Lecture#5: Representation Learning on Hyper-relational Knowledge Graphs

Joyce Jiyoung Whang

Department of AI Computing, KAIST

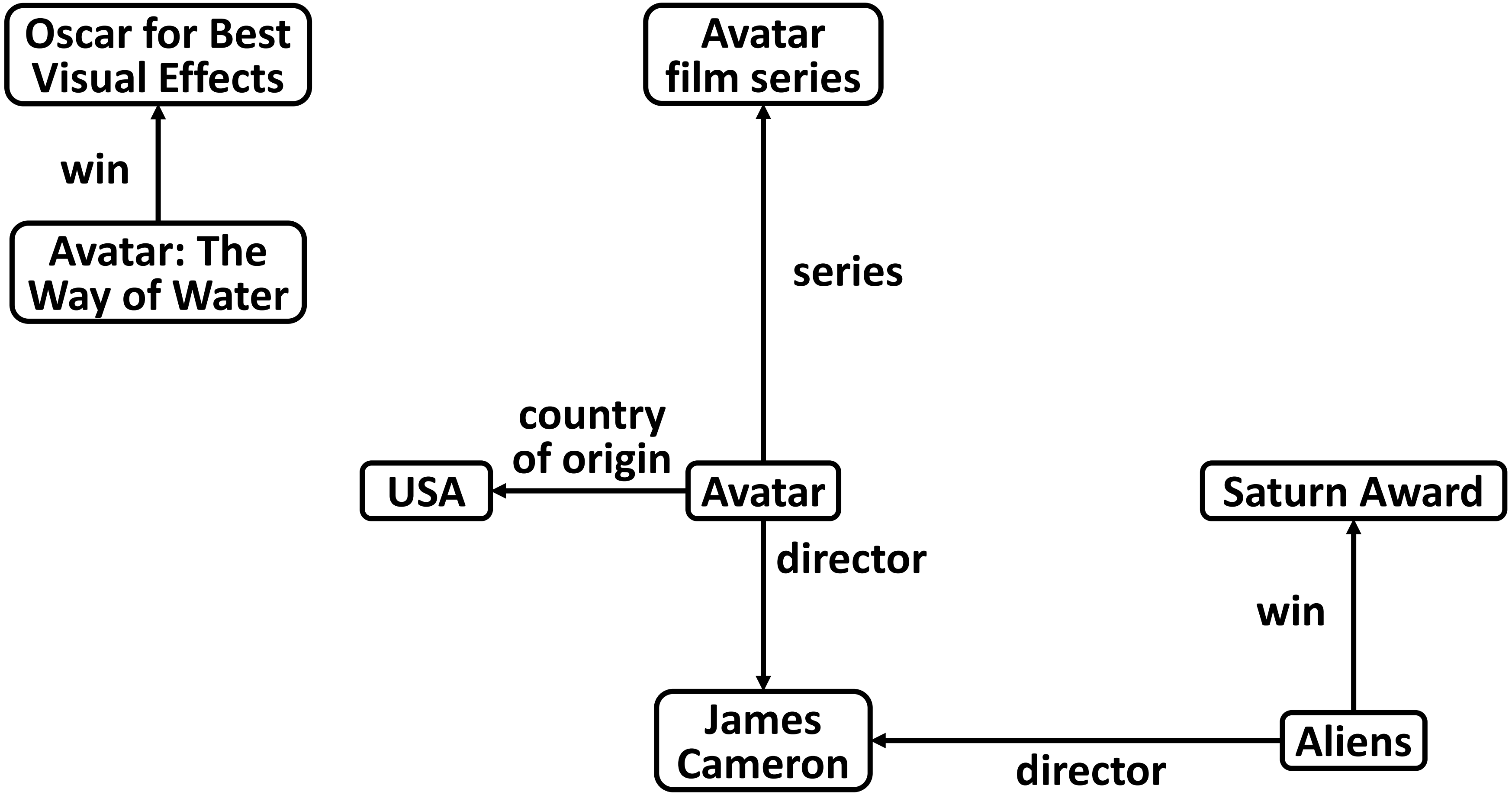
Knowledge Discovery with Knowledge Graphs: From Structural Embeddings to Generative Reasoning (KnowKG), KDD 2026 Tutorial

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

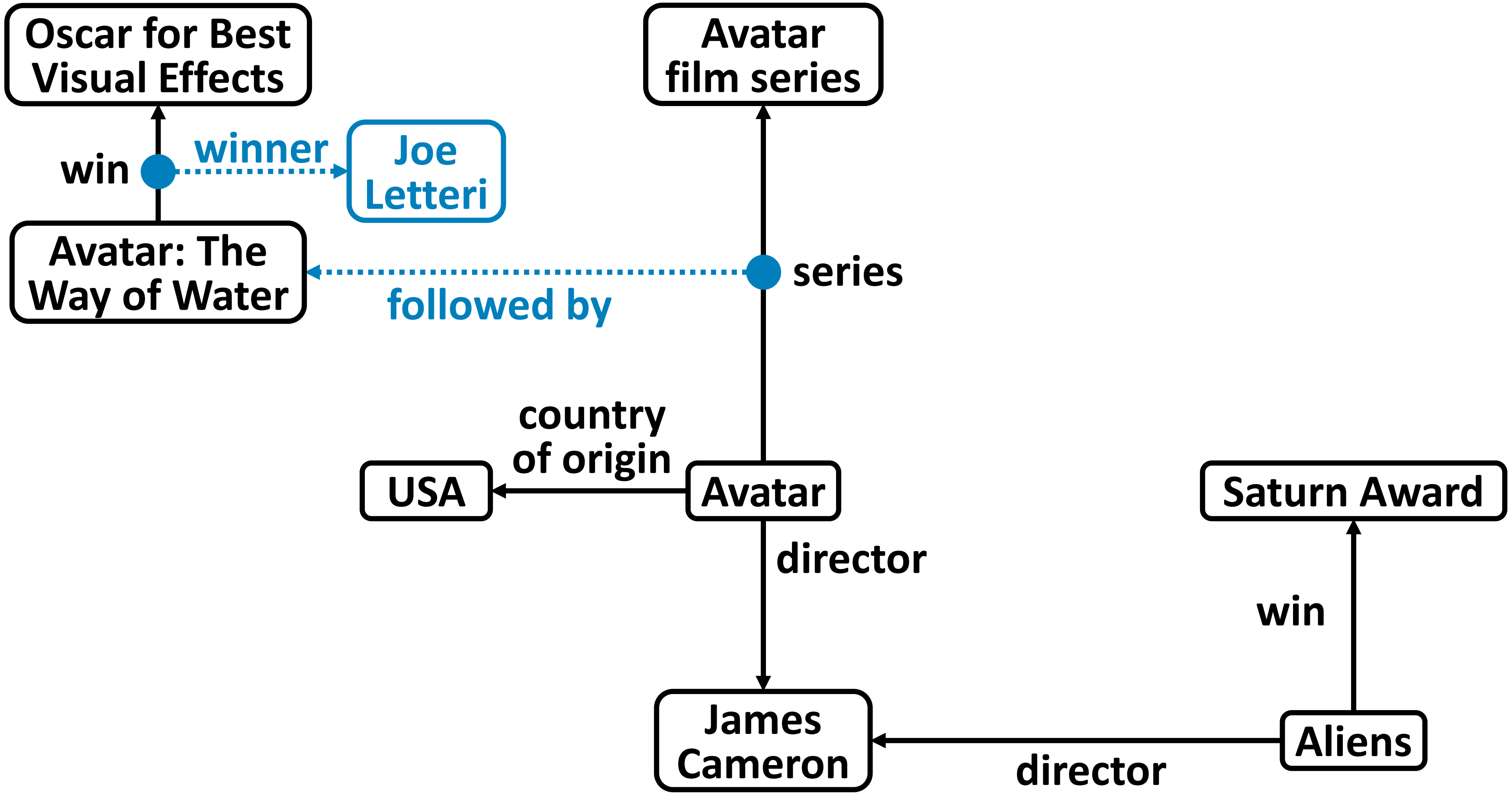
KAIST BDILab

BIG DATA INTELLIGENCE

01 Knowledge Graphs



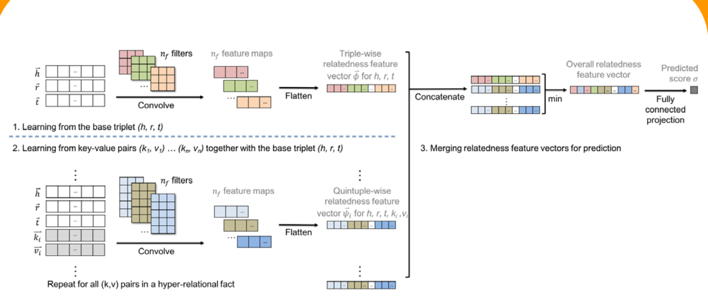
01 Hyper-relational Knowledge Graphs (HKGs)



01 Representation Learning on HKGs

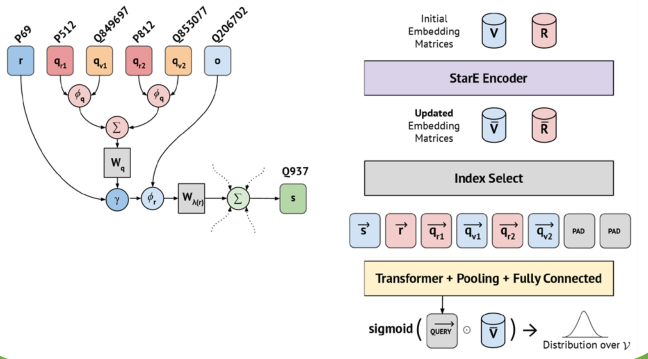
HINGE (TheWebConf 2020)

- A hyper-relational KG embedding model that directly learns from hyper-relational facts in a KG
- Captures the structural information of the KG encoded in the triplets
  - Also captures the correlation between each triplet and its associated qualifiers



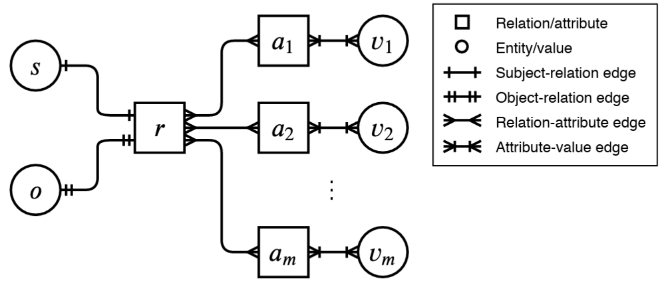
StarE (EMNLP 2020)

- A message passing based graph encoder that is capable of modeling hyper-relational KGs
- Can encode qualifiers along with the main triplet
- Demonstrates that existing benchmarks for evaluating LP on HKGs suffer from flaws



GRAN (ACL 2021 Findings)

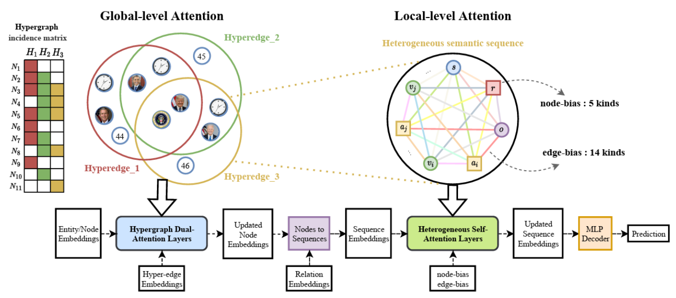
- Represents the structure of a hyper-relational fact as a small heterogeneous graph
  - Models the heterogeneous graph with edge-biased fully-connected attention
- Can fully model global and local dependencies in each fact



01 Representation Learning on HKGs

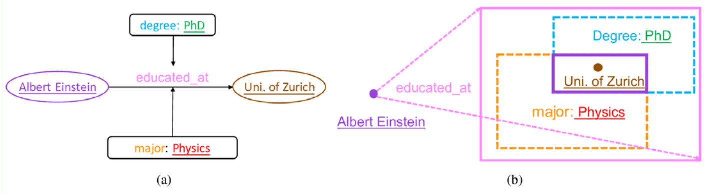
HAHE (ACL 2023)

- Represents the global structure of HKG as a hypergraph and the local structure as a semantic sequence
- Separately models the graphical structure of HKGs and the sequential structure inside facts
- Performs multi-position prediction in hyper-relational KGs



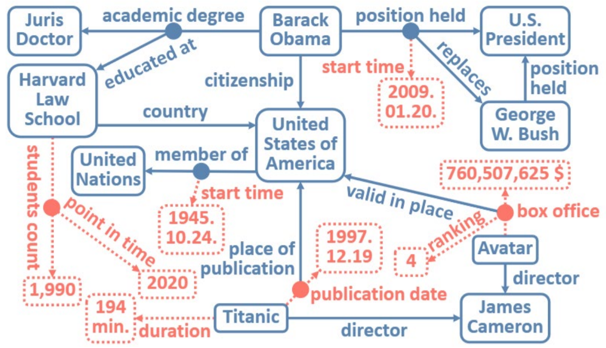
ShrinkE (ACL 2023)

- A geometric HKG embedding method aiming to explicitly model essential inference patterns of facts
- Models a primary triplet as a spatial-functional transformation from the head into a relation-specific box
  - Each qualifier shrinks the box to narrow down the answer set



HyNT (KDD 2023)

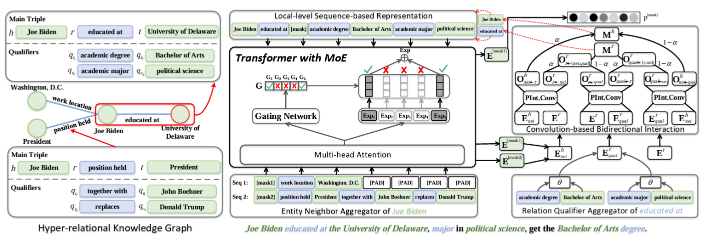
- A unified framework that learns representations of a HKG containing numeric literals in triplets/qualifiers
- Reduces the cost of transformers by learning compact representations of triplets and qualifiers
- Predicts missing entities, relations, and numeric values in a HKG



01 Representation Learning on HKGs

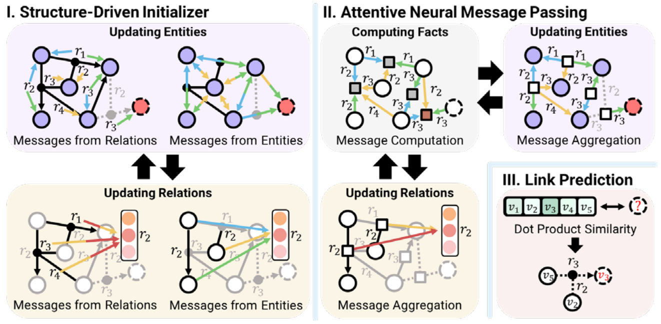
HyperFormer (CIKM 2023)

- A model that considers local-level sequential information
  - Encodes the content of the entities, relations, and qualifiers of a triplet
- Introduces a Mixture-of-Experts strategy to strengthen the representation capabilities



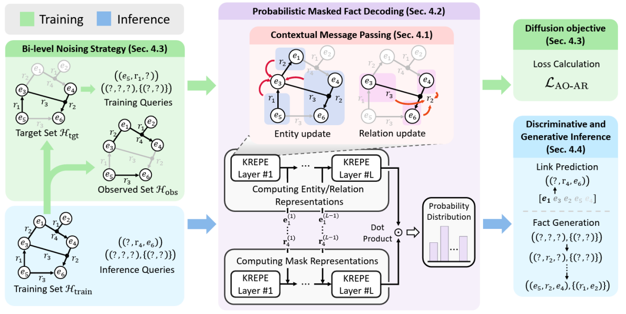
MAYPL (ICML 2025)

- Demonstrates that thoroughly leveraging the structure of an HKG is crucial for reasoning on HKGs
- The first structural representation learning method for HKGs that can be applied in both transductive and inductive learning settings



KREPE (ICML 2026)

- Proposes fact generation, a task of generating a valid fact from an arbitrarily masked query
- The first generative representation learning method for HKGs
  - models the joint conditional probability distributions of missing components in a query

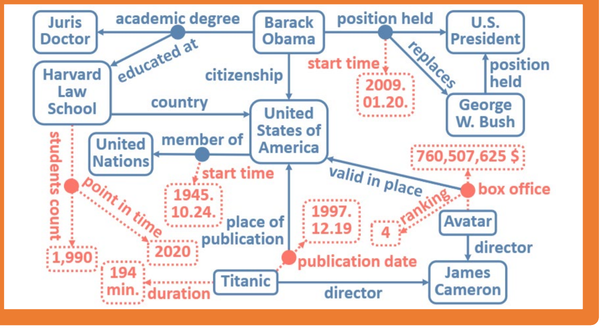


I

Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers

Chanyoung Chung‡, Jaejun Lee‡, and Joyce Jiyoung Whang*

KDD 2023

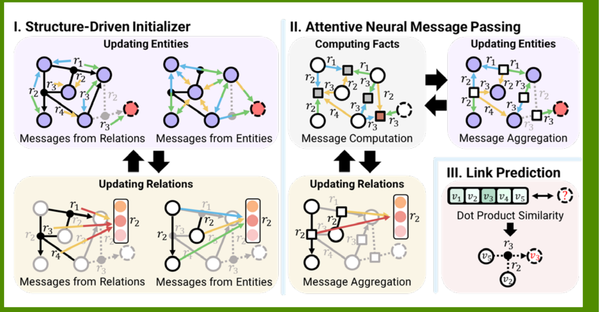


II

Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs

Jaejun Lee and Joyce Jiyoung Whang*

ICML 2025

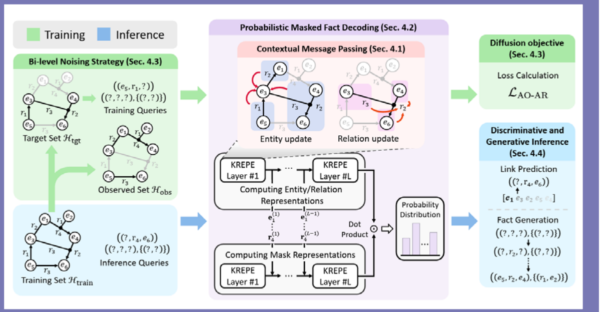


III

Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion

Jaejun Lee, Seheon Kim, and Joyce Jiyoung Whang*

ICML 2026

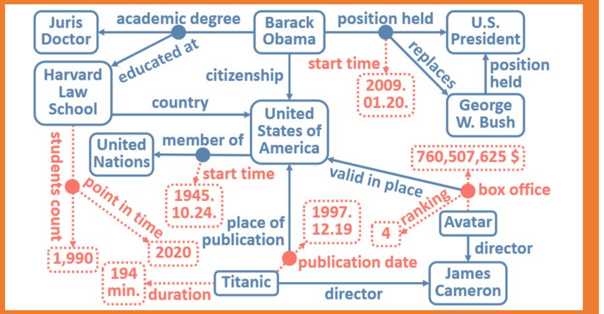


I

Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers

Chanyoung Chung‡, Jaejun Lee‡, and Joyce Jiyoung Whang*

KDD 2023

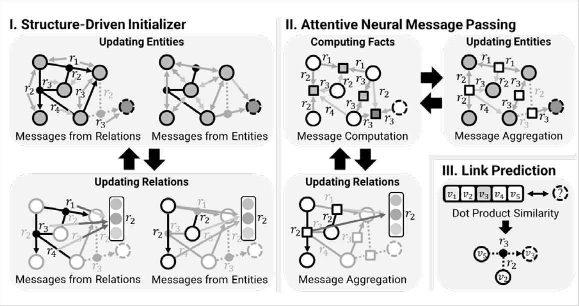


II

Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs

Jaejun Lee and Joyce Jiyoung Whang*

ICML 2025

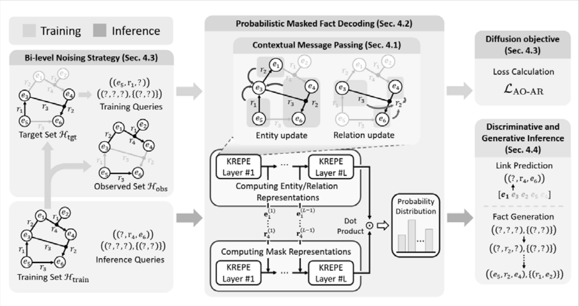


III

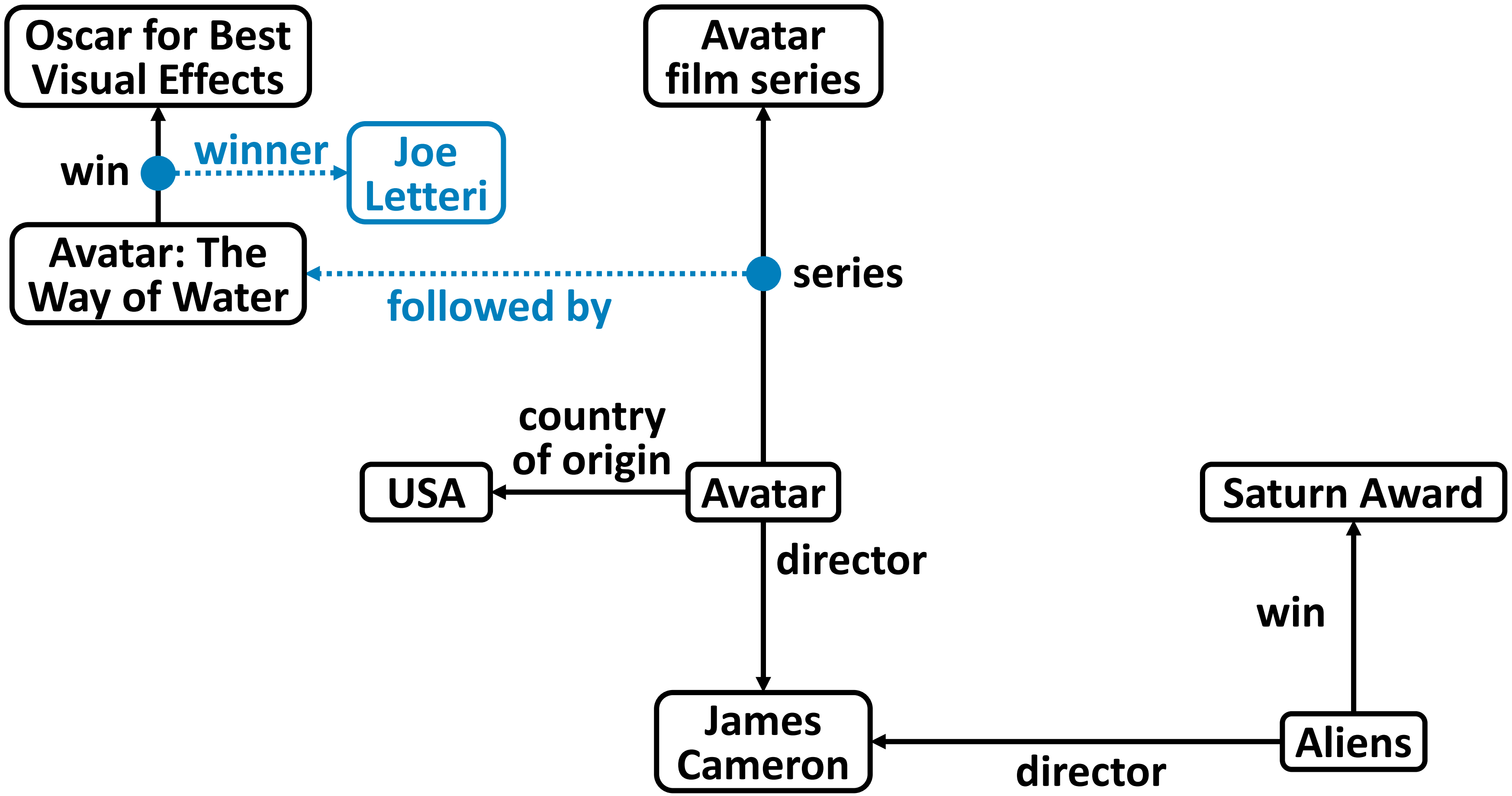
Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion

