Understanding DEKG-ILP: A New Method for Enhancing Knowledge Graph Link Prediction

Background

A Knowledge Graph (KG) is a structured representation of data through nodes (entities) and edges (relationships). Common knowledge graphs include Freebase, NELL, and DBpedia. They play a crucial role in applications such as information retrieval, recommendation systems, and question answering. An essential task in knowledge graphs is link prediction, which aims to predict missing links in the graph. This task has seen significant success, primarily due to recent advancements in knowledge graph embedding methods.

However, knowledge graphs are not static; they evolve dynamically. For instance, DBpedia saw around 200 new entities emerging daily from late 2015 to early 2016. Traditional embedding methods are inefficient in handling these new entities because they are unseen during training. Although retraining the entire graph can address this issue, it is time-consuming and computationally expensive in real-world applications. Therefore, Inductive Link Prediction (ILP) has become an important research direction, aiming to predict links for new entities in emerging knowledge graphs.

Why is the DEKG-ILP Model Needed?

Existing methods mainly focus on predicting enclosing links within new knowledge graphs. However, in many real-world applications, there are missing bridging links between new and original graphs that carry crucial evolutionary information. For example, in predicting drug interactions, bridging links may contain vital information. Current methods, such as Grail and TACT, fail to effectively predict bridging links in Disconnected Emerging Knowledge Graphs (DEKGs). The DEKG-ILP model was proposed to fill this gap.

(a) Original NBA KG   (b) Emerging KG of NBA 2008 draft
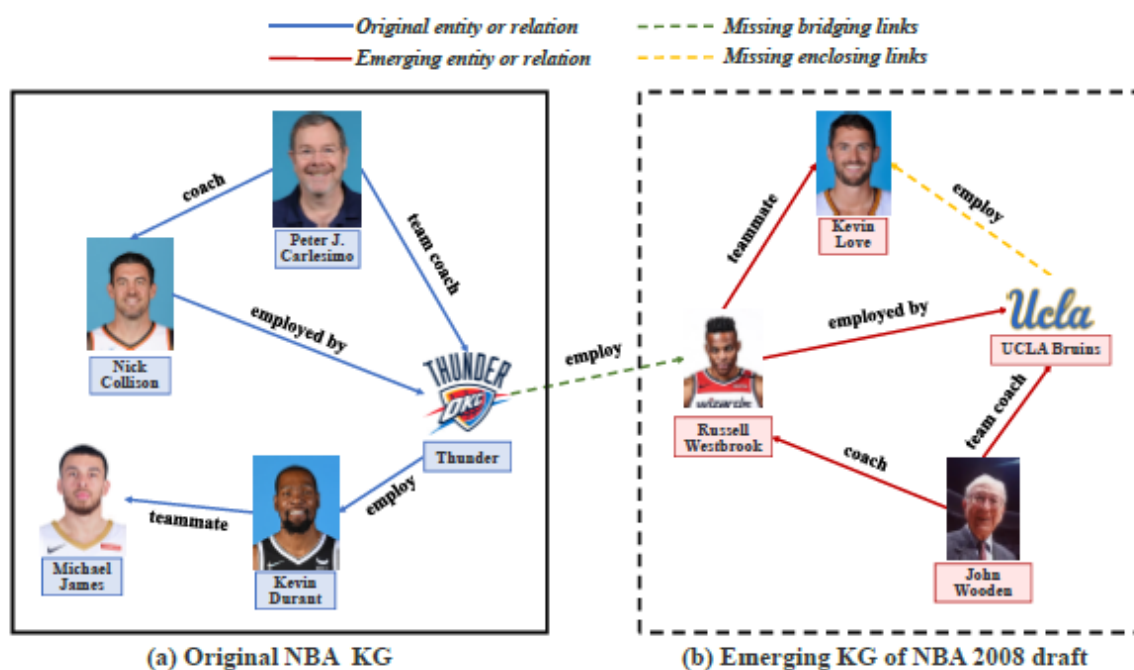
The image above illustrates an example of a DEKG and an original KG, explaining the concepts of bridging and enclosing links. The dashed lines indicate missing links.

