LUD lab Achievements Exhibition

--拔尖计划2.0线上书院主题活动周

September 20, 2022

Zhaoxuan Tan, Zhenyu Lei, Xinshun Feng, Wenqian Zhang, Hongrui Wang

LUD Lab, Xi’an Jiaotong University

luoundergradxjtu.github.io

What is the LUD Lab?

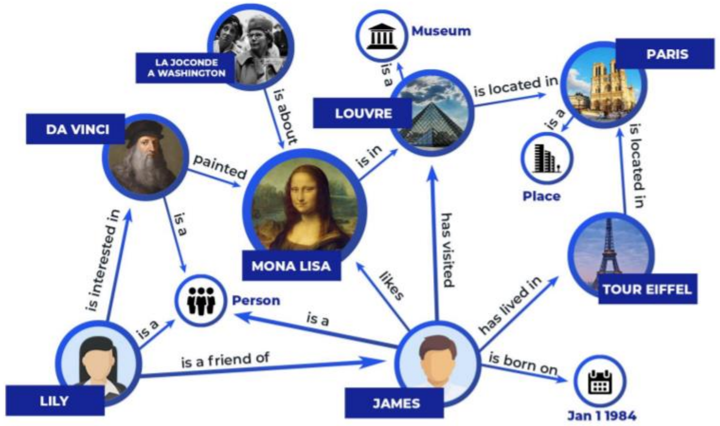
Prof. Minnan Luo

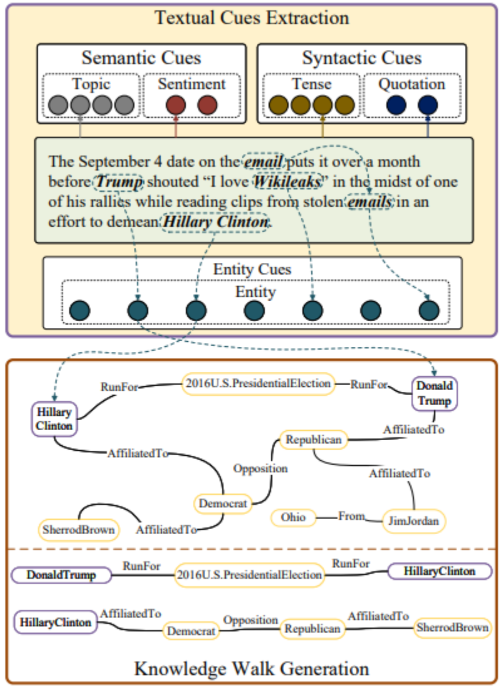
- **L**uo lab **U**ndergraduate **D**ivision
- bridge the gap between undergraduate studies and graduate research

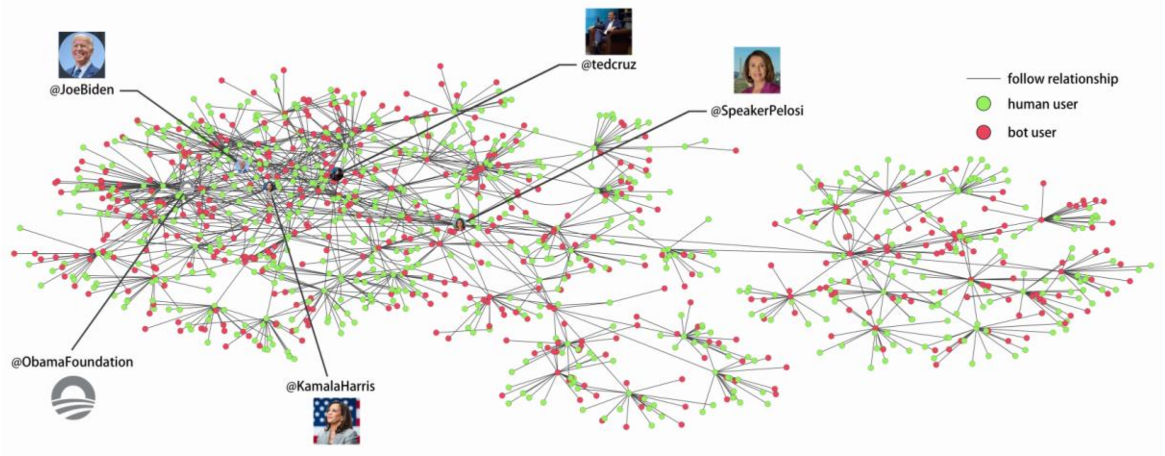
- 12 active members
  - 8 seniors + 2 juniors + 2 sophomore
- Top-notch publications
  - Published: [CIKM’21 a], [CIKM’21 b], [ASONAM’21], [AAAI’22], [NAACL’22], [CIKM’22], [NeurIPS’22]
  - Submission: [EMNLP’22], [AAAI’23]*4

Our Research Interests

- Knowledge Graphs
- Social Network Analysis
- Graph Neural Networks
- Natural Language Processing







Knowledge Graph

KRACL: Contrastive Learning with Graph Context Modeling for Sparse Knowledge Graph Completion

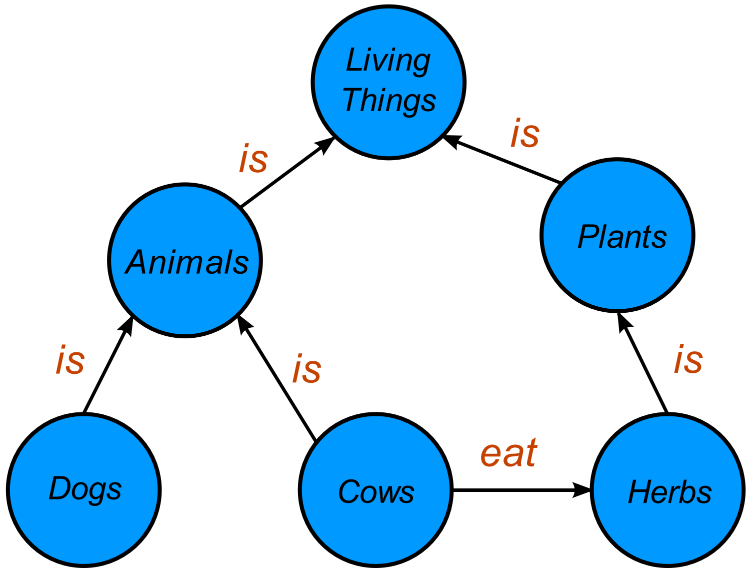
Zhaoxuan Tan

Director, LUD Lab, Xi'an Jiaotong University

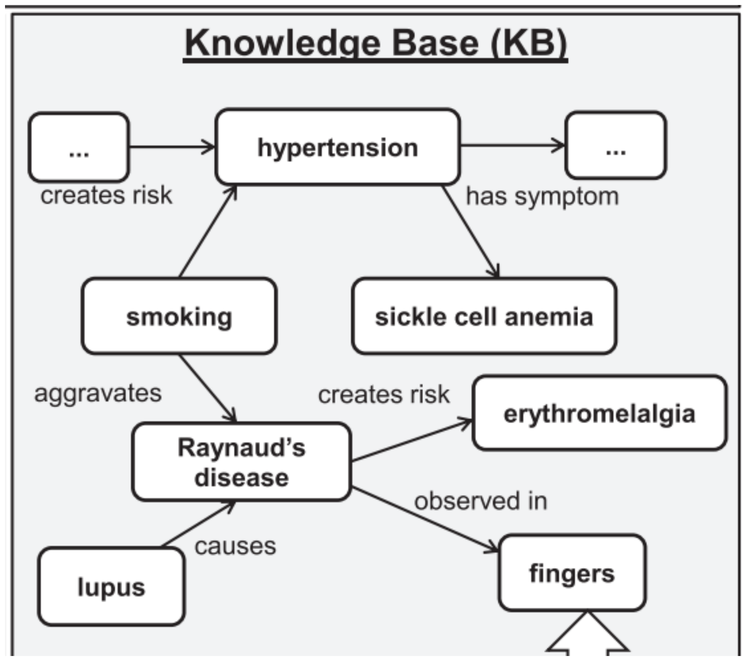
zhaoxuan.info

Knowledge Graphs (KGs)

