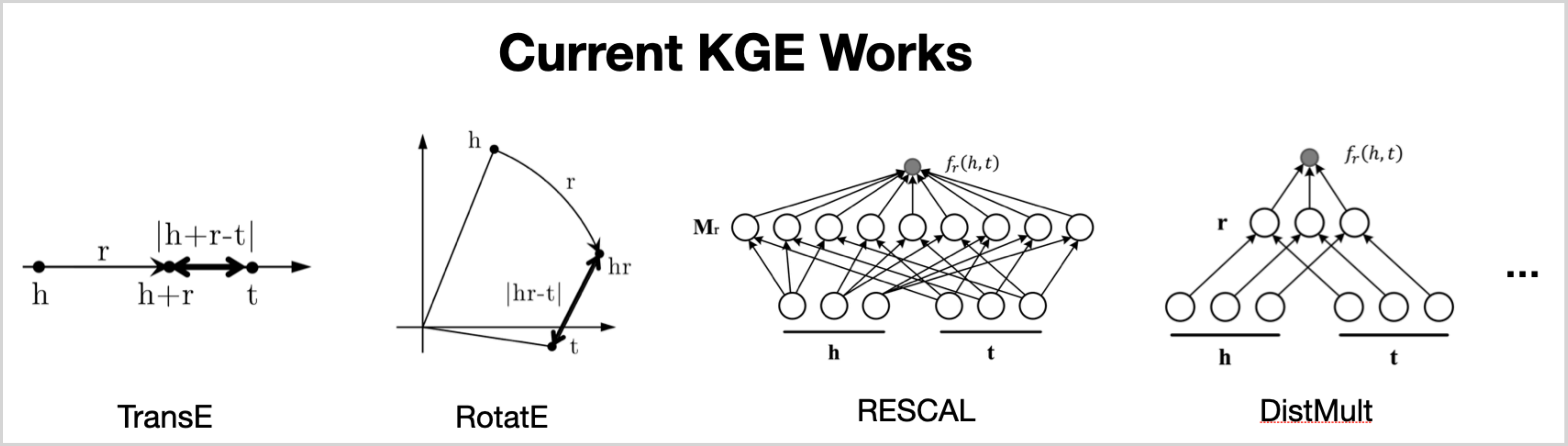


- Knowledge Graph Embedding (KGE): learning embedding for entities and relations in KG.
- Evaluation task: $(h, r, ?) \rightarrow t$ or $(?, r, t) \rightarrow h$, also known as Knowledge Graph Completion, Link Prediction (KG context).

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Knowledge Graph Embedding (KGE) Extrapolation

- Most KGE works gain great success on extrapolation scenarios, i.e. given an unseen triple (h, r, t), a trained model can still correctly predict $(h, r, ?) \rightarrow t$ or $(?, r, t) \rightarrow h$.
- Current KGE works mainly focus on the design of delicate triple modeling function $f(h, r, t)$, while lack the exploration of why / how the methods can extrapolate to unseen data.

