Inductive Relation Prediction from Relational Paths and Context with Hierarchical Transformers

Jiaang Li[1], Quan Wang[2*], Zhendong Mao[1]

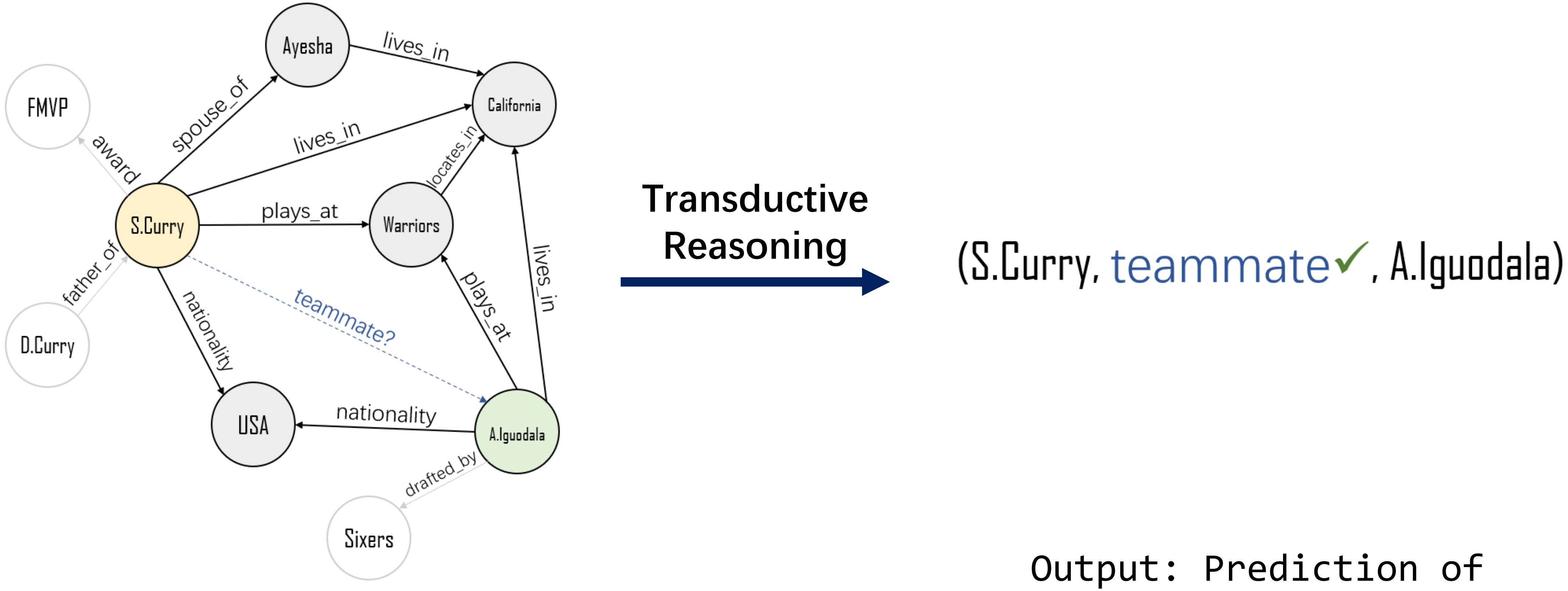
1 : University of Science and Technology of China, Hefei, Anhui, China;
2 : Beijing University of Posts and Telecommunications, Beijing, China