Jaejun Lee, Seheon Kim, and Joyce Jiyoung Whang*

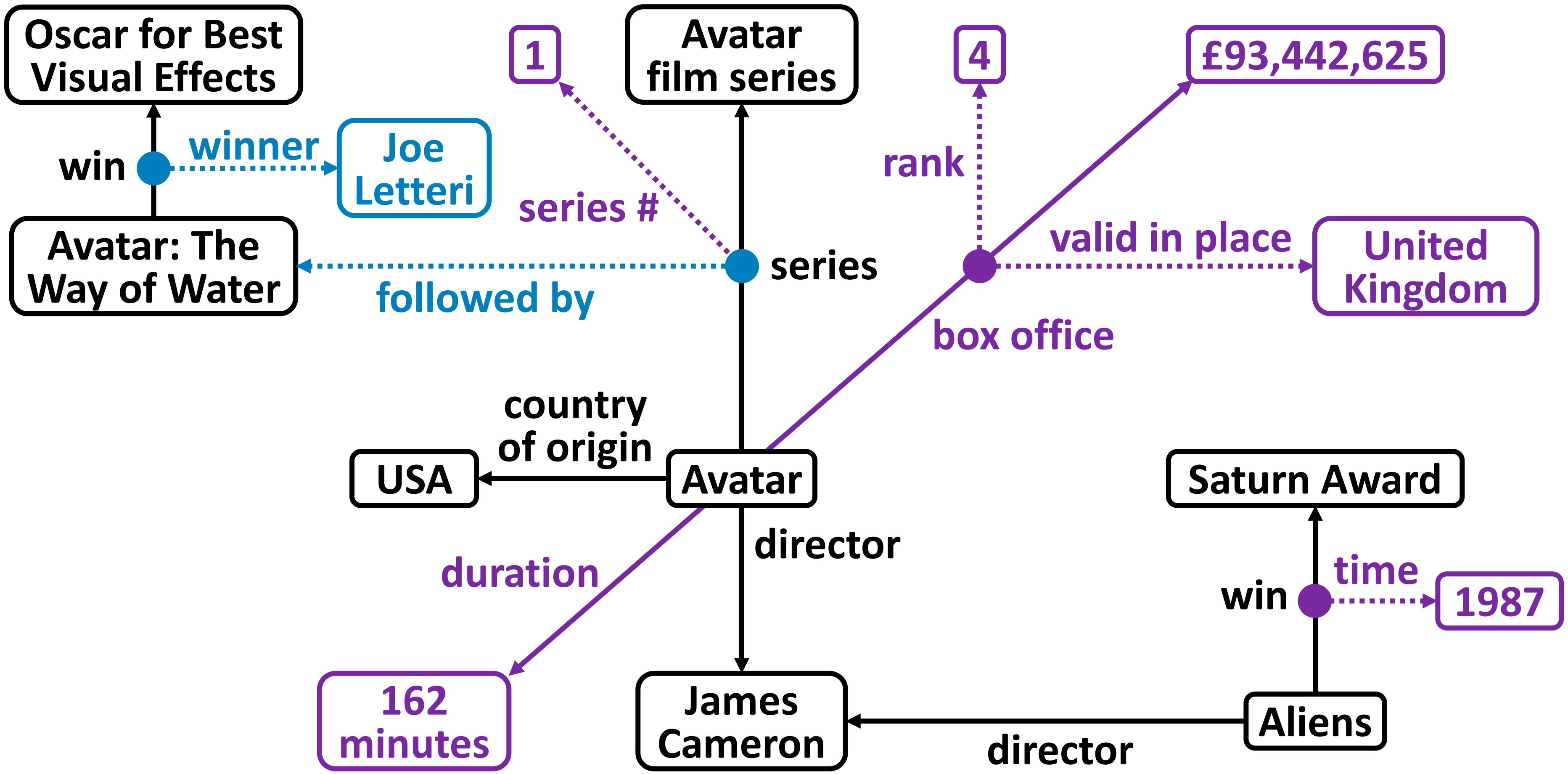
ICML 2026



02 Hyper-relational Knowledge Graphs



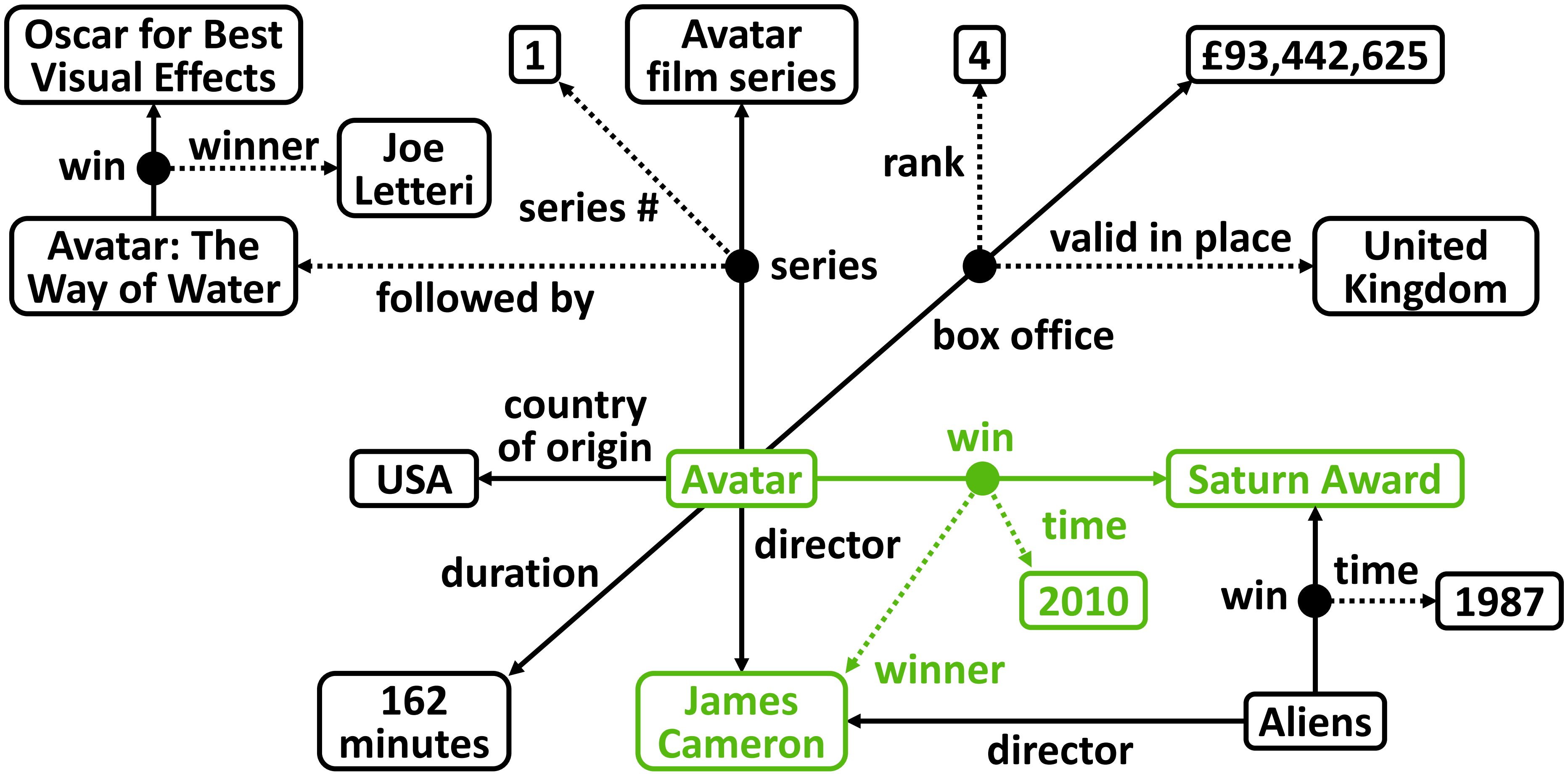
02 Hyper-relational and Numeric Knowledge Graphs



02 Hyper-relational and Numeric Knowledge Graphs

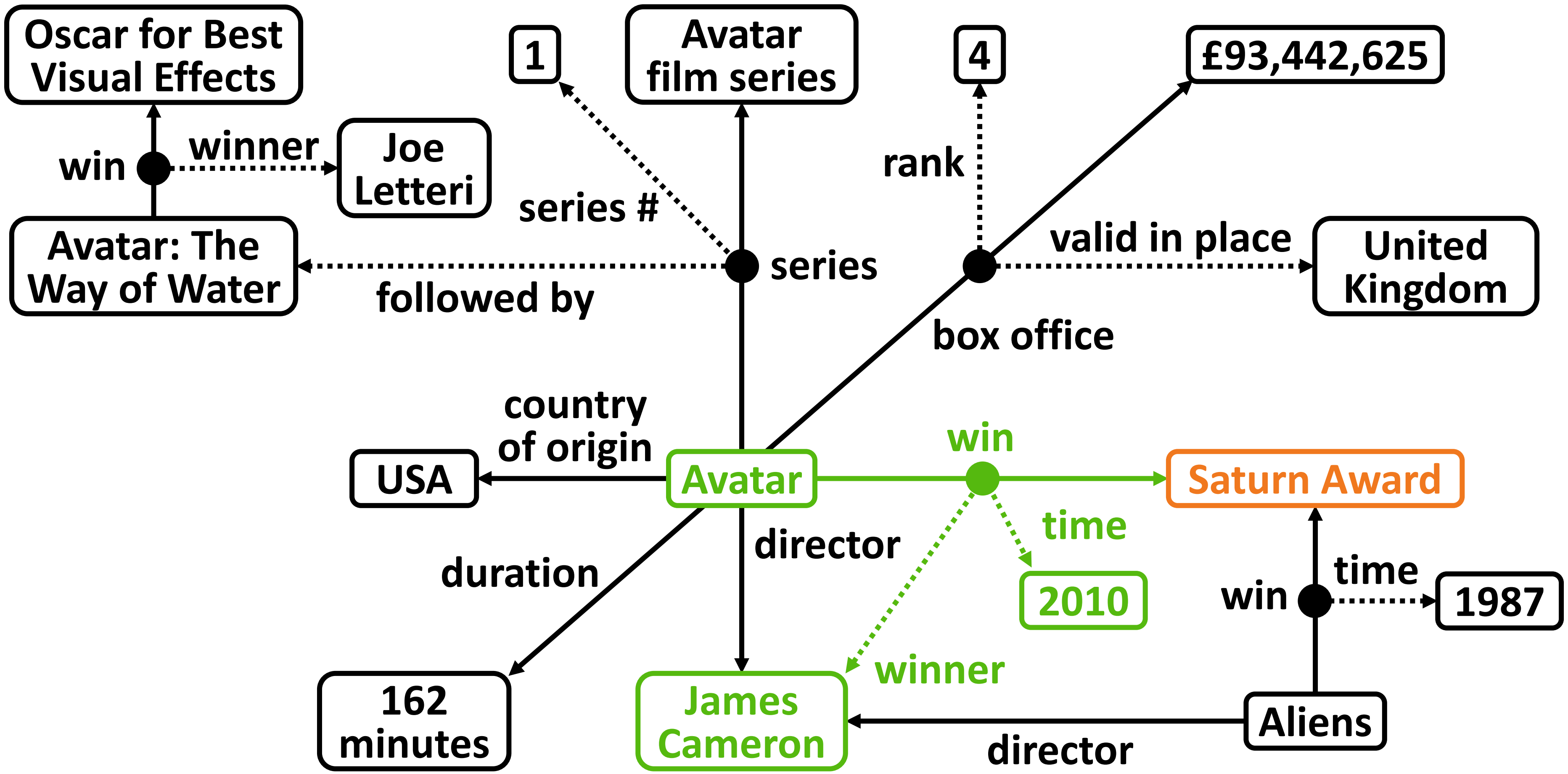
((Avatar, win, Saturn_Award), {(winner, James_Cameron), (time, 2010)})

Primary Triplet — Qualifier 1 — Qualifier 2



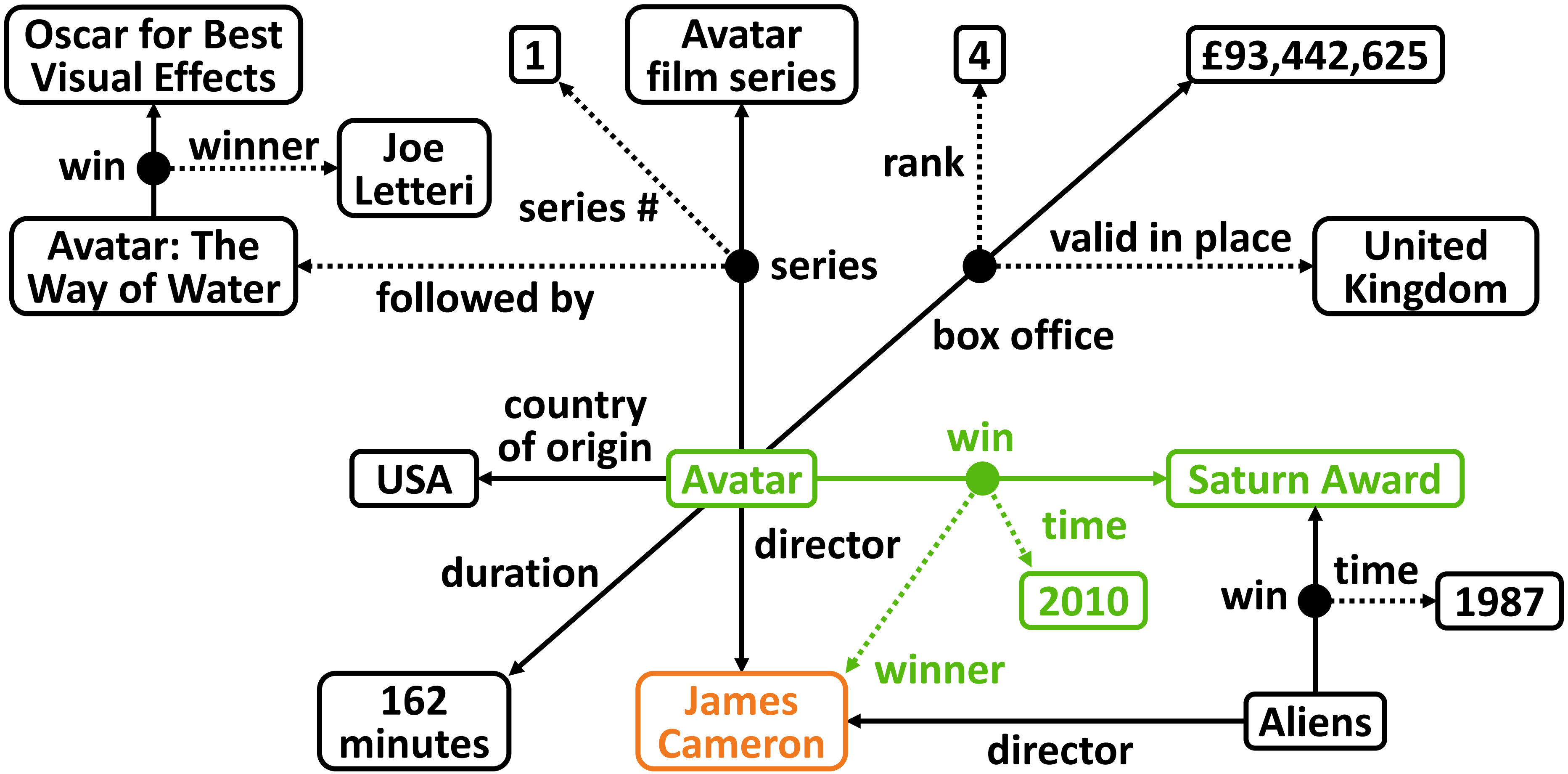
Link Prediction on HN-KGs

((Avatar, win, ?), {(winner, James_Cameron), (time, 2010)})



02 Link Prediction on HN-KGs

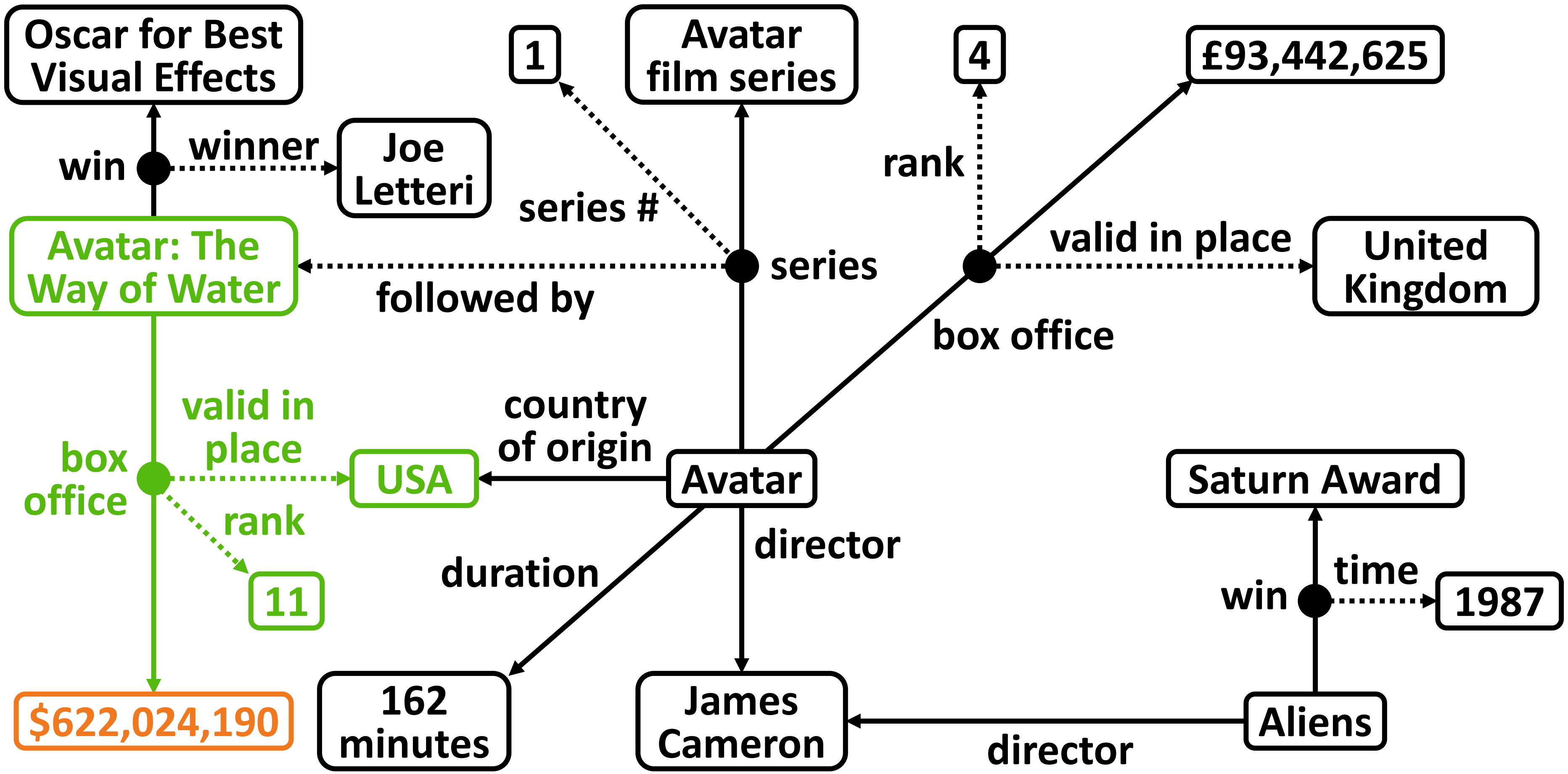
((Avatar, win, Saturn_Award), {(winner, ?), (time, 2010)})



02

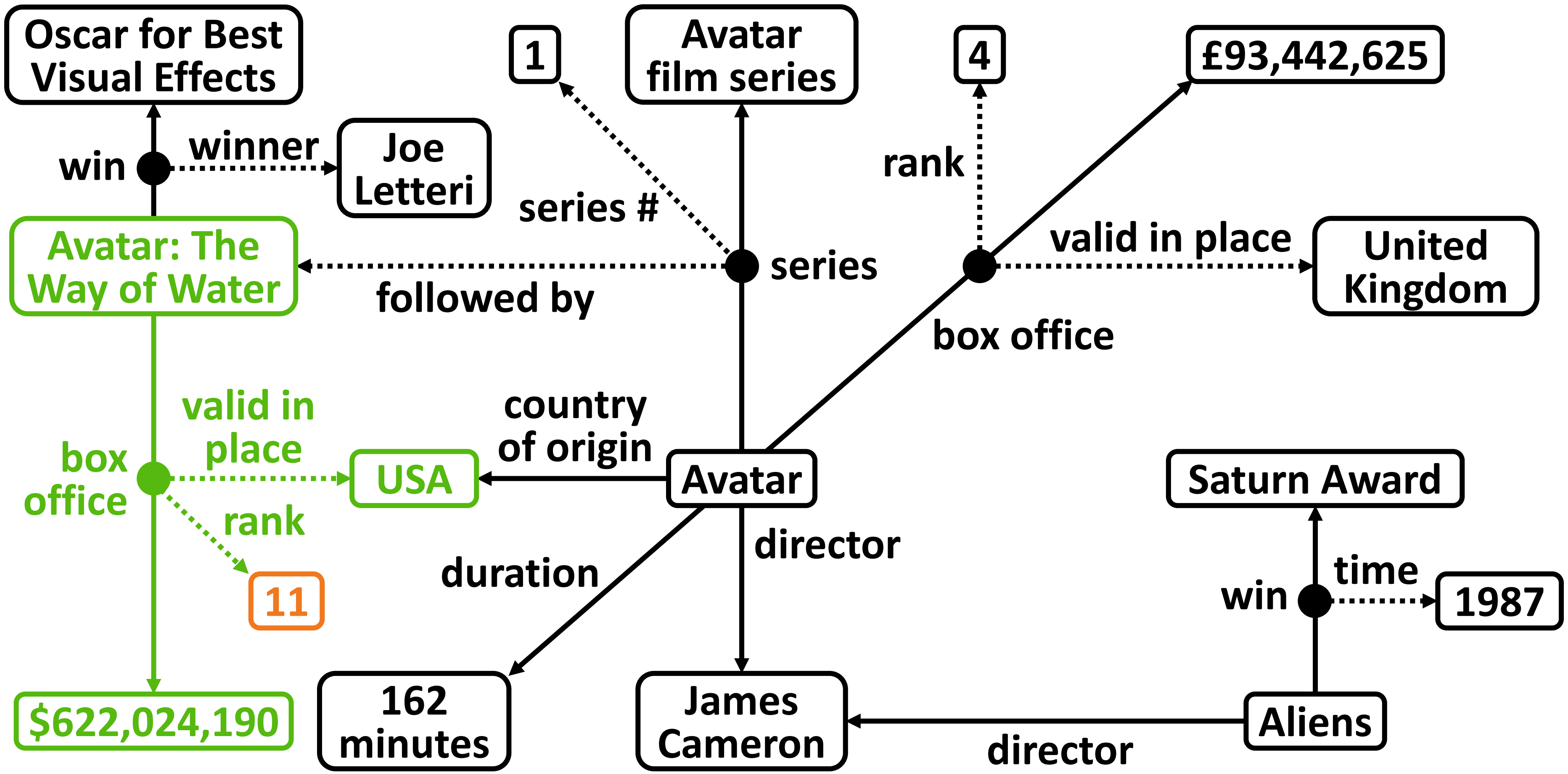
Numeric Value Prediction on HN-KGs

((Avatar:The_Way_of_Water, box_office, ?), {(rank, 11), (valid_in_place, USA)})



