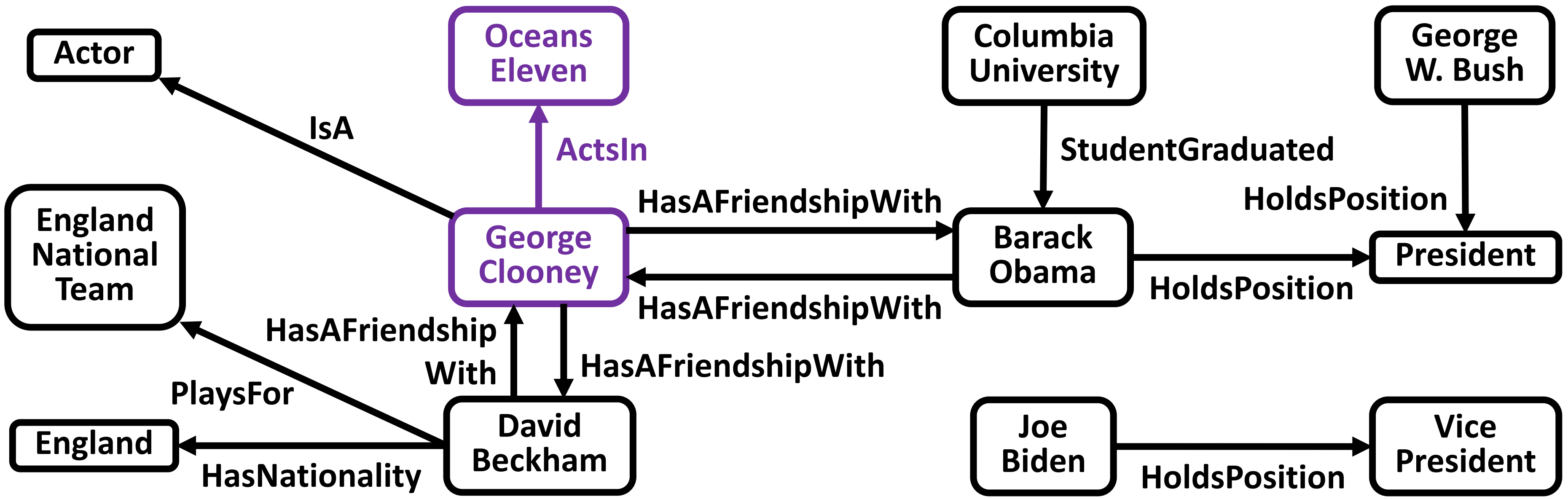
Learning Representations of Bi-Level Knowledge Graphs for Reasoning beyond Link Prediction

Chanyoung Chung and Joyce Jiyoung Whang*
School of Computing, KAIST
The 37th AAAI Conference on Artificial Intelligence (AAAI 2023)
* Corresponding Author

Knowledge Graphs

- **Knowledge Graphs** represent known facts using triplets.
  - **Knowledge Graph Embedding** represents entities and relations as feature vectors.



