Question: Is all the information necessary for reasoning on knowledge graphs? 🤔

ICLR

Less is More: One-shot Subgraph Reasoning on Large-scale Knowledge Graphs

Zhanke Zhou, Yongqi Zhang, Jiangchao Yao, Quanming Yao, Bo Han

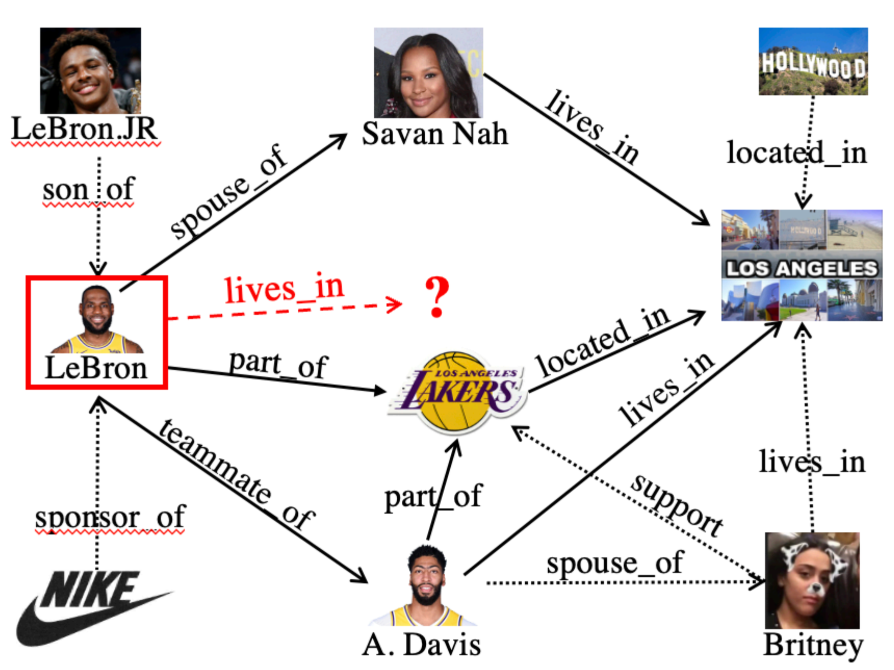
TMLR Trustworthy Machine Learning and Reasoning · 香港浸會大學 Hong Kong Baptist University · 清華大學 Tsinghua University · 上海交通大學 Shanghai Jiao Tong University · The Hong Kong University of Science and Technology (Guangzhou)

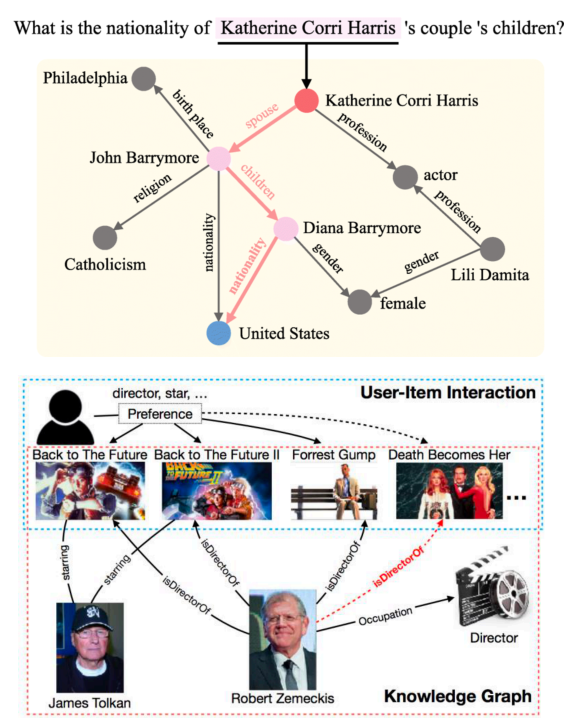
Paper Code Slides

Problem: Link Prediction on KG

KG reasoning with query: (LeBron, lives_in, ?)