jali@mail.ustc.edu.cn

Background

Transductive Relation Prediction

- **Definition: Predict missing relations from an incomplete KG**



Input: Incomplete KG

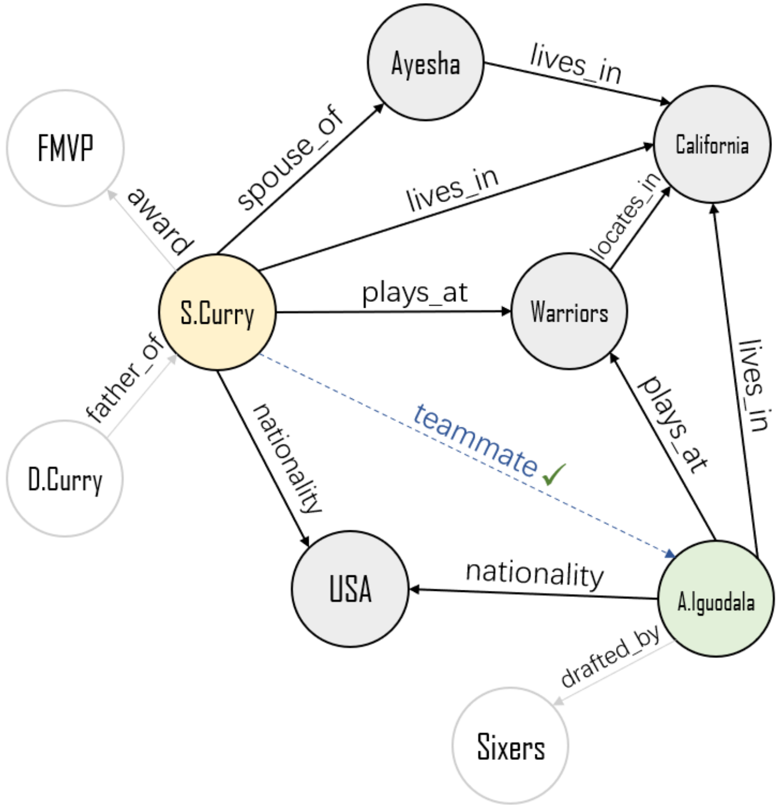
Output: Prediction of missing relations

Cannot generalize to new entities

Inductive Relation Prediction

- **Definition: Learn from one KG, and predict relations on a new one**

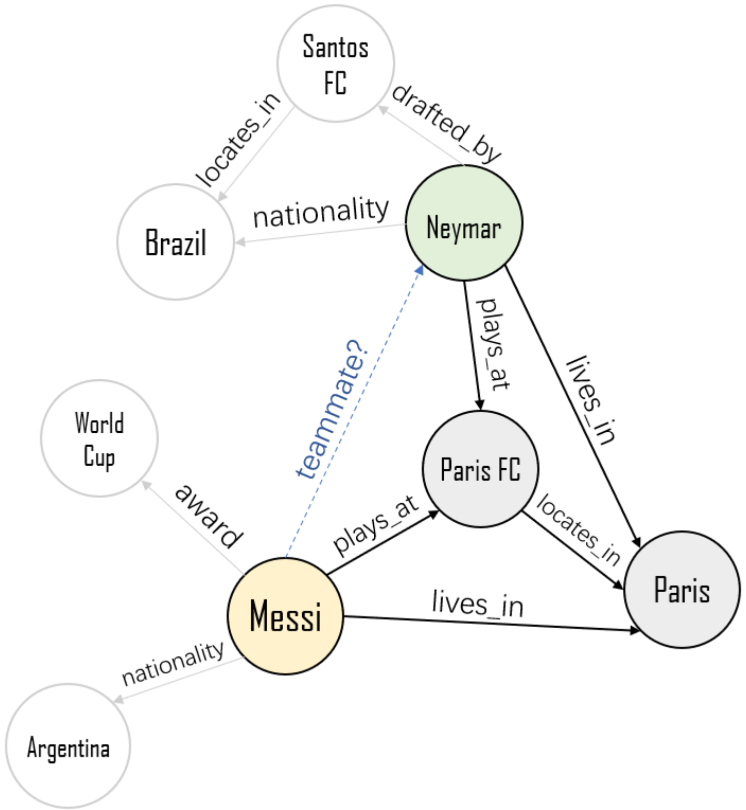
Training KG



Inductive Reasoning

```
Unseen Entities
& Seen Relations
```

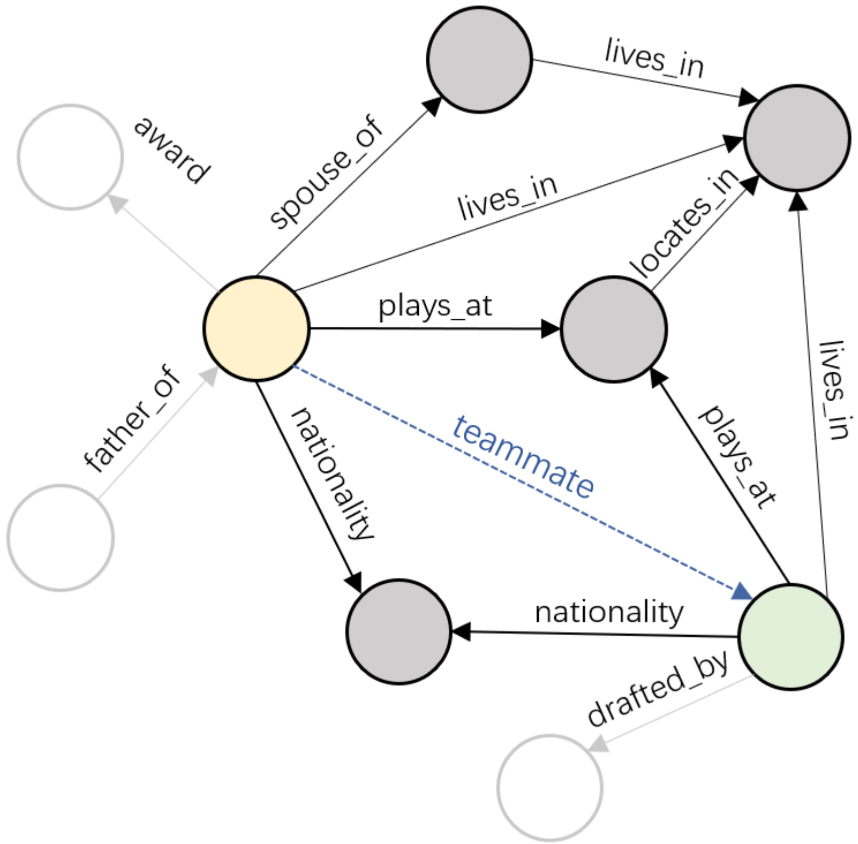
Inference KG



Generalize to KGs with entirely new entities

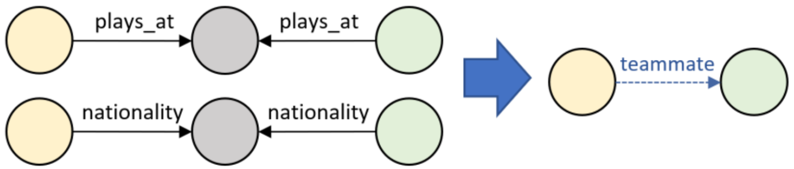
Existing Methods: Overview

- **Methods with *inductive* ability can be divided into two categories**



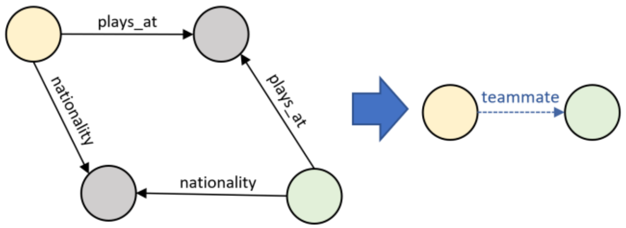
Rule-based/Path-based methods

- Use first-order logic rules or relational paths, which are related to each other