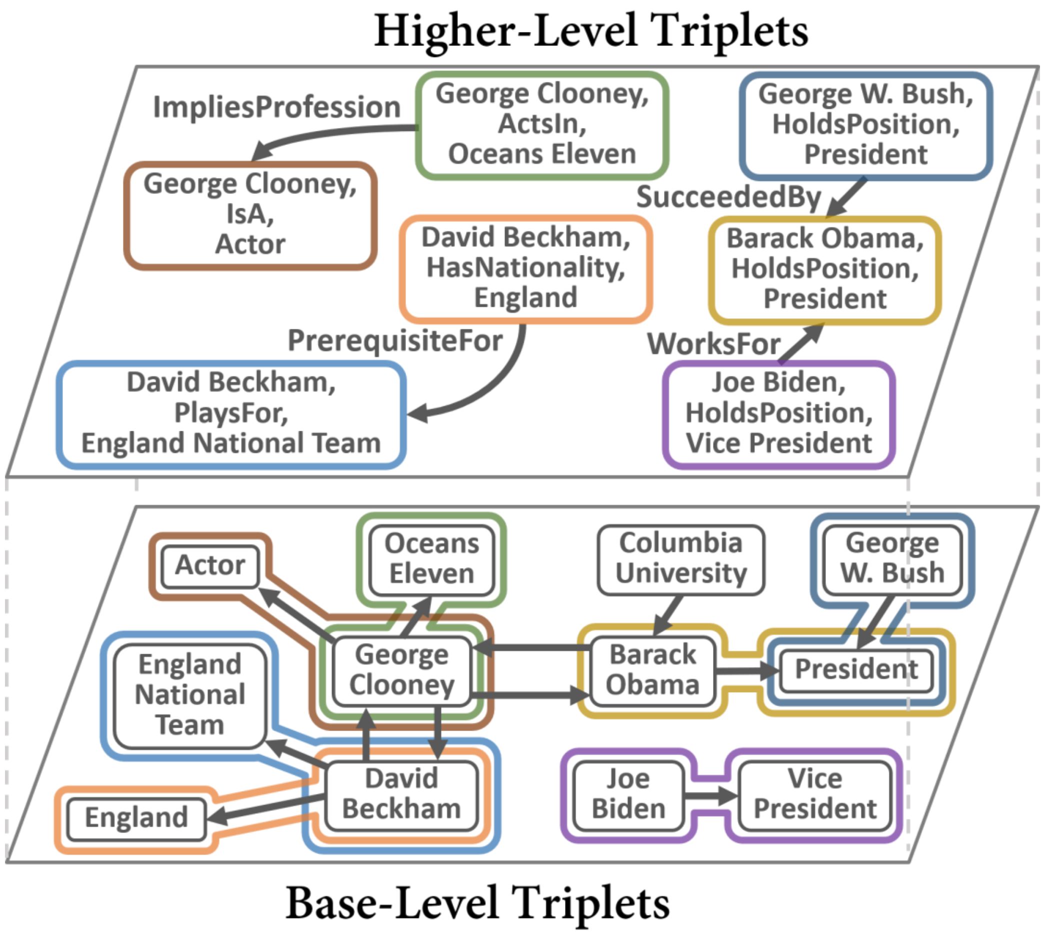
Relationships between Triplets

- **Each triplet can have a relationship with another triplet.**
  - T_1: (Joe Biden, HoldsPosition, Vice President)
  - T_2: (Barack Obama, HoldsPosition, President)

- **Higher-Level Triplets** represent the relationships between triplets.
  - **$\langle T_1$, WorksFor, $T_2\rangle$**
    - Joe Biden was a vice president when Barack Obama was a president.
  - Connects triplets using the **Higher-Level Relations**.

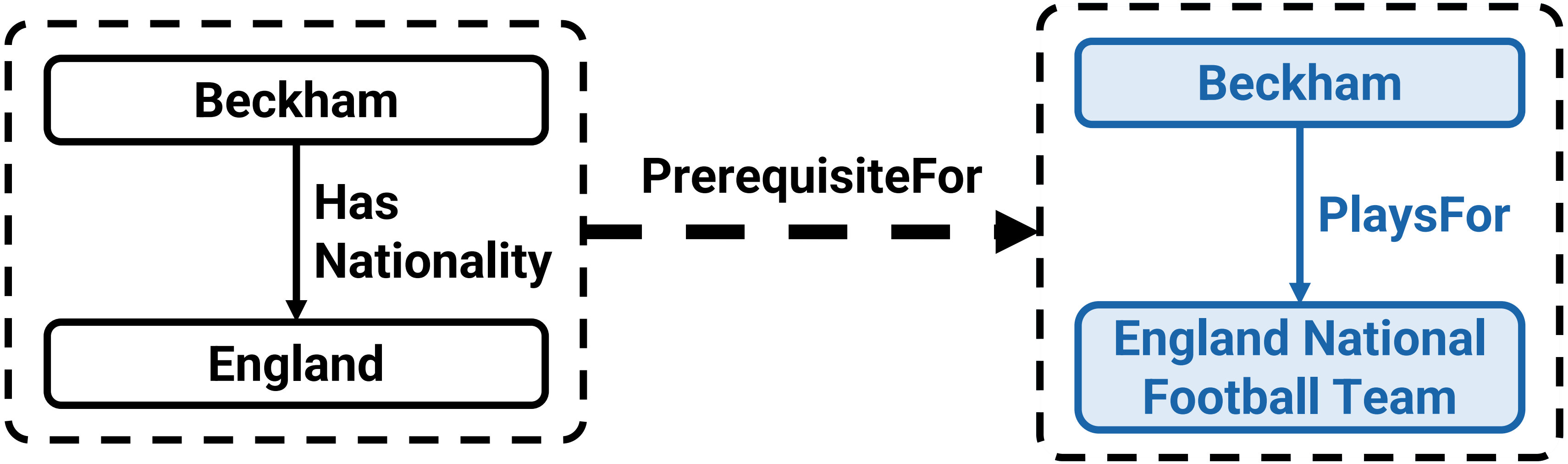
Bi-Level Knowledge Graphs

- **Base-level Triplets**
  - Relationships between entities
- **Higher-level Triplets**
  - Relationships between base-level triplets