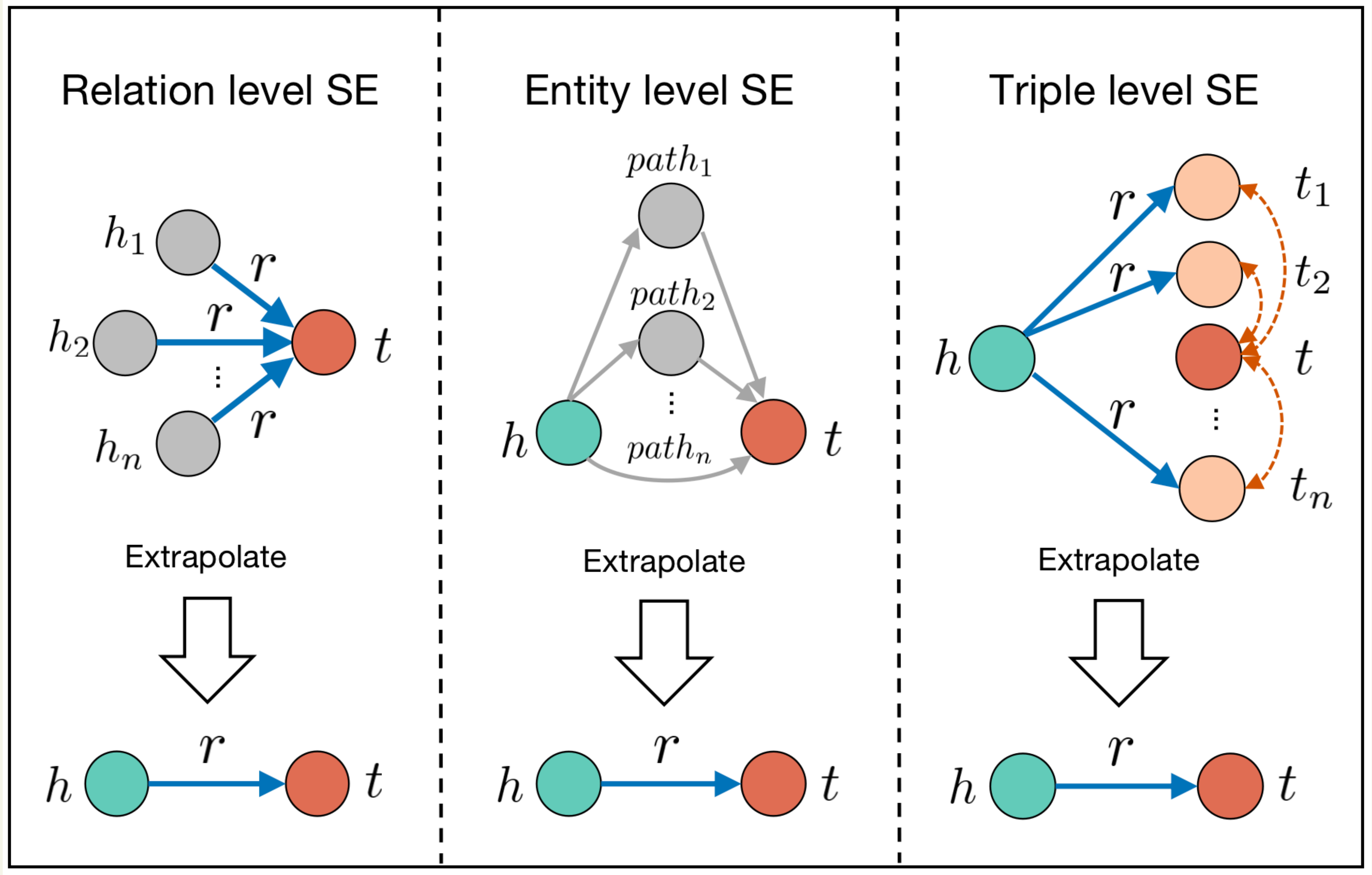


- In this work, we study the KGE extrapolation of two problems:
  - 1. How does KGE extrapolate to unseen data?
  - 2. How to design the KGE model with better extrapolation ability?

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

How does KGE extrapolate to unseen data?

- Prediction $(h, r, ?) \rightarrow t$ can be regarded as the semantic matching between $\mathrm{query}(h, r)$ and t.
- A good extrapolative matching means $\mathrm{query}(h, r)$ and t have obtained some semantic relatedness during training.
- The relatedness may come from following three levels:
  - Relation level: The co-occurrence between r and t in train set.
  - Entity level: The paths from h to t in train set.
  - Triple level: The similarity between t and other ground truth entity t' of $\mathrm{query}(h, r)$ in train set.



How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

How does KGE extrapolate to unseen data?

- Relation level: The high frequency of $\mathrm{query}(h_i, r) \rightarrow t$ in train set will make r contain the information to predict t. Intuitively this can be seen as entity type information.