02 Numeric Value Prediction on HN-KGs

((Avatar:The_Way_of_Water, box_office, $622,024,190), {(rank, ?), (valid_in_place, USA)})

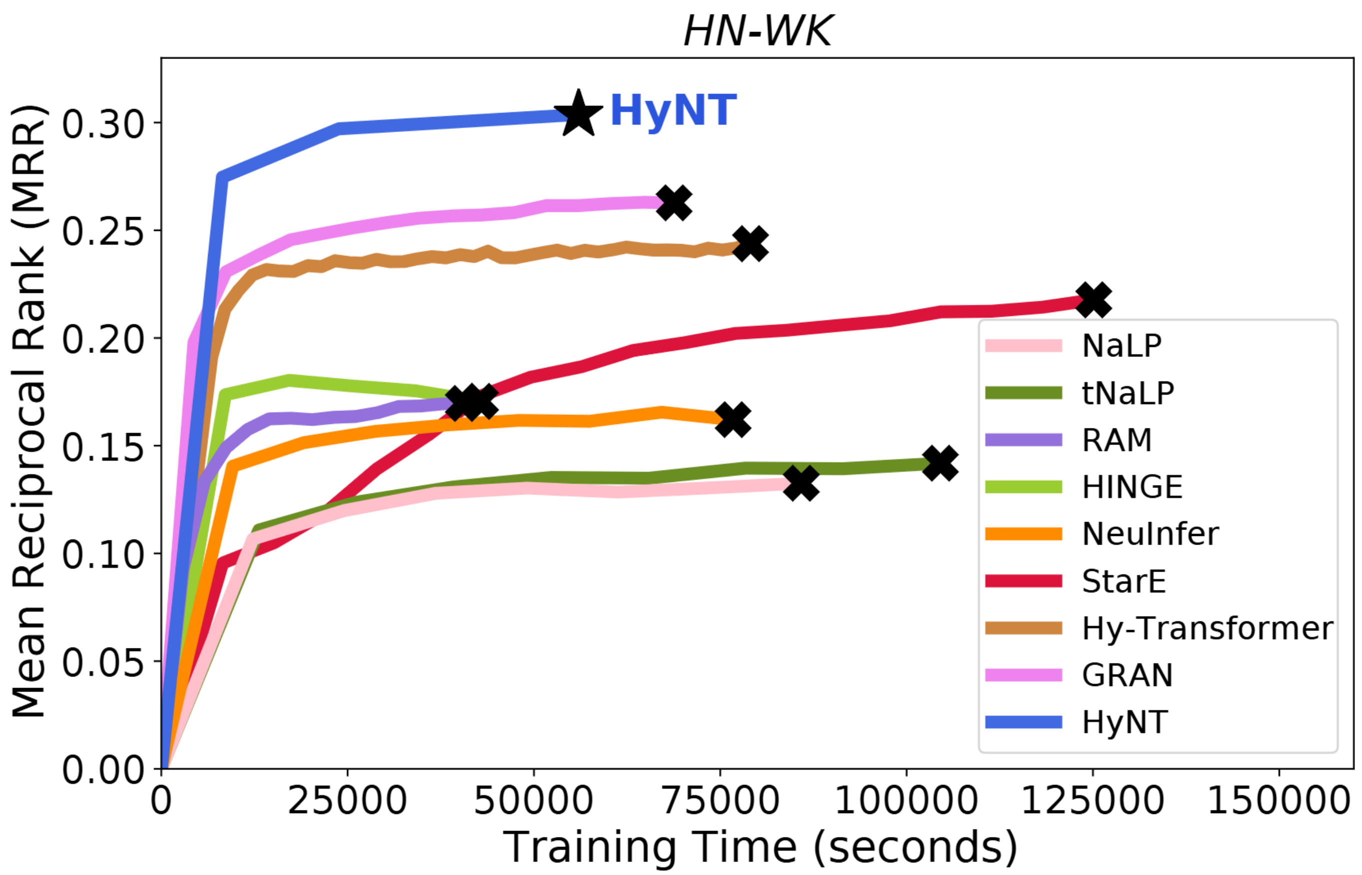


02 Contributions

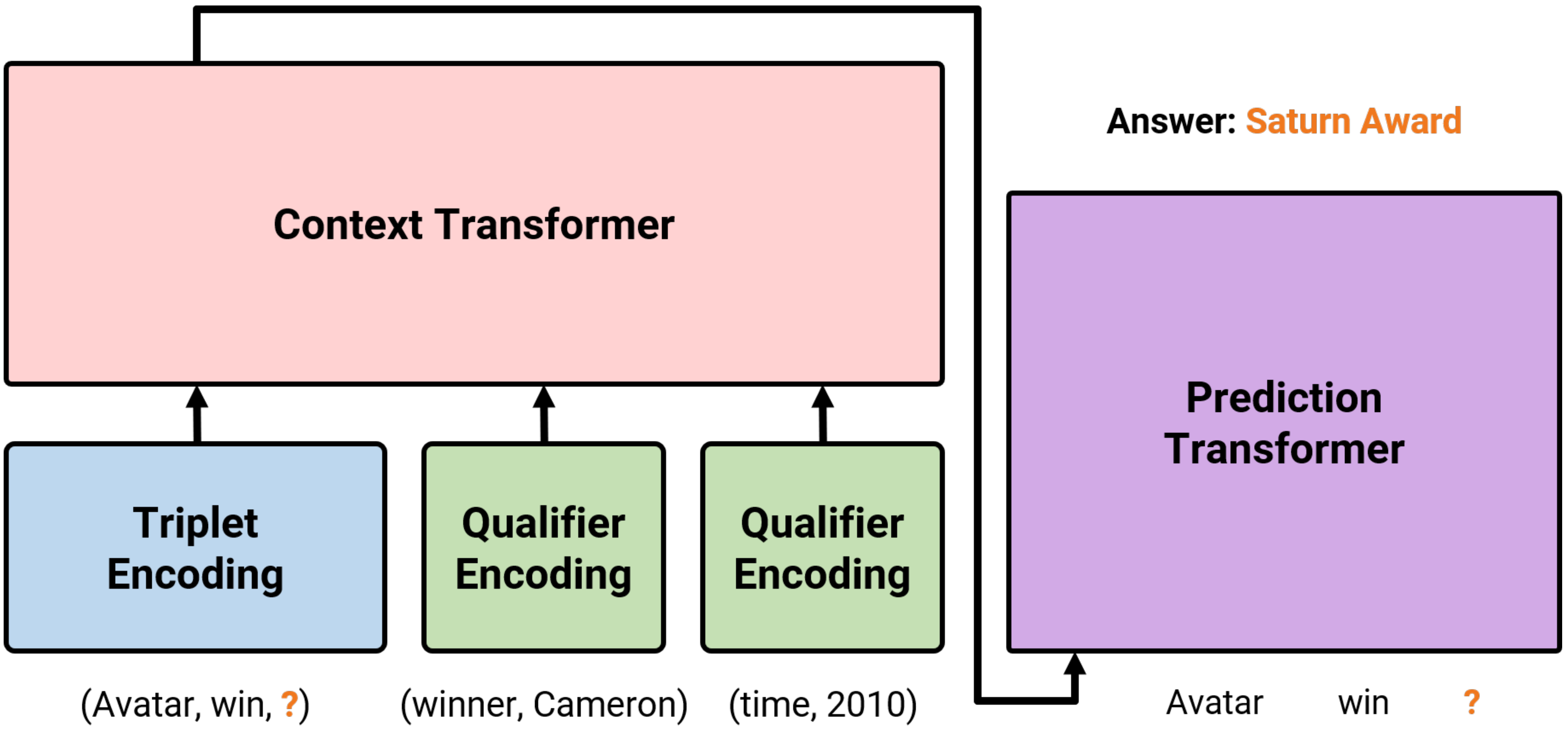
- Define **Hyper-relational and Numeric Knowledge Graphs**
  - Create 4 real-world HN-KG datasets
- Propose **HyNT**, **Hy**per-relational knowledge graph embedding with **N**umeric literals using **T**ransformers
  - Define a context transformer and a prediction transformer
  - Reduce the cost by learning compact representations of triplets and qualifiers
- HyNT significantly outperforms 12 different state-of-the-art methods for **link prediction**, **numeric value prediction**, and **relation prediction**

02 Contributions

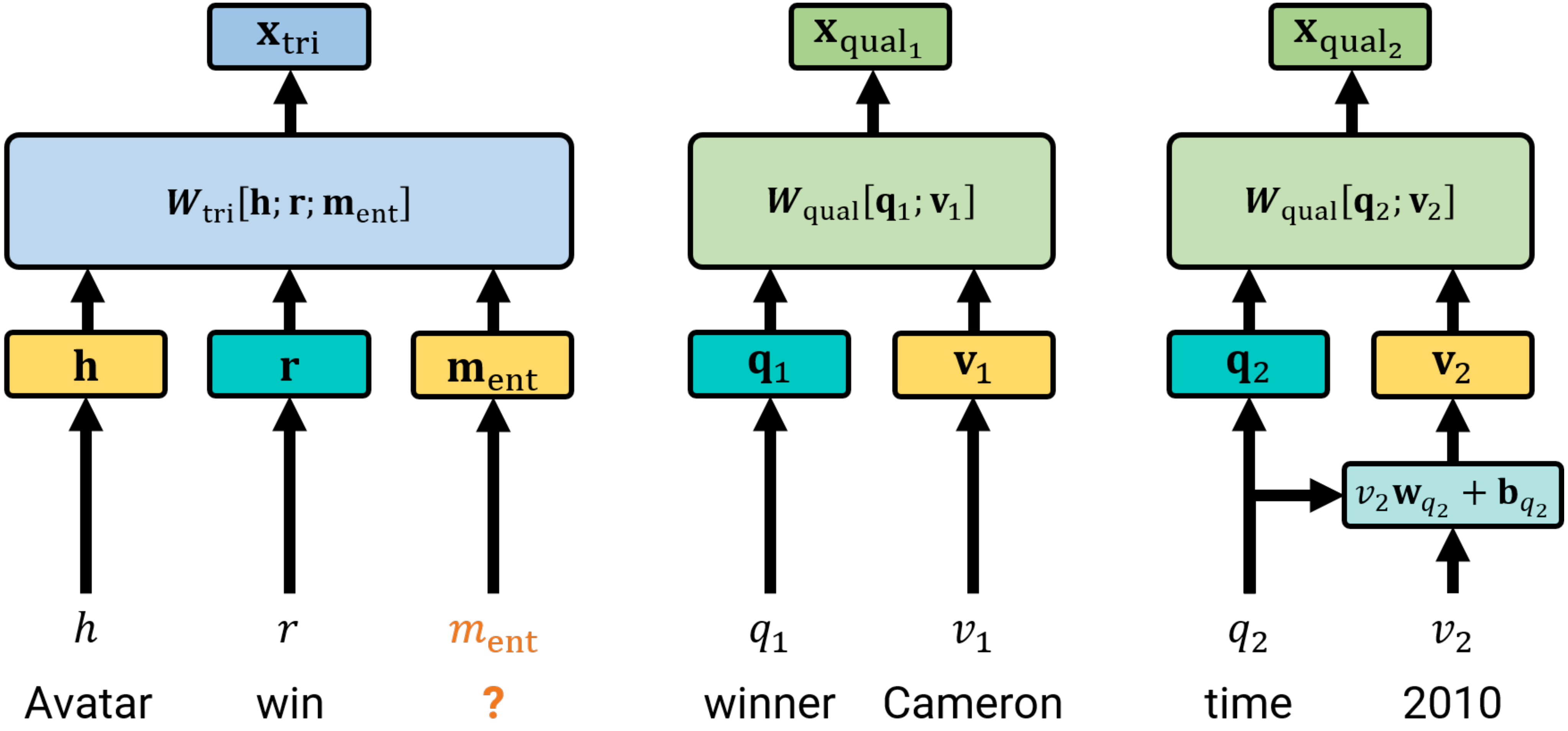
- Link Prediction Performance vs. Training Time



