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

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

BDILab BIG DATA INTELLIGENCE

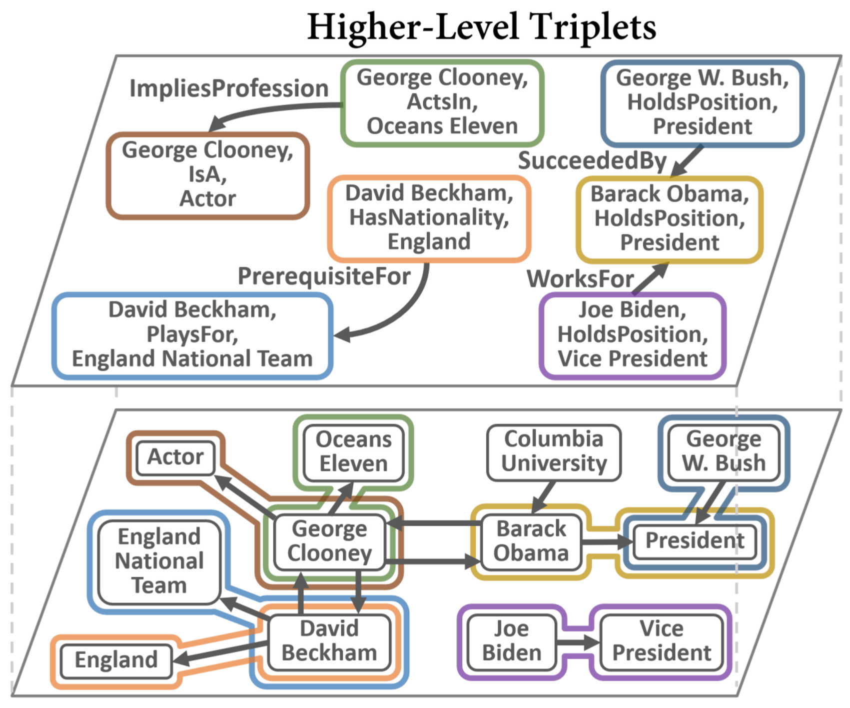
## Main Contributions

- Define a **Bi-Level Knowledge Graph**, which represents **the relationships between triplets** as well as the relationships between entities.
  - Create three real-world bi-level knowledge graphs: ***FBH***, ***FBHE***, ***DBHE***
- Propose a random-walk-based **data augmentation** strategy on Bi-level KGs.
- **BiVE** (embedding of **Bi**-le**V**el knowledg**E** graphs): embedding model which takes into account **base-level**, **higher-level** and **augmented** triplets.
- Propose two new tasks: **Triplet Prediction** and **Conditional Link Prediction**

## Bi-Level Knowledge Graph

- There can be meaningful relationships between triplets.
  - $T_1$: (Biden, HoldsPosition, Vice President), $T_2$: (Obama, HoldsPosition, President)
- **Base-Level Knowledge Graph**
  - $G = (\mathcal{V}, \mathcal{R}, \mathcal{E})$
  - $\mathcal{V}$: a set of entities
  - $\mathcal{R}$: a set of relations
  - $\mathcal{E}$: a set of base-level triplets
- **Higher-Level Triplets** $\mathcal{H}$
  - Relationships between triplets
  - e.g., $\langle T_1$, **WorksFor**, $T_2\rangle$
- **Bi-Level Knowledge Graph**
  - $\hat{G} = (\mathcal{V}, \mathcal{R}, \mathcal{E}, \hat{\mathcal{R}}, \mathcal{H})$
  - $\hat{\mathcal{R}}$: a set of **higher-level relations**
    - e.g., WorksFor