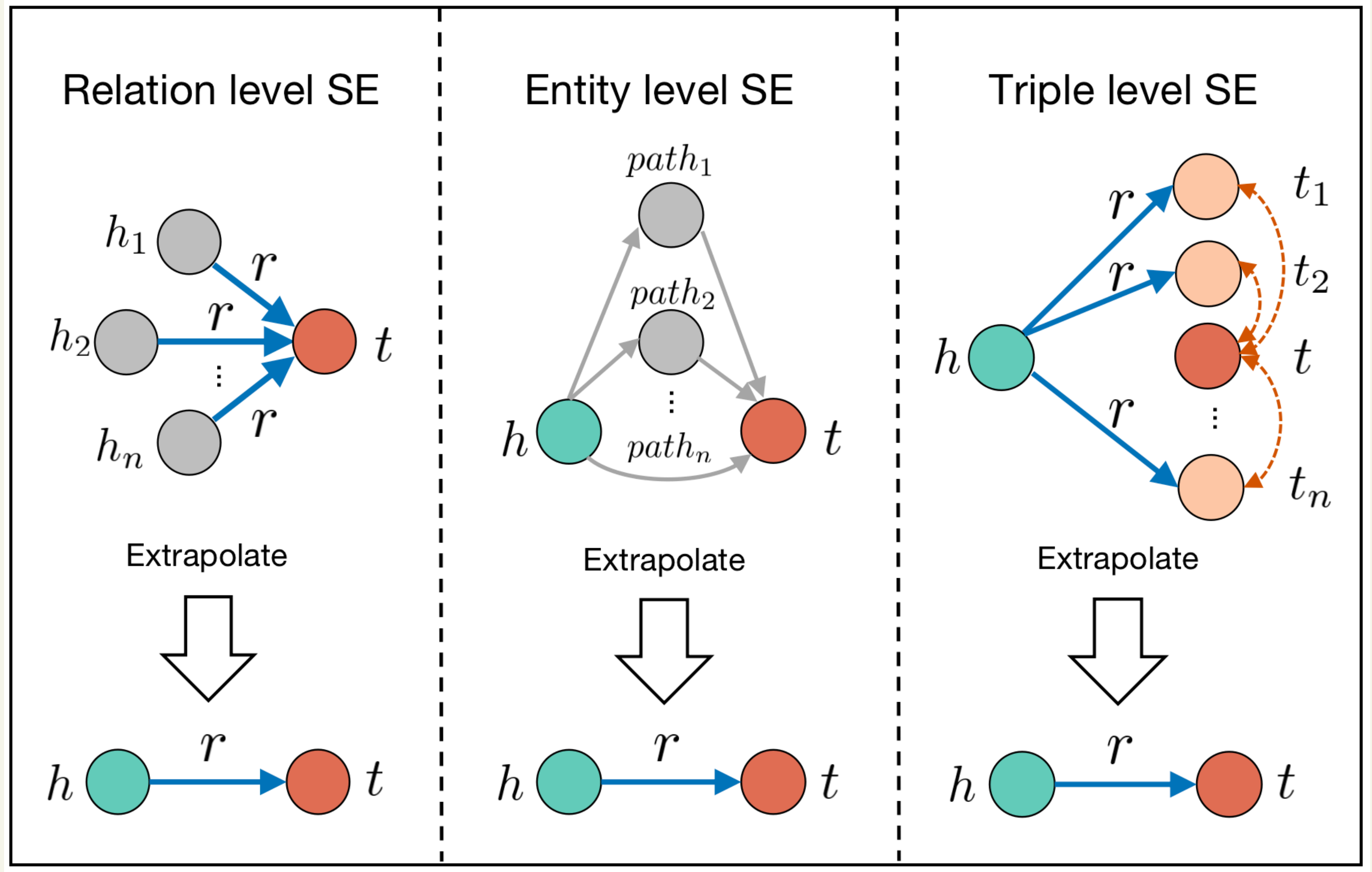
Example: $\mathrm{query}(h_i, born\_in) \rightarrow Florida$ *should be higher than* $Iron\_Man$ *, no matter what specific* h_i *is*

- Entity level: The observed queries or indirect queries from h to t in train set will close their semantic relevancy and provide evidences for other queries.

Example: $\mathrm{query}(h, is\_mother) \rightarrow e_1$ **,** $\mathrm{query}(e_1, is\_father) \rightarrow t \Rightarrow \mathrm{query}(h, is\_grandmother) \rightarrow t$

- Triple level: If the model has been trained for $\mathrm{query}(h, r) \rightarrow t'$, meanwhile t and t' share much similarity, it will be natural to extrapolate to $\mathrm{query}(h, r) \rightarrow t$.

- We name such relatedness as Semantic Evidences (SEs), to indicate the supporting semantic information they provide for extrapolation.



How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Measurement of SEs

- Relation level SE: the co-occurrence between r and t in train set.

$$S_{rel} = \left| \{(h_i, r, t) \mid (h_i, r, t) \in \mathcal{F}_{tr}\} \right|$$

- Entity level SE: the number of paths from h to t in train set.

$$S_{ent} = \left| \{(h, r_i, t) \mid (h, r_i, t) \in \mathcal{F}_{tr}\} \right| + \left| \{(h, r_i, e_k, r_j, t) \mid (h, r_i, e_k), (e_k, r_j, t) \in \mathcal{F}_{tr}\} \right|$$

- Triple level SE: the similarity between t and other ground truth entity t' in train set.

