Few-Shot Inductive Learning on Temporal Knowledge Graphs using Concept-Aware Information

Zifeng Ding[1,3], Jingpei Wu[2], Bailan He[1], Yunpu Ma[1,3], Zhen Han[1,3], Volker Tresp[1,3]
[1]Ludwig Maximilian University of Munich & [2]Technical University of Munich & [3]Siemens AG

New Entities Constantly Emerge in Temporal Knowledge Graphs (TKGs)

- Newly-emerged entities in TKGs exhibit long-tail distributions and have only few edges.
- Traditional TKG reasoning methods have no way to model newly-emerged, yet unseen entities.

What we do

- Propose the TKG few-shot out-of-graph (OOG) link prediction task.
- Propose a meta-learning-based model to solve the new task by effectively learning the inductive representations of unseen entities.
- Exploit entity concepts from temporal knowledge bases for boosting model performance.

GOAL: improving link prediction performance concerning newly-emerged entities in TKGs.

TKG Few-Shot Out-of-Graph Link Prediction

Given:

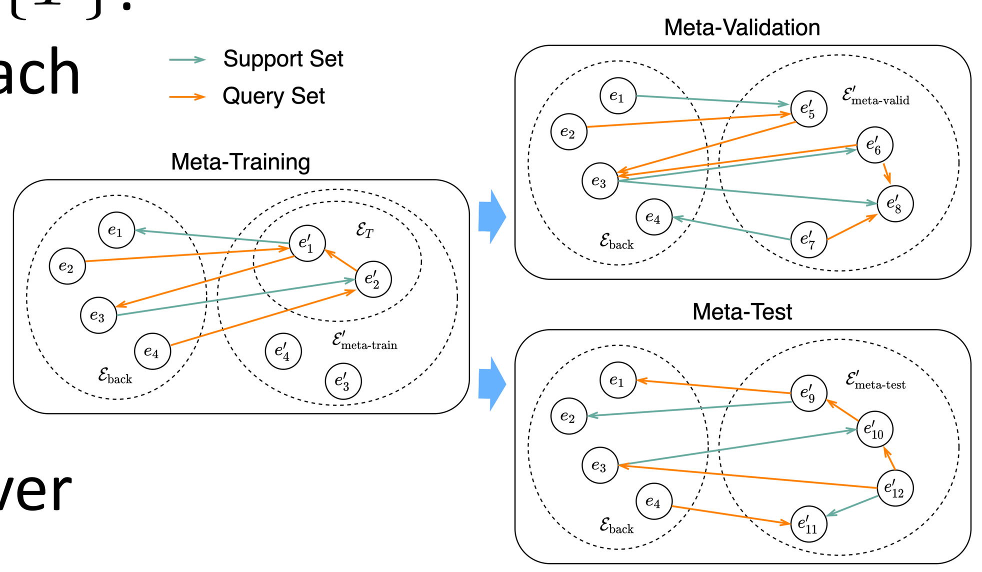
- A background TKG $\mathcal{G}_{\text{back}} \subseteq \mathcal{E}_{\text{back}} \times \mathcal{R} \times \mathcal{E}_{\text{back}} \times \mathcal{T}$.
- Unseen entities $\mathcal{E}' = \{e'\}, \mathcal{E}' \cap \mathcal{E}_{\text{back}} = \emptyset$.
- K (1 or 3) observed associated TKG facts for each e'.

TKG few-shot OOG link prediction aims to:

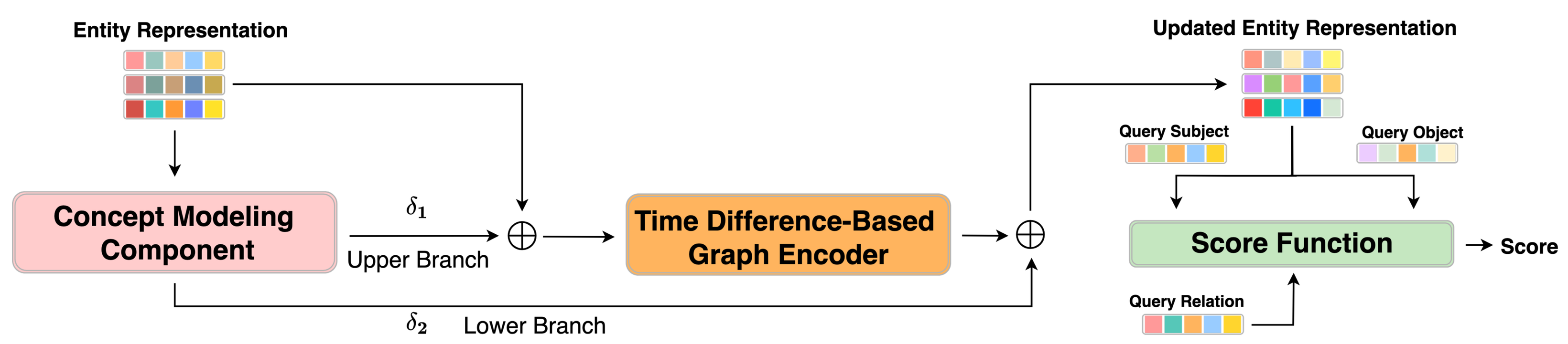
- Predict the missing entities in the link prediction queries derived from the unobserved TKG facts containing unseen entities, i.e., $(e', r_q, ?, t_q)$ or $(?, r_q, e', t_q), r_q \in \mathcal{R}, t_q \in \mathcal{T}$.

Formulate into Meta-Learning Problem

- Split $\mathcal{E}'$ into $\mathcal{E}'_{\text{meta-train}}, \mathcal{E}'_{\text{meta-valid}}, \mathcal{E}'_{\text{meta-test}}$.
- Meta-training over training tasks $\{T\}$.
  N unseen entities in each task, each with support set (observed) $\mathcal{S}_{e'}$, and query set (unobserved) $\mathcal{Q}_{e'}$.
  Maximize performance over $\mathcal{Q}_{e'}$ given $\mathcal{S}_{e'}$.
- Meta-validation and meta-test over $\mathcal{E}'_{\text{meta-valid}}, \mathcal{E}'_{\text{meta-test}}$.



FILT



FILT consists of

- Concept modeling component.
- Time difference-based graph encoder.
- Knowledge graph scoring function.

Concept Modeling Component

A TKG fact in the ICEWS temporal knowledge base

Event Date	Source Name	Source Sectors	Event Text	Target Name	Target Sectors
2014-09-09	Government (Iraq)	Government	Demand meeting	Al-Shabaab	Muslim
Timestamp	Subject	Subject Entity Concept	Relation	Object	Object Entity Concept

$$\mathbf{h}_c = \frac{1}{|\mathcal{N}_c|}\sum_{e \in \mathcal{N}_c} \mathbf{h}_e$$

Initialize concept representations

$$\mathbf{h}_c = \sum_{e_i \in \mathcal{N}_c} \alpha_c^{e_i} \mathbf{h}_{e_i}, \quad \alpha_c^{e_i} = \frac{\exp(\mathbf{h}_{e_i}^\top \mathbf{h}_c)}{\sum_{e_j \in \mathcal{N}_c} \exp(\mathbf{h}_{e_j}^\top \mathbf{h}_c)}$$