- Structured representation of commonsense and domain knowledge



commonsense



Domain-specific

KGs are incomplete

- Supervised (manually build)
  - Freebase
  - Wikidata

- Semi-supervised (human-in-the-loop)
  - NELL
  - Knowledge Vault

Freebase

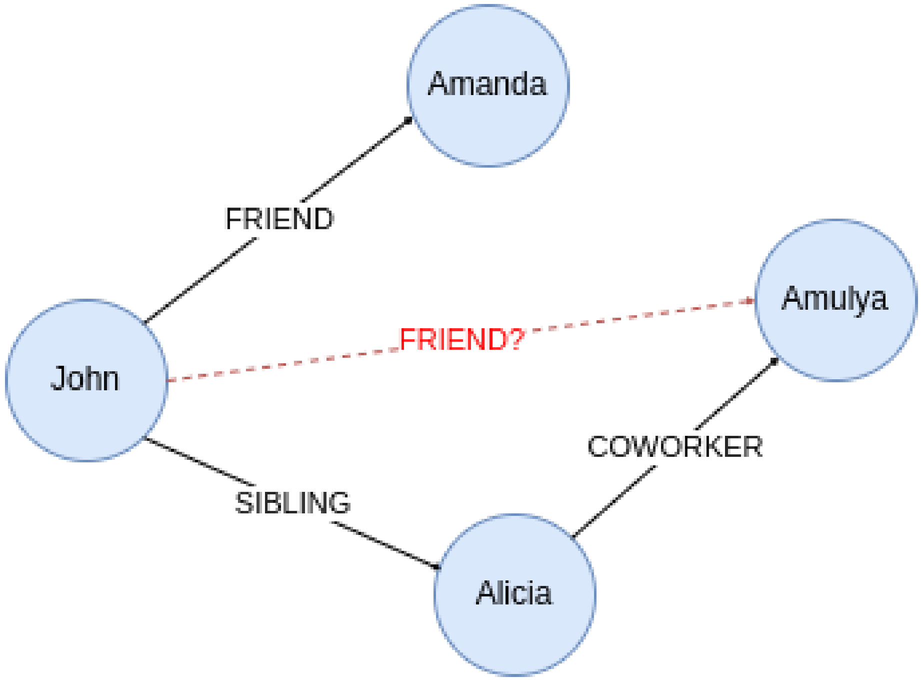
Wikidata

NELL

Knowledge Vault

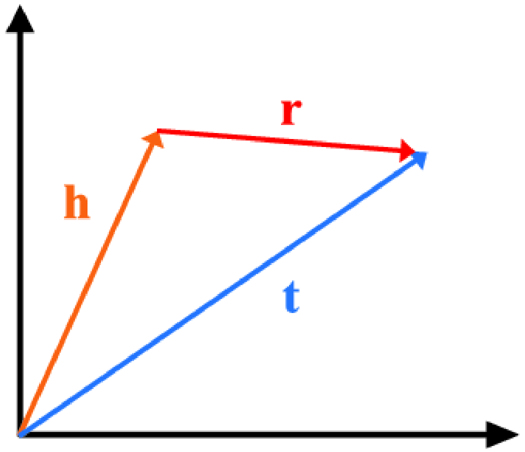
Knowledge Graph Completion

- Form: Given **(h, r, ?)**, predict **t**
- Link prediction
- 1 hop knowledge query

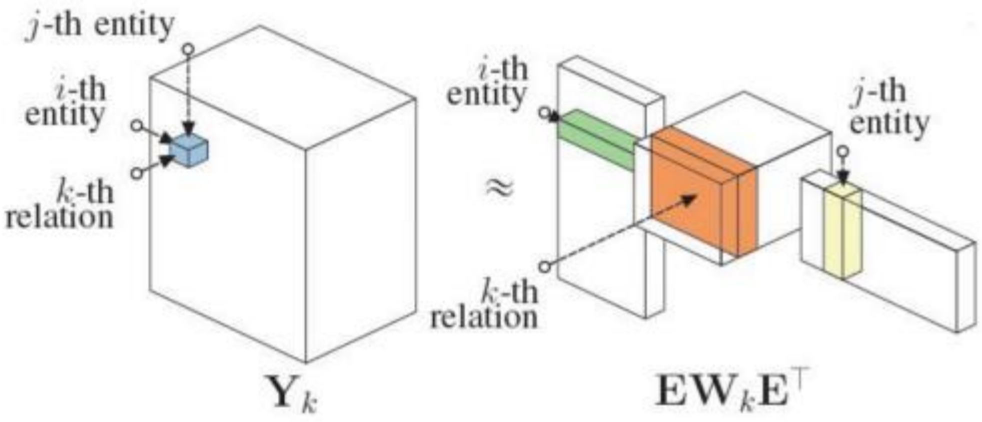


Related work

- Translation-based model
  - TransE
  - RotatE
- Factorization-based model
  - RESCAL
  - DistMult
- Neural-based model
  - ConvE
  - ConvKB
  - HittER