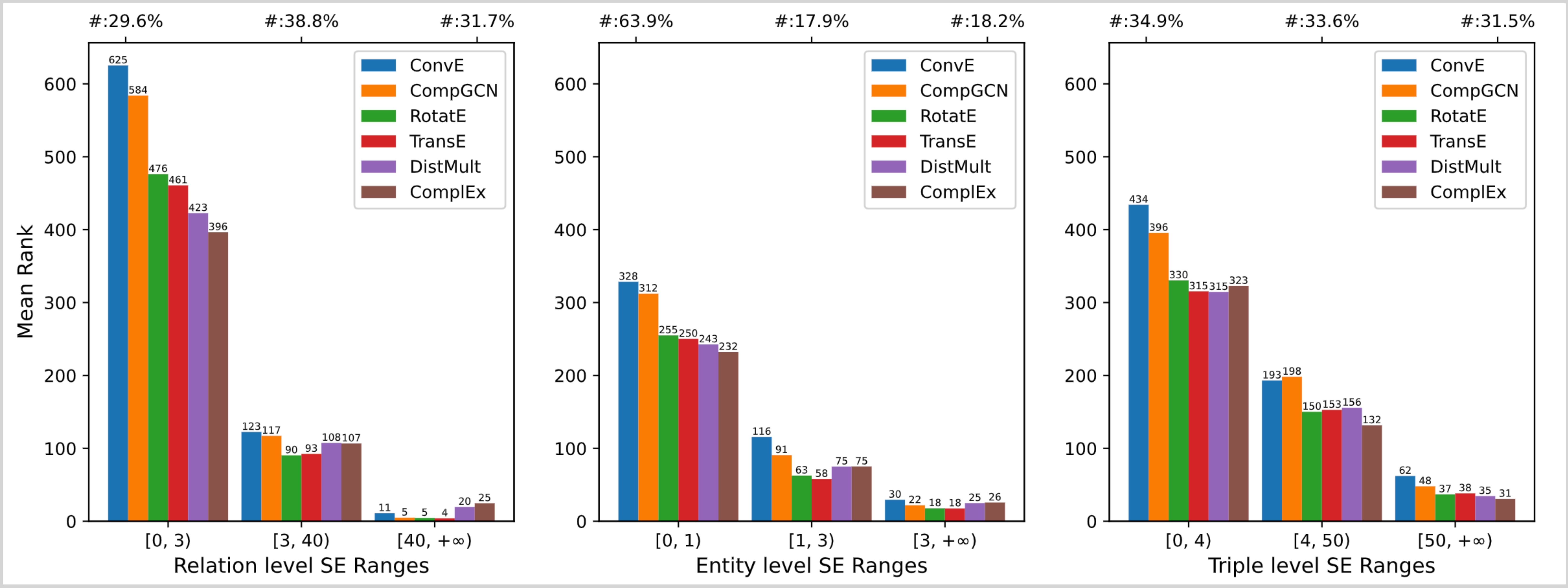
$$S_{tri} = \sum_{t'} \mathrm{Sim}(t, t'), (h, r, t') \in \mathcal{F}_{tr}$$

$$\mathrm{Sim}(t, t') = \left| \{(h_i, r_i) \mid (h_i, r_i, t) \in \mathcal{F}_{tr}\} \cap \{(h_i, r_i) \mid (h_i, r_i, t') \in \mathcal{F}_{tr}\} \right|$$

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Effectiveness Verification of SEs