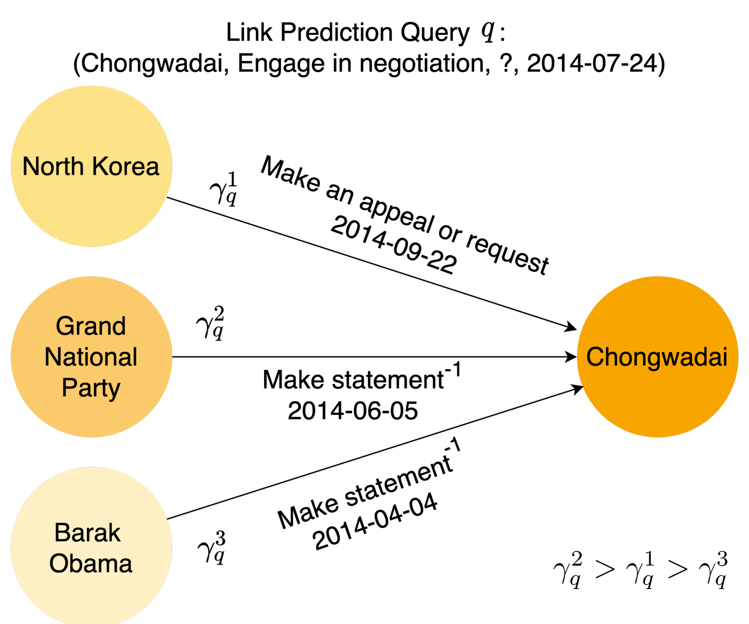
Correct concept representations

$$\mathbf{h}_e^{\mathcal{C}_e} = \sum_{c_i \in \mathcal{C}_e} \beta_e^{c_i} \mathbf{h}_{c_i}, \quad \beta_e^{c_i} = \frac{\exp(\mathbf{h}_{c_i}^\top \mathbf{h}_e)}{\sum_{c_j \in \mathcal{C}_e} \exp(\mathbf{h}_{c_j}^\top \mathbf{h}_e)}$$

Compute concept-aware information

Time-Difference-Based Graph Encoder

- Find temporal neighbors for unseen entities from support quadruples.
- We assume that the smaller the time difference is, the more important a temporal neighbor is.



$$\mathbf{h}_{(e', t_q)} = \sum_{(\tilde{e}_i, r_i, t_i) \in \mathcal{N}_{e'}} \gamma_q^i \mathbf{W_g}(\mathbf{h}_{\tilde{e}_i} \| \mathbf{h}_{r_i}),$$

$$\gamma_q^i = \frac{\exp(1/|t_q - t_i|)}{\sum_{(\tilde{e}_j, r_j, t_j) \in \mathcal{N}_{e'}} \exp(1/|t_q - t_j|)}$$

The Importance of a Temporal Neighbor

Parameter Learning

$$\mathcal{L} = \sum_{e' \in \mathcal{E}_T} \sum_{q^+ \in \mathcal{Q}_{e'}} \sum_{q^- \in \mathcal{Q}_{e'}^-} \max\{\theta - score(q^+) + score(q^-), 0\}$$

Query Quadruples | Negative Sample

Score computed with a knowledge graph scoring function ComplEx (Trouillon et al., 2016).

New Datasets for TKG OOG Link Prediction

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{E}'_{\text{meta-train}}\rvert$	$\lvert\mathcal{E}'_{\text{meta-valid}}\rvert$	$\lvert\mathcal{E}'_{\text{meta-test}}\rvert$	N_{back}	$N_{\text{meta-train}}$	$N_{\text{meta-valid}}$	$N_{\text{meta-test}}$
ICEWS14-OOG	7128	230	365	385	48	49	83448	5772	718	705
ICEWS18-OOG	23033	256	304	1268	160	158	444269	19291	2425	2373
ICEWS0515-OOG	10488	251	4017	647	80	82	448695	10115	1217	1228