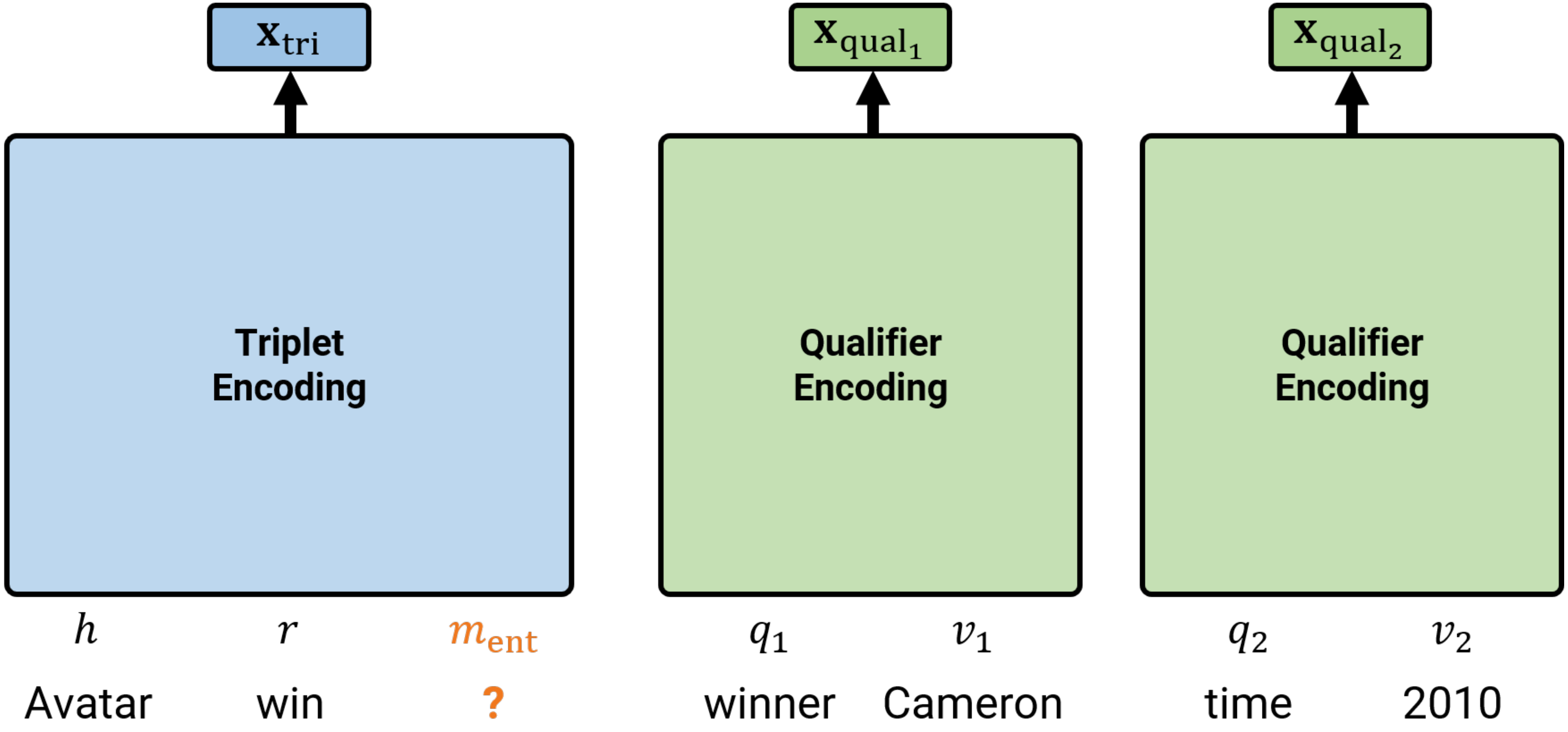
02 Overview of HyNT



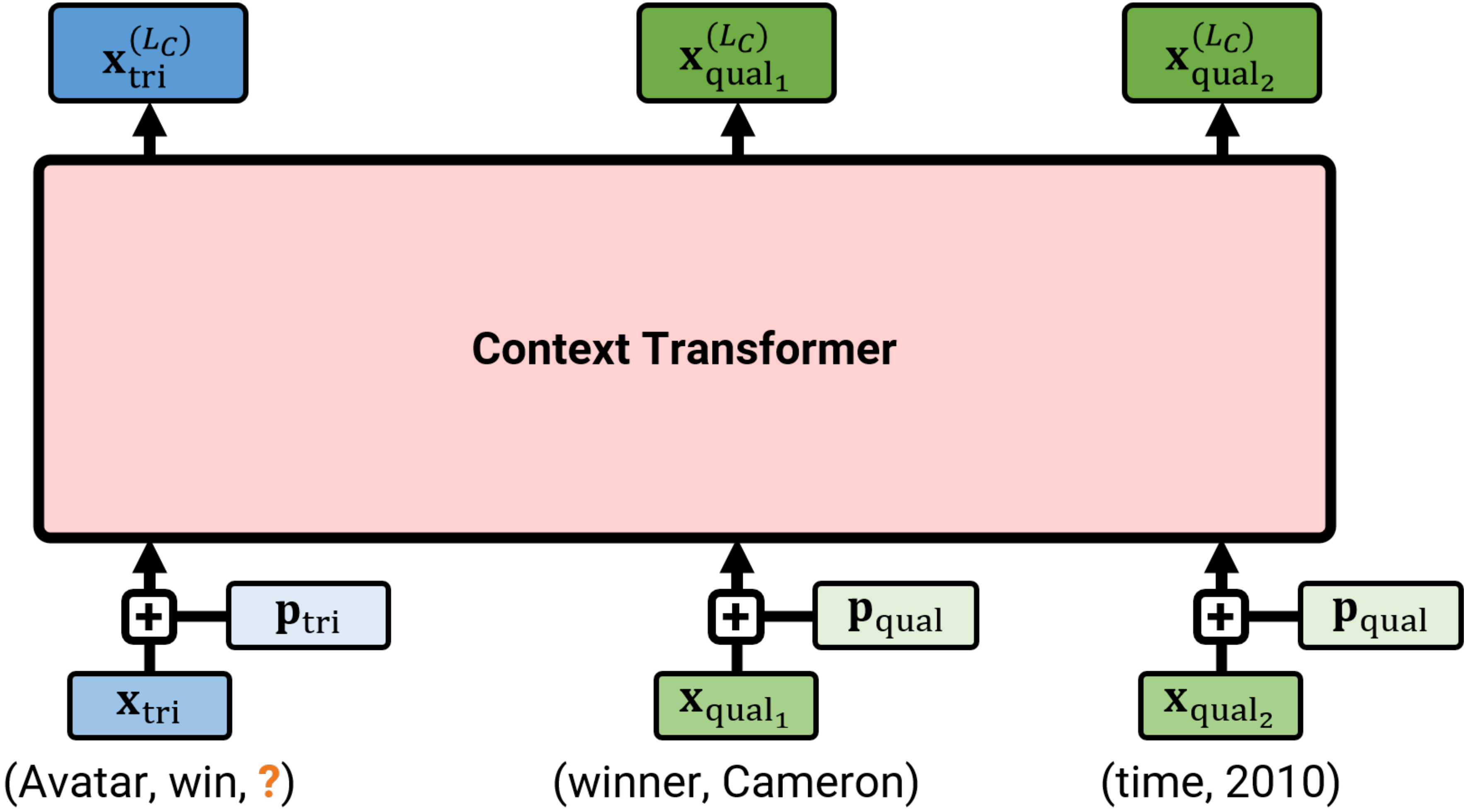
02 Triplet/Qualifier Encoding



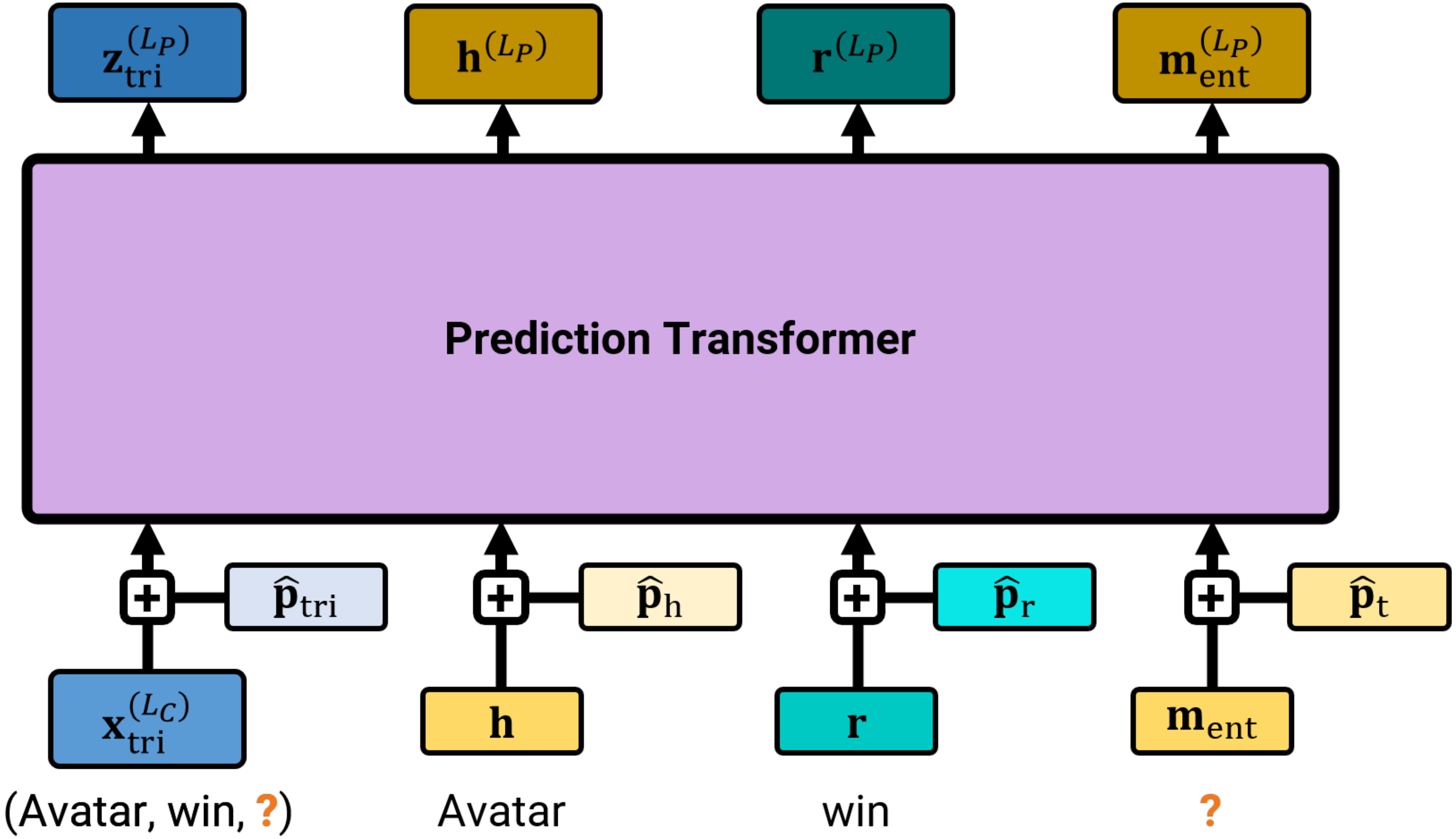
02 Triplet/Qualifier Encoding



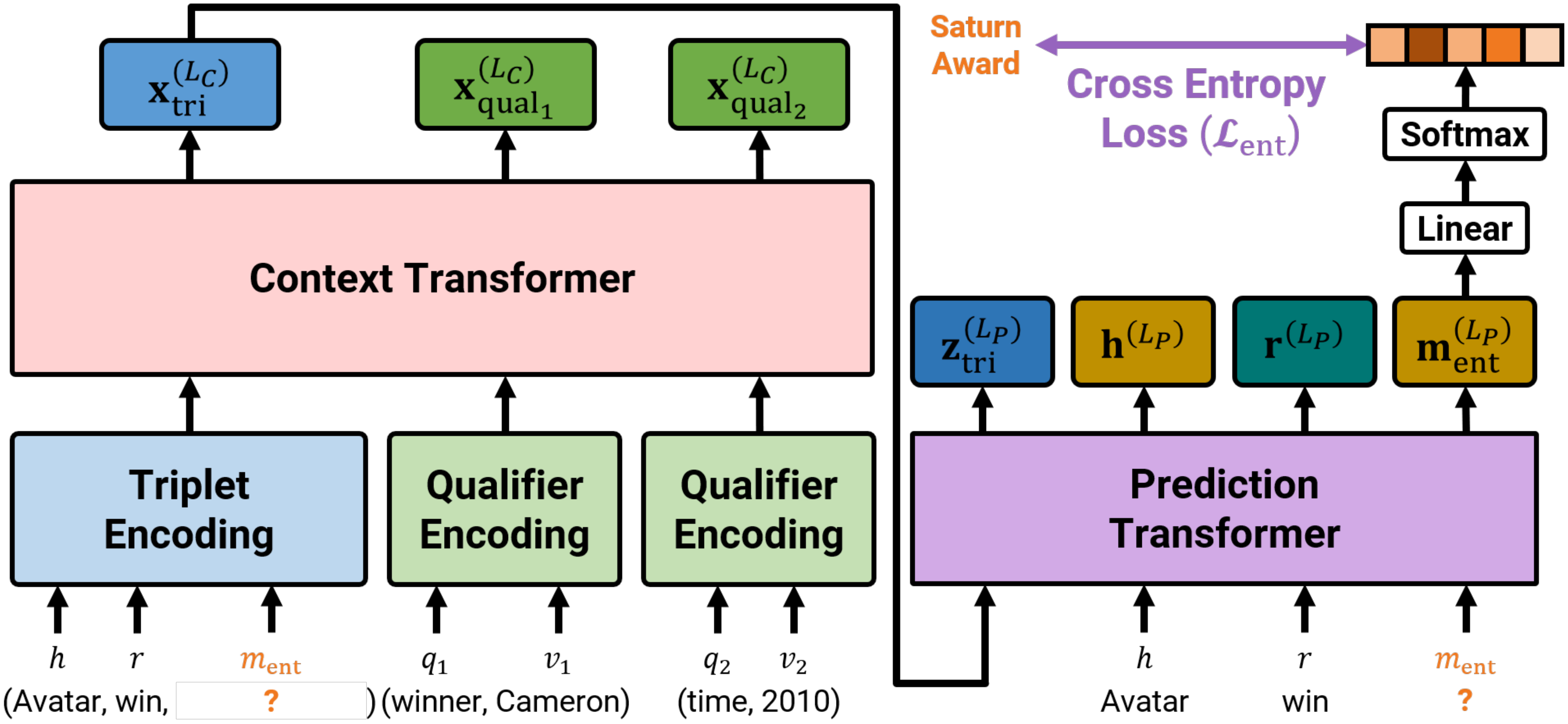
02 Context Transformer



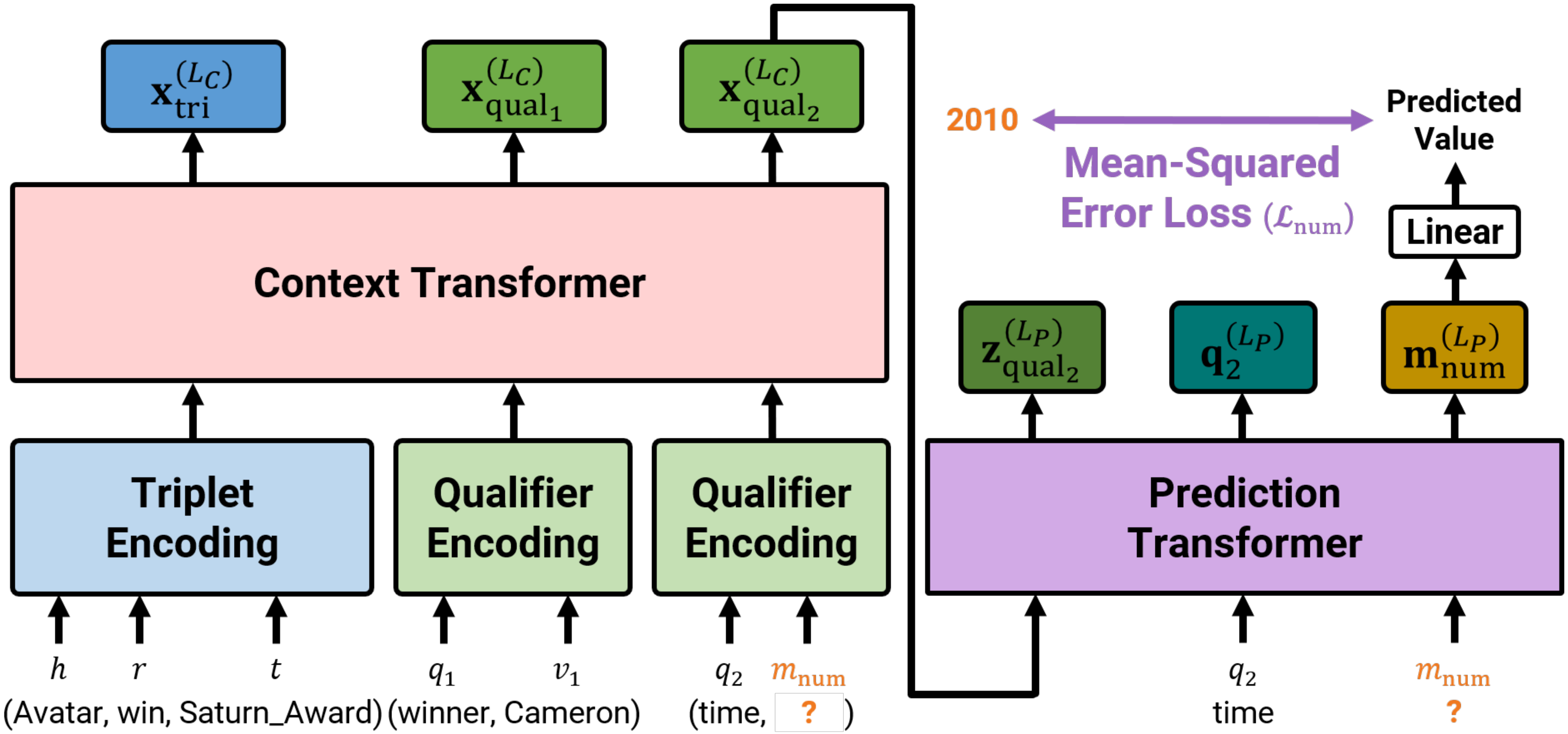
02 Prediction Transformer



02 Link Prediction using HyNT



02 Numeric Value Prediction using HyNT

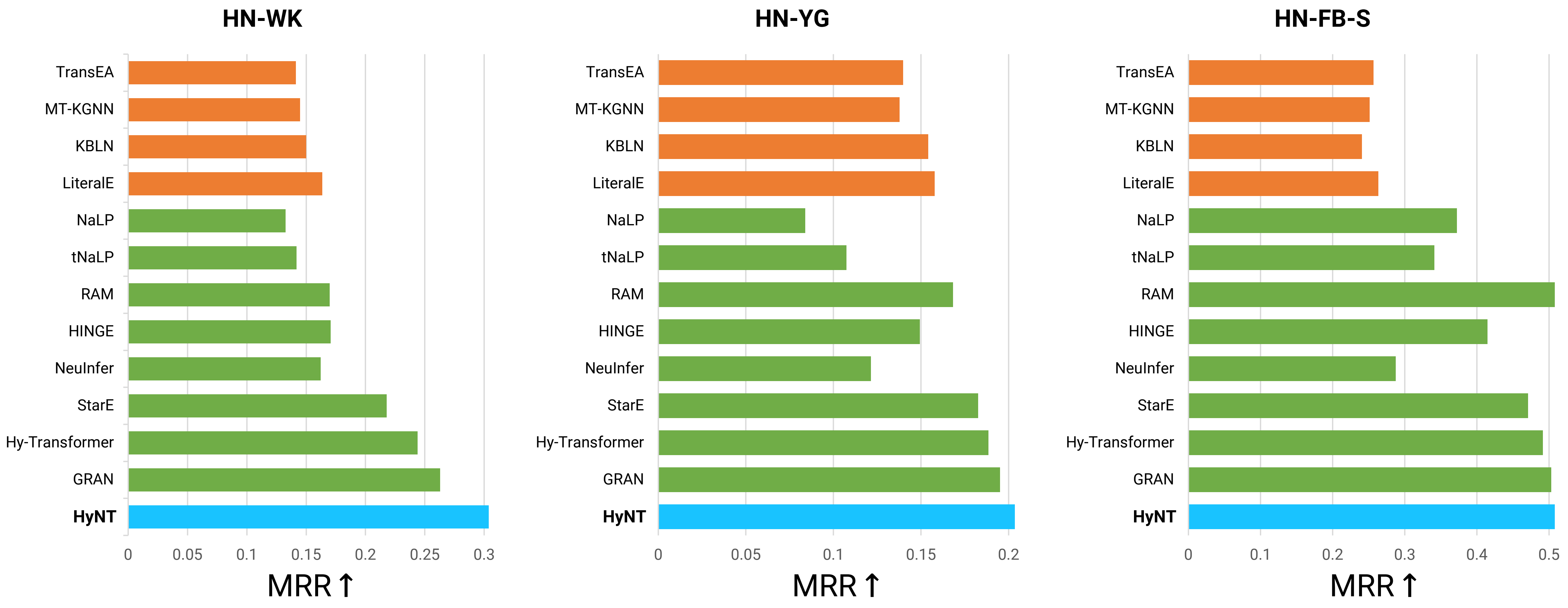


02 Experimental Results

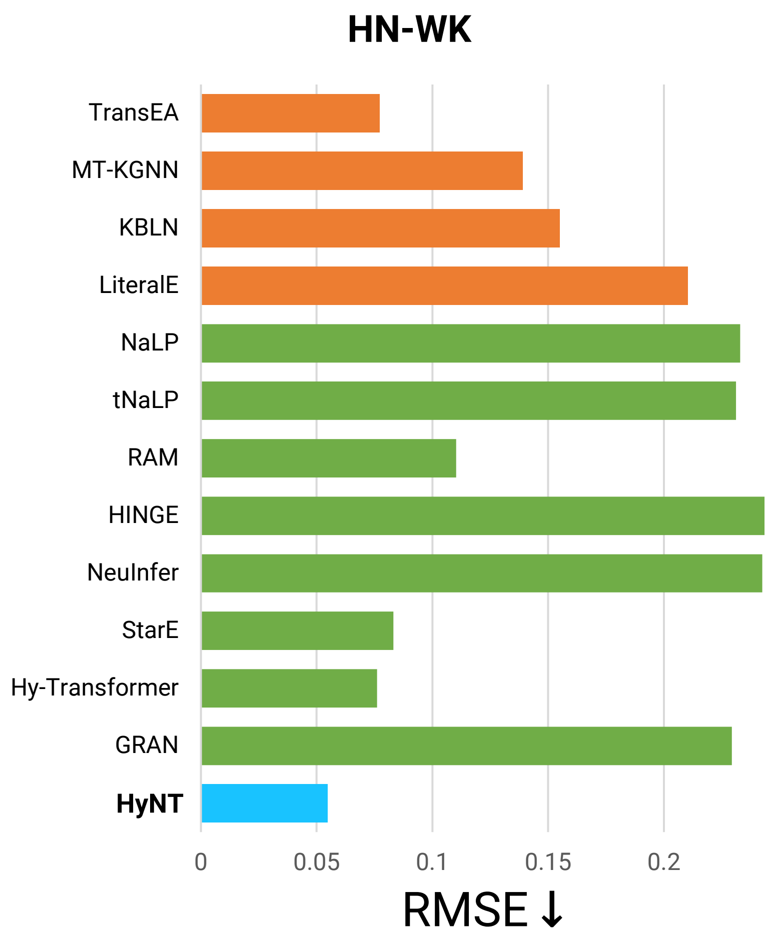
- Datasets
  - Based on Wikidata, YAGO, and Freebase
  - Create **4 Hyper-relational and Numeric Knowledge Graph (HN-KG)** datasets
    - HN-WK, HN-YG, HN-FB, HN-FB-S

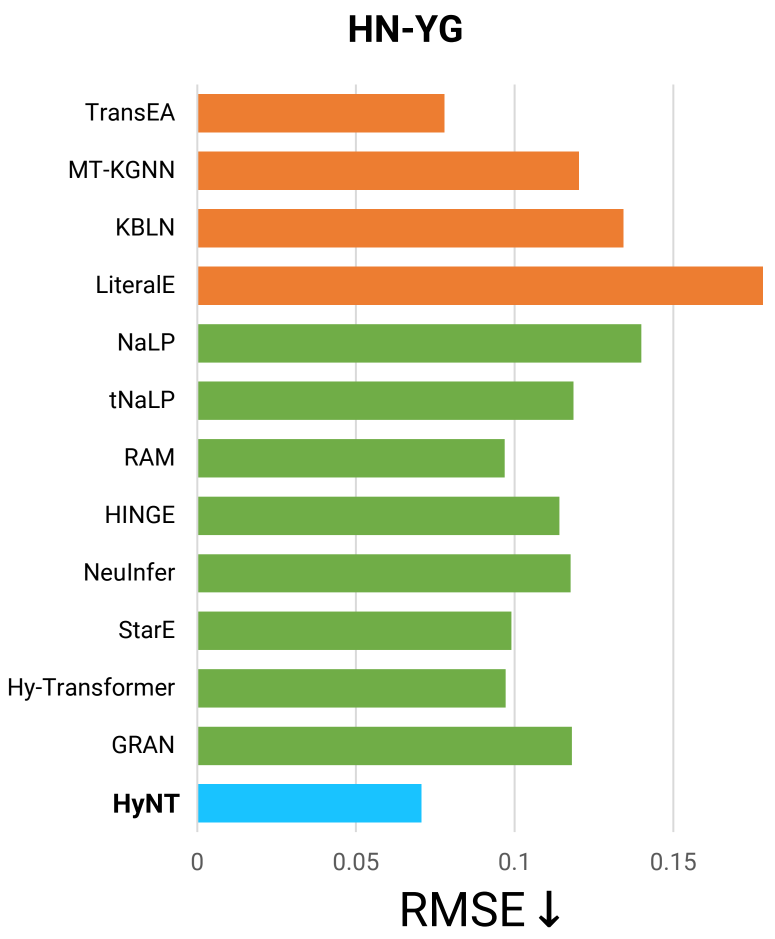
- Comparison with **12 baseline methods**
  - Methods for handling numeric literals
    - TransEA, MT-KGNN, KBLN, LiteralE
  - Methods for handling hyper-relational facts
    - NaLP, tNaLP, RAM, HINGE, NeuInfer, StarE, Hy-Transformer, GRAN

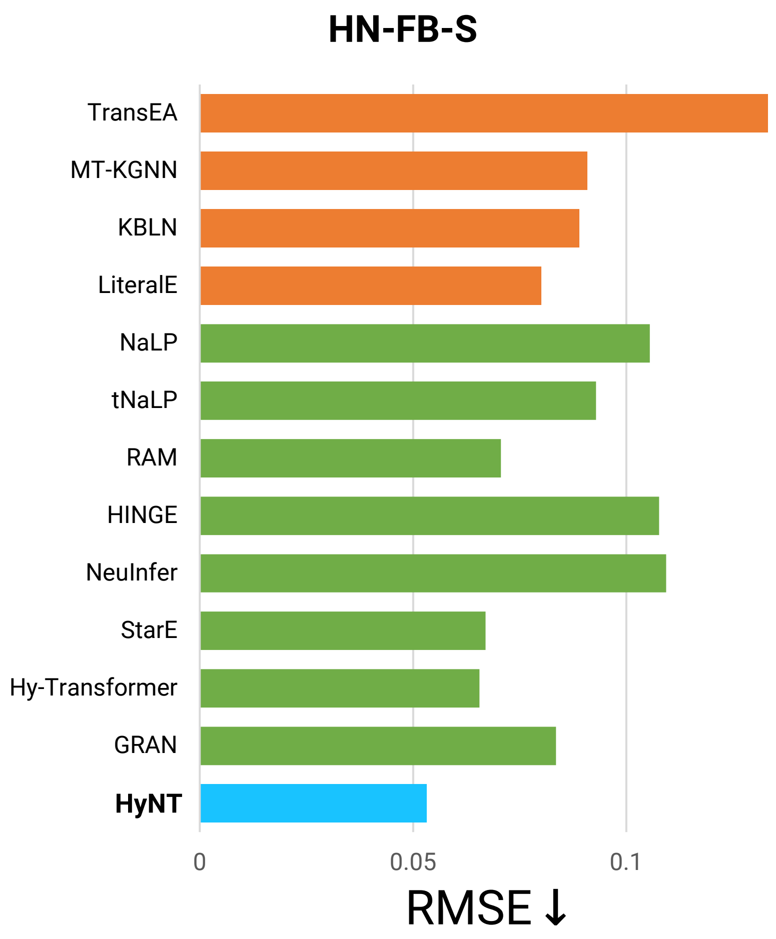
02 Link Prediction Results – Primary



02 Numeric Value Prediction Results – Primary







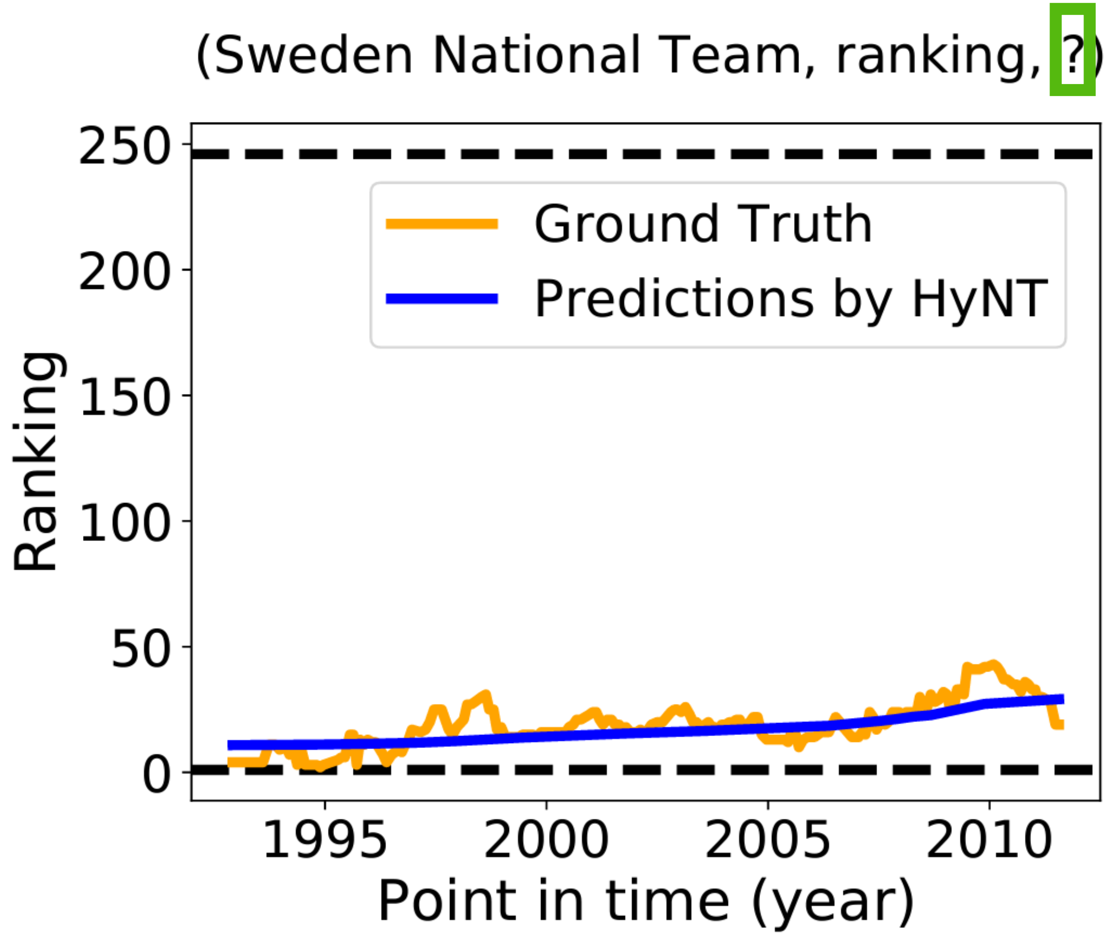
02 Visualization of the Predictions

Target Values

((Sweden National team, ranking, ?), {(point in time, 1995)})
((Sweden National team, ranking, ?), {(point in time, 1996)})
((Sweden National team, ranking, ?), {(point in time, 1997)})