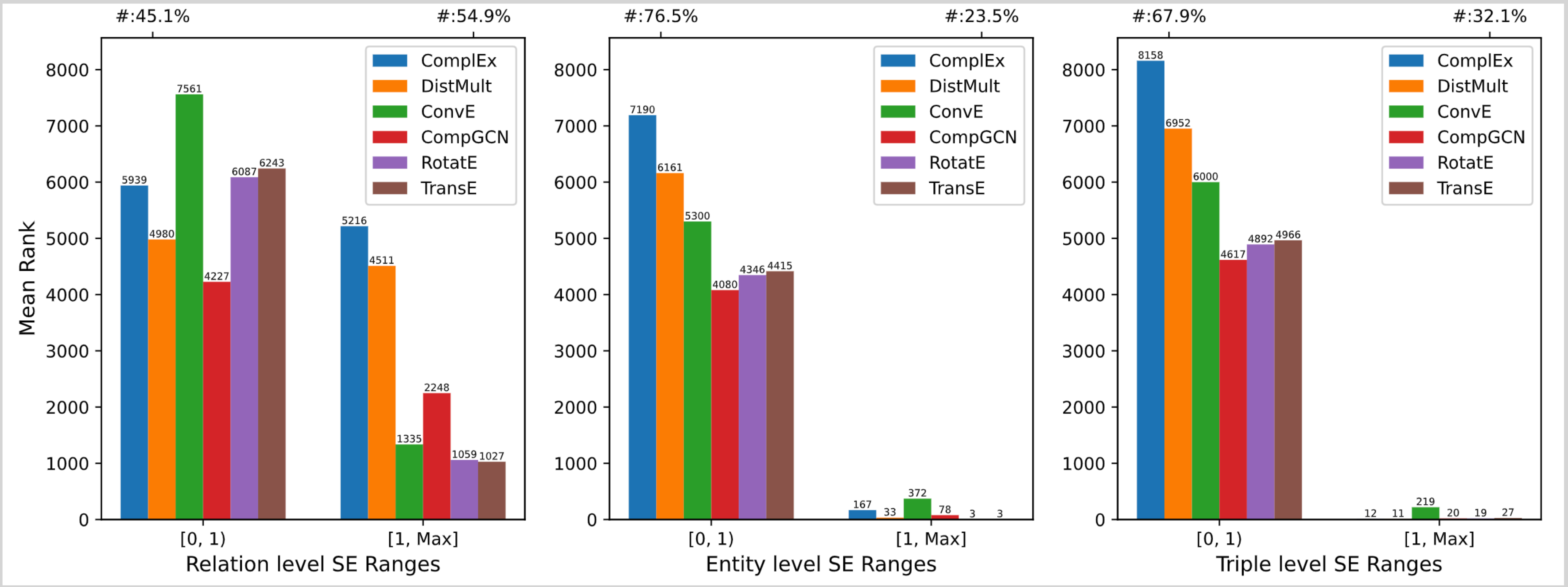
- Reproduce several typical KGE models (TransE, RotatE, DistMult…)
- Evaluate their extrapolation performance on test data with different SE strength.
- Results on FB15k-237 dataset:
  - x-axis: SE strength range (large number indicates abundant SE information)
  - y-axis: prediction rank of target entity (small number indicates good performance)
  - More abundant SE information, better extrapolation performance.



How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Effectiveness Verification of SEs

- The phenomenon also holds on WN18RR dataset.



- Note that different dataset may reveal a different focus for three SEs when extrapolating, because of the various data characteristics.