Subgraph-based methods

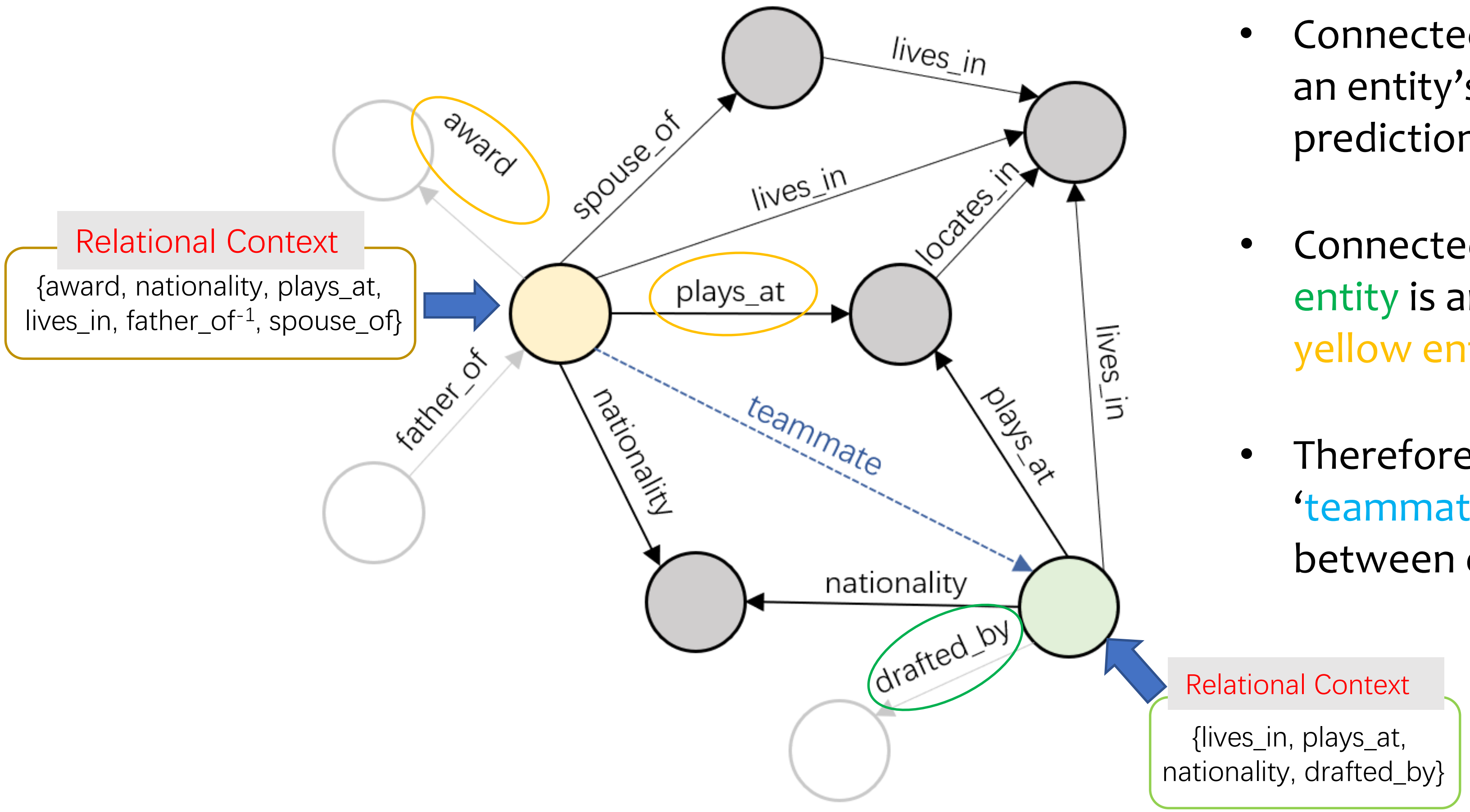
- Compose paths into a subgraph and mine more complex semantics from it