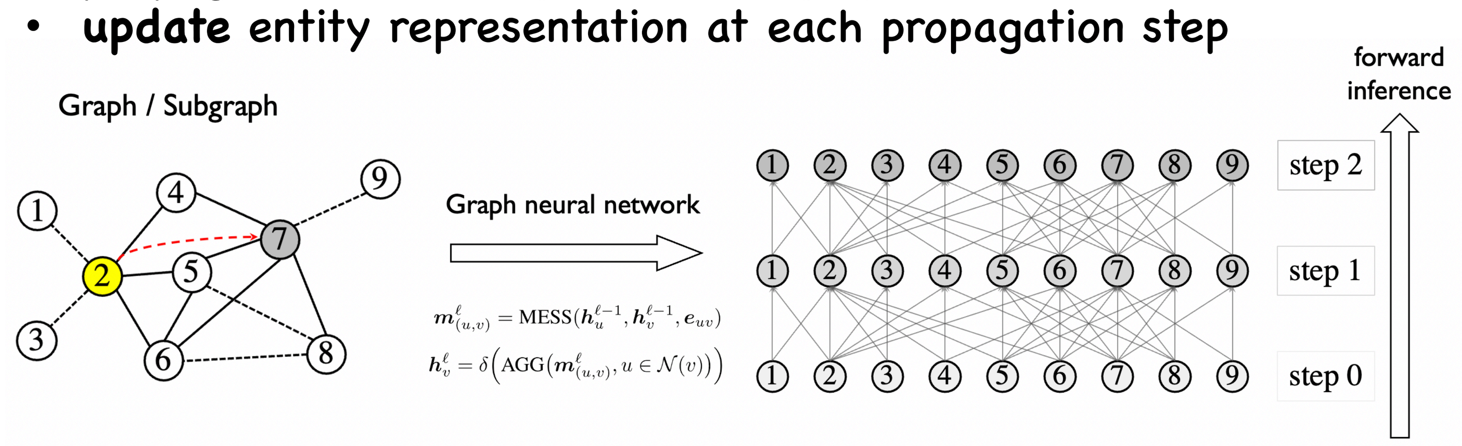
Applications: QA / RecSys



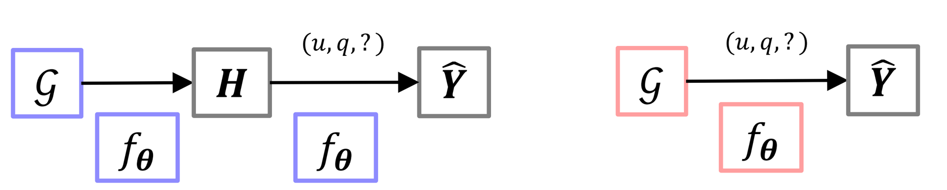


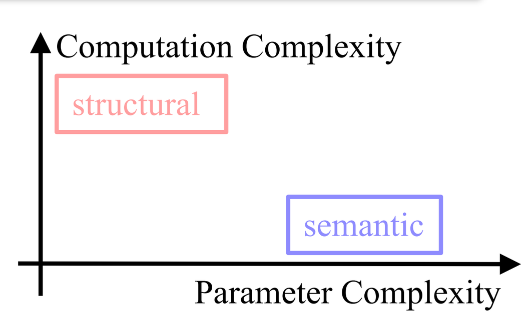
Structural models (e.g., GNNs) for KG reasoning

- **propagate** the message with the graph structure
- **update** entity representation at each propagation step



The Scalability Issue





semantic models structural models complexity comparison

semantic models (computation-efficient but parameter-expensive)

- $p(u,q,v)$ is measured by a scoring function with representations h_u, h_q, h_v

structural models (parameter-efficient but computation-expensive)

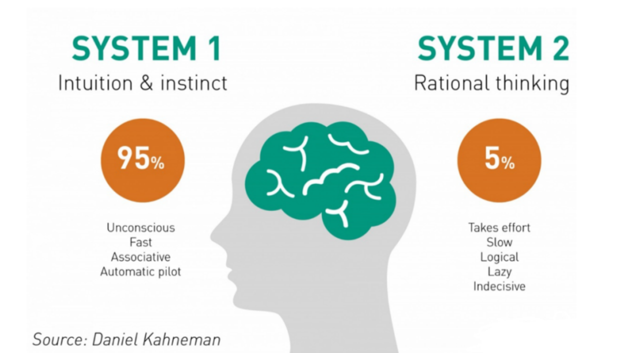
- learn the structures by leveraging the relational paths between u and v
- or use the graph structure for reasoning, capturing more complex semantics