TransE



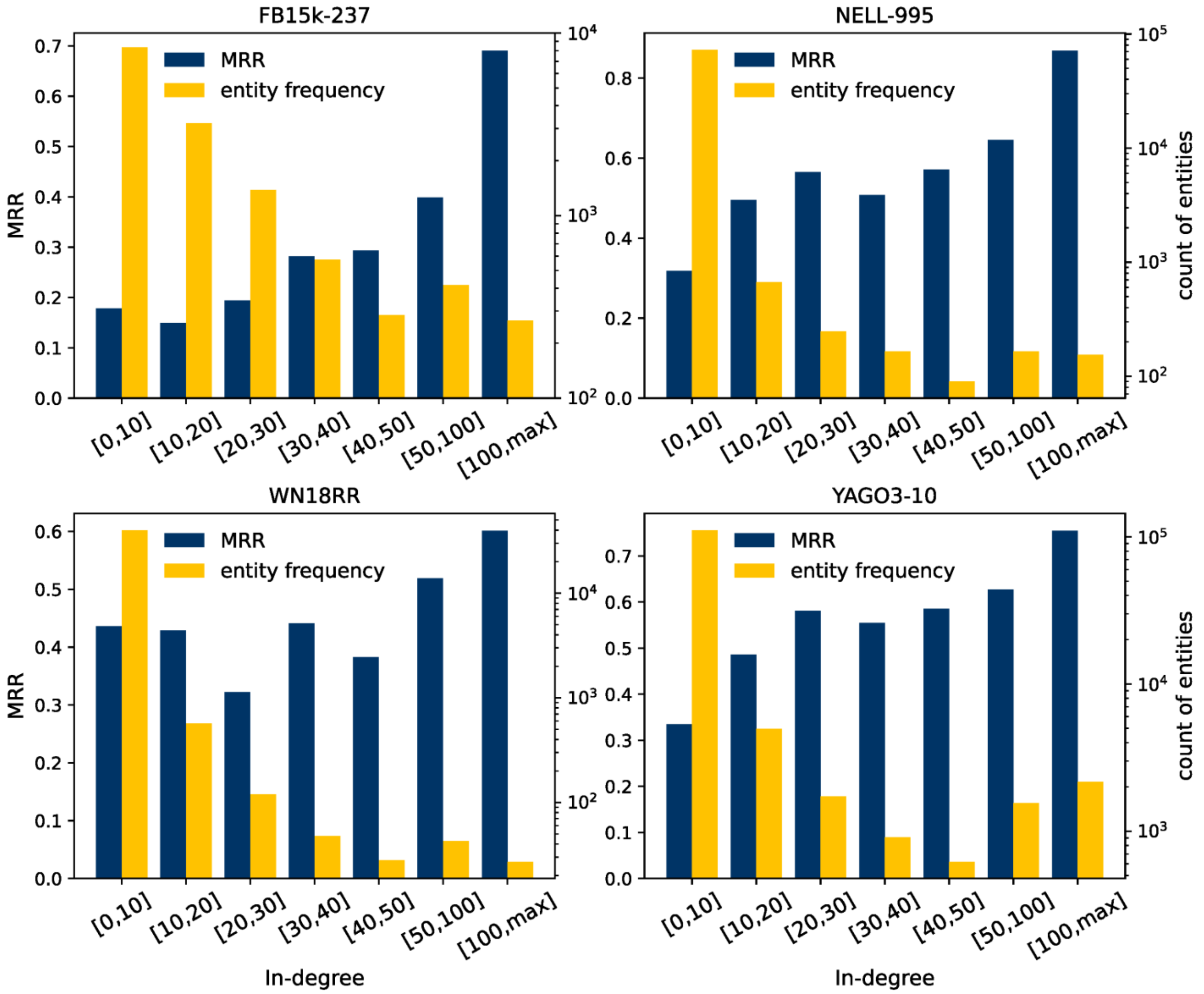
RESCAL

However..

- Predicting Entities rarely appear in KGs remains challenging

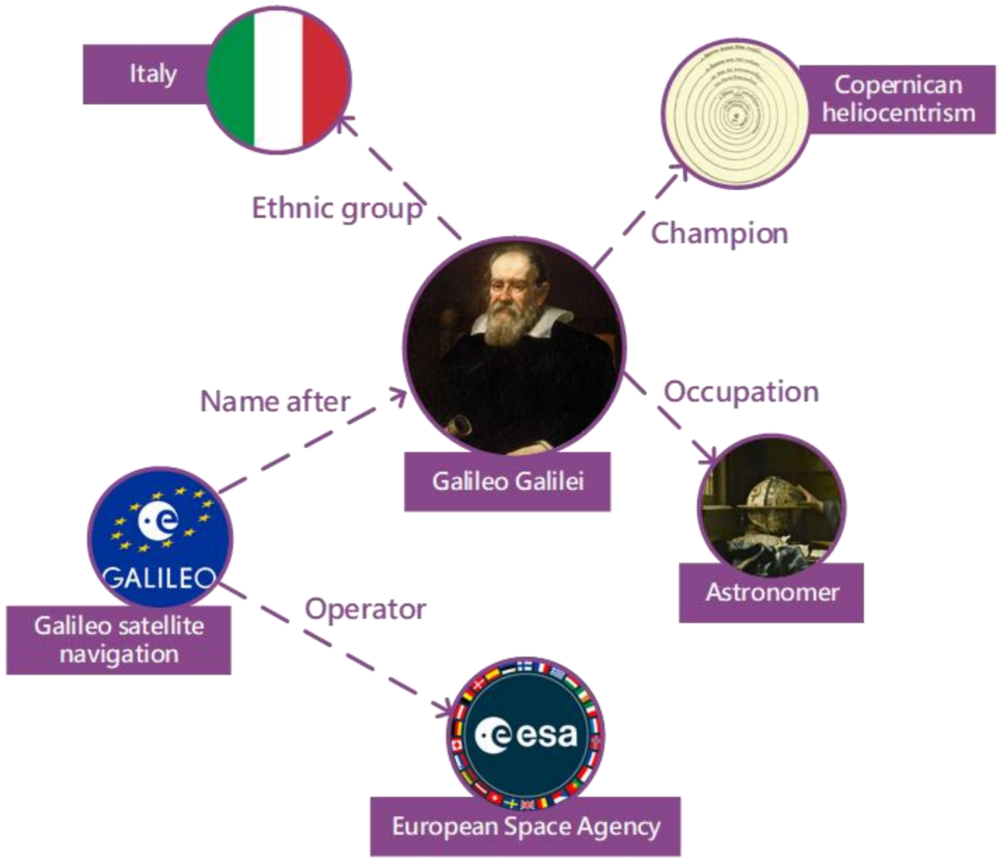
- 1. Common existence of sparse entities