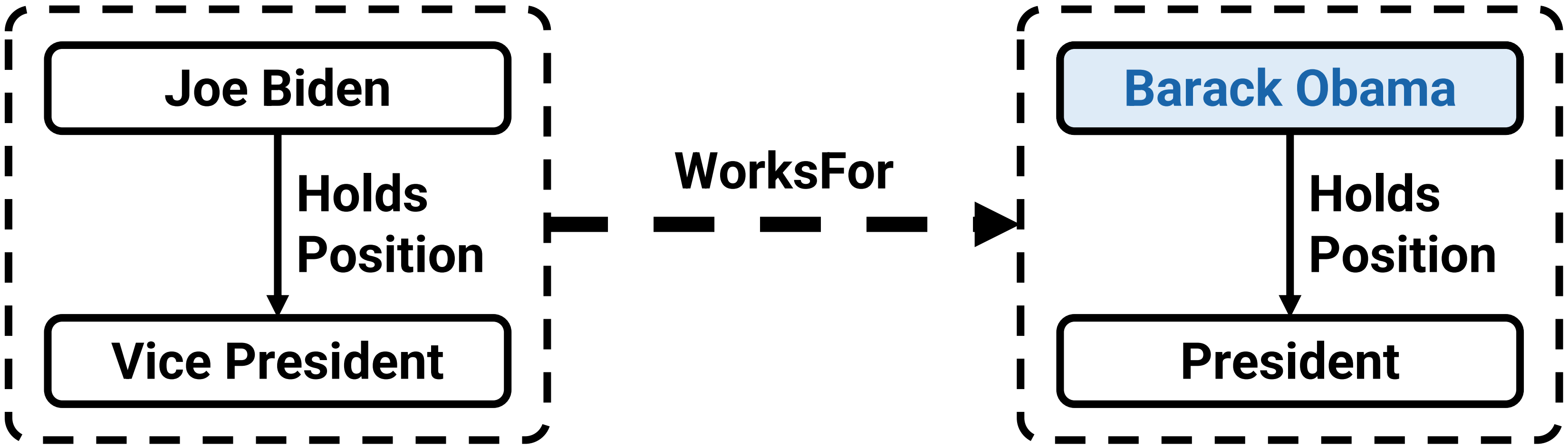
Reasoning beyond Link Prediction

- **Triplet Prediction**
  - Predicts a triplet that is likely to be connected to a given triplet.
  - ⟨(Beckham, HasNationality, England), PrerequisiteFor, **?**⟩
    - Answer: (Beckham, PlaysFor, England National Football Team)



Reasoning beyond Link Prediction

- **Conditional Link Prediction**
  - Predicts a missing entity in a triplet where another triplet is provided as a condition.
  - ⟨(Joe Biden, HoldsPosition, Vice President), WorksFor, (?, HoldsPosition, President)⟩
    - Answer: Barack Obama



Contributions

- Define **Bi-Level Knowledge Graphs** and create three real-world datasets
- Propose an efficient **data augmentation** strategy on a bi-level KG
- Develop **BiVE** (embedding of **Bi**-le**V**el knowledg**E** graphs)
- Two new tasks: **Triplet Prediction** and **Conditional Link Prediction**
  - BiVE outperforms 12 different state-of-the-art methods.

Bi-Level Knowledge Graphs

- A base-level knowledge graph $G = (\mathcal{V}, \mathcal{R}, \mathcal{E})$
  - $\mathcal{V}$: Set of entities, $\mathcal{R}$: Set of relations, $\mathcal{E}$: Set of triplets
- A set of **Higher-Level Triplets** $\mathcal{H} = \{\langle T_i, \hat{r}, T_j \rangle : T_i \in \mathcal{E}, \hat{r} \in \hat{\mathcal{R}}, T_j \in \mathcal{E}\}$
  - $\hat{\mathcal{R}}$: Set of **Higher-Level Relations**
- A **Bi-Level Knowledge Graph** $\hat{G} = (\mathcal{V}, \mathcal{R}, \mathcal{E}, \hat{\mathcal{R}}, \mathcal{H})$
  - Add higher-level triplets to the base-level KG by introducing the higher-level relations

Real-World Bi-Level KGs

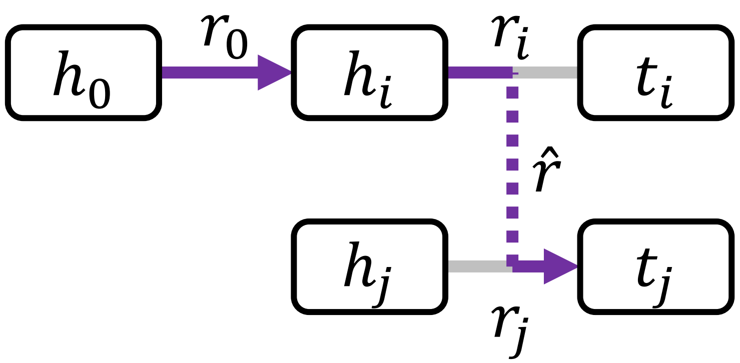
- Create three real-world bi-level knowledge graphs: ***FBH, FBHE, DBHE***
  - *FBH* and *FBHE* are based on *FB15K237* from *Freebase*.
  - *DBHE* is based on *DB15K* from *DBpedia*.
- Add the higher-level triplets to the base-level knowledge graphs
  - *FBHE* and *DBHE* contain some **externally-sourced knowledge**.
- Number of higher-level relations in the datasets
  - *FBH*: 6, *FBHE*: 10, *DBHE*: 8