Main Functions and Goals of the DEKG-ILP Model

The DEKG-ILP model aims to predict missing links in DEKGs, including both enclosing and bridging links. The model consists of two main modules:

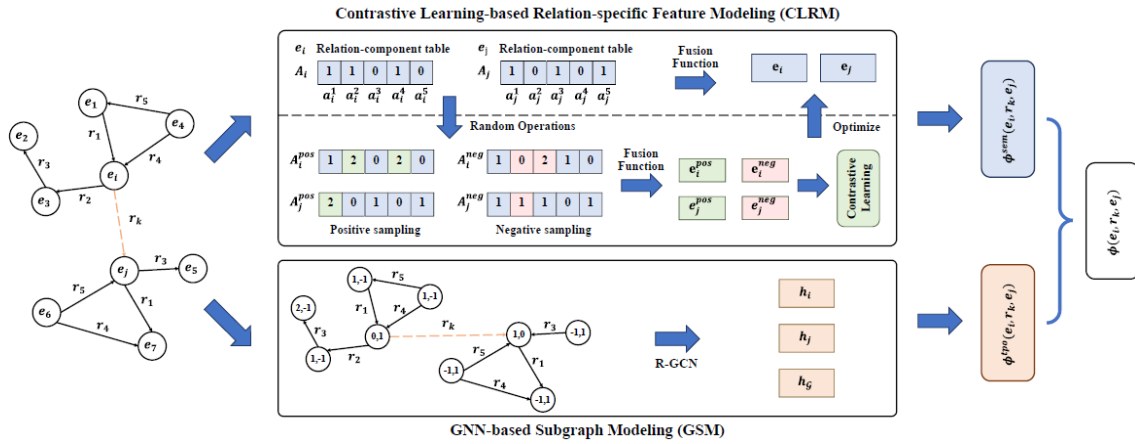1. **CLRM (Contrastive Learning-based Relation-specific Feature Modeling)**

   Extracts global, relation-based semantic features shared between original KGs and DEKGs.

2. **GSM (GNN-based Subgraph Modeling)**

   Extracts local topological information around each link.

Implementation of the DEKG-ILP Model

The diagram below shows the architecture of the DEKG-ILP model, including the workflows of the CLRM and GSM modules.

1. **CLRM Module**

   - **Relation-specific Feature Modeling**: Extracts and represents the semantic features of each relation, combining them to represent entities.

   - **Contrastive Learning**: Optimizes relation-specific features by generating positive and negative samples.

2. **GSM Module**

   - **Subgraph Extraction**: Extracts subgraphs around target links. Single subgraphs are used for enclosing links, while two disconnected subgraphs are used for bridging links.

   - **Node Labeling**: Labels each node in the subgraph based on their distance to the head and tail entities of the target link.

   - **GNN (R-GCN)**: Processes the subgraphs using relational graph convolutional networks to extract topological features.

3. **Comprehensive Scoring**

   Combines the semantic score from CLRM and the topological score from GSM to derive the final score for the target link.

Experimental Results

Experiments on multiple benchmark datasets (e.g., FB15k-237, NELL-995, and WN18RR) show that the DEKG-ILP model significantly outperforms existing methods in predicting both enclosing and bridging links. Notably, DEKG-ILP exhibits substantial improvements in bridging link prediction, demonstrating its effectiveness in handling DEKGs.

The image below compares the performance of the DEKG-ILP model with other baseline models across three datasets.