⋮

02 Visualization of the Predictions



Target Values

02 Conclusion & Future Work

- **H**yper-relational and **N**umeric **K**nowledge **G**raphs **(HN-KGs)**

- Propose **HyNT** to solve **link prediction**, **numeric value prediction**, and **relation prediction** on **HN-KGs**

- HyNT significantly outperforms 12 different state-of-the-art methods

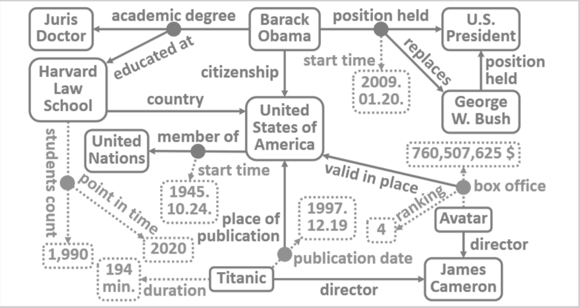
- Extend HyNT to **inductive learning scenarios**
  - New entities and relations appear at test time

I

Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers

Chanyoung Chung‡, Jaejun Lee‡, and Joyce Jiyoung Whang*

KDD 2023

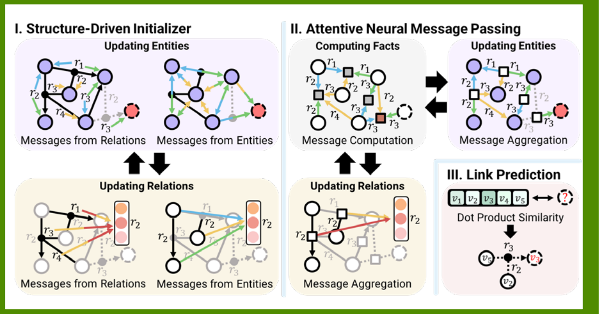


II

Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs

Jaejun Lee and Joyce Jiyoung Whang*

ICML 2025

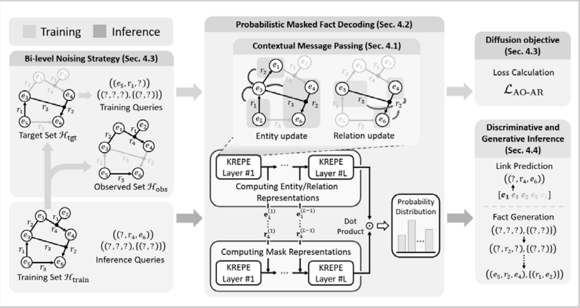


III

Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion

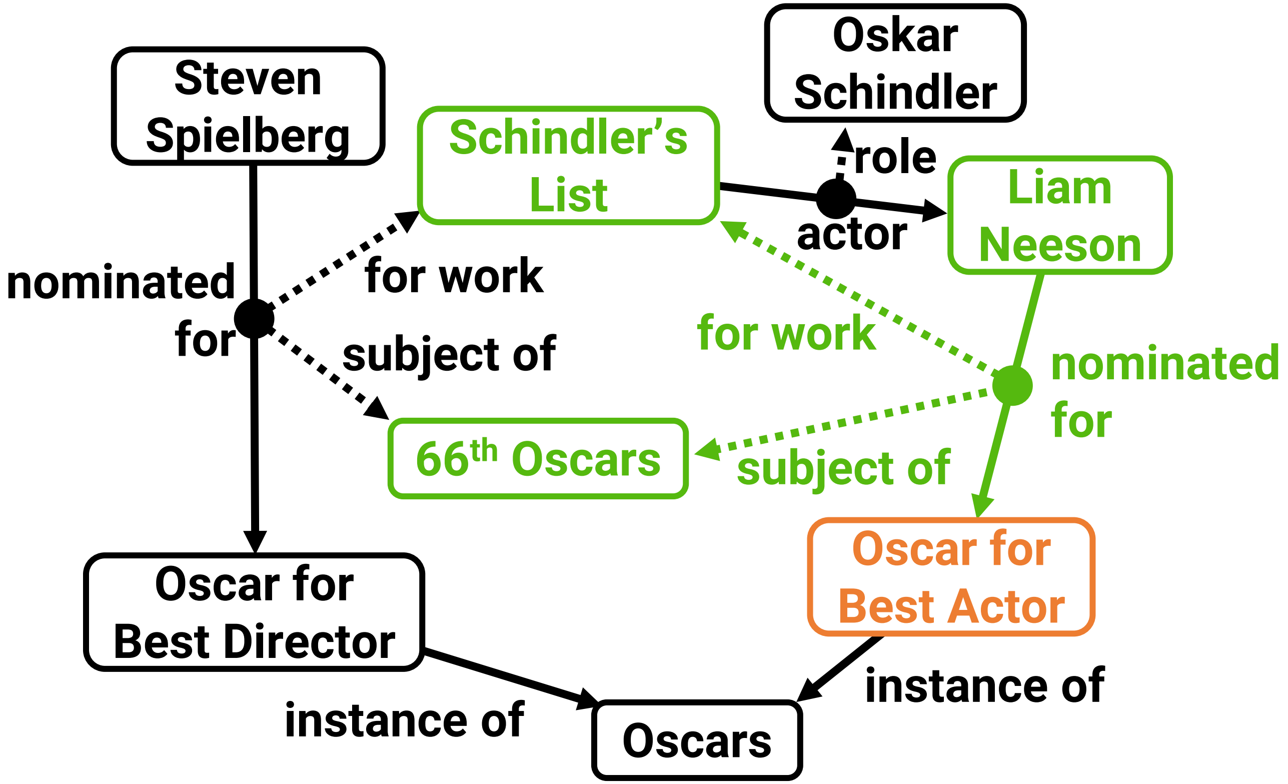
Jaejun Lee, Seheon Kim, and Joyce Jiyoung Whang*

ICML 2026



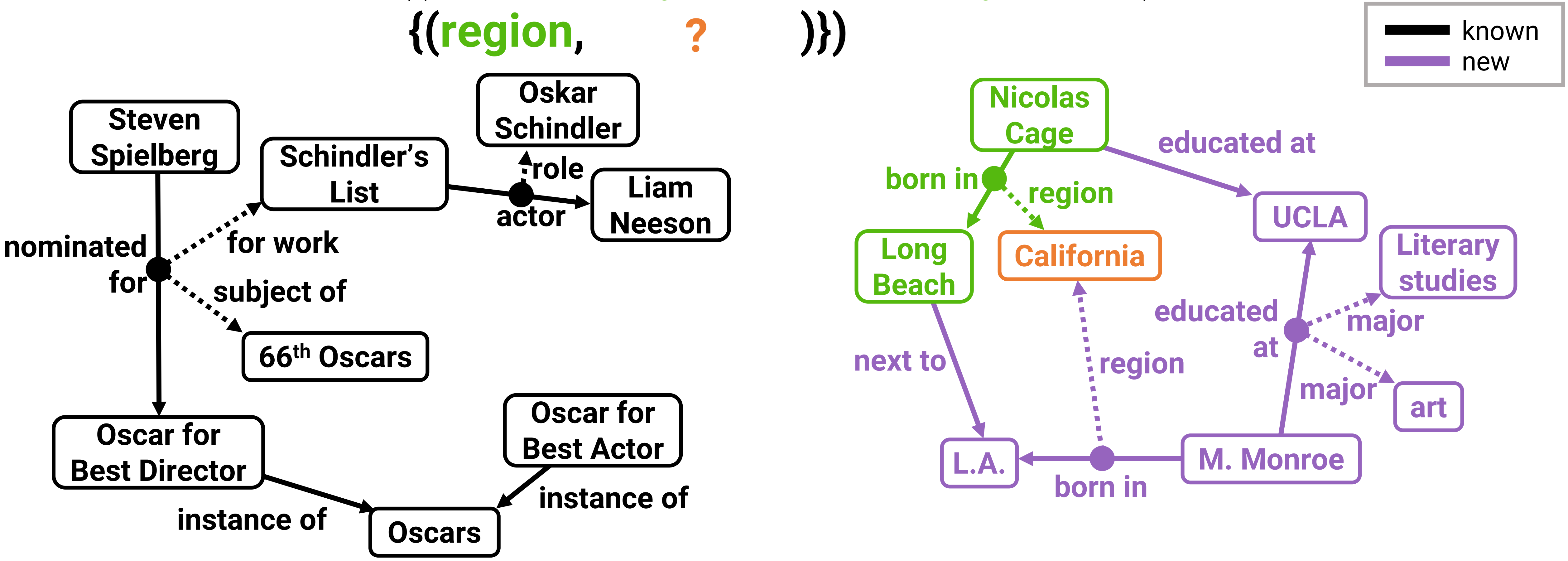
03 Transductive Link Prediction on HKGs

((Liam Neeson, nominated for, ?),
{(for work, Schindler's List), (subject of, 66^{th} Oscars)})



03 Inductive Link Prediction on HKGs

((Nicolas Cage, born in, Long Beach),
{(region, ?)})

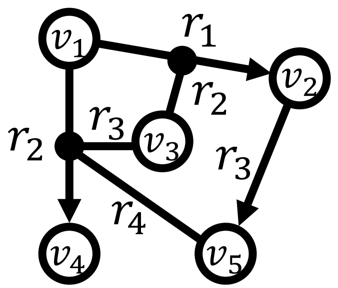


Training HKG

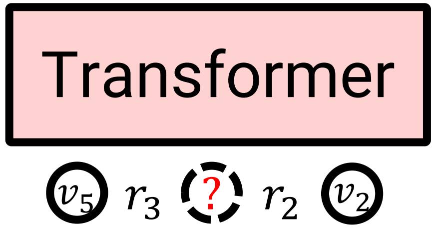
Inference HKG

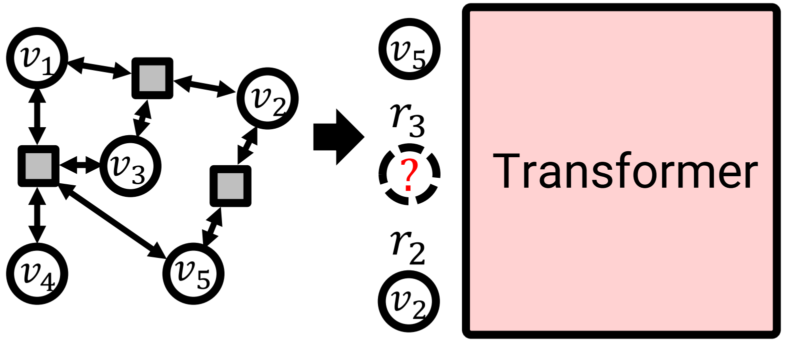
03 Existing HKG Methods

- Incorporate only limited structural information and fail to utilize HKG structures
  - Transformer-based methods: process each fact individually
  - GNN-based encoders: redundant / does not consider relations and positions of entities



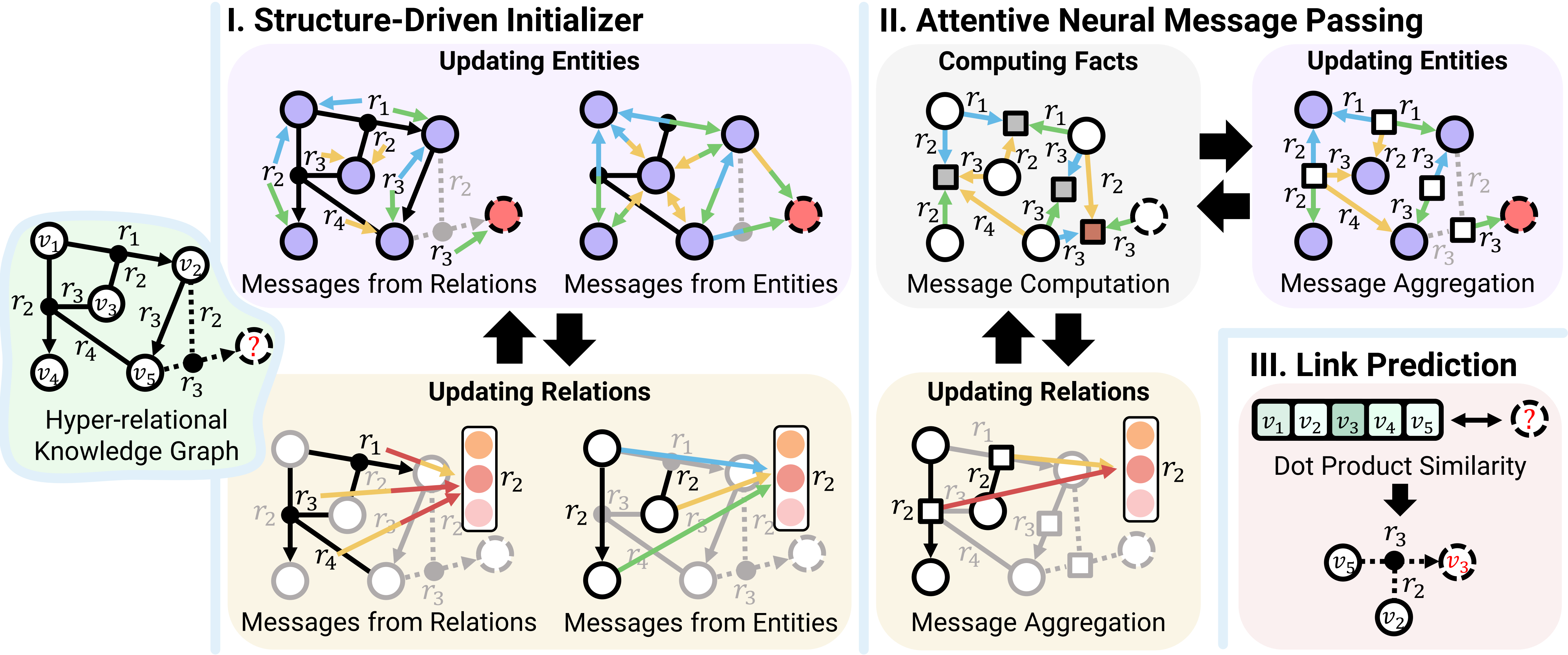
Hyper-relational Knowledge Graph





	Transformer-based Methods	Existing GNN-based Methods
Inductive Inference	X	X
Structure Utilization	X	△

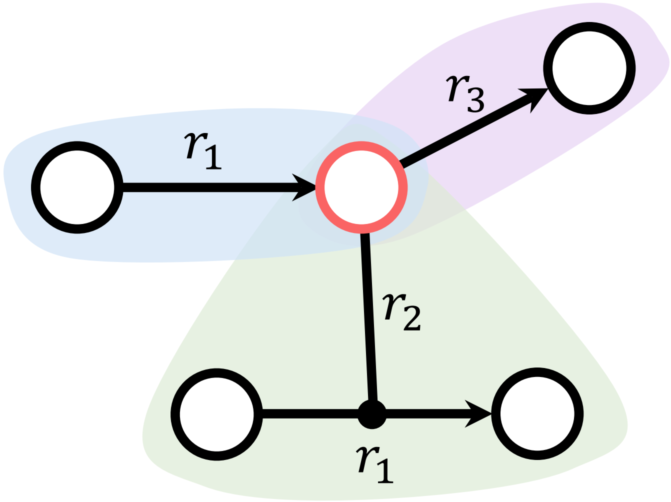
03 MAYPL: Message Passing Framework for Hyper-Relational Knowledge Graph Representation Learning



03 Structure-driven Initializer

- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

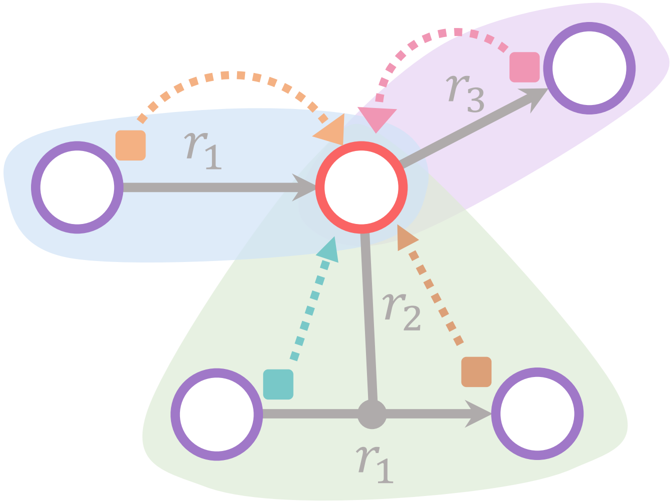
- Entity: Aggregate the messages of the co-occurring entities and incident relations
  - $\tilde{\mathbf{v}}^{(\tilde{l})} = \frac{1}{\sum_{u \in \mathcal{V}_v} |\mathcal{H}_u \cap \mathcal{H}_v|} \sum_{u \in \mathcal{V}_v} \sum_{h \in \mathcal{H}_u \cap \mathcal{H}_v} \widetilde{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(v)} \widetilde{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(u)} \tilde{\mathbf{u}}^{(l-1)} + \frac{1}{|\mathcal{R}_v|} \sum_{r \in \mathcal{R}_v} \widetilde{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$



03 Structure-driven Initializer

- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

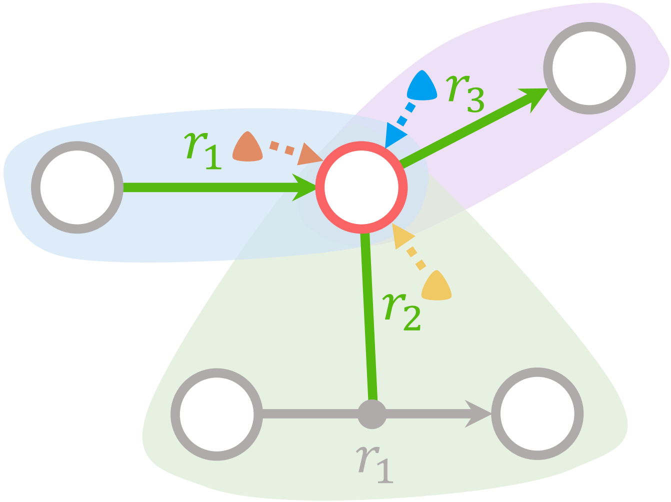
- Entity: Aggregate the messages of the **co-occurring entities** and incident relations
  - $\tilde{\mathbf{v}}^{(\tilde{l})} = \frac{1}{\sum_{u \in \mathcal{V}_v} |\mathcal{H}_u \cap \mathcal{H}_v|} \sum_{u \in \mathcal{V}_v} \sum_{h \in \mathcal{H}_u \cap \mathcal{H}_v} \widetilde{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(v)} \widetilde{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(u)} \tilde{\mathbf{u}}^{(l-1)} + \frac{1}{|\mathcal{R}_v|} \sum_{r \in \mathcal{R}_v} \widetilde{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$



03 Structure-driven Initializer

- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

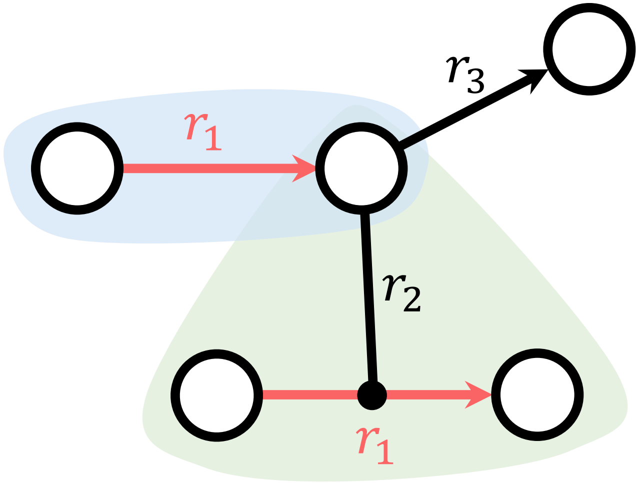
- Entity: Aggregate the messages of the co-occurring entities and **incident relations**
  - $\tilde{\mathbf{v}}^{(\tilde{l})} = \frac{1}{\sum_{u \in \mathcal{V}_v} |\mathcal{H}_u \cap \mathcal{H}_v|} \sum_{u \in \mathcal{V}_v} \sum_{h \in \mathcal{H}_u \cap \mathcal{H}_v} \widetilde{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(v)} \widetilde{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(u)} \tilde{\mathbf{u}}^{(l-1)} + \frac{1}{|\mathcal{R}_v|} \sum_{r \in \mathcal{R}_v} \widetilde{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$



03 Structure-driven Initializer

- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

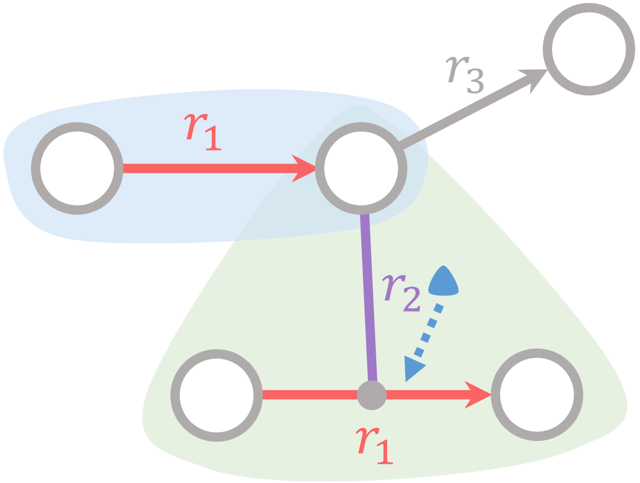
- Relation: Aggregate the messages of the co-occurring relations and incident entities
  - $\tilde{\mathbf{r}}^{(\tilde{l})} = \frac{1}{\sum_{y \in \mathcal{R}_r} |\mathcal{H}_y \cap \mathcal{H}_r|} \sum_{y \in \mathcal{R}_r} \sum_{h \in \mathcal{H}_y \cap \mathcal{H}_r} \widehat{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(r)} \widehat{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(y)} \tilde{\mathbf{y}}^{(l-1)} + \frac{1}{|\mathcal{V}_r|} \sum_{v \in \mathcal{V}_r} \widehat{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{v}}^{(\tilde{l}-1)}$