Real-World Bi-Level KGs

	Higher-Level Relation	Higher-Level Triplet
	$\hat{r}$	$\langle T_i, \hat{r}, T_j \rangle$
FBHE	PrerequisiteFor	T_i: (BAFTA Award, Nominates, The King's Speech) T_j: (The King's Speech, Wins, BAFTA Award)
	ImpliesProfession	T_i: (Liam Neeson, ActsIn, Love Actually) T_j: (Liam Neeson, IsA, Actor)
	WorksFor	T_i: (Joe Biden, HoldsPosition, Vice President) T_j: (Barack Obama, HoldsPosition, President)
	SucceededBy	T_i: (George W. Bush, HoldsPosition, President) T_j: (Barack Obama, HoldsPosition, President)
DBHE	ImpliesTimeZone	T_i: (Czech Republic, TimeZone, Central European) T_j: (Prague, TimeZone, Central European)
	NextAlmaMater	T_i: (Gerald Ford, StudiesIn, University of Michigan) T_j: (Gerald Ford, StudiesIn, Yale University)

Data Augmentation on a Bi-Level KG

- Propose a **Data Augmentation Strategy** based on **Random Walks**
- Randomly visit a neighbor by following a base-level or a higher-level triplet.
  - **Do not allow going back** to an entity that has already been visited.
- **Random Walk Path**
  - Sequence of visited entities, relations and higher-level relations



Random walk path: $(h_0, r_0, h_i, r_i, \hat{r}, r_j, t_j)$

Relation Sequence

- **Relation Sequence** (p_k)
  - Sequence of relations and higher-level relations extracted from a random walk path.

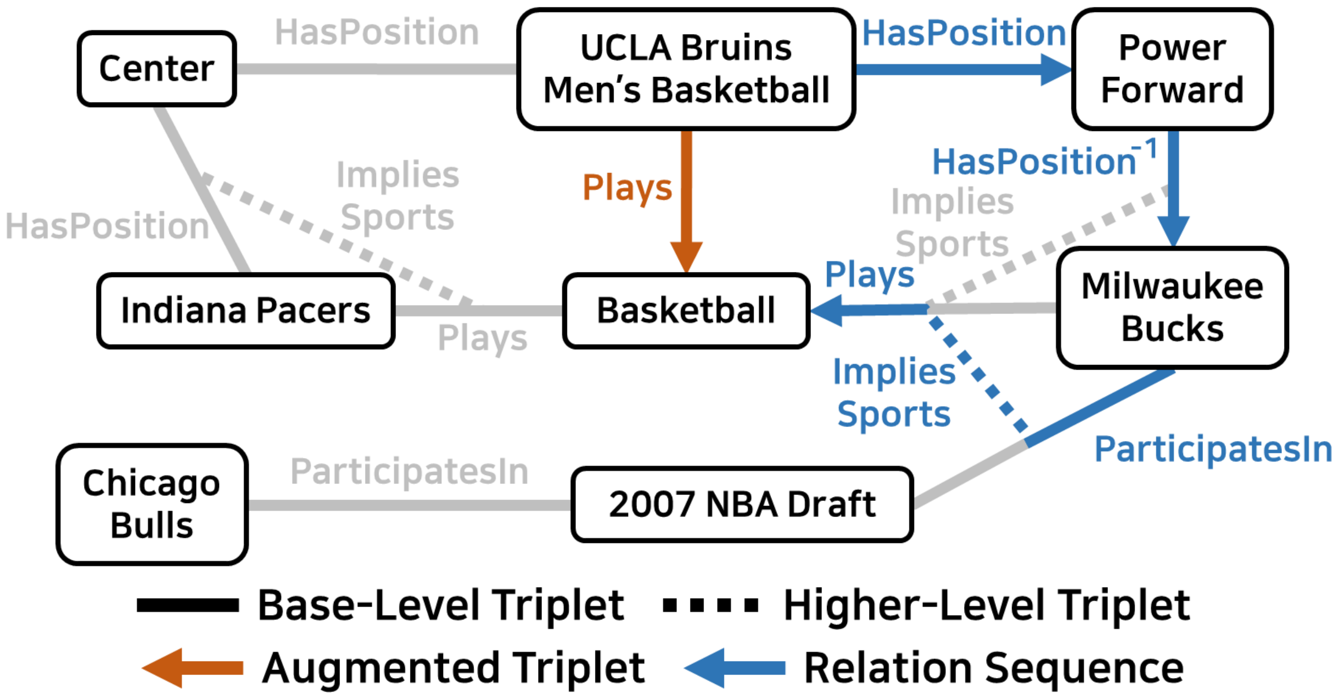
$$(h_0, \boldsymbol{r_0}, h_i, \boldsymbol{r_i}, \boldsymbol{\hat{r}}, \boldsymbol{r_j}, t_j) \rightarrow (r_0, r_i, \hat{r}, r_j)$$

Random Walk Path → Relation Sequence

- **Confidence Score** $c(p_k, r)$
  - Probability that the entity pair connected by p_k is also connected by r.

- Add the missing plausible triplets based on the confidence score.