- 2. Performance of sparse entities worse than that of frequent ones



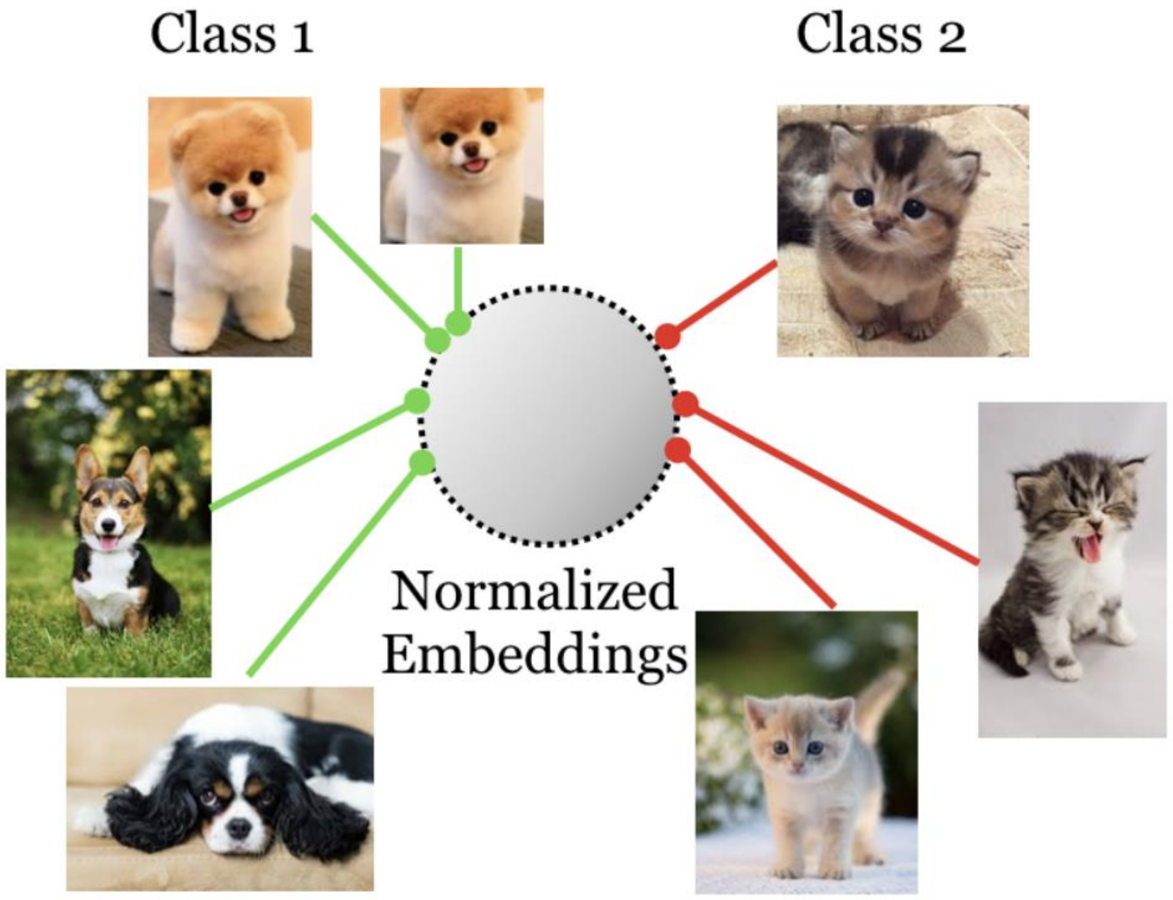
1/2 Intuition: Graph Context Modeling

- Context information in KGs
- Inductive bias in KG operator



2/2 Intuition: Contrastive Learning

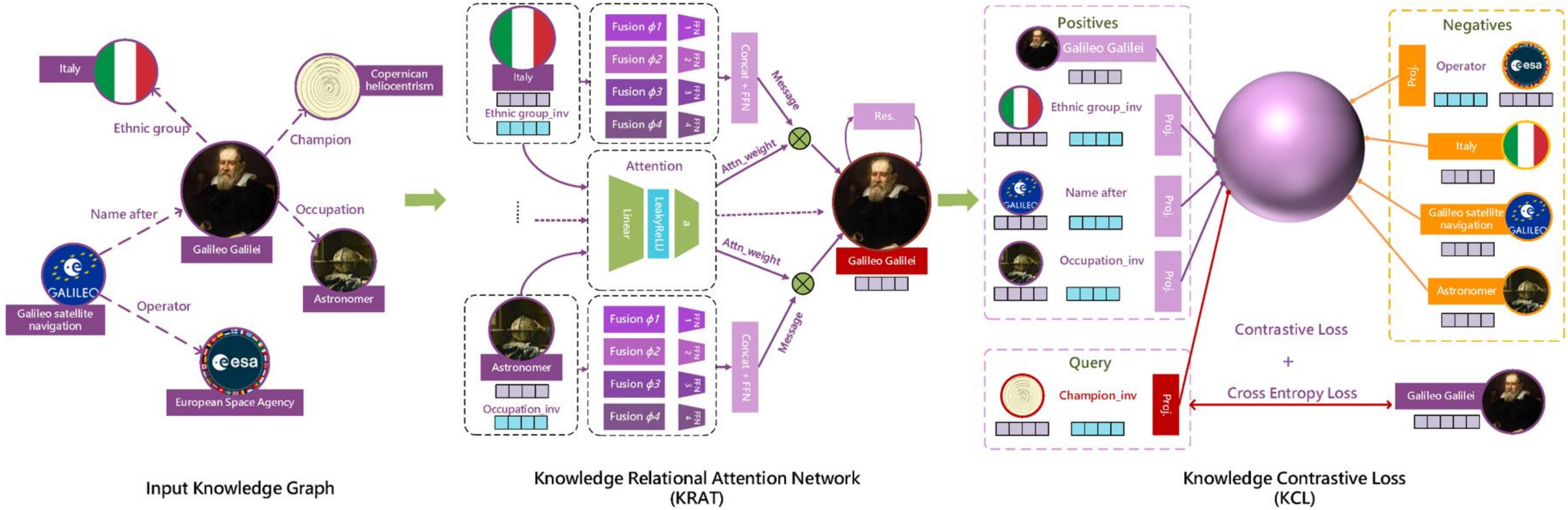
- Label-efficient
- Robust to noisy data



Supervised Contrastive

KRACL

- Contrastive Learning with Graph Context Modeling for Sparse Knowledge Graph Completion



1/2 Knowledge Relational Attention Network (KRAT)

- Inductive bias in message
  - **Subtraction(Sub)**: $\phi(\boldsymbol{h}_s, \boldsymbol{h_r}) = \boldsymbol{h}_s - \boldsymbol{h}_r$
  - **Multiplication(Mult)**: $\phi(\boldsymbol{h}_s, \boldsymbol{h_r}) = \boldsymbol{h}_s \cdot \boldsymbol{h}_r$
  - **Rotation(Rot)**: $\phi(\boldsymbol{h}_s, \boldsymbol{h_r}) = \boldsymbol{h}_s \circ \boldsymbol{h}_r$
  - **Circular-correlation(Corr)**: $\phi(\boldsymbol{h}_s, \boldsymbol{h_r}) = \boldsymbol{h}_s \star \boldsymbol{h}_r$

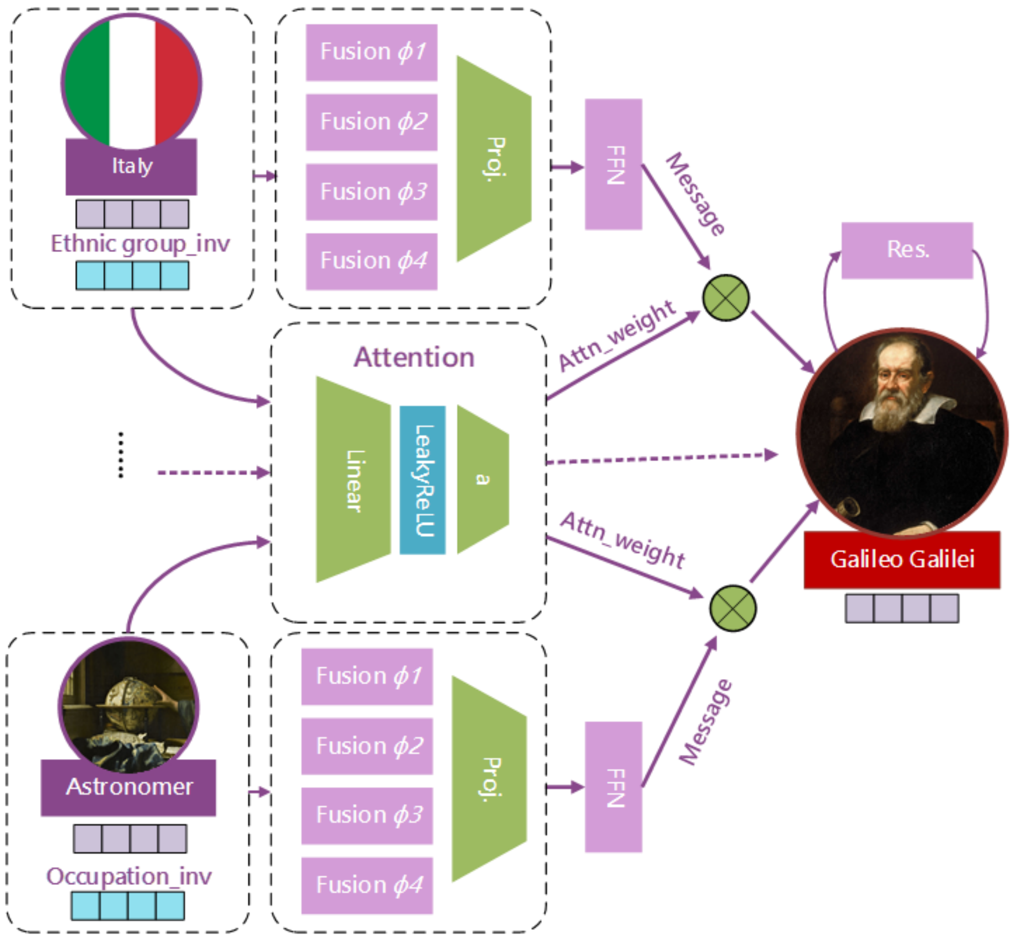
$$f(\boldsymbol{h}_s, \boldsymbol{h}_r) = \sigma\left(\left[\boldsymbol{W}_1^{(l)}\phi_1(\boldsymbol{h}_s, \boldsymbol{h}_r)||...||\boldsymbol{W}_n^{(l)}\phi_n(\boldsymbol{h}_s, \boldsymbol{h}_r)\right]\right)$$