Higher-Level Triplets




Base-Level Triplets

## Real-World Bi-Level Knowledge Graphs

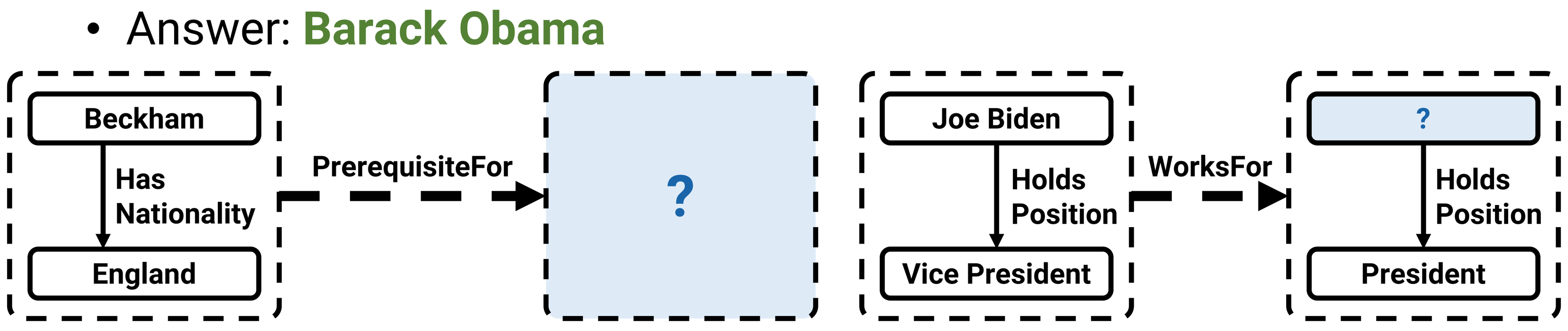
- Create three real-world bi-level knowledge graphs: ***FBH***, ***FBHE***, ***DBHE***
  - *FBH* and *FBHE* are based on ***FB15K237*** and *DBHE* is based on ***DB15K***.

| | $\hat{r}$ | $\langle T_i, \hat{r}, T_j\rangle$ |
|---|---|---|
| *FBHE* | PrerequisiteFor | $T_i$: (Beckham, HasNationality, England)<br>$T_j$: (Beckham, PlaysFor, England National Team) |
| | WorksFor | $T_i$: (Joe Biden, HoldsPosition, Vice President)<br>$T_j$: (Barack Obama, HoldsPosition, President) |

**Externally-sourced Knowledge**

## Triplet Prediction and Conditional Link Prediction

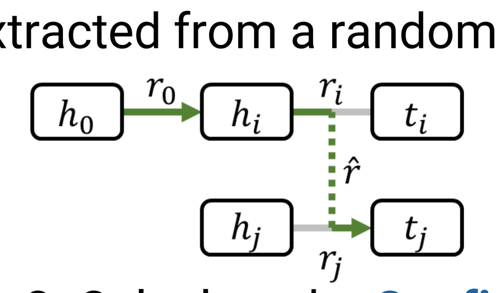
- **Triplet Prediction**
  - Predicts **a triplet** connected to a given triplet by a higher-level relation.
  - ⟨(Beckham, HasNationality, England), PrerequisiteFor, **?**⟩
    - Answer: **(Beckham, PlaysFor, England National Team)**
- **Conditional Link Prediction**
  - Predicts **a missing entity in a triplet** conditioned by another triplet.
  - ⟨(Joe Biden, HoldsPosition, Vice President), WorksFor, (**?**, HoldsPosition, President)⟩
    - Answer: **Barack Obama**




**Triplet Prediction** **Conditional Link Prediction**

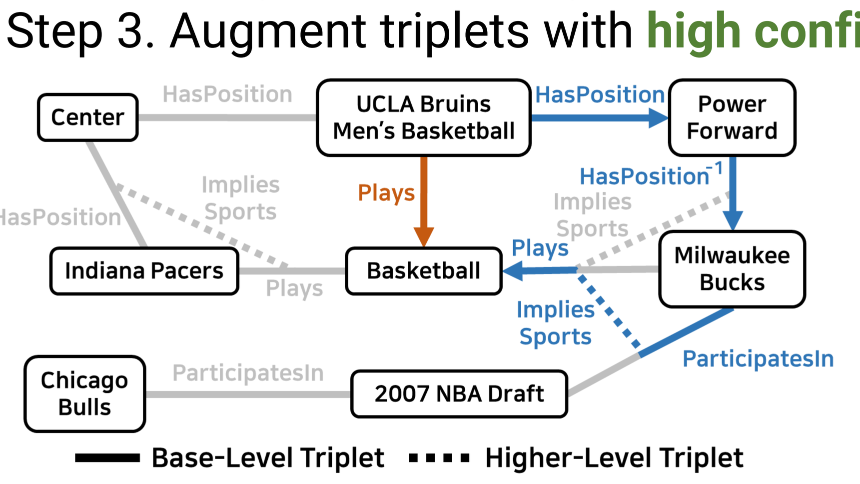
## Data Augmentation Strategy based on Random Walks

