Error Detection on Knowledge Graphs with Triple Embedding

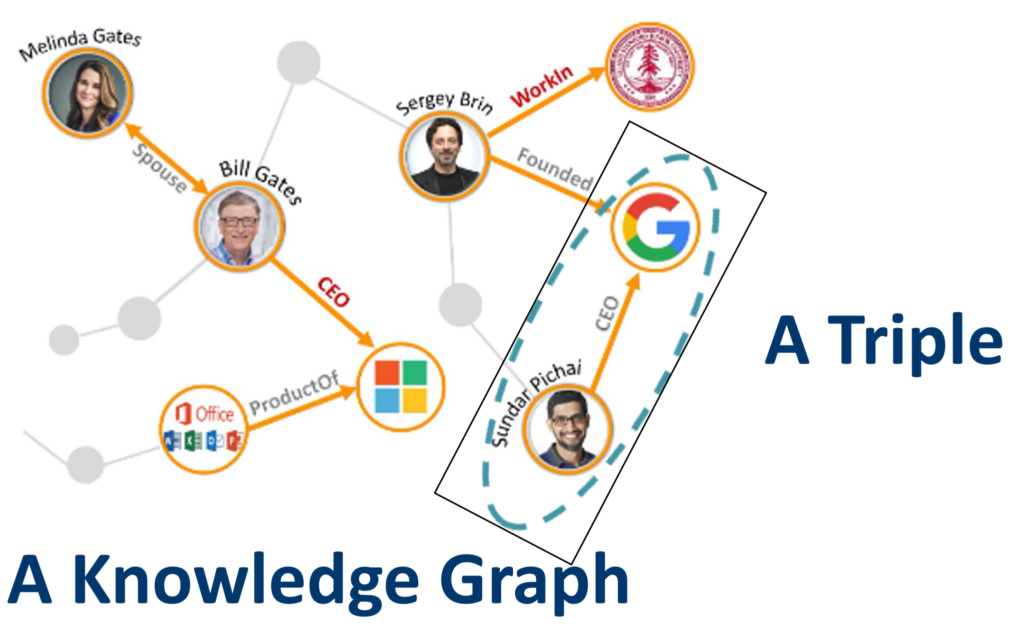
Yezi Liu∗, Qinggang Zhang†, Mengnan Du‡, Xiao Huang†, Xia Hu§

∗University of California Irvine, Email: yezil3@uci.edu
†Hong Kong Polytechnic University, Email: qinggangg.zhang, xiaohuang@comp.polyu.edu.hk
‡New Jersey Institute of Technology, Email: mengnan.du@njit.edu
§Rice University, Email: xia.hu@rice.edu

A Knowledge Graph

Errors are inevitably introduced when constructing Knowledge Graphs (KGs).

- **Noise in the sources;**
- **Imperfection of the acquisition methods.**



Background

- **Embedding-based error detection: rely on entity types, hard to obtain.**
- **Rule-mining-based: depend on the quality and limit by quantity of the rules.**
- **Build classifiers on KG features (entity categories, path features, out-degrees): need ground truth.**

Our solutions:

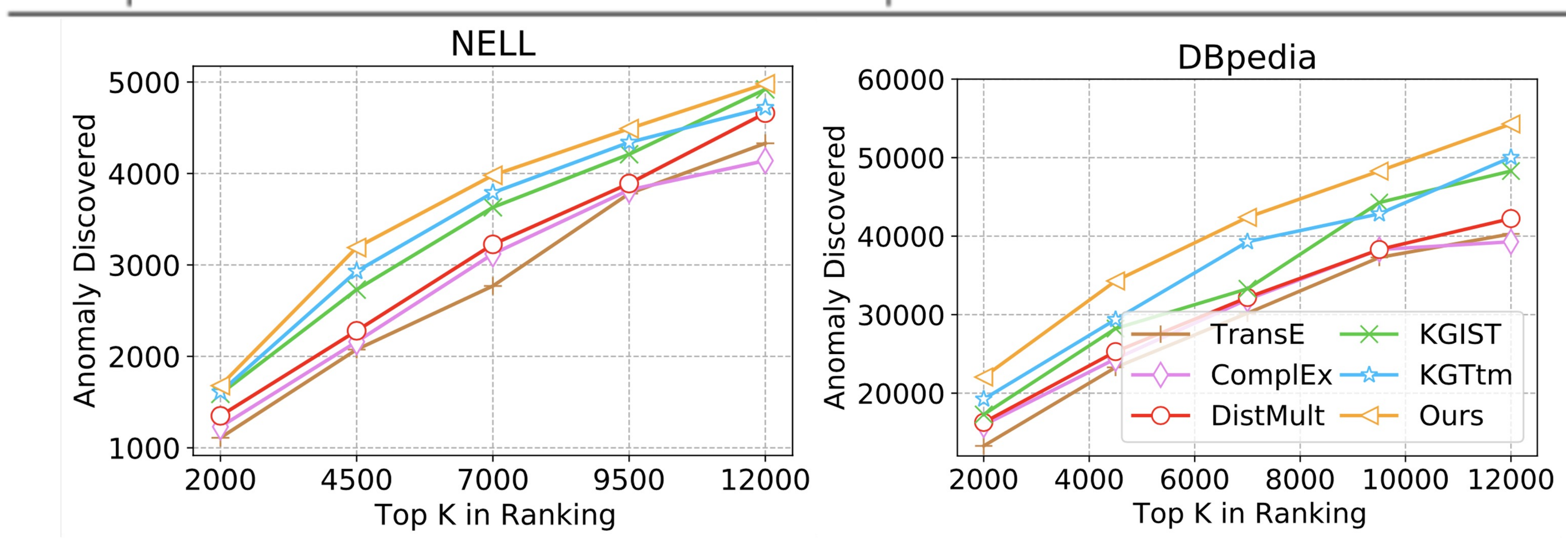
- **KG self-contained information;**
- **Generalizable algorithm for errors.**

Experiments

1. Effectiveness

TABLE II: Error detection results of Precision@K and Recall@K based on two datasets with error ratio $p = 5\%$.

Metric	Precision@K										Recall@K									
Dataset	NELL					DBpedia					NELL					DBpedia				
Top@K	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%
TransE	0.637	0.531	0.427	0.412	0.366	0.645	0.548	0.476	0.423	0.383	0.127	0.212	0.256	0.330	0.366	0.129	0.219	0.286	0.338	0.383
ComplEx	0.601	0.526	0.454	0.419	0.348	0.603	0.533	0.472	0.431	0.357	0.120	0.210	0.272	0.335	0.348	0.121	0.213	0.283	0.345	0.357
DistMult	0.631	0.532	0.472	0.423	0.401	0.662	0.539	0.489	0.438	0.420	0.126	0.213	0.283	0.338	0.401	0.132	0.216	0.293	0.350	0.420
KGIST	0.675	0.586	0.496	0.459	0.431	0.701	0.613	0.501	0.498	0.450	0.134	0.234	0.298	0.367	0.431	0.140	0.245	0.301	0.398	0.450
KGTtm	0.681	0.600	0.512	0.452	0.405	0.760	0.628	0.586	0.474	0.436	0.136	0.240	0.307	0.362	0.405	0.152	0.251	0.352	0.379	0.436
Ours	**0.738**	**0.623**	**0.538**	**0.477**	**0.435**	**0.844**	**0.729**	**0.632**	**0.557**	**0.497**	**0.148**	**0.249**	**0.323**	**0.382**	**0.436**	**0.169**	**0.292**	**0.379**	**0.445**	**0.497**



2. Efficiency

TABLE III: The running time for one iteration (in seconds).

	TransE	ComplEx	DistMult	KGIST	KGTtm	Ours
NELL	1	1	40	52	4	1
DBpedia	20	21	96	122	33	38

3. Ablation study

TABLE IV: Ablation Study on NELL with 5% ratio of errors.

	Precision@K					Recall@K				
Top@K	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%
TripleNet_Local	0.674	0.571	0.497	0.446	0.406	0.135	0.228	0.291	0.357	0.406
TripleNet_Global	0.714	0.619	0.526	0.464	0.422	0.143	0.247	0.315	0.371	0.422
TripleNet_GAT	0.738	0.623	0.538	0.477	0.435	0.148	0.249	0.323	0.382	0.436

4. Future Work

Using error-aware KG representation learning methods for guiding KG reasoning, question answering, etc.

The TripleNet Framework