- Attention weight

$$w_{sro} = \boldsymbol{a}^{(l)} LeakyReLU\left(\boldsymbol{W}_{att}^{(l)}\left[\boldsymbol{h}_s^{(l-1)}||\boldsymbol{h}_r^{(l-1)}||\boldsymbol{h}_o^{(l-1)}\right]\right)$$

$$\alpha_{sro} = softmax_{sr}(w_{sro})$$

$$= \frac{exp(w_{sro})}{\sum_{n\in\mathcal{N}_o}\sum_{p\in\mathcal{R}_{no}} exp(w_{npo})},$$



Knowledge Relational Attention Network (KRAT)

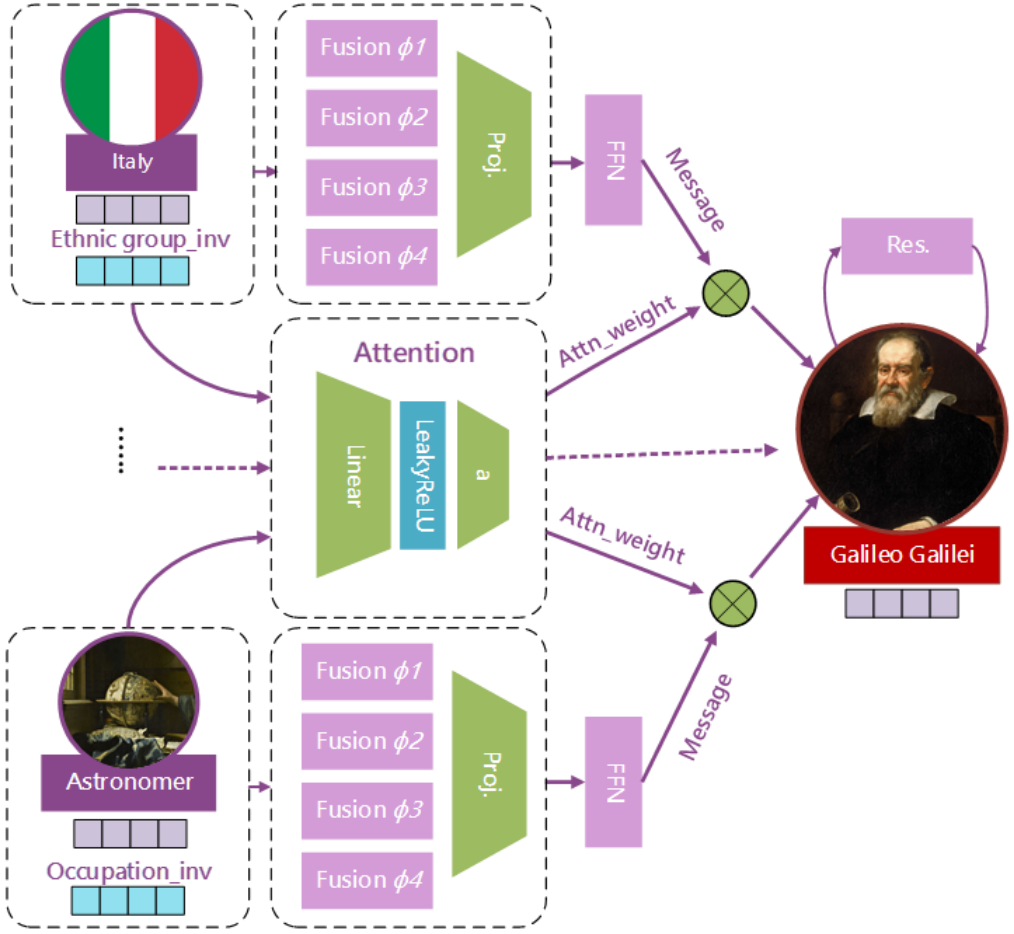
2/2 Knowledge Relational Attention Network (KRAT)

- Aggregation

$$h_o^{(l)} = \sigma\left(\sum_{(s,r)\in\mathcal{N}_o} \alpha_{sro} W_{agg}^{(l)} f(h_s, h_r) + W_{res}^{(l)} h_o^{(l-1)}\right)$$

- Relation update

$$h_r^{(l)} = W_{rel}^{(l)} \cdot h_r^{(l-1)}$$



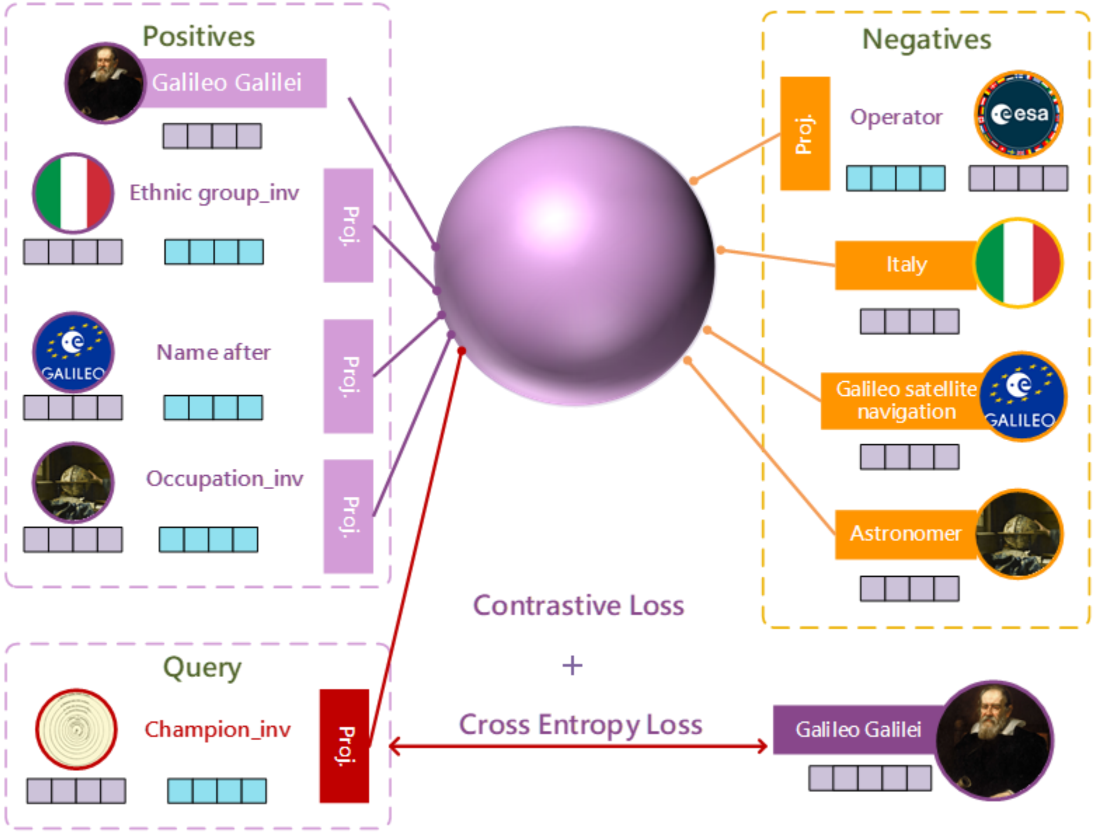
Knowledge Relational Attention Network (KRAT)