➔ f_θ acts on $\mathcal{G}$ to obtain $\hat{Y}$ of all entities

➔ The whole graph ($\mathcal{G}$), model (f_θ), and prediction ($\hat{Y}$) are coupled

How to efficiently and effectively conduct subgraph reasoning on KG? 🤔

Method: one-shot-subgraph reasoning



Design principle

- first to efficiently identify a subgraph (system1) ➔ sampler
- then effectively reason on the subgraph (system2) ➔ predictor

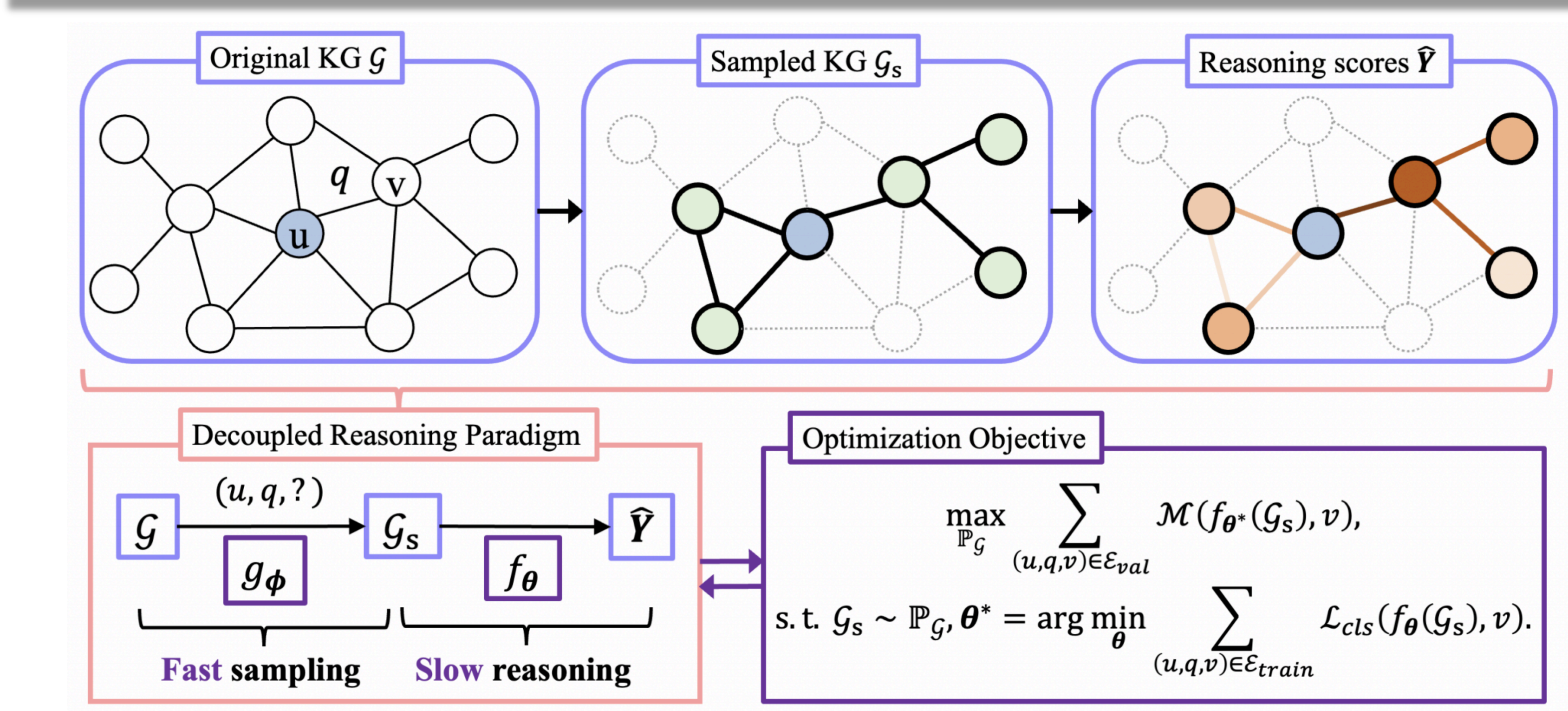
- Only partial knowledge stored in human brain is relevant to a question
- Generating candidates and then ranking the promising ones are common