03 Structure-driven Initializer

- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

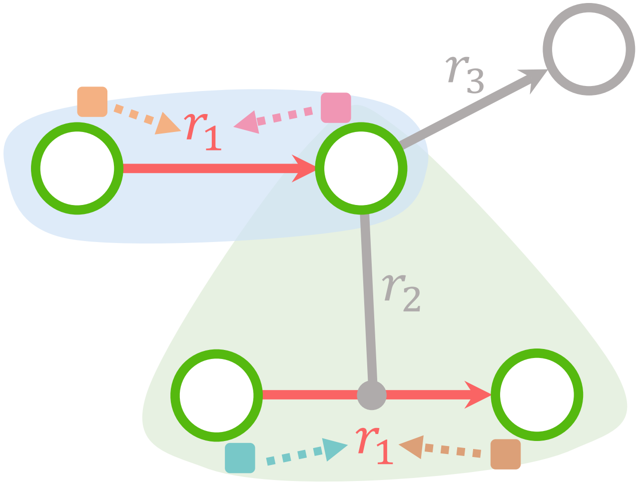
- Relation: Aggregate the messages of the **co-occurring relations** and incident entities
  - $\tilde{\mathbf{r}}^{(\tilde{l})} = \frac{1}{\sum_{y\in\mathcal{R}_r}|\mathcal{H}_y\cap\mathcal{H}_r|}\sum_{y\in\mathcal{R}_r}\sum_{h\in\mathcal{H}_y\cap\mathcal{H}_r}\widehat{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(r)}\widehat{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(y)}\tilde{\mathbf{y}}^{(l-1)} + \frac{1}{|\mathcal{V}_r|}\sum_{v\in\mathcal{V}_r}\widehat{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)}\tilde{\mathbf{v}}^{(\tilde{l}-1)}$



Structure-driven Initializer

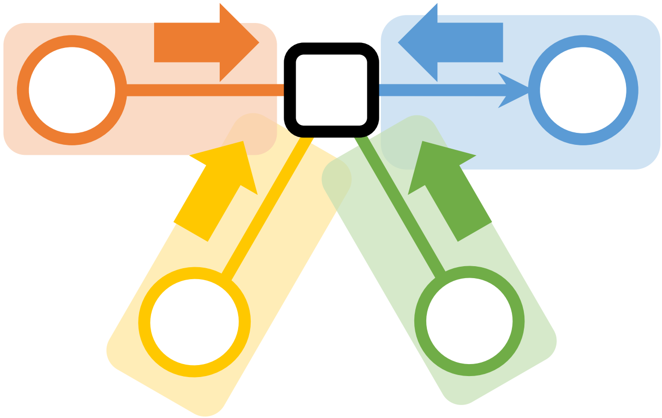
- Exploits the **interconnections**, **co-occurrence**, and **positions** of entities and relations

- Relation: Aggregate the messages of the co-occurring relations and **incident entities**
  - $\tilde{\mathbf{r}}^{(\tilde{l})} = \frac{1}{\sum_{y\in\mathcal{R}_r}|\mathcal{H}_y \cap \mathcal{H}_r|}\sum_{y\in\mathcal{R}_r}\sum_{h\in\mathcal{H}_y\cap\mathcal{H}_r}\widehat{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(r)}\widehat{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(y)}\tilde{\mathbf{y}}^{(l-1)} + \frac{1}{|\mathcal{V}_r|}\sum_{v\in\mathcal{V}_r}\widehat{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)}\tilde{\mathbf{v}}^{(\tilde{l}-1)}$



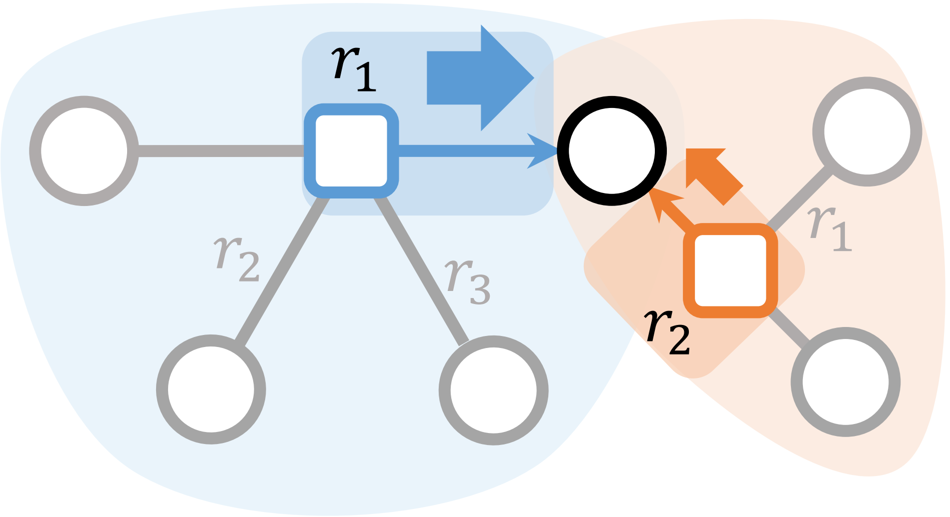
03 Attentive Neural Message Passing

- Updates entity and relation representations by attentively aggregating facts' messages
- Computes a fact's message by considering which entities and relations comprise the fact
  - Decompose each fact as a set of its relation-entity pairs
  - compute a pair's message by considering the **entities** and **relations** and their **positions** within the fact
    - $\mathbf{p}^{(l)} = \boldsymbol{P}_{\lambda_h(v)}^{(l)}\left(\left(\boldsymbol{W}_{\lambda_h(v)}^{(l)}\mathbf{v}^{(l-1)}\right) \odot \left(\boldsymbol{U}_{\lambda_h(v)}^{(l)}\mathbf{r}^{(l-1)}\right)\right)$
  - Compute a fact's message by aggregating messages of its **relation-entity pairs**
    - $\mathbf{h}^{(l)} = \frac{1}{|(\mathcal{R}\times\mathcal{V})_h|}\sum_{p\in(\mathcal{R}\times\mathcal{V})_h}\mathbf{p}^{(l)}$



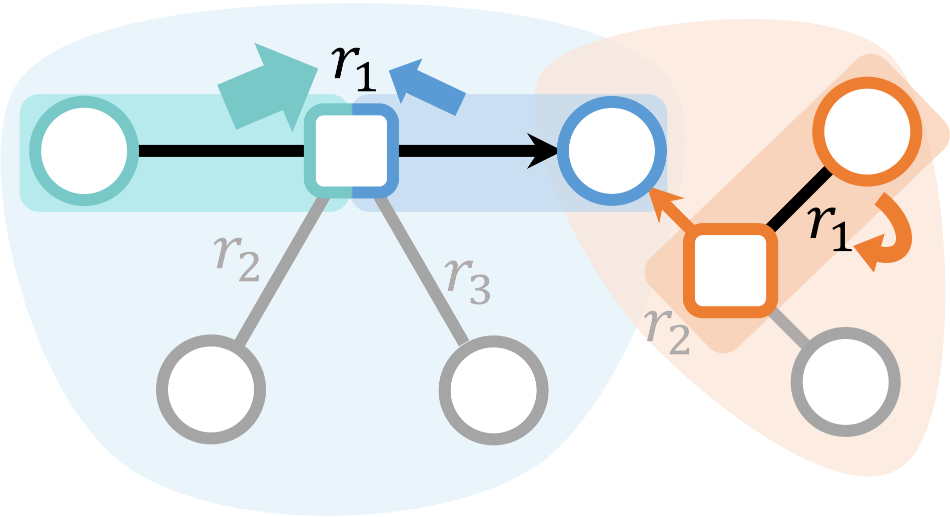
03 Attentive Neural Message Passing

- Updates an entity representation by considering which fact the entity belongs to
  - Considers pairs of incident relations and the corresponding facts
    - $\mathbf{q}^{(l)} = \boldsymbol{Q}^{(l)}_{\lambda_h(v)}\left(\mathbf{h}^{(l)} \odot \left(\boldsymbol{A}^{(l)}_{\lambda_h(v)}\mathbf{r}^{(l-1)}\right)\right)$
  - Attentively aggregate relation-fact pairs
    - $\mathbf{v}^{(l)} = \sum_{q\in(\mathcal{R}\times\mathcal{H})_v} \alpha^{(l)}_{q,v}\boldsymbol{B}^{(l)}\mathbf{q}^{(l)}$, where $\alpha^{(l)}_{q,v} = \frac{\exp(\mathbf{a}^{(l)}\cdot\sigma(\boldsymbol{Q}^{(l)}\mathbf{v}^{(l-1)}+\boldsymbol{K}^{(l)}\mathbf{q}^{(l)}))}{\sum_{k\in(\mathcal{R}\times\mathcal{H})_v}\exp(\mathbf{a}^{(l)}\cdot\sigma(\boldsymbol{Q}^{(l)}\mathbf{v}^{(l-1)}+\boldsymbol{K}^{(l)}\mathbf{k}^{(l)}))}$



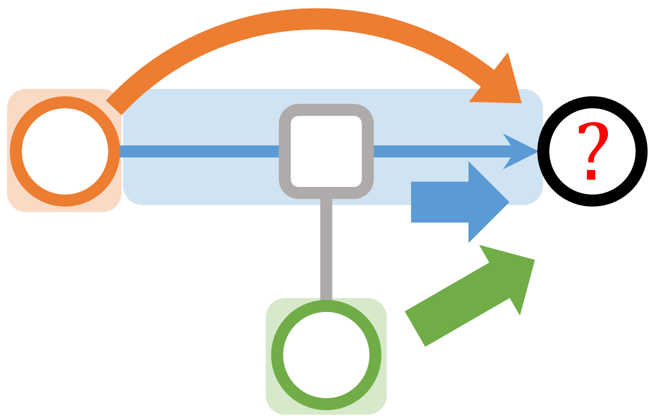
03 Attentive Neural Message Passing

- Updates a relation representation by considering which fact the relation belongs to
  - Considers pairs of incident entities and the corresponding facts
    - $\bar{\mathbf{q}}^{(l)} = \bar{\boldsymbol{Q}}^{(l)}_{\lambda_h(r)}\left(\mathbf{h}^{(l)} \odot \left(\bar{\boldsymbol{A}}^{(l)}_{\lambda_h(v)}\mathbf{v}^{(l-1)}\right)\right)$
  - Attentively aggregate relation-fact pairs
    - $\mathbf{r}^{(l)} = \sum_{\bar{q}\in(\mathcal{V}\times\mathcal{H})_r} \bar{\alpha}^{(l)}_{\bar{q},r}\bar{\boldsymbol{B}}^{(l)}\bar{\mathbf{q}}^{(l)}$, where $\bar{\alpha}^{(l)}_{\bar{q},r} = \frac{\exp(\bar{\mathbf{a}}^{(l)}\cdot\sigma(\bar{\boldsymbol{Q}}^{(l)}\mathbf{r}^{(l-1)}+\bar{\boldsymbol{K}}^{(l)}\bar{\mathbf{q}}^{(l)}))}{\sum_{\bar{k}\in(\mathcal{V}\times\mathcal{H})_r}\exp(\bar{\mathbf{a}}^{(l)}\cdot\sigma(\bar{\boldsymbol{Q}}^{(l)}\mathbf{r}^{(l-1)}+\bar{\boldsymbol{K}}^{(l)}\bar{\mathbf{k}}^{(l)}))}$



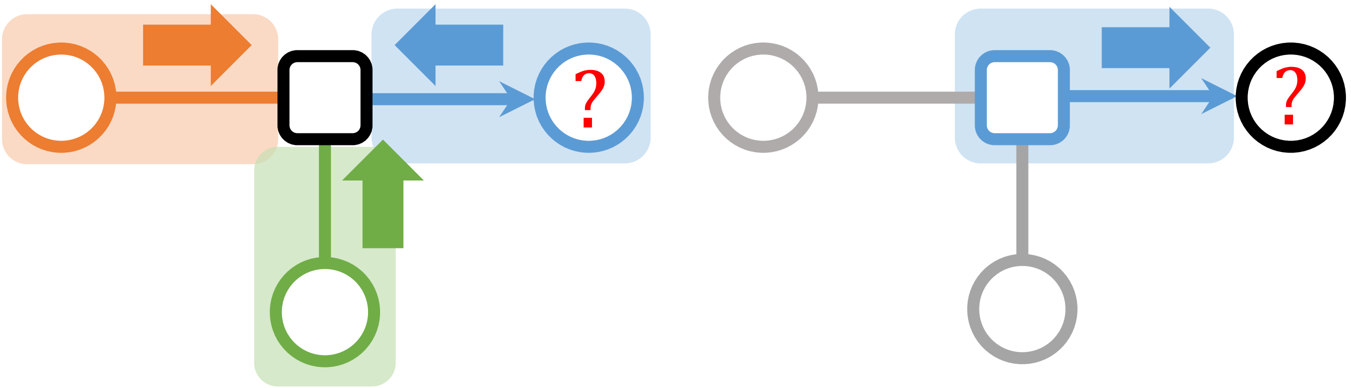
03 Link Prediction on HKGs

- Structure-driven initializer
  - $\tilde{\mathbf{x}}^{(\tilde{l})} = \sum_{v \in \mathcal{V}_x} \widetilde{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(x)} \widetilde{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(v)} \tilde{\mathbf{v}}^{(l-1)} + \widetilde{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$
- Attentive Neural Message Passing
  - $\mathbf{x}^{(l)} = \boldsymbol{B}^{(l)} \left(\boldsymbol{Q}^{(l)}_{\lambda_h(x)} \left(\mathbf{h}^{(l)} \odot \left(\boldsymbol{A}^{(l)}_{\lambda_h(x)} \mathbf{r}^{(l-1)} \right) \right) \right)$
- Compute the dot product similarity between $\mathbf{x}^{(L)}$ and each entity representation
  - MAYPL predicts the missing entity as the entity with the highest similarity



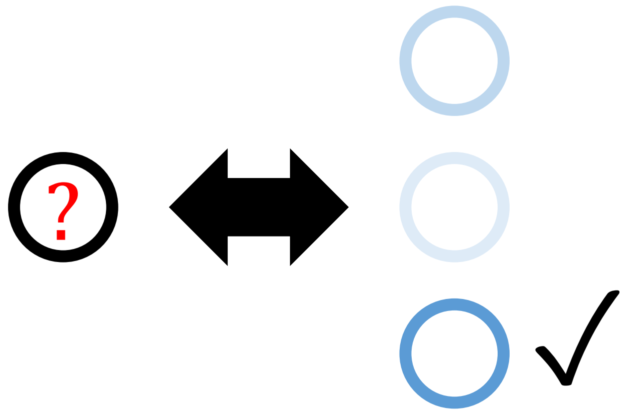
03 Link Prediction on HKGs

- Structure-driven initializer
  - $\tilde{\mathbf{x}}^{(\tilde{l})} = \sum_{v \in \mathcal{V}_x} \tilde{U}^{(\tilde{l})}_{\lambda_h(x)} \tilde{W}^{(\tilde{l})}_{\lambda_h(v)} \tilde{\mathbf{v}}^{(l-1)} + \tilde{A}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$
- Attentive Neural Message Passing
  - $\mathbf{x}^{(l)} = \boldsymbol{B}^{(l)} \left(\boldsymbol{Q}^{(l)}_{\lambda_h(x)} \left(\mathbf{h}^{(l)} \odot \left(\boldsymbol{A}^{(l)}_{\lambda_h(x)} \mathbf{r}^{(l-1)} \right) \right) \right)$
- Compute the dot product similarity between $\mathbf{x}^{(L)}$ and each entity representation
  - MAYPL predicts the missing entity as the entity with the highest similarity



03 Link Prediction on HKGs

- Structure-driven initializer
  - $\tilde{\mathbf{x}}^{(\tilde{l})} = \sum_{v\in\mathcal{V}_x} \widetilde{\boldsymbol{U}}^{(\tilde{l})}_{\lambda_h(x)} \widetilde{\boldsymbol{W}}^{(\tilde{l})}_{\lambda_h(v)} \tilde{\mathbf{v}}^{(l-1)} + \widetilde{\boldsymbol{A}}^{(\tilde{l})}_{\tau_r(v)} \tilde{\mathbf{r}}^{(\tilde{l}-1)}$
- Attentive Neural Message Passing
  - $\mathbf{x}^{(l)} = \boldsymbol{B}^{(l)}\left(\boldsymbol{Q}^{(l)}_{\lambda_h(x)}\left(\mathbf{h}^{(l)} \odot \left(\boldsymbol{A}^{(l)}_{\lambda_h(x)}\mathbf{r}^{(l-1)}\right)\right)\right)$
- Compute the dot product similarity between $\mathbf{x}^{(L)}$ and each entity representation
  - MAYPL predicts the missing entity as the entity with the highest similarity

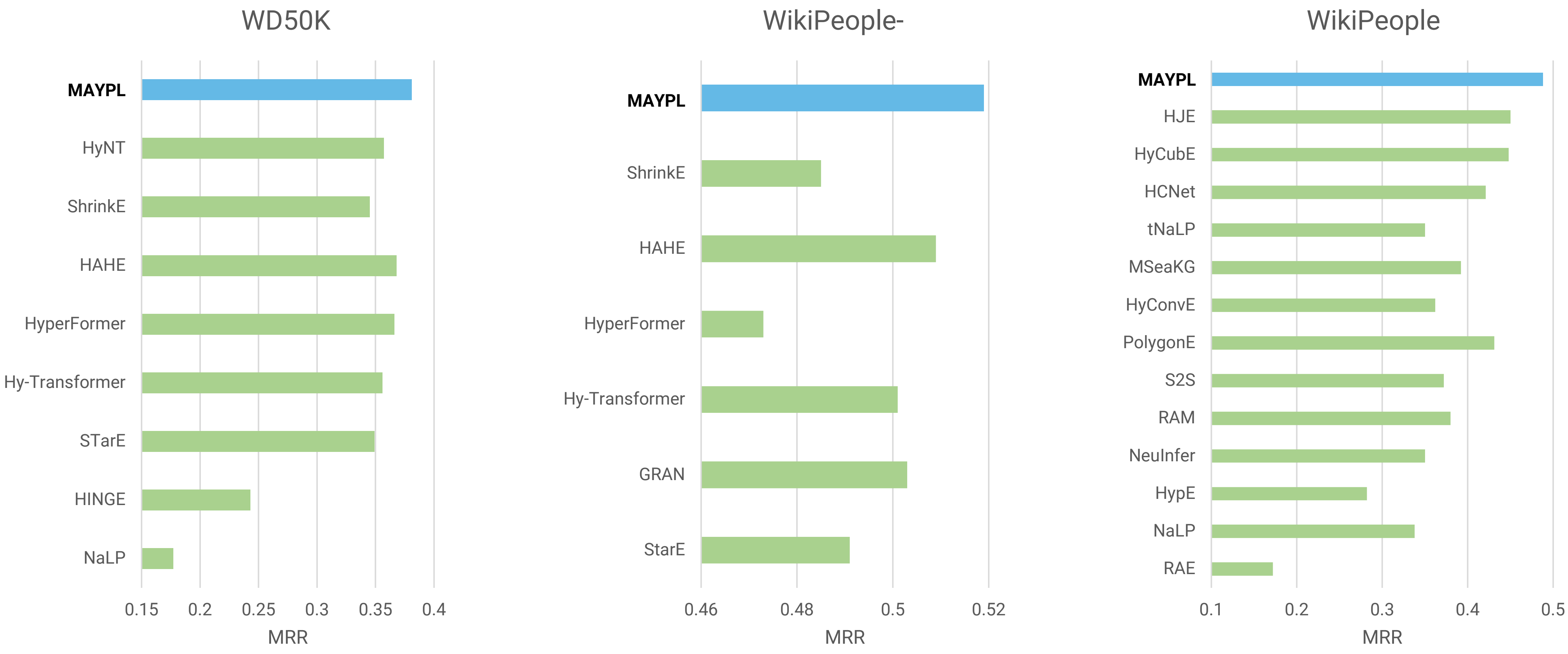


03 Experimental Results

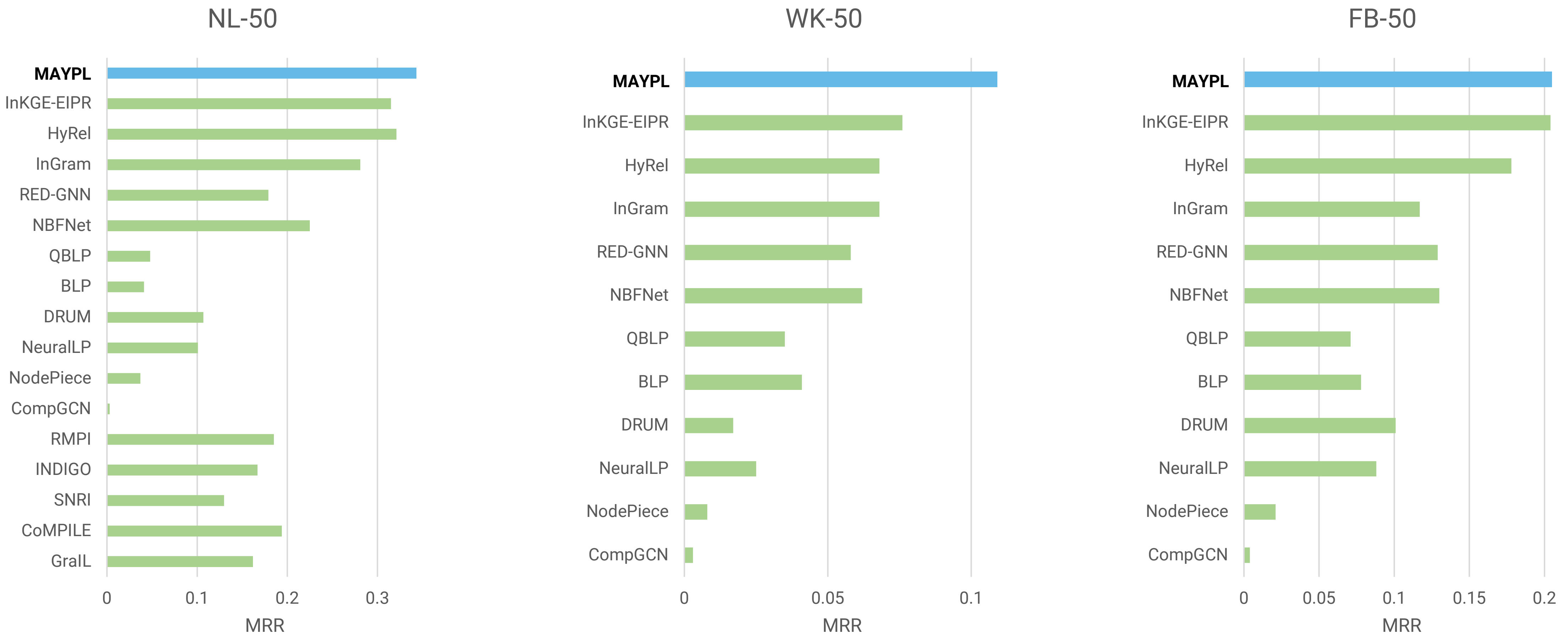
- Datasets
  - **3 Transductive HKG** datasets: WD50K, WikiPeople^{-}, WikiPeople
  - **12 Inductive KG** datasets from InGram (e.g., NL-100, WK-100, FB-100)
  - **4 Inductive HKG** datasets: WD20K(100)v1, WD20K(100)v2, WP-IND, MFB-IND

- Baselines
  - **41 knowledge graph completion methods**, compared with different baseline methods on different datasets