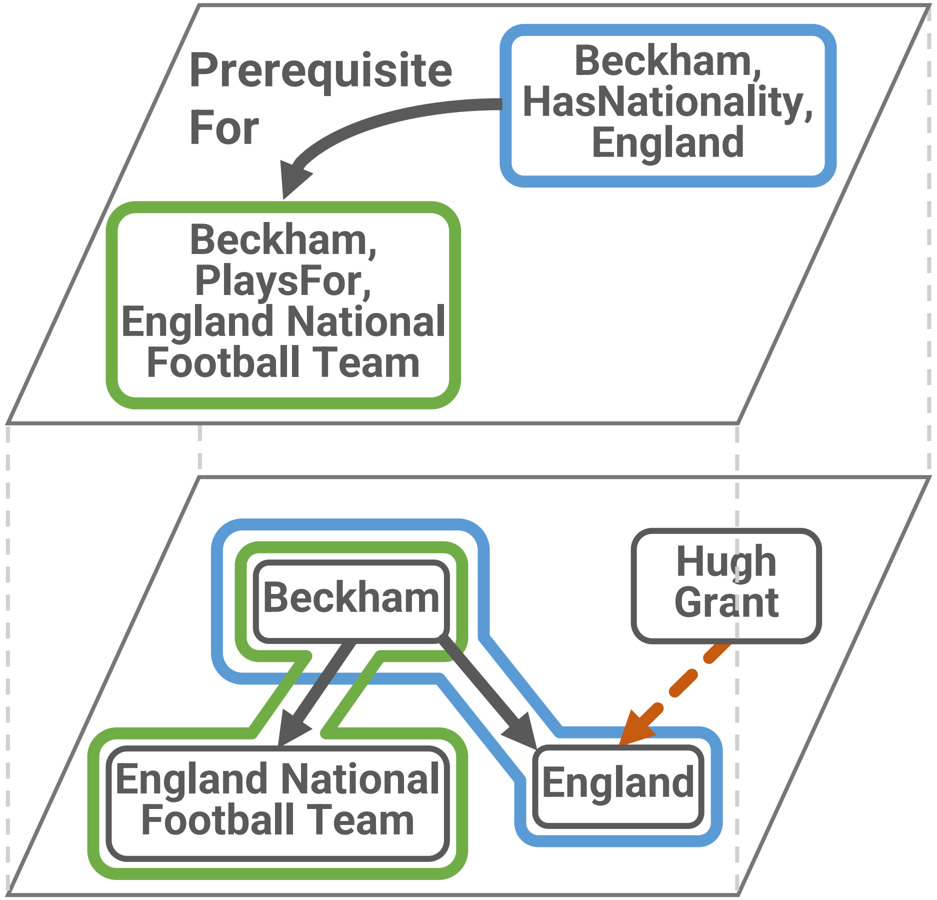
Example: Augmented Triplets



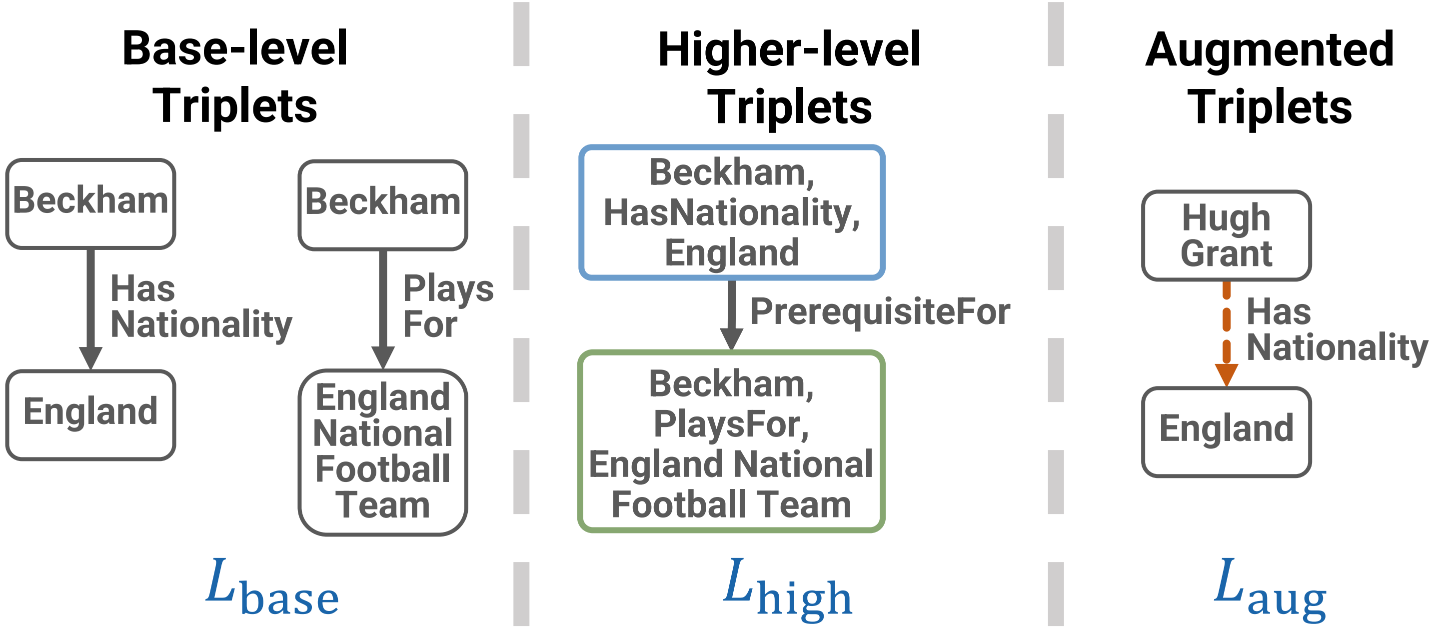
p_k: (HasPosition, Hasposition^{-1}, ParticipatesIn, ImpliesSports, Plays)