1/2 Knowledge Contrastive Loss

- Knowledge projection
  - TransE $z_{(s,r)} = h_s + h_r$
  - DistMult $z_{(s,r)} = h_s * h_r$
  - RotatE $z_{(s,r)} = h_s \circ h_r$
  - ConvE $\boldsymbol{z}_{(s,r)} = \sigma(vec(\sigma([\overline{\boldsymbol{h}_s}||\overline{\boldsymbol{h}_r}] * \omega))\boldsymbol{W}_p)$

- Contrastive loss

$$\mathcal{L}_{CL} = \sum_{o \in \mathcal{T}} \frac{-1}{|\mathcal{T}_o|} \sum_{\boldsymbol{z}_{(s,r)} \in \mathcal{T}_o} log \frac{exp(\boldsymbol{z}_{(s,r)} \cdot \boldsymbol{h}_o / \tau)}{\sum_{k \notin \mathcal{T}_o} exp(\boldsymbol{z}_k \cdot \boldsymbol{h}_o / \tau)}$$



2/2 Knowledge Contrastive Loss

- Scoring candidate entities

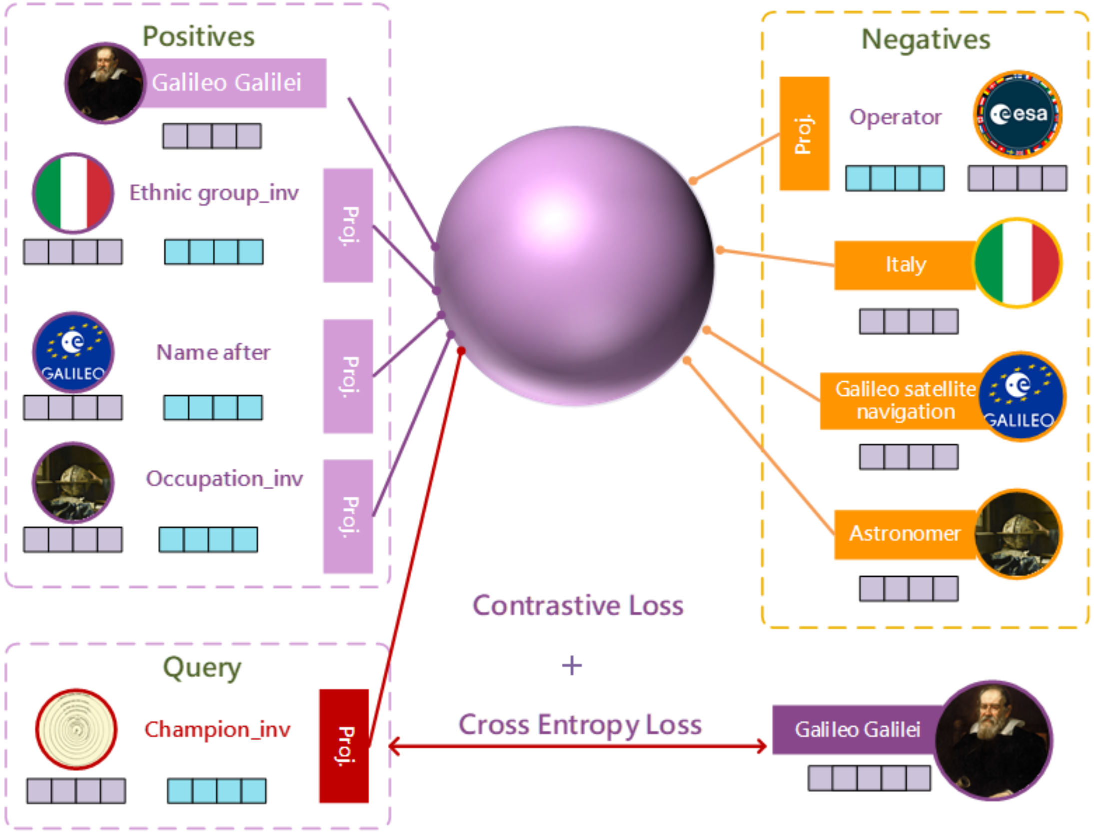
$$\hat{y}^{o}_{(s,r)} = \boldsymbol{z}_{(s,r)} \cdot \boldsymbol{h}_{o}^{T},$$

- Cross Entropy loss

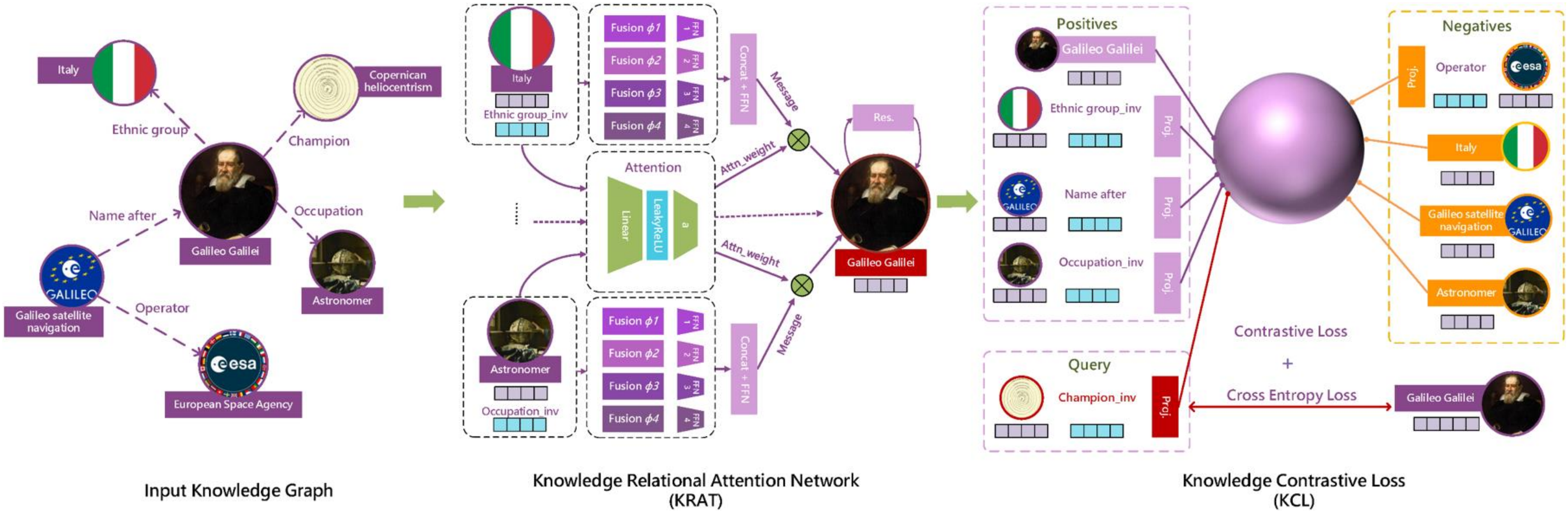
$$\mathcal{L}_{CE} = -\frac{1}{|\mathcal{T}|} \sum_{(s,r)\in\mathcal{T}} \sum_{o\in\mathcal{E}} y^{o}_{(s,r)} \cdot \log \hat{y}^{o}_{(s,r)}$$