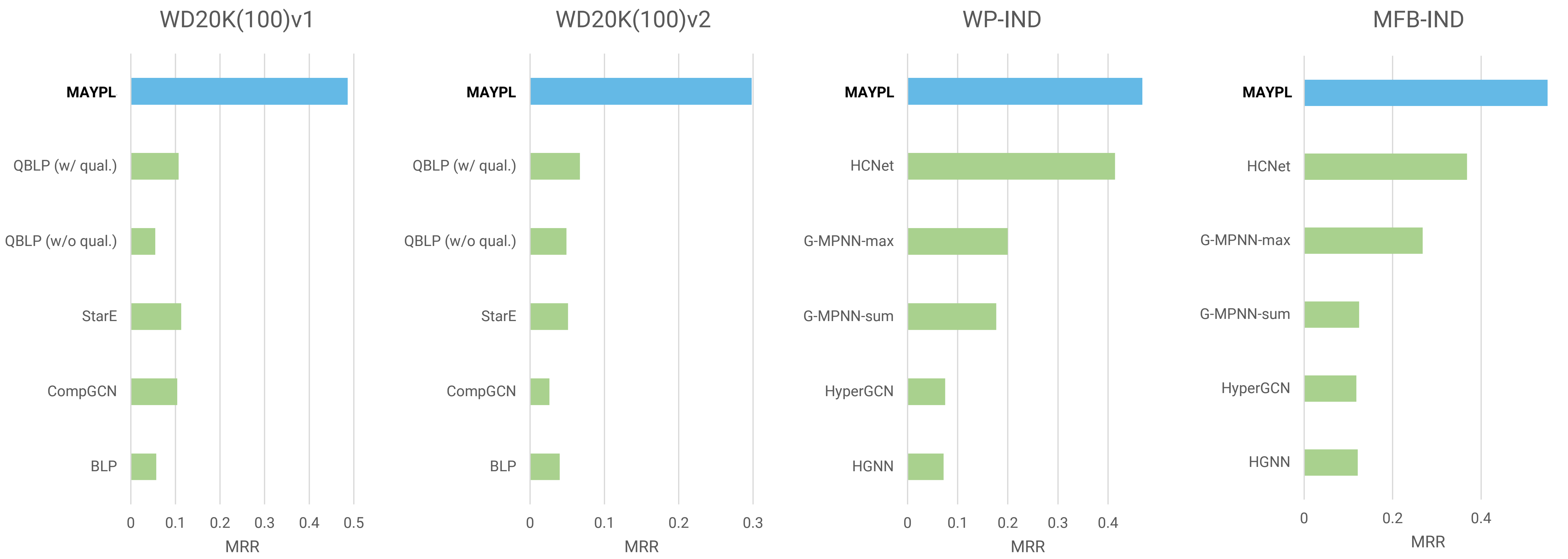
03 Transductive Link Prediction on HKGs



03 Inductive Link Prediction on KGs



03 Inductive Link Prediction on HKGs



03 Conclusion

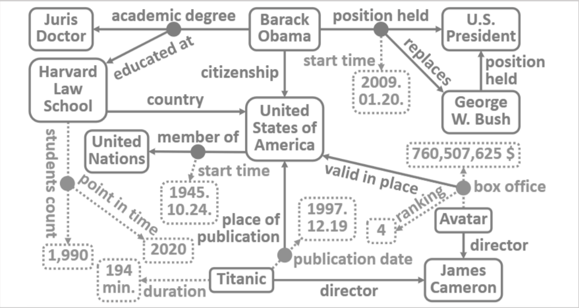
- Employing an **HKG's structure is crucial** for HKG reasoning
  - Thoroughly learning and exploiting the structure of an HKG is necessary and sufficient for learning representations on HKGs
- Propose **MAYPL**, the first structure-oriented representation learning method for HKGs
  - Can be applied in both transductive and inductive settings
- MAYPL can make inductive inferences with new entities and relations by **learning how to compute representations based solely on the structure** of a given HKG
- MAYPL outperforms **41 different methods on 19 benchmark datasets** in varied settings
  - Transductive link prediction on HKGs, inductive link prediction on KGs, and inductive link prediction on HKGs

I

Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers

Chanyoung Chung[‡], Jaejun Lee[‡], and Joyce Jiyoung Whang*

KDD 2023

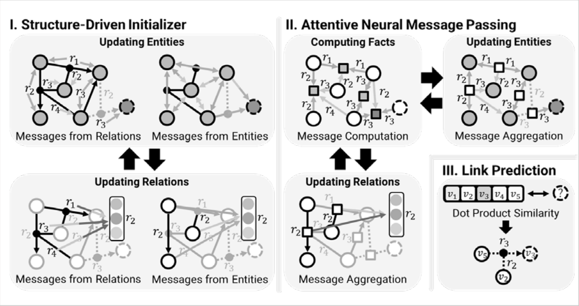


II

Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs

Jaejun Lee and Joyce Jiyoung Whang*

ICML 2025

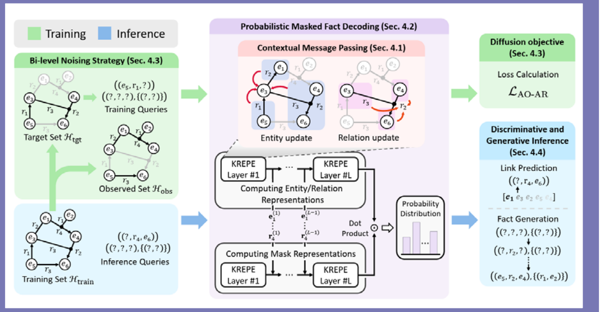


III

Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion

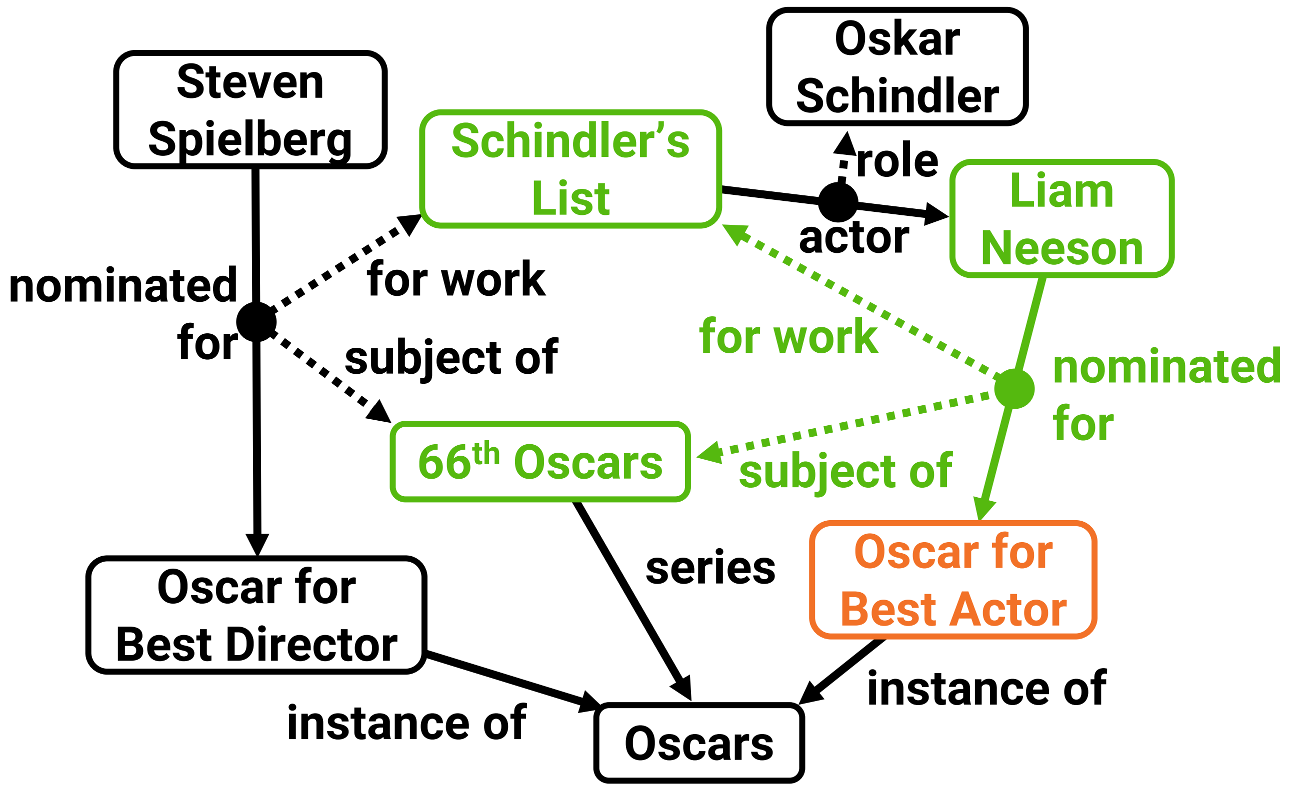
Jaejun Lee, Seheon Kim, and Joyce Jiyoung Whang*

ICML 2026



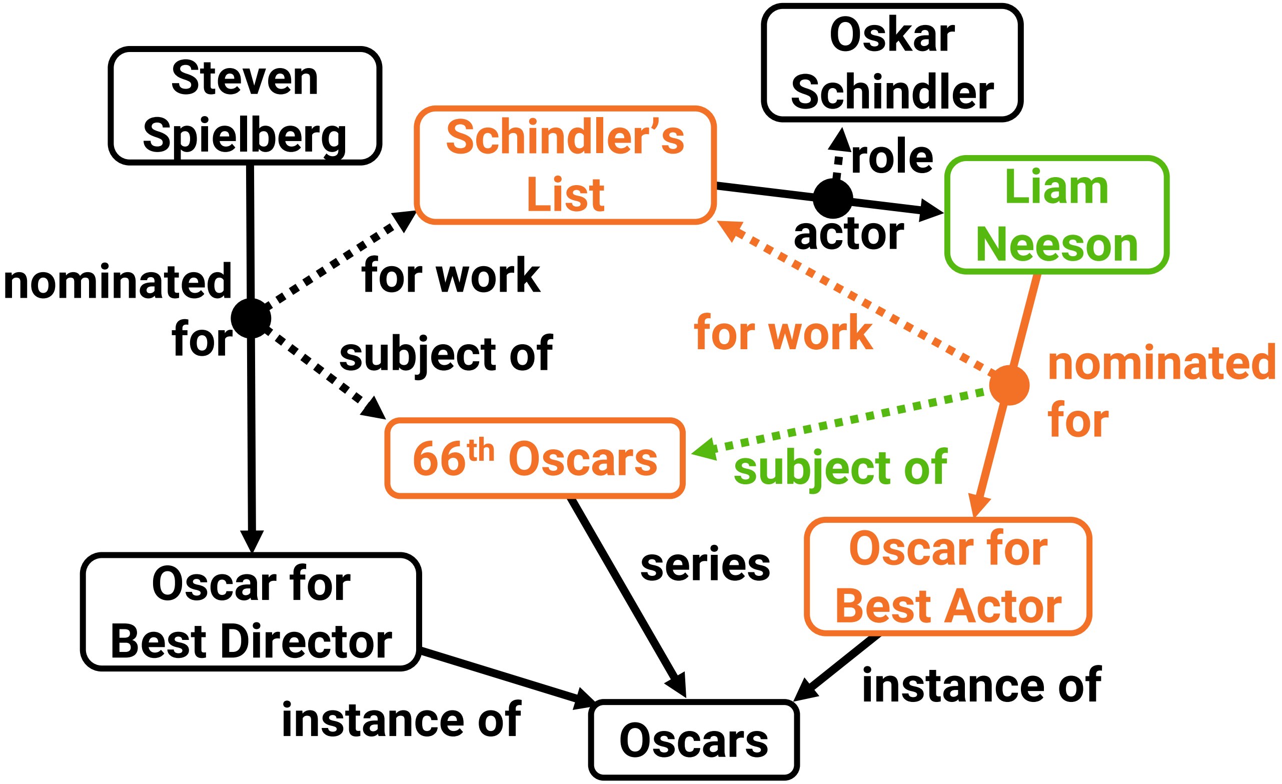
04 Link Prediction on HKGs

((Liam Neeson, nominated for, ?),
{(for work, Schindler's List), (subject of, 66th Oscars)})



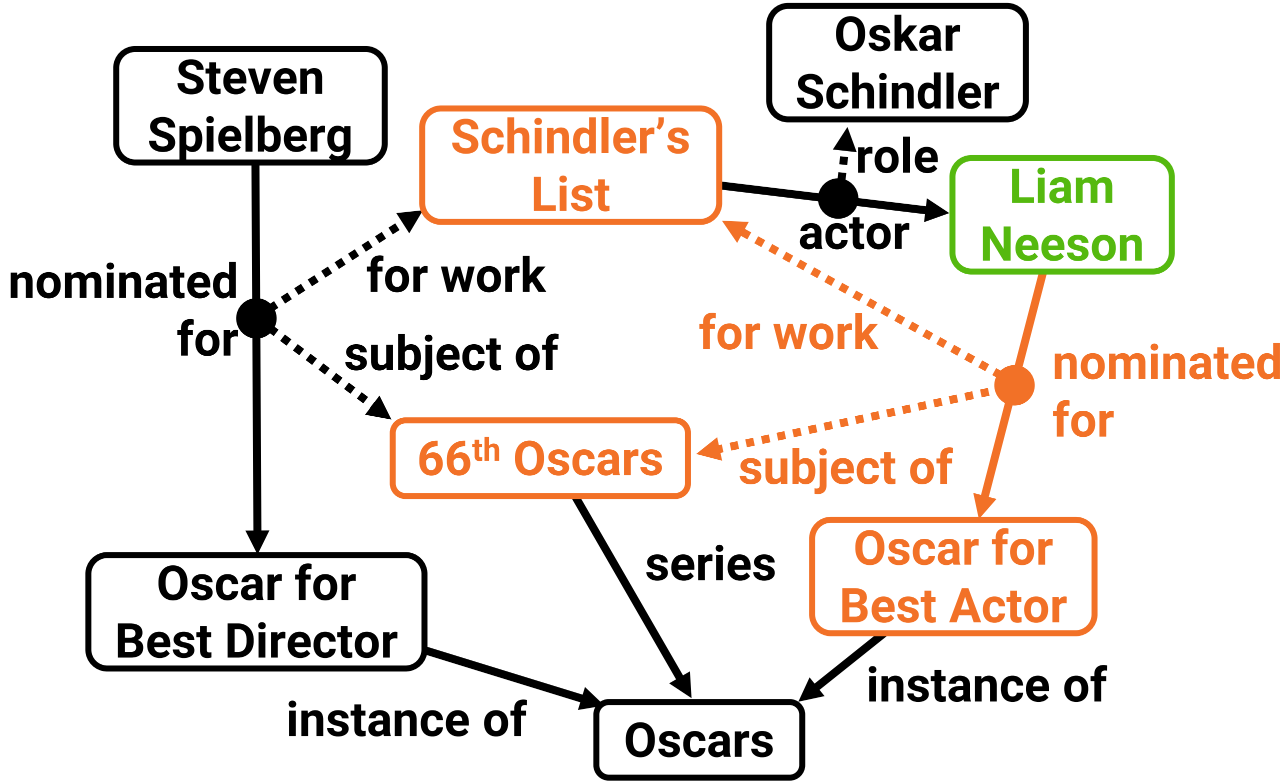
Fact Generation on HKGs

((Liam Neeson, ? , ?),
{(? , ?), (subject of, ?)})



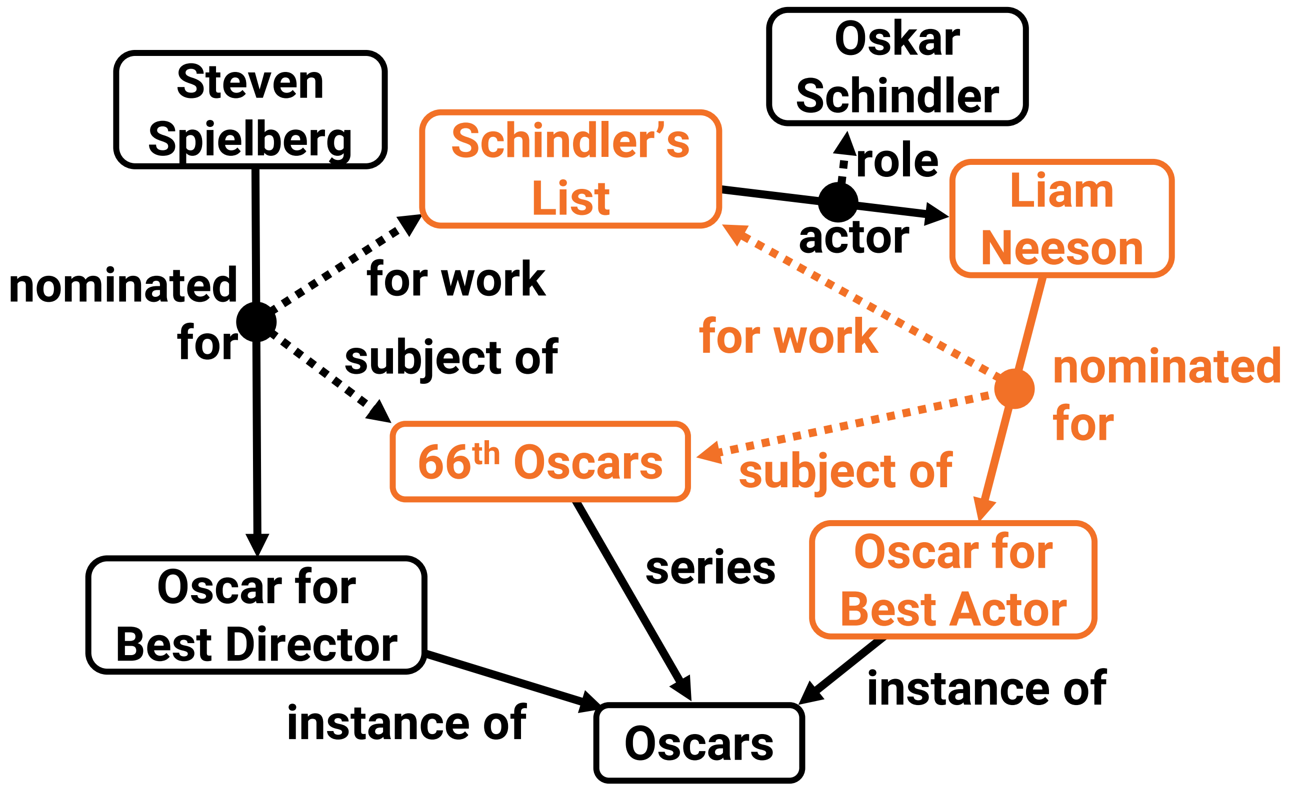
04 Fact Generation on HKGs

((Liam Neeson, ? , ?),
{(? , ?), (? , ?)})

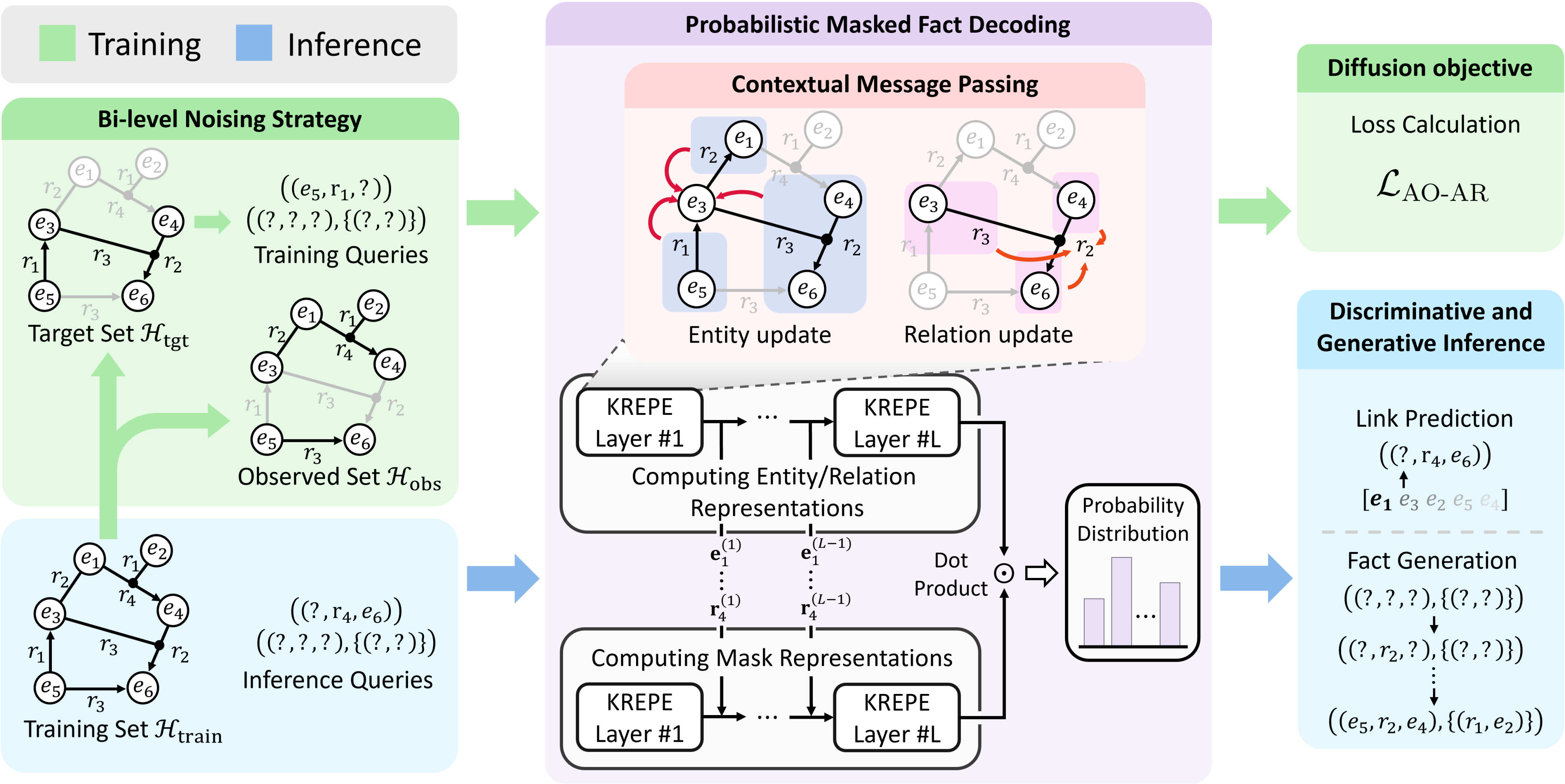


04 Fact Generation on HKGs

((? , ? , ?),
{(? , ?), (? , ?)})



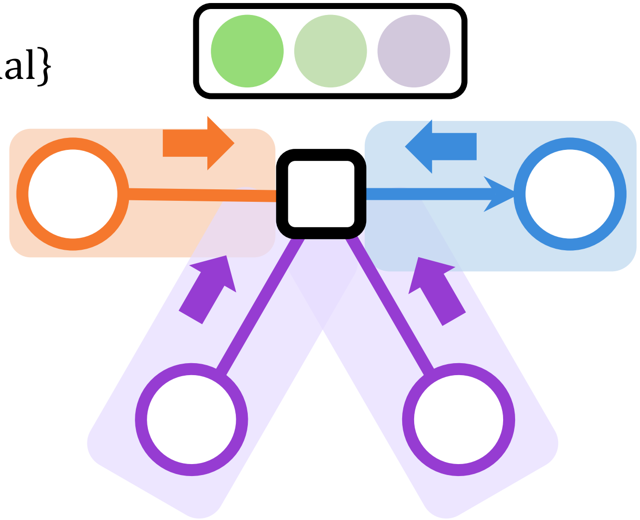
04 KREPE: Contextual HKG Representation Learning via Masked Discrete Diffusion