r: Plays

BiVE: Embedding of Bi-Level KGs

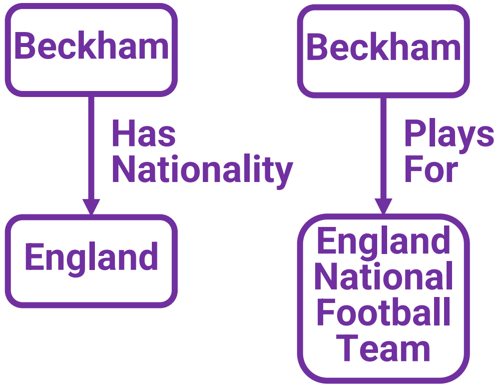


- BiVE considers three loss terms.



BiVE: Embedding of Bi-Level KGs

- Loss incurred by the **base-level triplets**
  - $L_{\text{base}} = \sum_{(h,r,t)\in\mathcal{E}_{\text{train}}} g\big(-f(\mathbf{h},\mathbf{r},\mathbf{t})\big) + \sum_{(h',r',t')\in\mathcal{E}'_{\text{train}}} g\big(f(\mathbf{h}',\mathbf{r}',\mathbf{t}')\big)$
    - $g(x) = \log(1+\exp(x))$
    - $\mathcal{E}'_{\text{train}}$ is a set of corrupted triplets.
    - We can use any knowledge graph embedding scoring function for $f(\cdot)$.

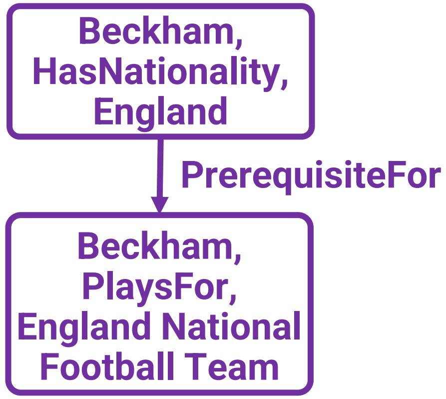


- We implement BiVE with two different scoring functions.
  - **BiVE-Q**: Uses scoring function of QuatE *
  - **BiVE-B**: Uses scoring function of BiQUE **

* QuatE: Zhang et al., Quaternion Knowledge Graph Embeddings, NeurIPS 2019
** BiQUE: Guo et al., BiQUE: Biquaternionic Embeddings of Knowledge Graphs, EMNLP 2021

BiVE: Embedding of Bi-Level KGs

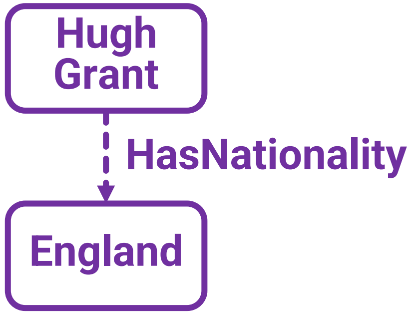
- Given $T_i = (h_i, r_i, t_i)$, let $\mathbf{T}_i$ denote an embedding vector of T_i.
  - Define $\mathbf{T}_i = \boldsymbol{W}[\mathbf{h}_i; \mathbf{r}_i; \mathbf{t}_i]$
    - $\boldsymbol{W}$ is a projection matrix.



- Loss incurred by the **higher-level triplets**
  - $L_{\text{high}} = \sum_{(T_i, \hat{r}, T_j) \in \mathcal{H}_{\text{train}}} g\left(-f\left(\mathbf{T}_i, \hat{\mathbf{r}}, \mathbf{T}_j\right)\right) + \sum_{\left(T_i', \hat{r}', T_j'\right) \in \mathcal{H}_{\text{train}}'} g\left(f\left(\mathbf{T}_i', \hat{\mathbf{r}}', \mathbf{T}_j'\right)\right)$
    - $\mathcal{H}_{\text{train}}'$ is a set of corrupted higher-level triplets.

BiVE: Embedding of Bi-Level KGs

- Loss incurred by the **augmented triplets**
  - $L_{\text{aug}} = \sum_{(h,r,t)\in\mathcal{S}} g\big(-f(\mathbf{h},\mathbf{r},\mathbf{t})\big) + \sum_{(h',r',t')\in\mathcal{S}'} g\big(f(\mathbf{h}',\mathbf{r}',\mathbf{t}')\big)$
    - $\mathcal{S}$ is a set of augmented triplets.
    - $\mathcal{S}'$ is a set of corrupted triplets.



- **Loss function of BiVE**
  - $L_{\text{BiVE}} = L_{\text{base}} + \lambda_1 \cdot L_{\text{high}} + \lambda_2 \cdot L_{\text{aug}}$
    - λ_1 and λ_2 control the importance of L_{high} and L_{aug}, respectively.