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

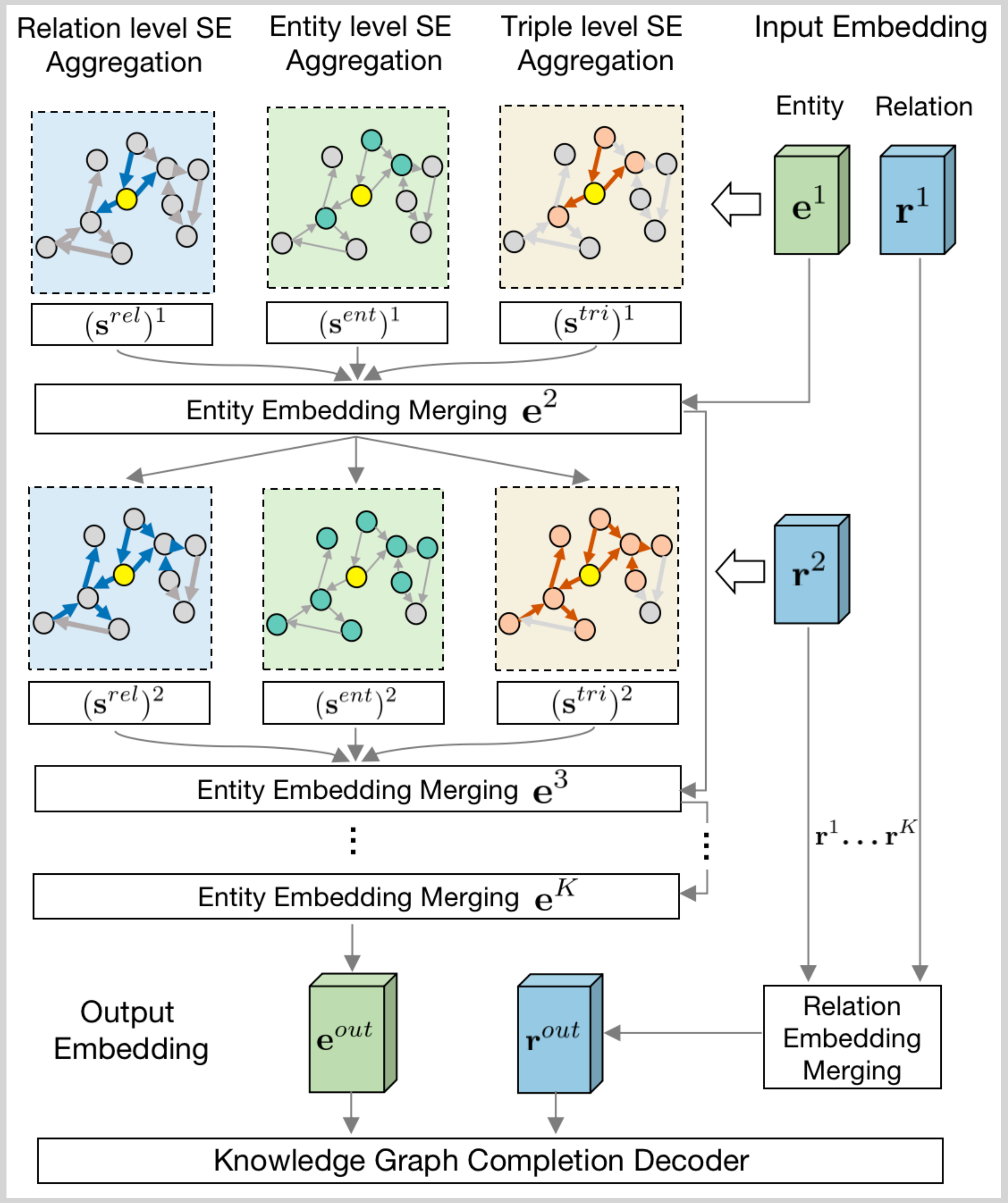
Qualitative Case Study

	Extrapolating Cases	Observed Evidences in Train Set	SE-GNN Rank	TransE/RotatE/DistMult ComplEx/ConvE Rank
Relation level SE	(San Diego, travel_month, ?) → December	(Miami, travel_month, December) (Zurich, travel_month, December) (Melbourne, travel_month, December) ...	**1**	[**1**, **1**, **1**, **1**, **1**]
	(The Orphanage (film), language, ?) → Spanish	(Midnight in Barcelona (film), language, Spanish) (Todo sobre mi madre (film), language, Spanish) (The Treasure of the Sierra Madre (film), language, Spanish)...	**1**	[2, **1**, 2, **1**, 2]
	(Amy Pohler, gender, ?) → Female	(Erika Christensen, gender, Female) (Nancy Snyderman, gender, Female) (Olivia Williams, gender, Female) ...	**1**	[**1**, **1**, 2, **1**, **1**]
Entity level SE	(Robert Downey Jr, live_in, ?) → New York City	(Robert Downey Jr, friendship, Jenifer Aniston, live_in, New York City) (Robert Downey Jr, place_of_birth, Manhattan, administrative_parent, New York City) (Robert Downey Jr , spouse, Sarah Jessica Parker, live_in, New York City) ...	**2**	[**2**, **2**, **2**, 7, 7]
	(England, contains, ?) → Watford	(Englaland, contains, Hertfordshire, contains, Watford) (Englaland, second_level_divisions, Hertfordshire, contains, Watford) (Englaland, administrative_parent, UK of GB and NI, contains, Watford) ...	**5**	[28. 37, 15, 9, 21]
	(Amber Riley, profession, ?) → Theatre actress	(Amber Riley, meanwile_award, Diana Agron, profession, Theatre actress) (Amber Riley, meanwile_award, Jessalyn Gilsig, profession, Theatre actress) (Amber Riley, meanwile_award, Naya Rivera, profession, Theatre actress) ...	**1**	[**1**, **1**, **1**, **1**, **1**]
Triple level SE	(Freshman Program, major, ?) → Computer Science	(Freshman Program, major, Mathmatics) & Mathmatics ~ Computer Science (Freshman Program, major, Electrical Eng.) & Electrical Eng. ~ Computer Science (Freshman Program, major, Chemical Science) & Chemical Science ~ Computer Science	**1**	[7, 8, 2, 2, **1**]
	(Dorothy Fields, profession, ?) → Lyricist	(Dorothy Fields, profession, Songwriting) & Songwriting ~ Lyricists (Dorothy Fields, profession, Scenario Writer) & Scenario Writer ~ Lyricists (Dorothy Fields, profession, Dramatist) & Dramatist ~ Lyricists	**1**	[**1**, **1**, **1**, **1**, **1**]
	(RoboCop (film), genre, ?) → Thriller film	(RoboCop (film), genre, Action movie) & Action movie ~ Thriller film (RoboCop (film), genre, Murder mystery) & Murder mystery ~ Thriller film (RoboCop (film), genre, Superhero movie) & Superhero movie ~ Thriller film	**1**	[2, 2, 3, 5, **1**]