- Final objective

$$\mathcal{L} = \mathcal{L}_{CL} + \mathcal{L}_{CE}.$$



Quick recap



Italy
Ethnic group
Copernican heliocentrism
Champion
Galileo Galilei
Name after
Occupation
Astronomer
Galileo satellite navigation
Operator
European Space Agency
Italy
Ethnic group_inv
Fusion $\phi 1$
Fusion $\phi 2$
Fusion $\phi 3$
Fusion $\phi 4$
Concat + FFN
Message
Attention
Linear
LeakyReLU
a
Attn_weight
Res.
Galileo Galilei
Astronomer
Occupation_inv
Positives
Galileo Galilei
Ethnic group_inv
Name after
Occupation_inv
Proj.
Negatives
Operator
Italy
Galileo satellite navigation
Astronomer
Contrastive Loss
+
Cross Entropy Loss
Query
Champion_inv
Galileo Galilei
Input Knowledge Graph
Knowledge Relational Attention Network (KRAT)
Knowledge Contrastive Loss (KCL)

1/4 Experiments

- Sparse KGs

Model	WN18RR					NELL-995				
	MRR	MR	H@10	H@3	H@1	MRR	MR	H@10	H@3	H@1
TransE	.243	2300	.532	.441	.043	.401	2100	.501	.472	.344
DistMult	.444	7000	.504	.47	.412	.485	4213	.61	.524	.401
ComplEx	.449	7882	.53	.469	.409	.482	4600	.606	.528	.399
RotatE	.494	4046	.571	.510	.455	.483	2582	.565	.514	.435
ConvE	.456	4464	.531	.47	.419	.491	3560	.613	.531	.403
HypER	.493	4687	.549	.503	.464	.540	1763	.657	.580	.471
TuckER	.470	-	.526	.482	.443	.520	2330	.624	.561	.455
R-GCN	.123	6700	.207	.137	.08	.12	7600	.188	.126	.082
KBAT	.412	1921	.554	-	-	.319	3683	.474	.370	.233
CompGCN	.481	3113	.548	.492	.448	.534	1246	.644	**.607**	.466
HAKE	.497	-	.582	.516	.452	.508	5836	.613	.557	.442
GC-OTE	.491	-	.583	.511	.442	.538	837	.657	.576	.469
HittER	.503	-	.584	.516	.462	-	-	-	-	-
DisenKGAT	.506	4135	.590	.522	.462	.547	882	.666	.598	.474
GIE	.491	-	.575	.505	.452	.474	2218	.596	.504	.408
CAKE	-	-	-	-	-	.543	**433**	.655	.583	.477
KRACL	**.527**	**1388**	**.613**	**.547**	**.482**	**.563**	716	**.672**	.602	**.495**

- Dense KGs

Model	FB15k-237					Kinship				
	MRR	MR	H@10	H@3	H@1	MRR	MR	H@10	H@3	H@1
TransE	.294	357	.465	-	-	.211	38.9	.470	.252	.093
DistMult	.241	254	.419	.263	.155	.48	7.9	.708	.491	.377
ComplEx	.247	339	.428	.275	.158	.823	2.48	.971	.899	.733
RotatE	.338	177	.533	.375	.241	.738	2.9	.954	.827	.617
ConvE	.325	244	.501	.356	.237	.772	3.0	.950	.858	.665
HypER	.341	250	.520	.376	.252	.868	1.96	.981	.935	.790
TuckER	.355	152	.541	.390	.262	.885	1.67	.986	.948	.816
R-GCN	.248	339	.428	.275	.158	.109	25.9	.239	.088	.03
KBAT	.156	392	.305	.167	.085	.637	3.41	.955	.757	.470
CompGCN	.355	197	.535	.390	.264	.810	2.26	.977	.892	.709
HAKE	.346	-	.542	.381	.250	.802	2.38	.968	.881	.704
GC-OTE	.361	-	.550	.396	.267	.832	2.05	.984	.917	.735
HittER	**.373**	-	**.558**	**.409**	**.279**	-	-	-	-	-
DisenKGAT	.368	179	.553	.407	.275	.832	1.96	.986	.914	.737
GIE	.362	-	.552	.401	.271	.664	3.43	.927	.770	.520
CAKE	.321	170	.515	.355	.226	-	-	-	-	-
KRACL	.360	**150**	.548	.395	.266	**.895**	**1.48**	**.991**	**.970**	**.817**

2/4 Experiments

- Entity indegree analysis

In-degree	RotatE		ConvE		CompGCN		KRACL	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
$[0, 10]$	.178	.309	.186	.338	.198	.348	**.232**	**.394**
$[10, 20]$	.149	.294	.154	.299	.156	.296	**.181**	**.335**
$[20, 30]$	.194	.381	.199	.386	.198	.370	**.218**	**.405**
$[30, 40]$	.282	.497	.287	.485	.280	.476	**.307**	**.501**
$[40, 50]$	.294	.547	.297	.516	.298	.520	**.328**	**.552**
$[50, 100]$	.399	.681	.403	.675	.400	.663	**.434**	**.702**
$[100, max]$	.691	.929	.714	**.936**	.674	.905	**.716**	.932