Definition 1 (One-shot-subgraph Link Prediction on Knowledge Graphs). *Instead of directly predicting on the original graph $\mathcal{G}$, the prediction procedure is decoupled to two-fold: (1) one-shot sampling of a query-dependent subgraph and (2) prediction on this subgraph. The prediction pipeline becomes*

$$\mathcal{G} \xrightarrow{g_\phi,(u,q)} \mathcal{G}_s \xrightarrow{f_\theta} \hat{\boldsymbol{Y}}, \quad (1)$$

where the sampler g_ϕ generates only one subgraph $\mathcal{G}_s$ (satisfies $|\mathcal{V}_s| \ll |\mathcal{V}|, |\mathcal{E}_s| \ll |\mathcal{E}|$) conditioned on the given query $(u,q,?)$. Based on subgraph $\mathcal{G}_s$, the predictor f_θ outputs the final predictions $\hat{\boldsymbol{Y}}$.

Implementation



The three key steps of one-shot-subgraph reasoning are

1. generate the sampling distribution

   Non-parametric indicator: $\boldsymbol{p}^{(k+1)} \leftarrow \alpha\cdot\boldsymbol{s} + (1-\alpha)\cdot\boldsymbol{D}^{-1}\boldsymbol{A}\cdot\boldsymbol{p}^{(k)}$,

2. extract a subgraph with top entities and edges

   Entity Sampling: $\mathcal{V}_s \leftarrow \text{TopK}\big(\mathcal{V},\ \boldsymbol{p},\ K{=}r^q_{\mathcal{V}}\times|\mathcal{V}|\big)$,

   Edge Sampling: $\mathcal{E}_s \leftarrow \text{TopK}\big(\mathcal{E},\ \{\boldsymbol{p}_x\cdot\boldsymbol{p}_o : x,o\in\mathcal{V}_s, (x,r,o)\in\mathcal{E}\},\ K{=}r^q_{\mathcal{E}}\times|\mathcal{E}|\big)$.

3. inference on the subgraph and get the final prediction

   Indicating: $\boldsymbol{h}^0_o \leftarrow \mathbb{1}(o=u)$,

   Propagation: $\boldsymbol{h}^{l+1}_o \leftarrow \text{DROPOUT}\Big(\text{ACT}\Big(\text{AGG}\big\{\text{MESS}(\boldsymbol{h}^l_x, \boldsymbol{h}^l_r, \boldsymbol{h}^l_o) : (x,r,o)\in\mathcal{E}_s\big\}\Big)\Big)$

Experiments

Table 1: Empirical results of WN18RR, NELL-995, YAGO3-10 datasets. Best performance is indicated by the **bold face** numbers, and the underline means the second best. "–" means unavailable results. "H@1" and "H@10" are short for Hit@1 and Hit@10 (in percentage), respectively.

type	models	WN18RR MRR↑	WN18RR H@1↑	WN18RR H@10↑	NELL-995 MRR↑	NELL-995 H@1↑	NELL-995 H@10↑	YAGO3-10 MRR↑	YAGO3-10 H@1↑	YAGO3-10 H@10↑
Semantic Models	ConvE	0.427	39.2	49.8	0.511	44.6	61.9	0.520	45.0	66.0
	QuatE	0.480	44.0	55.1	0.533	46.6	64.3	0.379	30.1	53.4
	RotatE	0.477	42.8	57.1	0.508	44.8	60.8	0.495	40.2	67.0
Structural Models	MINERVA	0.448	41.3	51.3	0.513	41.3	63.7	–	–	–
	DRUM	0.486	42.5	58.6	0.532	46.0	**66.2**	0.531	45.3	67.6
	RNNLogic	0.483	44.6	55.8	0.416	36.3	47.8	0.554	50.9	62.2
	CompGCN	0.479	44.3	54.6	0.463	38.3	59.6	0.489	39.5	58.2
	DPMPN	0.482	44.4	55.8	0.513	45.2	61.5	0.553	48.4	67.9
	NBFNet	0.551	49.7	**66.6**	0.525	45.1	63.9	0.550	47.9	68.3
	RED-GNN	0.533	48.5	62.4	0.543	47.6	65.1	0.559	48.3	68.9
	one-shot-subgraph	**0.567**	**51.4**	**66.6**	**0.547**	**48.5**	65.1	**0.606**	**54.0**	**72.1**