Scoring Functions of BiVE

- **Triplet Prediction**
  - Given an incomplete higher-level triplet $\langle T_i, \hat{r}, ?\rangle$,
    - BiVE computes $F_{\text{tp}}(X) = f(\mathbf{T}_i, \hat{\mathbf{r}}, \mathbf{X})$ for every base-level triplet $X \in \mathcal{E}_{\text{train}}$.
    - Rank base-level triplets based on the calculated scores.

- **Conditional Link Prediction**
  - Given an incomplete higher-level triplet $\langle T_i, \hat{r}, (h_j, r_j, ?)\rangle$,
    - BiVE compute $F_{\text{clp}}(x) = f(\mathbf{h}_j, \mathbf{r}_j, \mathbf{x}) + \lambda_1 \cdot f(\mathbf{T}_i, \hat{\mathbf{r}}, \boldsymbol{W}[\mathbf{h}_j; \mathbf{r}_j; \mathbf{x}])$ for every entity $\mathrm{x} \in \mathcal{V}$.
    - Rank entities based on the calculated scores.

Experimental Settings

- **Datasets**
  - Three real-world bi-level KGs: ***FBH, FBHE, DBHE***
  - Split $\mathcal{E}$ and $\mathcal{H}$ into training, validation, test sets with a ratio of 8:1:1.

	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\hat{\mathcal{R}}\rvert$	$\lvert\mathcal{H}\rvert$	$\lvert\hat{\mathcal{E}}\rvert$
FBH	14,541	237	310,117	6	27,062	33,157
FBHE	14,541	237	310,117	10	34,941	33,719
DBHE	12,440	87	68,296	8	6,717	8,206

$\lvert\hat{\mathcal{E}}\rvert$ is the number of base-level triplets involved in the higher-level triplets

Experimental Settings

- Baselines: **12 different state-of-the-art methods**
  - **String matching**: ASER
  - **Path finding**: MINERVA, Multi-Hop
  - **Rule or logic**: AnyBURL, Neural-LP, DRUM
  - **Relation path**: PTransE, RPJE
  - **Latent space**: TransD, ANALOGY, QuatE, BiQUE

- We repeat experiments ten times for each method.

Results of Triplet Prediction

	FBH			*FBHE*			*DBHE*		
	MR(↓)	MRR(↑)	Hit@10(↑)	MR(↓)	MRR(↑)	Hit@10(↑)	MR(↓)	MRR(↑)	Hit@10(↑)
ASER	74541.7	0.011	0.015	57916.0	0.050	0.070	18157.6	0.042	0.075
MINERVA	109055.1	0.093	0.113	85571.5	0.220	0.300	20764.3	0.177	0.221
Multi-Hop	108731.7	0.105	0.117	83643.8	0.255	0.311	20505.8	0.191	0.230
Neural-LP	115016.6	0.070	0.073	90000.4	0.238	0.274	21130.5	0.170	0.209
DRUM	115016.6	0.069	0.073	90000.3	0.261	0.274	21130.5	0.166	0.209
AnyBURL	108079.6	0.096	0.108	83136.8	0.191	0.252	20530.8	0.177	0.214
PTransE	111024.3	0.069	0.071	86793.2	0.249	0.274	18888.7	0.158	0.195
RPJE	113082.0	0.070	0.072	89173.1	0.267	0.274	20290.4	0.166	0.206
TransD	74277.3	0.052	0.104	52159.4	0.238	0.280	16698.1	0.116	0.189
ANALOGY	93383.4	0.072	0.107	60161.5	0.286	0.318	18880.0	0.150	0.211
QuatE	145603.8	0.103	0.114	94684.4	0.101	0.209	26485.0	0.157	0.179
BiQUE	81687.5	0.104	0.115	61015.2	0.135	0.205	19079.4	0.163	0.185
BiVE-Q	18.7	0.748	0.853	33.1	0.531	0.683	56.6	0.315	0.523
BiVE-B	19.7	0.731	0.837	27.9	0.555	0.718	4.7	0.644	0.914

Best baseline
Second-best
Best

Results of Conditional Link Prediction

	FBH			*FBHE*			*DBHE*		
	MR(↓)	MRR(↑)	Hit@10(↑)	MR(↓)	MRR(↑)	Hit@10(↑)	MR(↓)	MRR(↑)	Hit@10(↑)
ASER	1183.9	0.251	0.316	970.7	0.289	0.382	1893.5	0.225	0.348
MINERVA	3817.8	0.328	0.415	3018.5	0.407	0.492	2934.1	0.362	0.433
Multi-Hop	1878.2	0.421	0.578	1447.3	0.443	0.615	1012.3	0.442	0.652
Neural-LP	185.9	0.433	0.648	146.2	0.466	0.716	32.2	0.517	0.756
DRUM	262.7	0.394	0.555	207.6	0.413	0.620	49.0	0.470	0.732
AnyBURL	228.5	0.380	0.563	166.0	0.418	0.607	81.7	0.403	0.594
PTransE	214.8	0.440	0.686	167.0	0.516	0.752	19.3	0.505	0.780
RPJE	212.5	0.440	0.686	159.0	0.528	0.753	19.3	0.504	0.779
TransD	190.1	0.300	0.496	165.6	0.363	0.529	35.5	0.436	0.708
ANALOGY	341.0	0.182	0.291	113.3	0.409	0.581	279.1	0.140	0.253
QuatE	163.7	0.346	0.494	1546.4	0.124	0.189	551.6	0.208	0.309
BiQUE	111.0	0.423	0.641	90.1	0.387	0.617	29.5	0.378	0.677
BiVE-Q	7.0	0.752	0.906	11.0	0.698	0.839	12.5	0.606	0.828
BiVE-B	6.6	0.762	0.911	12.8	0.696	0.834	3.2	0.801	0.958