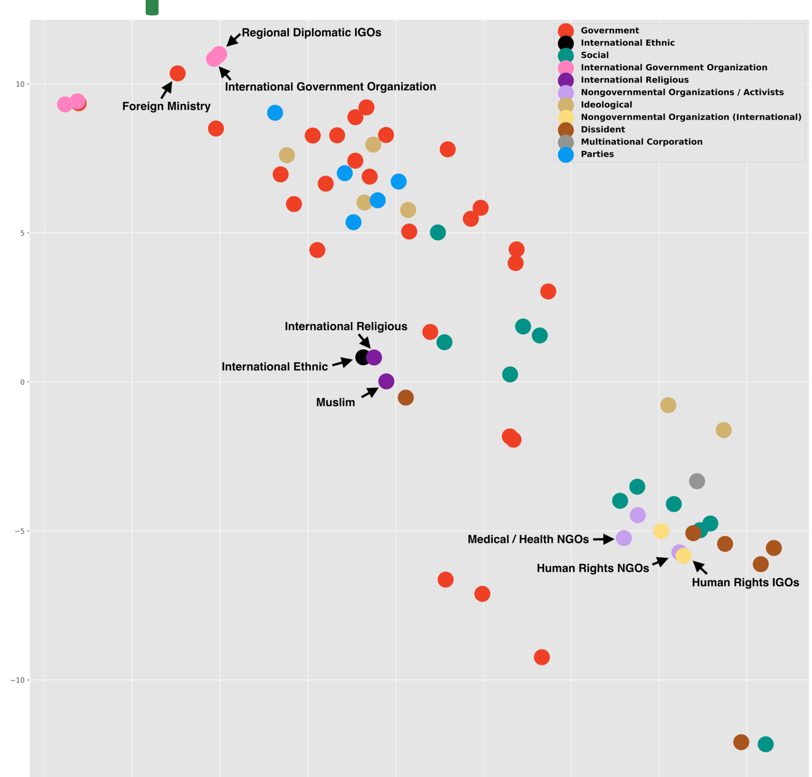
1-Shot and 3-Shot Results

Datasets	ICEWS14-OOG							
	MRR		H@1		H@3		H@10	
Model	1-S	3-S	1-S	3-S	1-S	3-S	1-S	3-S
ComplEx	.048	.046	.018	.014	.045	.046	.099	.089
BiQUE	.039	.035	.015	.014	.041	.030	.073	.066
TNTComplEx	.043	.044	.015	.016	.033	.042	.102	.096
TeLM	.032	.035	.012	.009	.021	.023	.063	.077
TeRo	.009	.010	.002	.002	.005	.002	.015	.020
MEAN	.035	.144	.013	.054	.032	.145	.082	.339
LAN	.168	.199	.050	.061	.199	.255	**.421**	**.500**
GEN	.231	.234	.162	.155	.250	.284	.378	.389
FILT	**.278**	**.321**	**.208**	**.240**	**.305**	**.357**	.410	.475

Datasets	ICEWS18-OOG							
	MRR		H@1		H@3		H@10	
Model	1-S	3-S	1-S	3-S	1-S	3-S	1-S	3-S
ComplEx	.039	.044	.031	.026	.048	.042	.085	.093
BiQUE	.029	.032	.022	.021	.033	.037	.064	.073
TNTComplEx	.046	.048	.023	.026	.043	.044	.087	.082
TeLM	.049	.019	.029	.001	.045	.013	.084	.054
TeRo	.007	.006	.003	.001	.006	.003	.013	.006
MEAN	.016	.101	.003	.014	.012	.114	.043	.283
LAN	.077	.127	.018	.025	.067	.165	.199	.344
GEN	.171	.216	.112	.137	.189	.252	.289	.351
FILT	**.191**	**.266**	**.129**	**.187**	**.209**	**.298**	**.316**	**.417**

Datasets	ICEWS0515-OOG							
	MRR		H@1		H@3		H@10	
Model	1-S	3-S	1-S	3-S	1-S	3-S	1-S	3-S
ComplEx	.077	.076	.045	.048	.074	.071	.129	.120
BiQUE	.075	.083	.044	.049	.072	.077	.130	.144
TNTComplEx	.034	.037	.014	.012	.031	.036	.060	.071
TeLM	.080	.072	.041	.034	.077	.072	.138	.151
TeRo	.012	.023	.000	.010	.008	.017	.024	.040
MEAN	.019	.148	.003	.039	.017	.175	.052	.384
LAN	.171	.182	.081	.068	.180	.191	.367	.467
GEN	.268	.322	.185	.231	**.308**	.362	**.413**	.507
FILT	**.273**	**.370**	**.201**	**.299**	.303	**.391**	.405	**.516**

Visualization of Concept Representations



Concepts bearing the same label tend to form a cluster.

Clusters having similar semantic meanings tend to be close to each other.

Concept modeling component learns the semantics of entity concepts.