Table 2: Empirical results of two OGB datasets (Hu et al., 2020) with regard to official leaderboards.

type	models	OGBL-BIOKG Test MRR↑	OGBL-BIOKG Valid MRR↑	OGBL-BIOKG #Params↓	OGBL-WIKIKG2 Test MRR↑	OGBL-WIKIKG2 Valid MRR↑	OGBL-WIKIKG2 #Params↓
Semantic Models	TripleRE	0.8348	0.8360	469,630,002	0.5794	0.6045	500,763,337
	AutoSF	0.8309	0.8317	93,824,000	0.5458	0.5510	500,227,800
	PairRE	0.8164	0.8172	187,750,000	0.5208	0.5423	500,334,800
	ComplEx	0.8095	0.8105	187,648,000	0.4027	0.3759	1,250,569,500
	DistMult	0.8043	0.8055	187,648,000	0.3729	0.3506	1,250,569,500
	RotatE	0.7989	0.7997	187,597,000	0.4332	0.4353	1,250,435,750
	TransE	0.7452	0.7456	187,648,000	0.4256	0.4272	1,250,569,500
Structural Models	**one-shot-subgraph**	**0.8430**	**0.8435**	**976,801**	**0.6755**	**0.7080**	**6,831,201**

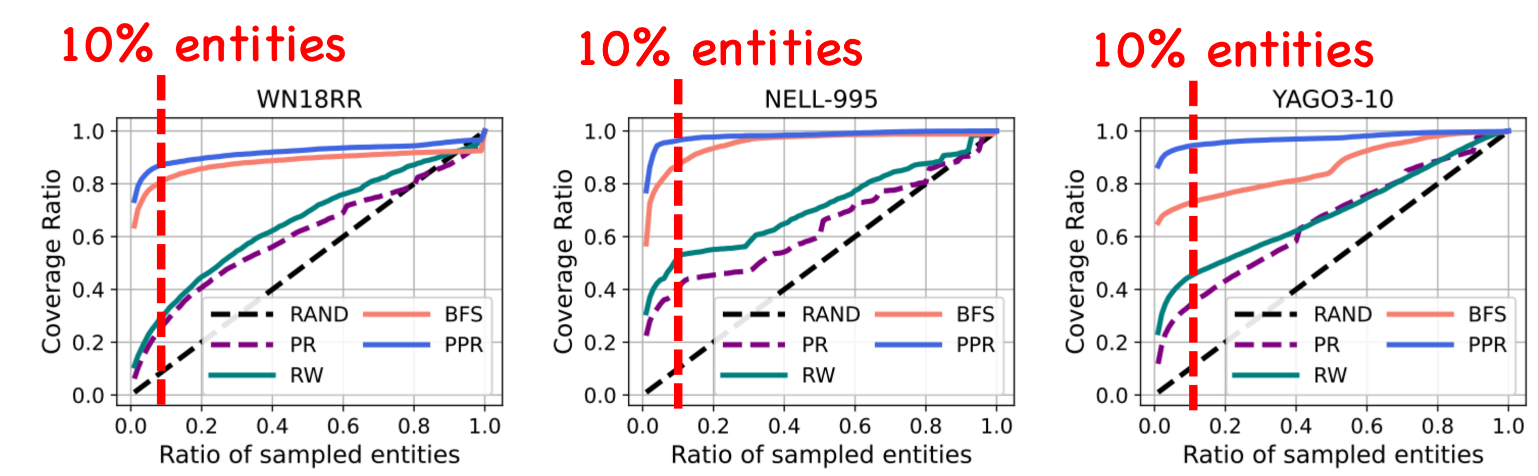


Table 3: Coverage Ratio of different heuristics. **Bold face** numbers indicate the best results in column.