How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

How to design the KGE model with better extrapolation ability?

- Because of no awareness of extrapolation factors, current KGE models capture SE information mainly through an implicit and insufficient way, which limits their extrapolation ability.
- For more extrapolative knowledge representation, we propose SE-GNN (Semantic Evidence aware Graph Neural Networks), where each SE is explicitly captured from corresponding neighbor pattern, and sufficiently modeled by multi-layer neighbor aggregation.
- Relation SE aggregation: $\mathbf{s}_i^{rel} = \sigma\left(\sum_{(e_j, r_j)\in\mathcal{N}_i} \alpha_{ij}^{rel} W^{rel} \mathbf{r}_j\right)$
- Entity SE aggregation: $\mathbf{s}_i^{ent} = \sigma\left(\sum_{(e_j, r_j)\in\mathcal{N}_i} \alpha_{ij}^{ent} W^{ent} \mathbf{e}_j\right)$
- Triple SE aggregation: $\mathbf{s}_i^{tri} = \sigma\left(\sum_{(e_j, r_j)\in\mathcal{N}_i} \alpha_{ij}^{tri} W^{tri} \varphi(\mathbf{e}_j, \mathbf{r}_j)\right)$
- Merging with original embedding: $\mathbf{e}_i' = \mathbf{e}_i + \mathbf{s}_i^{rel} + \mathbf{s}_i^{ent} + \mathbf{s}_i^{tri}$



How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Experiment Results

- Knowledge Graph Completion

Models	FB15k-237					WN18RR				
	MRR	MR	H@1	H@3	H@10	MRR	MR	H@1	H@3	H@10
Translational Distance										
TransE (Bordes et al. 2013)†	.330	173	.231	.369	.528	.223	3380	.014	.401	.529
RotatE (Sun et al. 2019)	.338	177	.241	.375	.533	.476	3340	.428	.492	.571
PaiRE (Chao et al. 2021)	.351	160	.256	.387	.544	-	-	-	-	-
Semantic Matching										
DistMult (Yang et al. 2015)†	.308	173	.219	.336	.485	.439	4723	.394	.452	.533
ComplEx (Trouillon et al. 2016)†	.323	165	.229	.353	.513	.468	5542	.427	.485	.554
TuckER (Balazevic and Allen 2019)	.358	-	.266	.394	.544	.470	-	.443	.482	.526
ConvE (Dettmers et al. 2018)	.325	244	.237	.356	.501	.430	4187	.400	.440	.520
InteractE (Vashishth et al. 2020a)	.354	172	.263	-	.535	.463	5202	.430	-	.528
PROCRUSTES (Peng et al. 2021)	.345	-	.249	.379	.541	.474	-	.421	.502	.569
GNN-based										
R-GCN (Schlichtkrull et al. 2018)	.248	-	.151	-	.417	-	-	-	-	-
KBGAT (Nathani et al. 2019)‡	.157	270	-	-	.331	.412	**1921**	-	-	.554
SACN (Shang et al. 2019)	.350	-	.260	.390	.540	.470	-	.430	.480	.540
A2N (Bansal et al. 2019)	.317	-	.232	.348	.486	.450	-	.420	.460	.510
CompGCN (Vashishth et al. 2020b)	.355	197	.264	.390	.535	.479	3533	.443	.494	.546
SE-GNN (ours)	**.365**	**157**	**.271**	**.399**	**.549**	**.484**	3211	**.446**	**.509**	**.572**