TABLE III
MAIN RESULT ON *EQ*, *MB*, *ME* OF FB15k-237, NELL-995, AND WN18RR

Datasets	Models	EQ				MB				ME			
		MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
FB15k-237	TransE	0.241	0.169	0.264	0.337	0.210	0.143	0.249	0.310	0.197	0.123	0.218	0.317
	RotatE	0.089	0.021	0.095	0.192	0.101	0.025	0.125	0.233	0.094	0.022	0.105	0.209
	ConvE	0.102	0.033	0.105	0.227	0.097	0.031	0.105	0.200	0.119	0.030	0.114	0.228
	GEN	0.093	0.032	0.089	0.196	0.109	0.030	0.130	0.241	0.101	0.027	0.110	0.207
	RuleN	0.265	0.237	0.267	0.268	0.212	0.186	0.239	0.240	0.402	0.360	0.443	0.447
	Grail	0.279	0.216	0.323	0.342	0.226	0.164	0.259	0.281	0.456	0.378	0.523	0.569
	TACT	0.227	0.130	0.316	0.401	0.186	0.101	0.249	0.339	0.311	0.222	0.382	0.463
	DEKG-ILP	0.508	0.351	0.693	0.841	0.535	0.396	0.693	0.832	0.634	0.512	0.785	0.891
NELL-995	TransE	0.083	0.023	0.079	0.156	0.234	0.161	0.263	0.345	0.158	0.088	0.177	0.269
	RotatE	0.118	0.021	0.170	0.331	0.090	0.027	0.089	0.179	0.091	0.014	0.094	0.187
	ConvE	0.098	0.035	0.108	0.172	0.102	0.033	0.088	0.194	0.100	0.026	0.102	0.179
	GEN	0.091	0.029	0.084	0.181	0.132	0.105	0.156	0.239	0.127	0.058	0.142	0.223
	RuleN	0.234	0.197	0.230	0.258	0.300	0.249	0.336	0.340	0.452	0.371	0.508	0.510
	Grail	0.193	0.109	0.231	0.393	0.307	0.233	0.352	0.423	0.411	0.305	0.518	0.614
	TACT	0.156	0.071	0.221	0.328	0.223	0.125	0.305	0.420	0.292	0.165	0.428	0.557
	DEKG-ILP	0.353	0.218	0.489	0.631	0.468	0.301	0.694	0.830	0.532	0.380	0.727	0.842
WN18RR	TransE	0.133	0.073	0.135	0.353	0.164	0.105	0.167	0.369	0.160	0.065	0.126	0.389
	RotatE	0.161	0.038	0.271	0.476	0.142	0.030	0.215	0.458	0.135	0.028	0.197	0.407
	ConvE	0.111	0.051	0.124	0.308	0.122	0.049	0.165	0.333	0.129	0.049	0.115	0.327
	GEN	0.158	0.101	0.147	0.372	0.153	0.098	0.170	0.365	0.148	0.051	0.119	0.352
	RuleN	0.382	0.342	0.402	0.410	0.252	0.232	0.241	0.466	0.335	0.312	0.334	0.434
	Grail	0.401	0.320	0.473	0.613	0.261	0.179	0.344	0.513	0.341	0.243	0.424	0.611
	TACT	0.442	0.328	0.578	0.593	0.335	0.231	0.455	0.472	0.321	0.224	0.431	0.475
	DEKG-ILP	0.471	0.350	0.607	0.701	0.359	0.240	0.480	0.625	0.378	0.245	0.534	0.685

Here are some explanations for the contents of the figure:

Datasets

- **FB15k-237**
- **NELL-995**
- **WN18RR**

Models

- **TransE**: An embedding method based on translation.
- **RotatE**: An embedding method modeling rotation in complex space.
- **ConvE**: An embedding method based on convolutional neural networks.
- **GEN**: A method based on graph neural networks.
- **RuleN**: A method based on rule mining.
- **Grail**: A subgraph reasoning method specifically designed for DEKGs.
- **TACT**: A subgraph reasoning method focused on relation prediction.
- **DEKG-ILP**: The newly proposed model in this paper.

Evaluation Metrics

- **MRR (Mean Reciprocal Rank)**
- **Hits@1**
- **Hits@5**
- **Hits@10**

Future Research Directions and Impact

1. **Further Model Optimization**: Explore more effective contrastive learning strategies or more powerful graph neural network structures.

2. **Application Expansion**: Apply DEKG-ILP to more real-world scenarios, such as predicting drug interactions in medical knowledge graphs.

3. **Cross-Graph Applications**: Study how to predict links between knowledge graphs in different domains to enhance cross-domain knowledge integration and utilization.

The DEKG-ILP model not only addresses the limitations of existing methods but also provides an effective approach to handling complex and dynamically evolving knowledge graphs. It holds broad application prospects and research value.

Conclusion

The DEKG-ILP model combines global and local information to effectively predict missing bridging and enclosing links in knowledge graphs, thereby enhancing the completeness and reasoning capability of knowledge graphs. This model stands out among existing methods and offers new ideas and directions for future knowledge graph research and applications.

[Paper Link](#)