Use only the connections between entities

Existing Methods: Issue

- **Connections between entities may ignore valuable information**



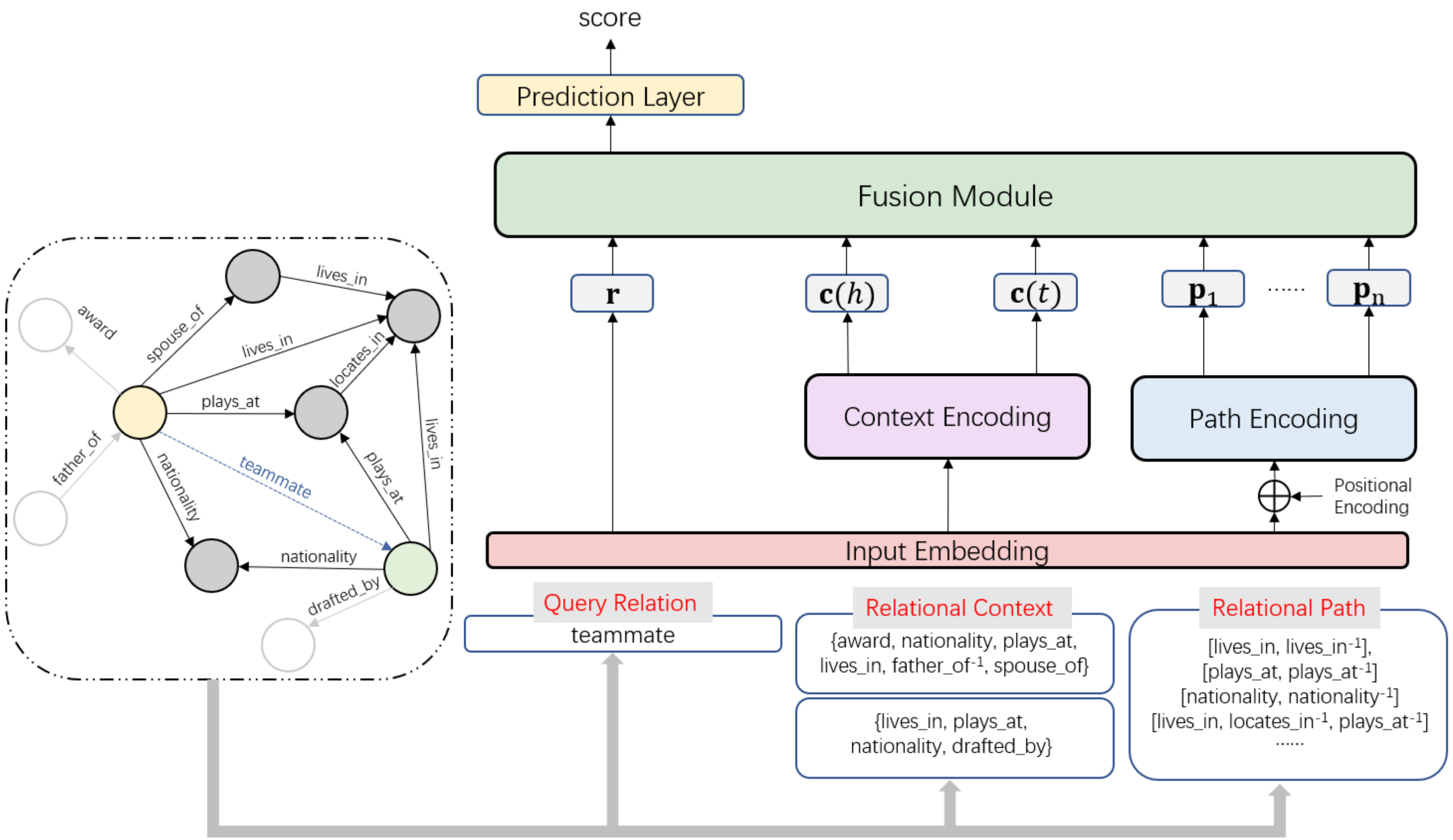
- Connected relations (**Relational Context**) indicate an entity's attribute, valuable for relation prediction but missed in connections
- Connected relation 'drafted_by' suggests green entity is an athlete. 'award' & 'plays_at' suggests yellow entity is highly likely to be an athlete, too.
- Therefore, it's more likely to be a missing relation 'teammate' between these two entities, than between others.

ISSUE: Ignore each entity's own attributes

Method

Method Overview

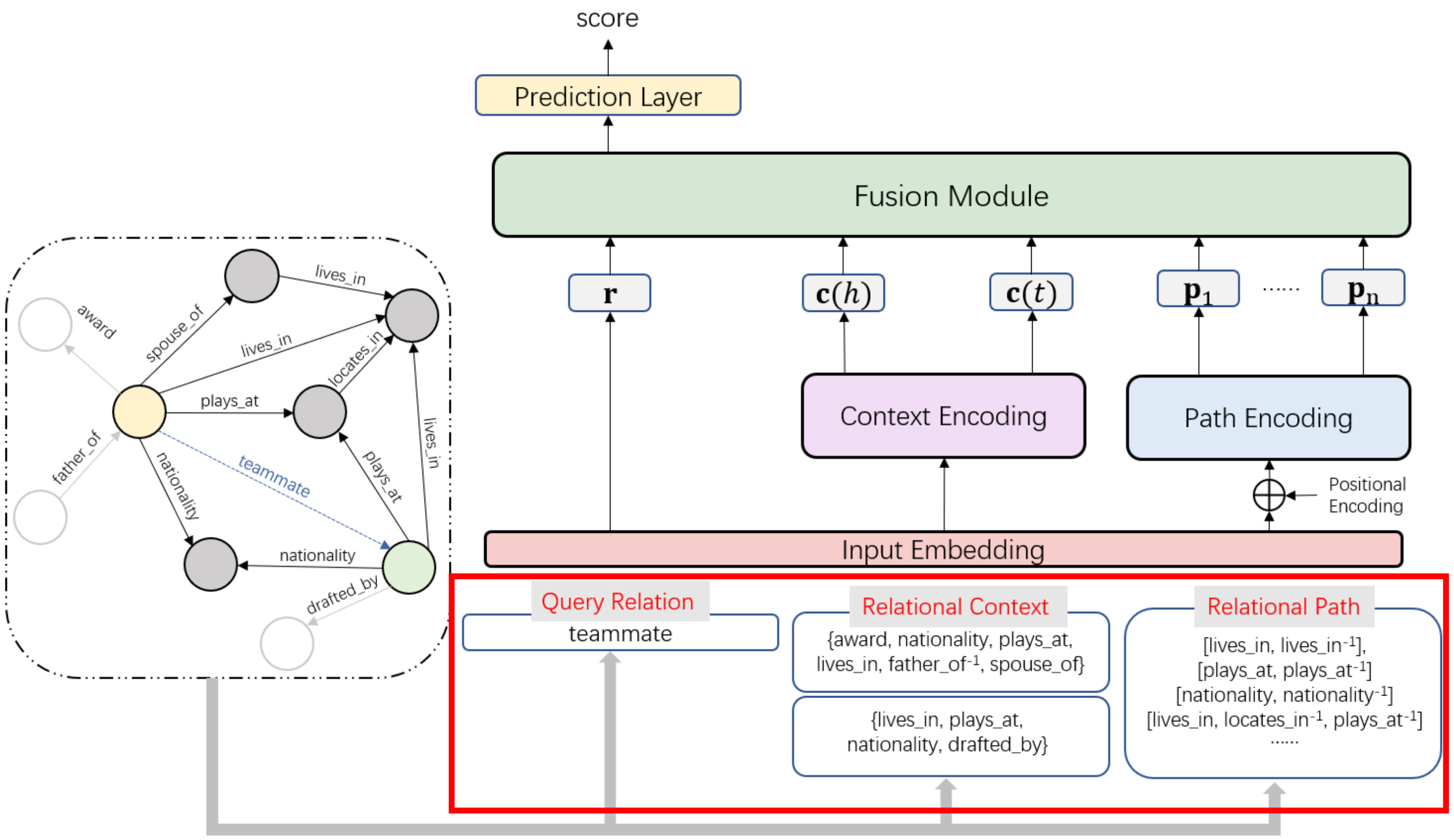
- We propose **REPORT: RE**lational **P**aths and c**O**ntext with hie**R**archical **T**ransformers