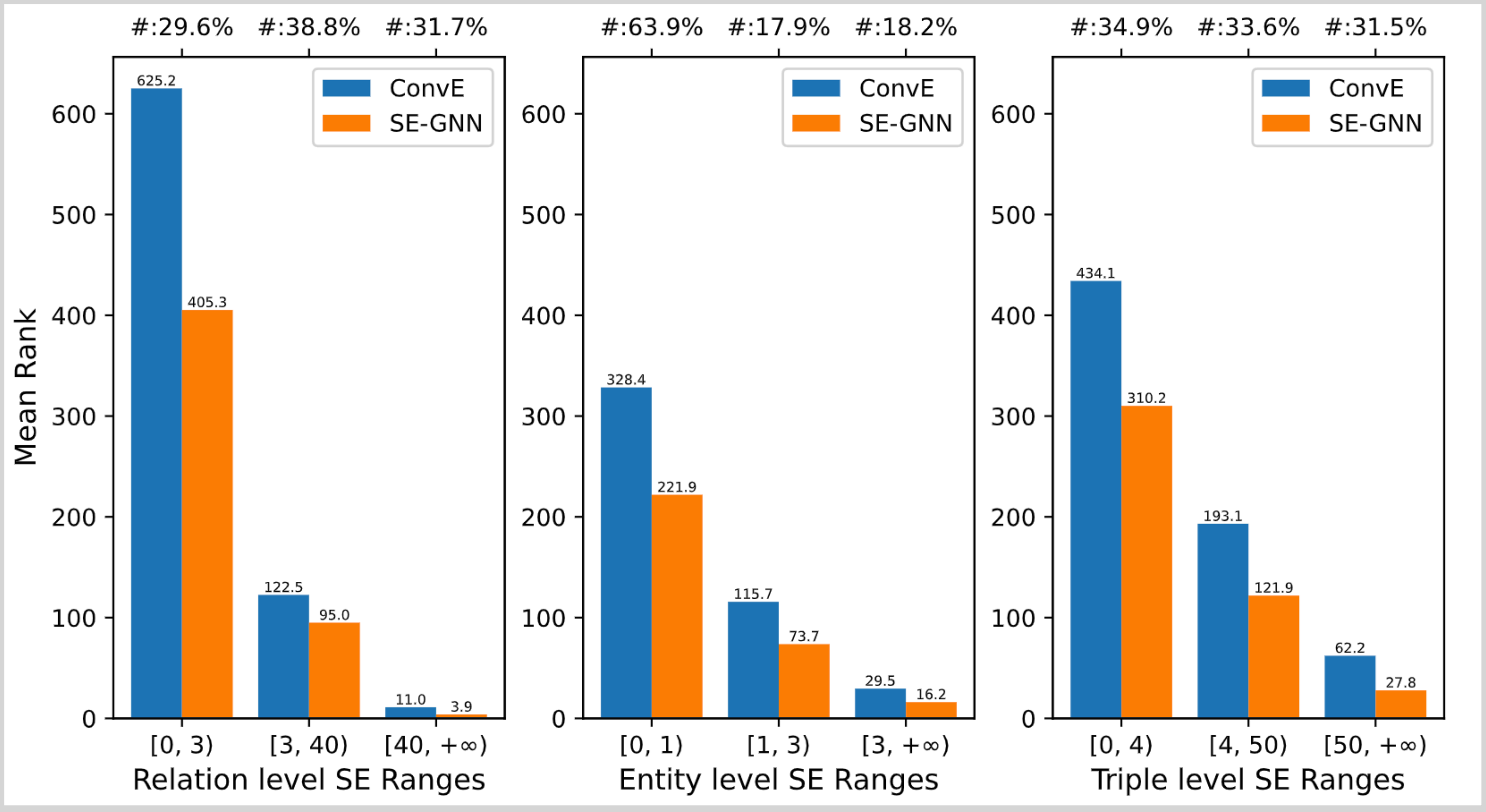
How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View (AAAI'22)

Modeling Efficiency of Semantic Evidences

- To control the variables, we compare the results with ConvE, which is the decoder used in SE-GNN, so the only difference is the SE aggregation encoder.

Ablation Study

- Remove only one SE modeling part and simultaneously remove two of them.



Models	FB15k-237			
	MRR	MR	H@1	H@10
SE-GNN	**.365**	**157**	**.271**	**.549**
w/o relation SE	.361	168	.264	.542
w/o triple SE	.359	173	.262	.537
w/o entity SE	.360	172	.265	.539
w/o relation & entity SE	.357	179	.257	.532
w/o relation & triple SE	.355	181	.254	.535
w/o entity & triple SE	.352	185	.249	.525

Thanks for listening

Q & A