- Step 1. Perform **Random Walks** by following base-level/higher-level triplets.
  - **Random Walk Path**: Sequence of entities, relations, higher-level relations
  - **Relation Sequence** $p_k$: Sequence of relations and higher-level relations extracted from a random walk path




Random walk path $w = h_0, r_0, h_i, r_i, \hat{r}, r_j, t_j$
Relation sequence $p_k = r_0, r_i, \hat{r}, r_j$
$(h_0, r_0, h_i, r_i, \hat{r}, r_j, t_j) \rightarrow (h_0, p_k, t_j)$

- Step 2. Calculate the **Confidence Score** $c(p_k, r)$.
  - Probability that the entity pair connected by $p_k$ is also connected by $r$
- Step 3. Augment triplets with **high confidence scores**.




**Relation sequence** $p_k$:
(HasPosition, HasPosition$^{-1}$, ParticipatesIn, ImpliesSports, Plays)

**Relation** $r$: Plays

**Augmented triplet:**
(UCLA Bruins Men's Basketball, Plays, Basketball)

## BiVE: embedding of Bi-leVel knowledgE graphs

- Loss incurred by the **base-level triplets** ($L_{\text{base}}$)
  - $L_{\text{base}} = \sum_{(h,r,t)\in\mathcal{E}_{\text{train}}} g(-f(\mathbf{h},\mathbf{r},\mathbf{t})) + \sum_{(h',r',t')\in\mathcal{E}'_{\text{train}}} g(f(\mathbf{h}',\mathbf{r}',\mathbf{t}'))$
    - $g(x) = \log(1+\exp(x))$
  - We can use any knowledge graph embedding scoring function for $f(\cdot)$.
    - **BiVE-Q**: uses the scoring function of QuatE
    - **BiVE-B**: uses the scoring function of BiQUE
- Loss incurred by the **higher-level triplets** ($L_{\text{high}}$)
  - $L_{\text{high}} = \sum_{(T_i,\hat{r},T_j)\in\mathcal{H}_{\text{train}}} g(-f(\mathbf{T}_i,\hat{\mathbf{r}},\mathbf{T}_j)) + \sum_{(T'_i,\hat{r}',T'_j)\in\mathcal{H}'_{\text{train}}} g(f(\mathbf{T}'_i,\hat{\mathbf{r}}',\mathbf{T}'_j))$
    - $\mathbf{T}_i = \boldsymbol{W}[\mathbf{h}_i;\mathbf{r}_i;\mathbf{t}_i]$ is the embedding vector of $T_i = (h_i, r_i, t_i)$.
- Loss incurred by the **augmented triplets** ($L_{\text{aug}}$)
  - $L_{\text{aug}} = \sum_{(h,r,t)\in\mathcal{S}} g(-f(\mathbf{h},\mathbf{r},\mathbf{t})) + \sum_{(h',r',t')\in\mathcal{S}'} g(f(\mathbf{h}',\mathbf{r}',\mathbf{t}'))$
    - $\mathcal{S}$ is the set of augmented triplets.
- **Loss Function of BiVE**: $L_{\text{BiVE}} = L_{\text{base}} + \lambda_1 \cdot L_{\text{high}} + \lambda_2 \cdot L_{\text{aug}}$
- To solve **a triplet prediction problem** $\langle T_i, \hat{r}, ?\rangle$, compute $F_{\text{tp}} = f(\mathbf{T}_i, \hat{\mathbf{r}}, \mathbf{X})$ for every base-level triplet $X \in \mathcal{E}_{\text{train}}$.
- To solve **a conditional link prediction problem** $\langle T_i, \hat{r}, (h_j, r_j, ?)\rangle$, compute $F_{\text{clp}} = f(\mathbf{h}_j, \mathbf{r}_j, \mathbf{x}) + \lambda_1 \cdot f(\mathbf{T}_i, \hat{\mathbf{r}}, \boldsymbol{W}[\mathbf{h}_j;\mathbf{r}_j;\mathbf{x}])$ for every $x \in \mathcal{V}$.