- Use **relational paths** to capture connections between entities
- Use **relational context** to reflect entity attributes
- Aggregate different types of input by **hierarchical Transformers** Hierarchically composing different information sources is a divide-and-conquer strategy, thus is efficient.

REPORT: RElational Paths and cOntext with hieRarchical Transformers

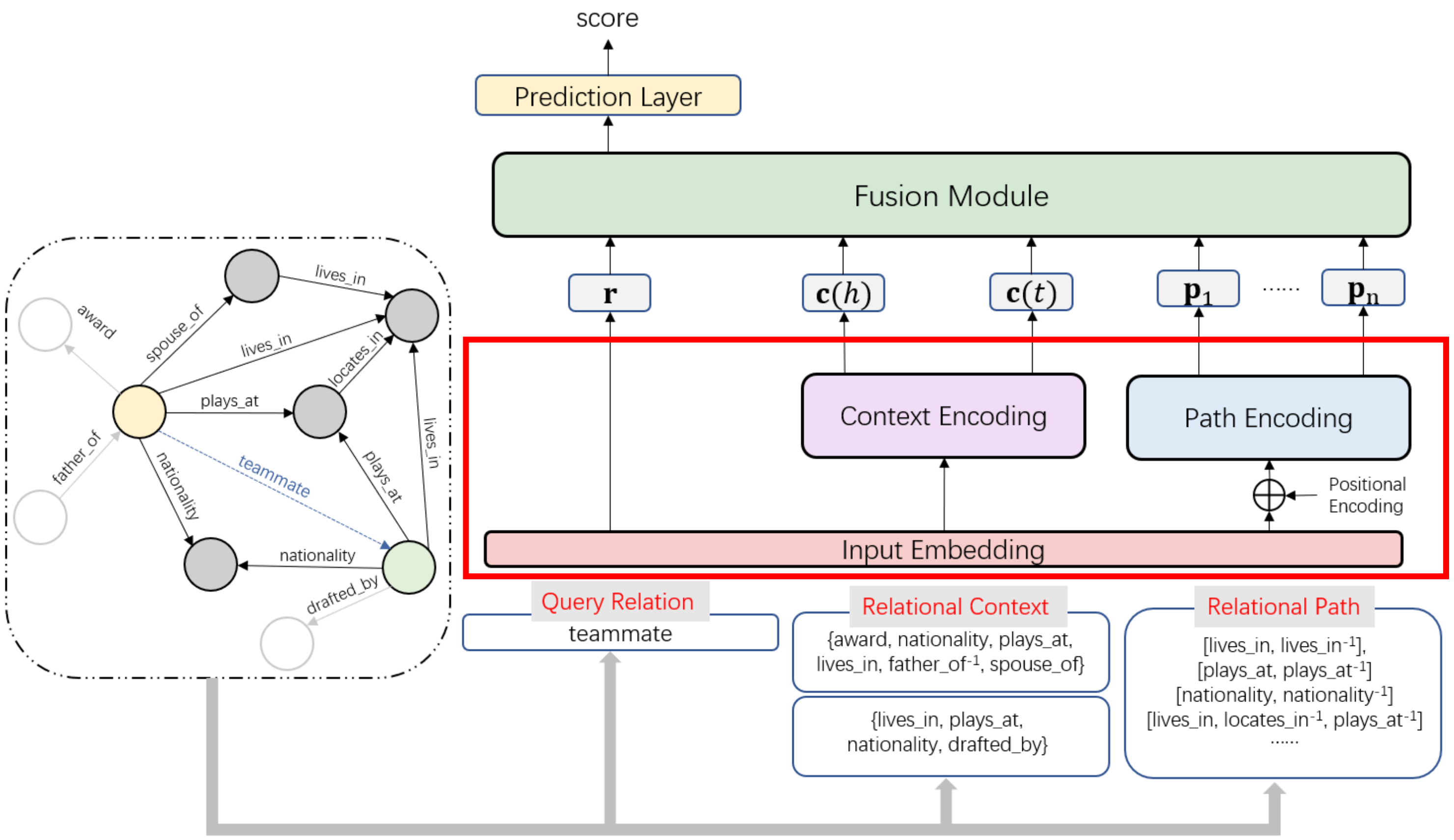
- Three types of input



- **Relational path:** A sequence of relations taken from a path linking two entities.
- **Relational context:** A set of an entity's connected relations.
- **Query relation:** indicate the missing link to be predicted.