heuristics	WN18RR $r^q_{\mathcal{V}}=0.1$	WN18RR $r^q_{\mathcal{V}}=0.2$	WN18RR $r^q_{\mathcal{V}}=0.5$	NELL-995 $r^q_{\mathcal{V}}=0.1$	NELL-995 $r^q_{\mathcal{V}}=0.2$	NELL-995 $r^q_{\mathcal{V}}=0.5$	YAGO3-10 $r^q_{\mathcal{V}}=0.1$	YAGO3-10 $r^q_{\mathcal{V}}=0.2$	YAGO3-10 $r^q_{\mathcal{V}}=0.5$
Random Sampling (RAND)	0.100	0.200	0.500	0.100	0.200	0.500	0.100	0.200	0.500
PageRank (PR)	0.278	0.407	0.633	0.405	0.454	0.603	0.340	0.432	0.694
Random Walk (RW)	0.315	0.447	0.694	0.522	0.552	0.710	0.449	0.510	0.681
Breadth-first-searching (BFS)	0.818	0.858	0.898	0.872	0.935	0.982	0.728	0.760	0.848
Personalized PageRank (PPR)	**0.876**	**0.896**	**0.929**	**0.965**	**0.977**	**0.987**	**0.943**	**0.957**	**0.973**

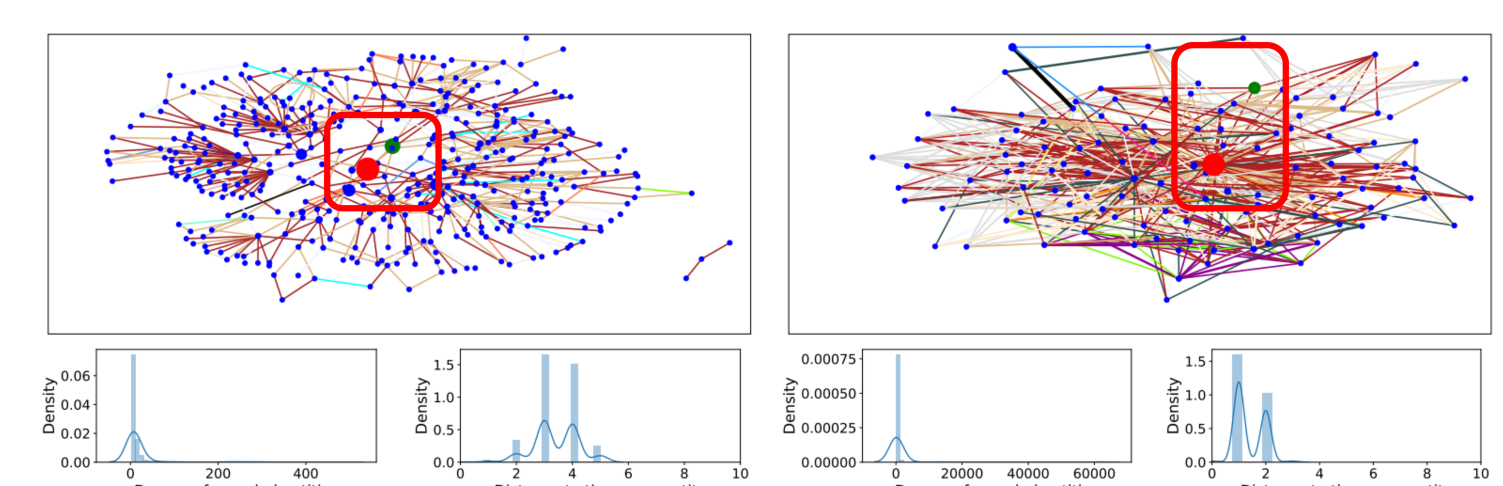


Figure 5: Exemplar subgraphs sampled from WN18RR (left) and YAGO3-10 (right). The red and green nodes indicate the query entity and answer entity. The colors of the edges indicate relation types. The bottom distributions of degree and distance show the statistical properties of each subgraph.