## Experimental Settings

- Statistics of the datasets used in the experiments

| | $\lvert\mathcal{V}\rvert$ | $\lvert\mathcal{R}\rvert$ | $\lvert\mathcal{E}\rvert$ | $\lvert\hat{\mathcal{R}}\rvert$ | $\lvert\mathcal{H}\rvert$ | $\lvert\hat{\mathcal{E}}\rvert$ |
|---|---|---|---|---|---|---|
| *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 |

No. of base-level triplets involved in the higher-level triplets

- **12 baseline methods:** ASER, MINERVA, Multi-Hop, AnyBURL, Neural-LP, DRUM, PTransE, RPJE, TransD, ANALOGY, QuatE, BiQUE

## Experimental Results

- **Results of Triplet Prediction (TP)**

| | *FBH* | | *FBHE* | | *DBHE* | |
|---|---|---|---|---|---|---|
| | MR | Hit@10 | MR | Hit@10 | MR | Hit@10 |
| Best-baseline | 74277.3 | 0.117 | 52159.4 | 0.318 | 16698.1 | 0.230 |
| BiVE-Q | **18.7** | **0.853** | 33.1 | 0.683 | 56.6 | 0.523 |
| BiVE-B | 19.7 | 0.837 | **27.9** | **0.718** | **4.7** | **0.914** |

- **Results of Conditional Link Prediction (CLP)**

| | *FBH* | | *FBHE* | | *DBHE* | |
|---|---|---|---|---|---|---|
| | MR | Hit@10 | MR | Hit@10 | MR | Hit@10 |
| Best-baseline | 111.0 | 0.686 | 90.1 | 0.753 | 19.3 | 0.780 |
| BiVE-Q | 7.0 | 0.906 | **11.0** | **0.839** | 12.5 | 0.828 |
| BiVE-B | **6.6** | **0.911** | 12.8 | 0.834 | **3.2** | **0.958** |

- BiVE outperforms 12 different baseline methods on both **TP** and **CLP**.
- BiVE shows comparable base-level link prediction results to baselines.
- **Examples of Conditional Link Prediction**
  - ⟨(Joe Jonas, IsA, **?**), ImpliesProfession, (Joe Jonas, IsA, Actor)⟩
    - Prediction made by BiVE-Q: **Voice Actor**
  - ⟨(Joe Jonas, IsA, **?**), ImpliesProfession, (Joe Jonas, IsA, Musician)⟩
    - Prediction made by BiVE-Q: **Singer-songwriter**

## Analysis on the Augmented Triplets

- Examples of the augmented triplets in *FBHE* and *DBHE*

| 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) |
| Program$^{-1}$, Program, Language | Language | 0.70 | (David Copperfield, Language, English) |
| IsPartOf, IsPartOf, **ImpliesLocation**, IsPartOf | IsPartOf | 0.76 | (San Pedro, IsPartOf, California) |
| IsProducedBy$^{-1}$, IsProducedBy, **ImpliesProfession**, IsA | IsA | 0.73 | (Jim Wilson, IsA, Film Producer) |

- Statistics of the Augmented Triplets

| | *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 the augmented triplets contained in either $\mathcal{E}_{\text{valid}}$ or $\mathcal{E}_{\text{test}}$

- Our augmented triplets include many **ground-truth** triplets that are missing.

## Conclusion & Future Work

- We define a **bi-level knowledge graph** by introducing the higher-level relations between triplets, and **BiVE** successfully incorporates the structures of the **base-level**, the **higher-level** and the **augmented** triplets.
- Our method can contribute to advancing many knowledge-based applications, including **Conditional QA** and **Multi-Hop QA**.