REPORT: REIational Paths and cOntext with hieRarchical Transformers

- Bottom layers of hierarchical Transformers



- Each path and context is transformed and encoded by encoding modules

Encode Path

Input: [PCLS] +relational path sequence

$$\mathbf{x}_j^0 = \mathbf{ele}_j^{p_i} + \mathbf{pos}_j$$

$$\mathbf{x}_j^l = \mathrm{Transformer}(\mathbf{x}_j^{l-1}), \quad l = 1, \cdots, L_p$$

Output: hidden state of [PCLS]

Encode Context

Input: [HCLS]\[TCLS]+relational context set

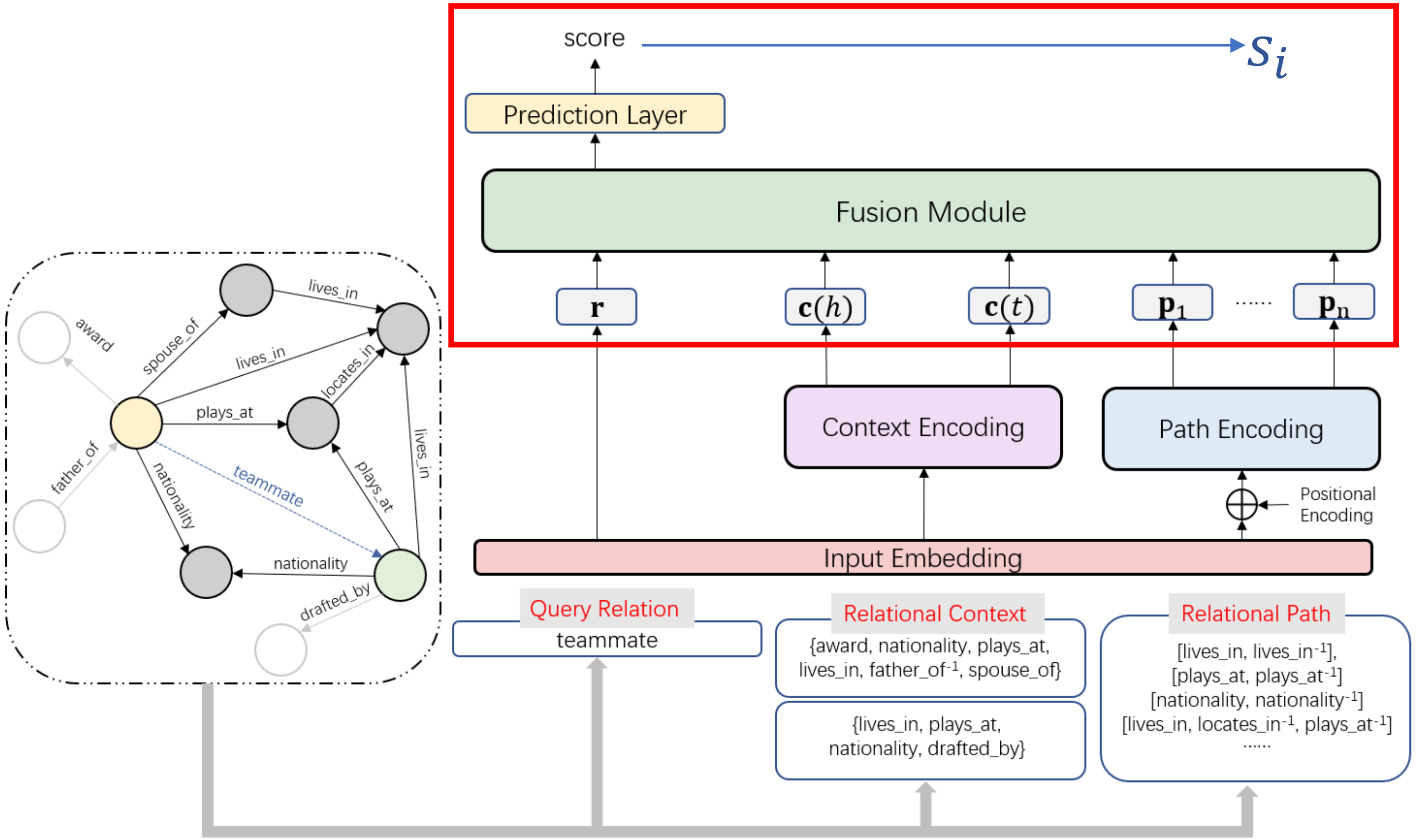
$$\mathbf{y}_j^0 = \mathbf{ele}_j^h$$

$$\mathbf{y}_j^l = \mathrm{Transformer}(\mathbf{y}_j^{l-1}), \quad l = 1, \cdots, L_c$$

Output: hidden state of [HCLS]\[TCLS]