Best baseline
Second-best
Best

Example: Conditional Link Prediction

- Example 1
  - ⟨(Joe Jonas, IsA, **?**), ImpliesProfession, (Joe Jonas, IsA, Actor)⟩
    - Answer: **Voice Actor**
  - ⟨(Joe Jonas, IsA, **?**), ImpliesProfession, (Joe Jonas, IsA, Musician)⟩
    - Answer: **Singer-songwriter**
- Example 2
  - ⟨(Saturn Award, Nominates, Avatar), PrerequisiteFor, (Avatar, Wins, **?**)⟩
    - Answer: **Saturn Award**
  - ⟨(Academy Awards, Nominates, Avatar), PrerequisiteFor, (Avatar, Wins, **?**)⟩
    - Answer: **Academy Awards**

Results of Base-Level Link Prediction

- **Our BiVE models show comparable results to the baseline methods.**

- BiVE models have the extra capability of dealing with the **Triplet Prediction** and **Conditional Link Prediction** tasks.

	FBH Hit@10(↑)	*FBHE* Hit@10(↑)	*DBHE* Hit@10(↑)
ASER	0.323	0.323	0.197
MINERVA	0.339	0.339	0.297
Multi-Hop	0.500	0.500	0.404
Neural-LP	0.486	0.486	0.357
DRUM	0.490	0.490	0.359
AnyBURL	0.526	0.526	0.364
PTransE	0.333	0.333	0.277
RPJE	0.368	0.368	0.341
TransD	0.527	0.527	0.423
ANALOGY	0.486	0.486	0.323
QuatE	0.581	0.581	0.440
BiQUE	0.583	0.583	0.446
BiVE-Q	0.584	0.584	0.438
BiVE-B	0.586	0.586	0.444

Second-best

Best

Example: Augmented Triplets

- **Our augmented triplets include many ground-truth triplets that were missing in the training set.**

Relation Sequence p_k	Relation r	$c(p_k, r)$	Examples of the Augmented Triplets
Plays, Plays^{-1}, **ImpliesSports**, HasPosition	HasPosition	0.78	(Bayer 04 Leverkusen, HasPosition, Forward)
IsPartOf, IsPartOf, **ImpliesLocation**, IsPartOf	IsPartOf	0.76	(San Pedro, IsPartOf, California)

$c(p_k, r)$ is the confidence score of the pair (p_k, r)

	FBH	*FBHE*	*DBHE*
No. of augmented triplets	16,601	17,463	2,026
$\lvert\mathcal{S} \cap \mathcal{E}_{\text{valid}}\rvert + \lvert\mathcal{S} \cap \mathcal{E}_{\text{test}}\rvert$	5,237	5,380	316

No. of augmented triplets contained in either $\mathcal{E}_{\text{valid}}$ or $\mathcal{E}_{\text{test}}$

Conclusion

- Define a **Bi-Level Knowledge Graph**
- Propose a **Random-Walk based Data Augmentation Strategy** on bi-level KGs
- **BiVE** takes into account the structures of the base-level triplets, the higher-level triplets and the augmented triplets.
- BiVE significantly outperforms state-of-the-art methods in terms of the two newly defined tasks: **Triplet Prediction** and **Conditional Link Prediction**
- Our method can contribute to advancing many knowledge-based applications, including **Conditional QA** and **Multi-Hop QA**.

Thanks!

Our datasets and codes are available at:

https://github.com/bdi-lab/BiVE

You can find us at:

{chanyoung.chung, jjwhang}@kaist.ac.kr

https://bdi-lab.kaist.ac.kr

BDILab
BIG DATA INTELLIGENCE

Differences between BiVE and Rule-based Methods

- **Rule-based Methods**
  - Only consider the relationships between entities
  - Consider the first-order-logic-like rules between connected entities
    - e.g., $\forall x, y, z: (x, r_1, y) \wedge (y, r_2, z) \Rightarrow (x, r_3, z)$
- **BiVE**
  - Considers the relationships between entities and the **relationships between triplets**
  - Entities are **not necessarily connected** by the base-level triplets, and the patterns are **not restricted to the first-order-logic-like formula**
    - e.g., $(x, r_1, y) \overset{\hat{r}}{\Rightarrow} (p, r_2, q)$