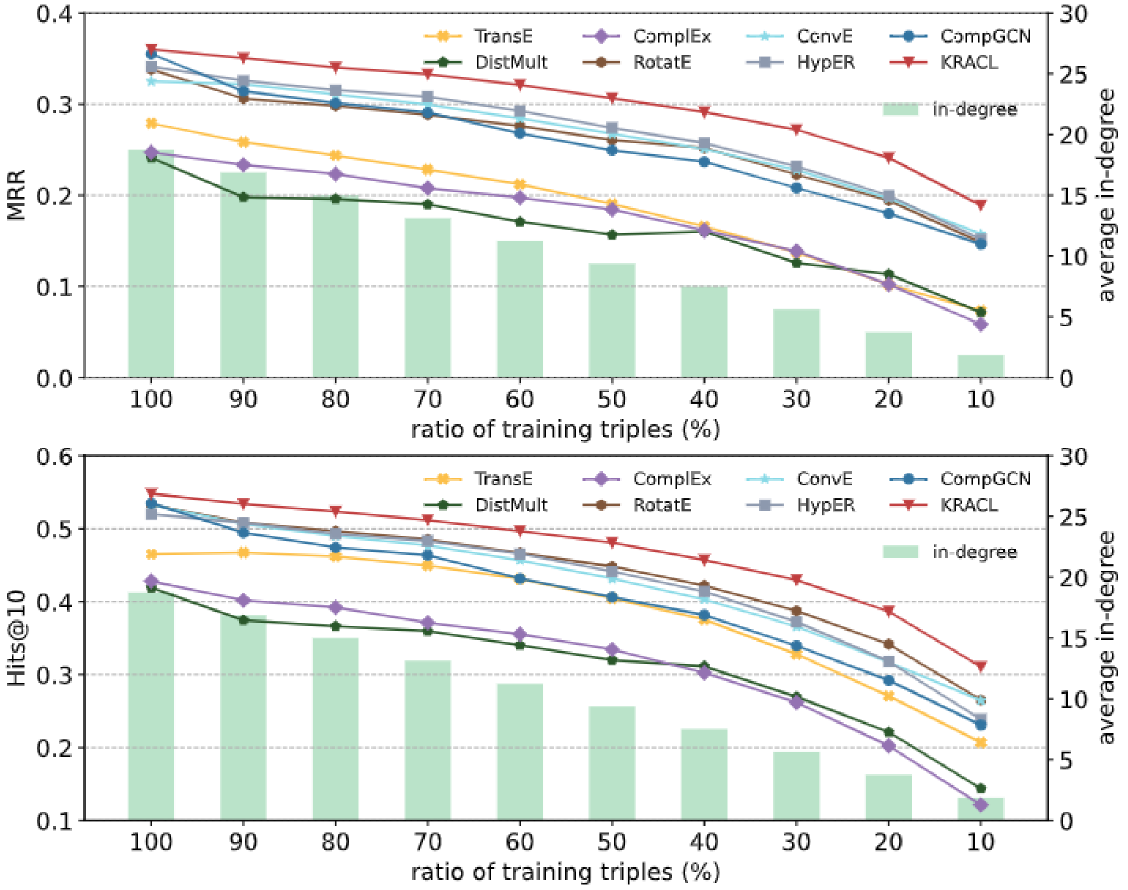
- Ablation study

Model	WN18RR		NELL-995	
	MRR	H@3	MRR	H@3
w/o KRAT	.509	.522	.543	.589
w/o attention	.504	.521	.543	.583
w/o res.	.518	.532	.551	.593
w/o $\mathcal{L}_{CL}$	.502	.514	.496	.541
w/o $\mathcal{L}_{CE}$	.495	.531	.542	.586
$BCELoss$	.469	.478	.507	.547
KRACL	**.527**	**.547**	**.563**	**.602**

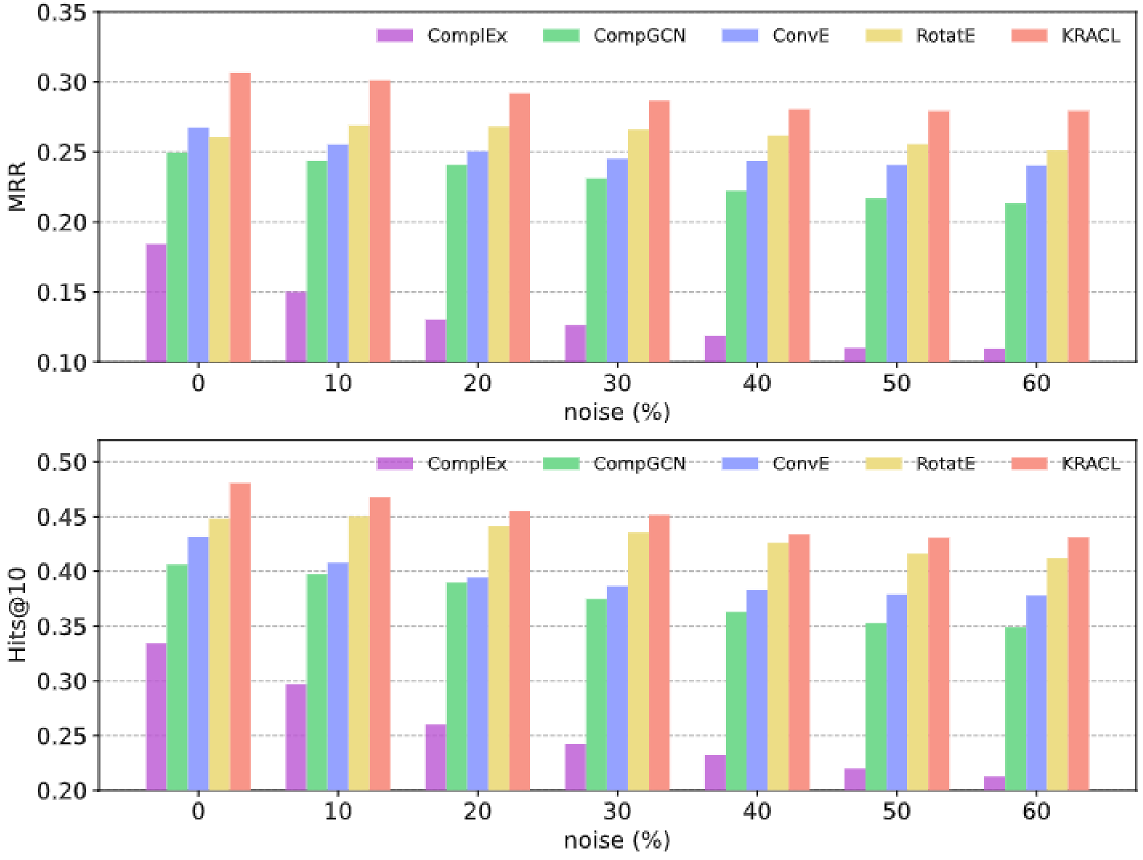
Table 5: Results of ablation study of the proposed KRACL on the WN18RR and NELL-995 dataset. $BCELoss$ denotes replacing the KCL loss with binary cross entropy loss.

3/4 Experiments

- Sparsity study



- Robustness against noise



4/4 Experiments

- A more powerful message function is needed

Dec./Proj. (=X) →	TransE		DistMult		RotatE		ConvE	
Methods ↓	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
X	.279	.441	.241	.419	.338	.533	.325	.501
X+R-GCN	.281	.469	.324	.499	.295	.457	.342	.525
X+W-GCN	.264	.444	.324	.504	.272	.430	.244	.525
X+CompGCN (Sub)	.335	.514	.336	.513	.290	.453	.352	.530
X+CompGCN (Mult)	.337	.515	**.338**	**.518**	.296	.456	.353	.532
X+CompGCN (Rot)	.271	.447	.289	.448	.296	.461	.325	.506
X+CompGCN (Corr)	.336	.518	.335	.514	.294	.459	.355	.535
X+KRAT (Sub)	.334	.519	.333	.512	.332	.512	.355	.541
X+KRAT (Mult)	.332	.513	.331	.510	.334	.511	.356	.540
X+KRAT (Rot)	.332	.512	.331	.508	.334	.513	.351	.538
X+KRAT (Corr)	.333	.518	.334	.512	.332	.509	.353	.538
X+KRAT (All operators)	**.340**	**.524**	**.338**	.517	**.339**	**.522**	**.360**	**.548**

Resources

- Paper: https://arxiv.org/abs/2208.07622
- Code: https://github.com/TamSiuhin/KRACL

Thank you!

Zhaoxuan Tan

tanzhaoxuan@stu.xjtu.edu.cn

zhaoxuan.info