REPORT: RElational Paths and cOntext with hieRarchical Transformers

- Top layers of hierarchical Transformers



- Fusion module aggregates information from bottom layers for prediction

Input: representations of (1) query relation (2) relational context of head and tail (3) relational paths

Fuse Information & Make Predictions

$$\mathbf{h}_j^l = \mathrm{Transformer}(\mathbf{h}_j^{l-1}), \quad l = 1, \cdots, L_f$$

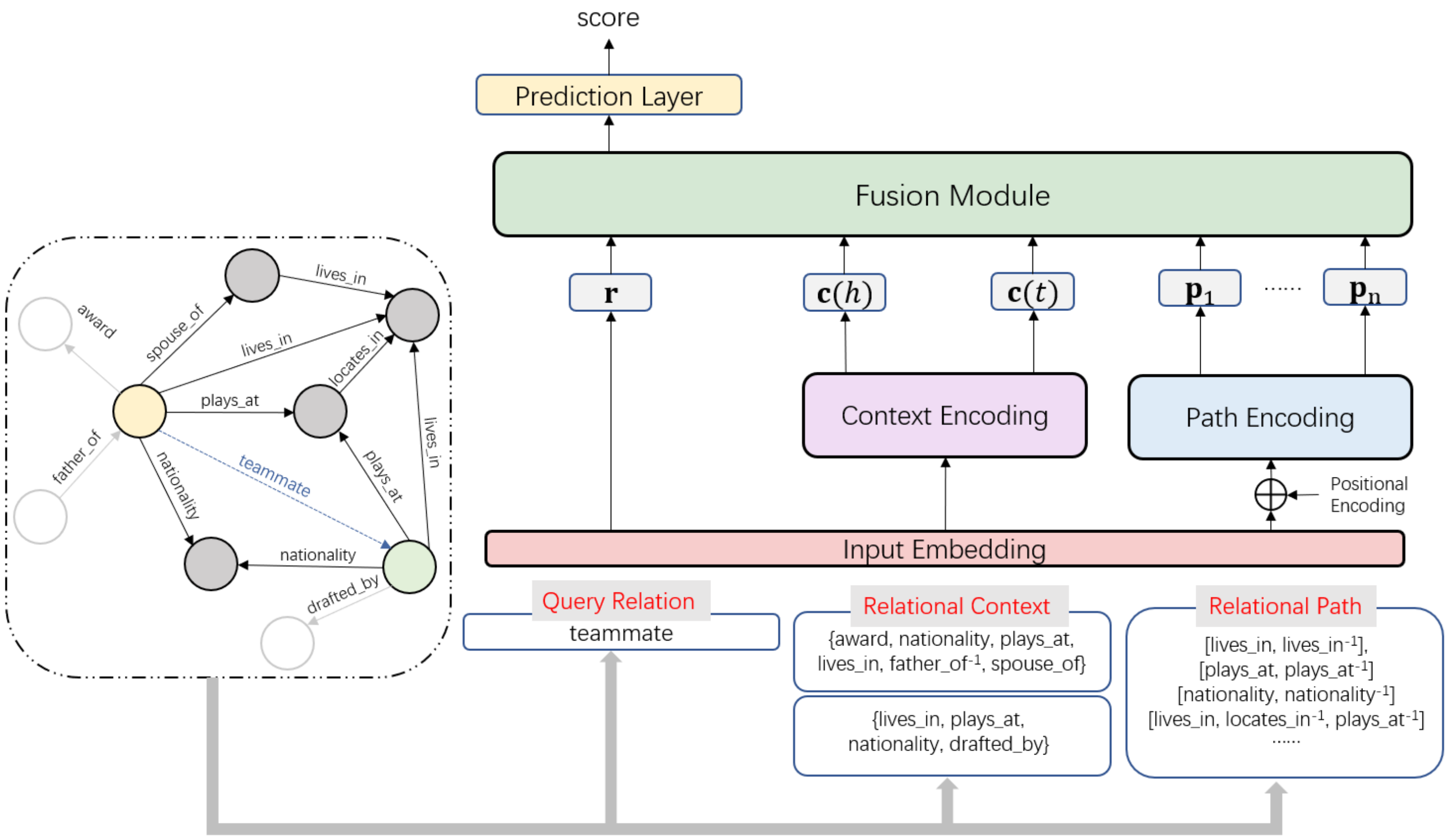
$$s = \mathrm{sigmoid}(\mathbf{W}_2(\mathrm{GELU}(\mathbf{W}_1 \mathbf{h}_0^{L_f} + \mathbf{b}_1)) + \mathbf{b}_2)$$

Output: score of current fact

- REPORT outputs a score s_i for each fact i

Model Training

- REPORT: RElational Paths and cOntext with hieRarchical Transformers



- To generate higher scores for positive triples and lower scores for negative ones. REPORT is trained with Binary Cross-Entropy (BCE) Loss:

$$\mathcal{L} = -\sum_{f_i \in \mathcal{F}^+ \cup \mathcal{F}^-} (y_i \log s_i + (1 - y_i) \log(1 - s_i))$$

- Negative samples are constructed by replacing the heads or tails of positive triplets with randomly selected entities.

$$\mathcal{F}^- = \{(h', r, t) \; or \; (h, r, t') \notin \mathcal{F}^+ | (h, r, t) \in \mathcal{F}^+\}$$

Experiments

Experimental Settings

Datasets

- We use **two benchmark datasets,** each containing **four subsets** with varying sizes sampled from WN18RR and FB15k-237, proposed in GraIL(Teru et al. 2020).
- Each subset has a training KG, and an inference KG with unseen entities and seen relations

		WN18RR				FB15K-237			
		#Relation	#Entity	#Query	#Fact	#Relation	#Entity	#Query	#Fact
v1	train	9	2,746	630	5,410	180	1,594	489	4,245
	infer	8	922	188	1,618	142	1,093	205	1,993
v2	train	10	6,954	1,838	15,262	200	2,608	1,166	9,739
	infer	10	2,757	441	4,011	172	1,660	478	4,145
v3	train	11	12,078	3,097	25,901	215	3,668	2,194	17,986
	infer	11	5,084	605	6,327	183	2,501	865	7,406
v4	train	9	3,861	934	7,940	219	4,707	3,352	27,203
	infer	9	7,084	1,429	12,334	200	3,051	1,424	11,714

In *train*:
#Query: #triplets for validation in training KG;
#Fact: #triplets in the background KG (training KG).

In *infer*:
#Query: #triplets for inference in inference KG;
#Fact: #triplets in the background KG (inference KG).

Quantitative Results

- We report mean reciprocal rank (**MRR**) and **Hits@10,** averaged in five runs.
- Best results are bold, second best ones are underlined.

Model	WN18RR								FB15k-237							
	v1		v2		v3		v4		v1		v2		v3		v4	
	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR
NeuraLP	74.37	71.74	68.93	68.54	46.18	44.23	67.13	67.14	52.92	46.13	58.94	51.85	52.90	48.70	55.88	49.54
DRUM	74.37	72.46	68.93	68.82	46.18	44.96	67.13	67.27	52.92	47.55	58.73	52.78	52.90	49.64	55.88	50.43
RuleN	80.85	79.15	78.23	77.82	53.39	51.53	71.59	71.65	49.76	45.97	77.82	69.08	87.69	**73.68**	85.60	**74.19**
GraIL	82.45	80.45	78.68	78.13	58.43	54.11	73.41	73.84	64.15	48.56	81.80	62.54	82.83	70.35	89.29	70.60
CoMPILE	83.60	78.28	79.82	79.61	60.69	53.97	75.49	75.34	67.64	50.52	82.98	65.54	84.67	66.95	87.44	63.69
TACT	84.04	—	81.63	—	67.97	—	76.56	—	65.76	—	83.56	—	85.20	—	88.69	—
RPC-IR	85.11	—	81.63	—	62.40	—	76.35	—	67.56	—	82.53	—	84.36	—	89.22	—
ConGLR	85.64	—	**92.93**	—	70.74	—	**92.90**	—	68.29	—	85.98	—	88.61	—	89.31	—
REPORT	**88.03**	**80.95**	85.83	**82.01**	**72.31**	**58.38**	81.46	**77.43**	**71.69**	**53.22**	**88.91**	**70.62**	**91.62**	71.51	**92.28**	71.28

Achieve better results than SOTA baselines on most subsets

Ablation Studies

- Ablation results (Hits@10) when discarding relational context or relational paths from the input

	WN18RR				FB15k-237			
	v1	v2	v3	v4	v1	v2	v3	v4
REPORT	88.03	85.83	72.31	81.46	71.69	88.91	91.62	92.28
w\o context	83.78	81.63	63.31	76.35	61.22	79.81	77.86	72.93
w\o path	27.66	31.29	38.51	28.90	38.54	59.94	41.39	35.50

Both types of information are effective for inductive relation prediction

Case Studies

- Attention scores in fusion module's last layer indicate different elements' contribution scores.
- Elements with higher scores are more important for prediction.

Query Fact	Component	Score
(University of Arizona, field_of_study, Finance)	[people/institution^{-1}, people/institution, field_of_study]	0.222
	[people/institution^{-1}, people/study, field_of_study^{-1}, field_of_study]	0.188
	[field_of_study, study/students_majoring, field_of_study^{-1}, field_of_study]	0.061
(Sony BMG Music Entertainment, music/artist,Christina Aguilera)	[music/artist, music_genere/artist^{-1}, music_genere/artist]	0.187
	tail:{celebrity/dated^{-1},music_genere/artist^{-1},celebrity/canoodled^{-1}, people/place_lived,award/award_winner^{-1}}	0.148
	[headquarters_location, vacation_choice_of, nominated_with, celebrity/canoodled]	0.120
(Grammy Award for Best Female Pop Vocal ,award/ceremony, 52nd Grammy Awards-US)	[award/award_winner, ceremony/award_winner^{-1}]	0.282
	[award_nominee/award^{-1},ceremony/award_winner^{-1}]	0.158
	[award_nominee/award^{-1}, nominated_with, award/award_winner^{-1}]	0.086
(Sardina, administrative_division_of, Italy)	[administrative_parent]	0.368
	[location_contain^{-1}]	0.326
	head:{administrative_parent,location_contain^{-1},vacation_choice_of}	0.221
(Broadcast Film Critics Association Award for Best Film, award_nomination, Walk the Line)	[award_nomination, award_nomination^{-1}, award_nomination]	0.331
	[award_nomination, award_wining_work, award_nomination]	0.266
	[award_nomination, film_country, film_country^{-1}]	0.119

REPORT is interpretable by providing most important elements for prediction

Conclusion

- ✓ The first work that involves relational context to supplement connection information in IRP;
- ✓ A new framework is proposed for aggregating context and paths in a hierarchical fashion;
- ✓ Consistent improvements across eight subsets of two benchmarks;
- ✓ Explanations can be automatically generated for prediction results;

Thank you !