04 Contextual Message Passing

- Updates entity and relation representations by aggregating messages from the facts they belong to while explicitly excluding their own information from those facts' messages
  - Forces the representations to rely solely on the surrounding context in the facts

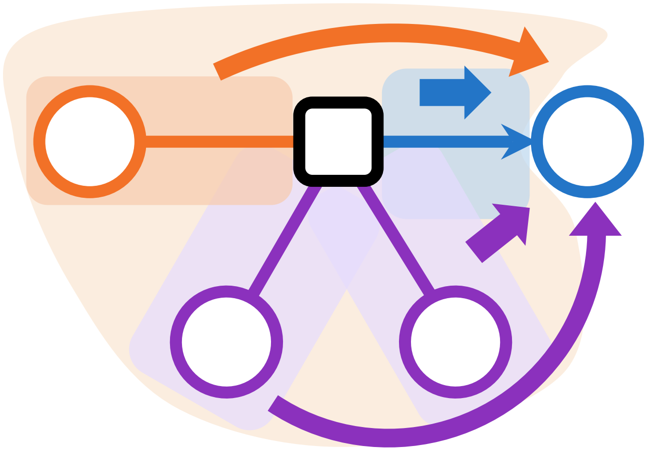
- The representation of a fact $\xi = ((h, r, t), \mathcal{Q}_\xi)$ is obtained via the sum of the representations of its constituent relation-entity pairs
  - A fact is decomposed into a set of its constituent relation-entity pairs
  - The representation of a relation-entity pair (p, e) with role $\rho \in \{\text{head}, \text{tail}, \text{qual}\}$ is computed by: $\mathbf{z}_{(p,e)}^{(l)} = \boldsymbol{W}_\rho^{(l)}[\mathbf{p}^{(l-1)}; \mathbf{e}^{(l-1)}]$
  - $\mathbf{z}_\xi^{(l)} = \mathbf{z}_{(r,h)}^{(l)} + \mathbf{z}_{(r,h)}^{(l)} + \sum_{(k,v)\in\mathcal{Q}_\xi} \mathbf{z}_{(k,v)}^{(l)}$



04 Contextual Message Passing

- Updates entity and relation representations by **aggregating messages from the facts they belong to** while explicitly **excluding their own information** from those facts' messages
  - Forces the representations to rely solely on the surrounding **context** in the facts

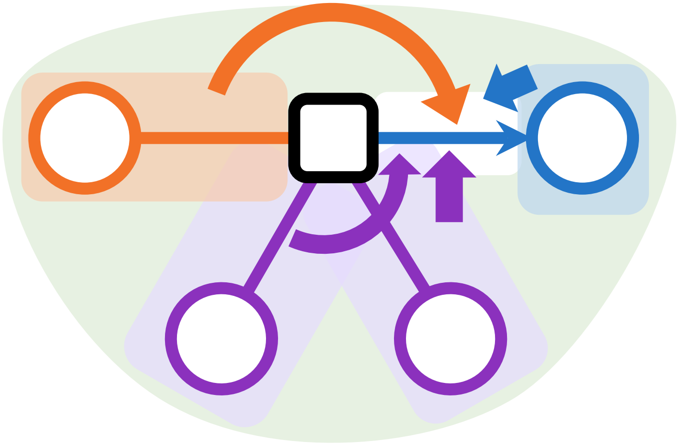
- To update a component within a fact, we compute a **context message** that encodes **all the other components in the fact**
  - The context messages from fact ξ to its components e and p in the constituent air (p, e) are computed by:
    - $\mathbf{m}_{\xi \to (p,e)}^{(l)} = \mathrm{MLP}^{(l)}\left(\frac{\mathbf{z}_{\xi}^{(l)} - \mathbf{z}_{(p,e)}^{(l)}}{n_{\xi}+1}\right)$
    - $\mathbf{m}_{\xi \to e}^{(l)} = \boldsymbol{W}_{\mathrm{ENT}}^{(l)}\left[\mathbf{m}_{\xi \to (p,e)}^{(l)}; \mathbf{p}^{(l-1)}\right]$
    - $\mathbf{m}_{\xi \to p}^{(l)} = \boldsymbol{W}_{\mathrm{REL}}^{(l)}\left[\mathbf{m}_{\xi \to (p,e)}^{(l)}; \mathbf{e}^{(l-1)}\right]$



04 Contextual Message Passing

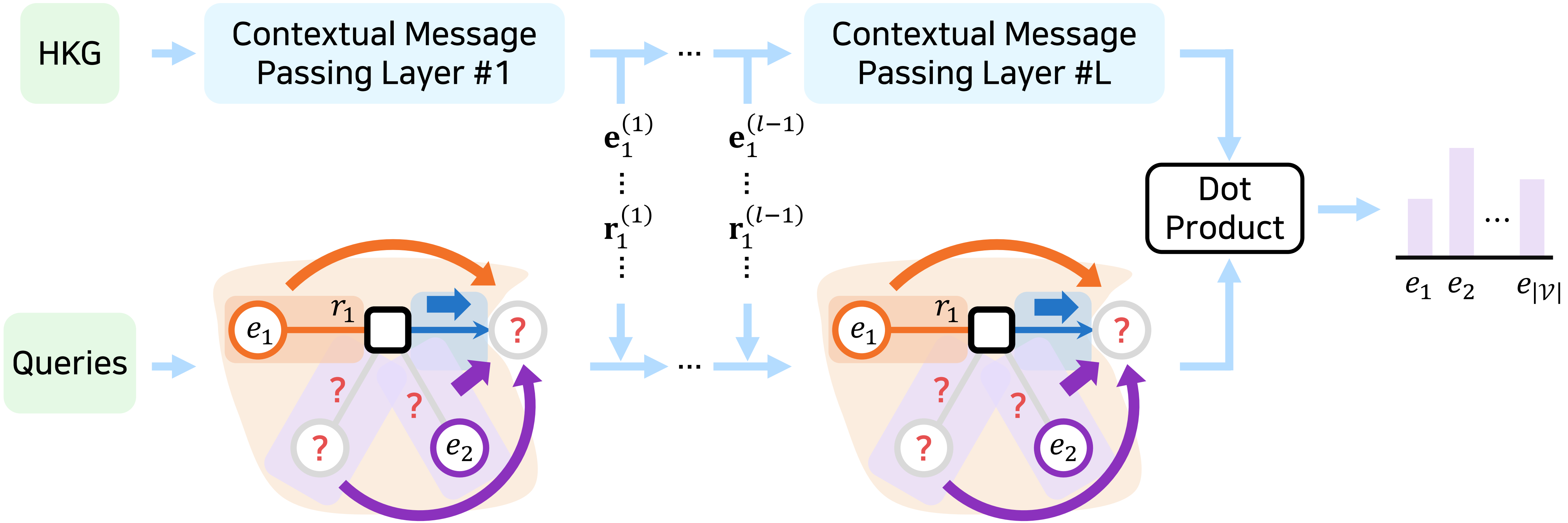
- Updates entity and relation representations by **aggregating messages from the facts they belong to** while explicitly **excluding their own information** from those facts' messages
  - Forces the representations to rely solely on the surrounding **context** in the facts

- To update a component within a fact, we compute a **context message** that encodes **all the other components in the fact**
  - The context messages from fact ξ to its components e and p in the constituent air (p, e) are computed by:
    - $\mathbf{m}_{\xi \to (p,e)}^{(l)} = \mathrm{MLP}^{(l)}\left(\frac{\mathbf{z}_{\xi}^{(l)} - \mathbf{z}_{(p,e)}^{(l)}}{n_{\xi}+1}\right)$
    - $\mathbf{m}_{\xi \to e}^{(l)} = \boldsymbol{W}_{\mathrm{ENT}}^{(l)}\left[\mathbf{m}_{\xi \to (p,e)}^{(l)}; \mathbf{p}^{(l-1)}\right]$
    - $\mathbf{m}_{\xi \to p}^{(l)} = \boldsymbol{W}_{\mathrm{REL}}^{(l)}\left[\mathbf{m}_{\xi \to (p,e)}^{(l)}; \mathbf{e}^{(l-1)}\right]$



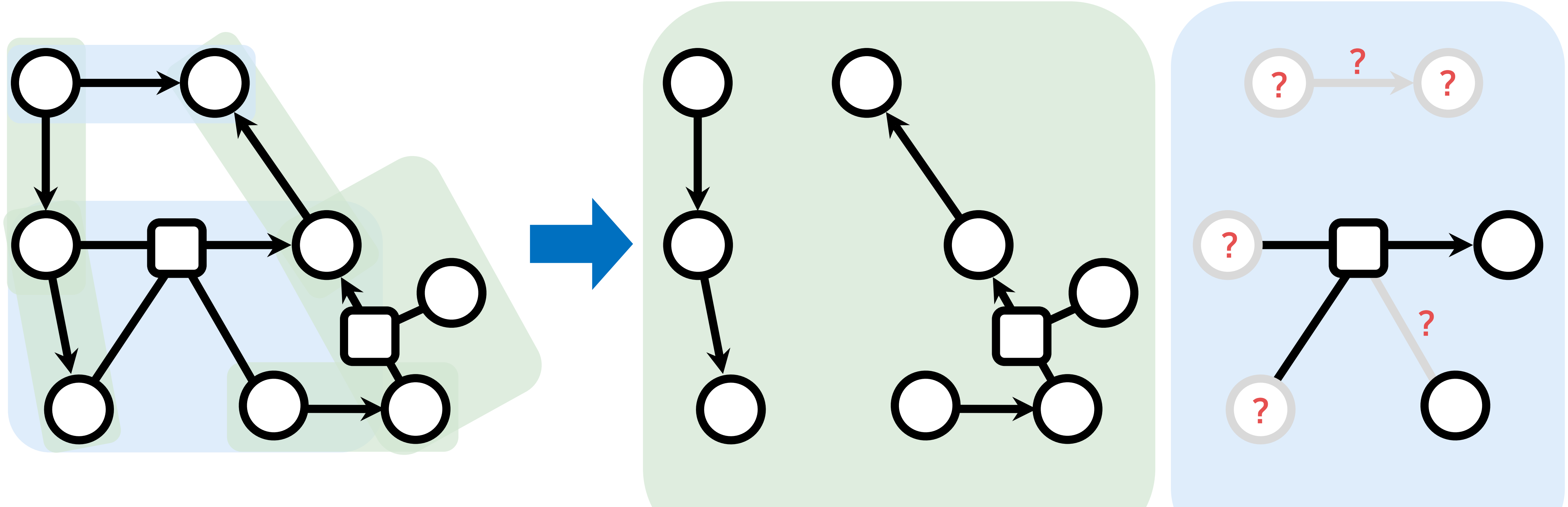
04 Probabilistic Masked Fact Decoding

- Treats each masked entity and relation as a vanilla entity and relation, respectively, and updates their representations via contextual message passing
  - A masked fact ζ is used only to update its masked components



04 Bi-level Noising Strategy

- Simulates diverse conditional generation scenarios by varying both the observed structure and the query during training



Training Set

Observed set

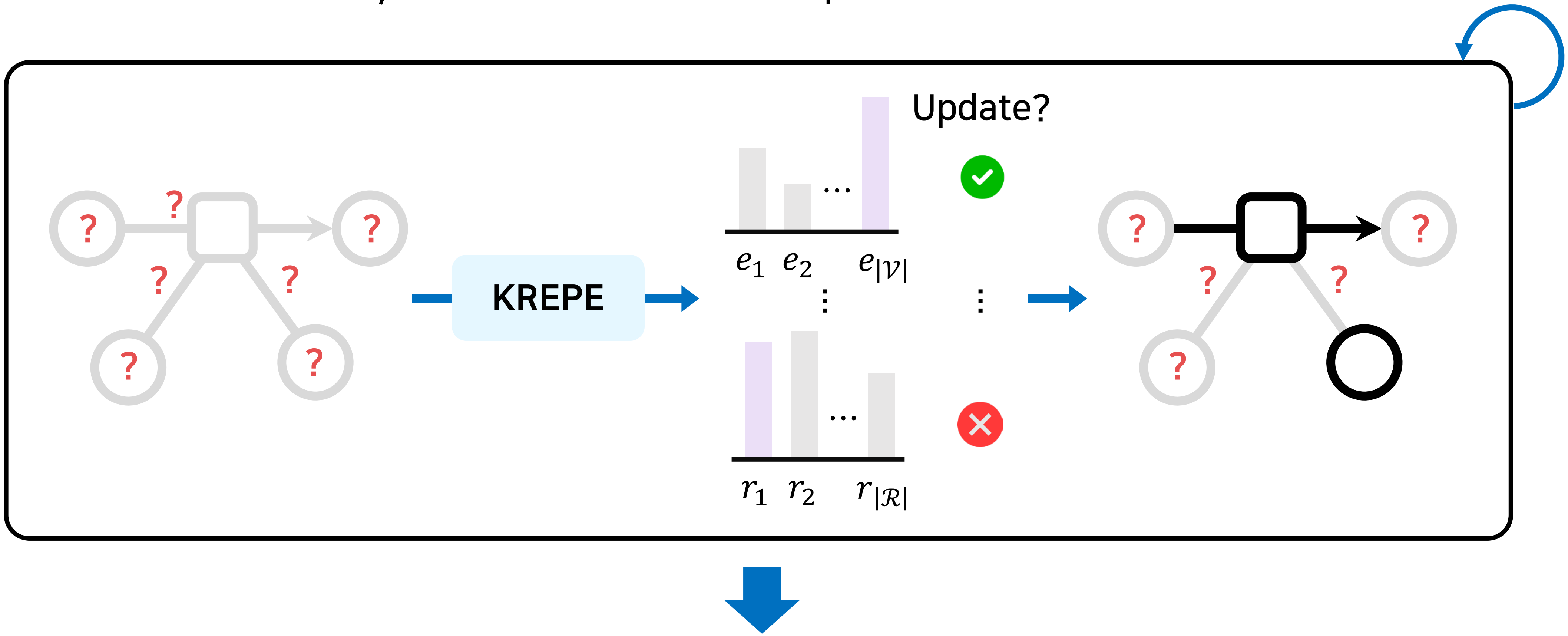
Training Queries

04 Discriminative and Generative Inference

- Link prediction: leverages the computed probability distribution to rank the candidates

04 Discriminative and Generative Inference

- Fact Generation: Iteratively decodes the masked components



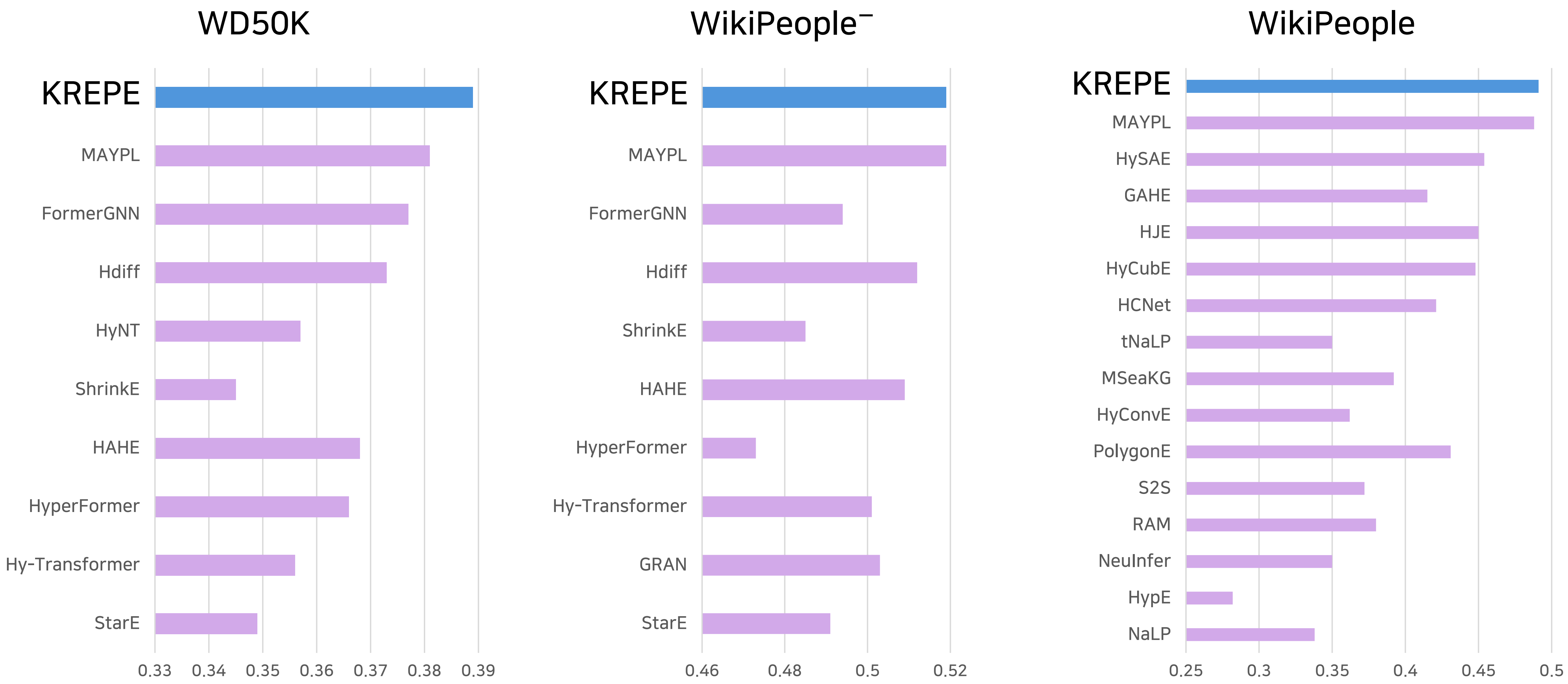
((Gigi, awarded, Oscars Best Adapted Screenplay), {(subject of, 31st Oscars), (winner, A. J. Lerner)})

04 Experimental Results

- Datasets
  - 3 HKG datasets: WD50K, WikiPeople$^-$, WikiPeople
- Baselines
  - 25 link prediction methods and seven competitive fact generation baselines
    - Fact generation baselines include two variations of discriminative HKG models and five LLM-based approaches
- Fact Generation Settings
  - Scratch: generating a valid fact without any known component
  - Targeted: requiring the generation of a fact containing a single given component called a target
  - Arbitrary Masking: generating a fact from an arbitrarily masked input

04

Link Prediction Results



04 Fact Generation Baselines

- Iterative Prediction (based on an HKG method)
  - Missing components are randomly initialized, and sequentially masked and replaced by an HKG model
- Gibbs Sampling (based on an HKG method)
  - Missing components are randomly initialized, and iteratively masked and replaced by resampling from the distribution predicted by an HKG model
- Re-ranking (based on an HKG method and an LLM)
  - Missing components are initialized using a training fact that matches the query. Each missing component is sequentially masked, and an LLM selects the best candidate from the top-5 predictions of an HKG model

04 Fact Generation Baselines

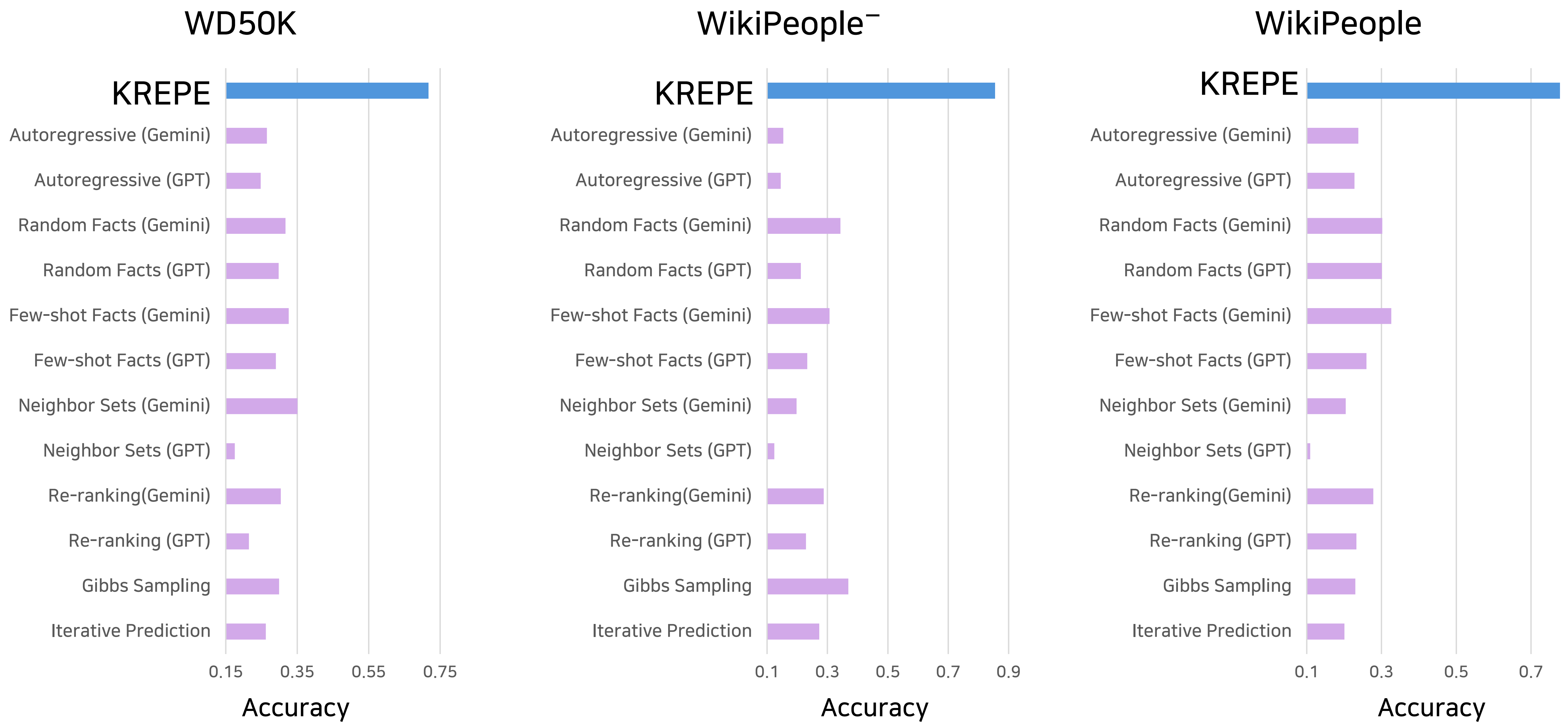
- Neighbor Sets (based on an LLM)
  - An LLM generates each fact prompted by the neighbors of known components in the query

- Few-shot Facts (based on an LLM)
  - An LLM generates each fact prompted by 30 training facts containing at least one known component in the query

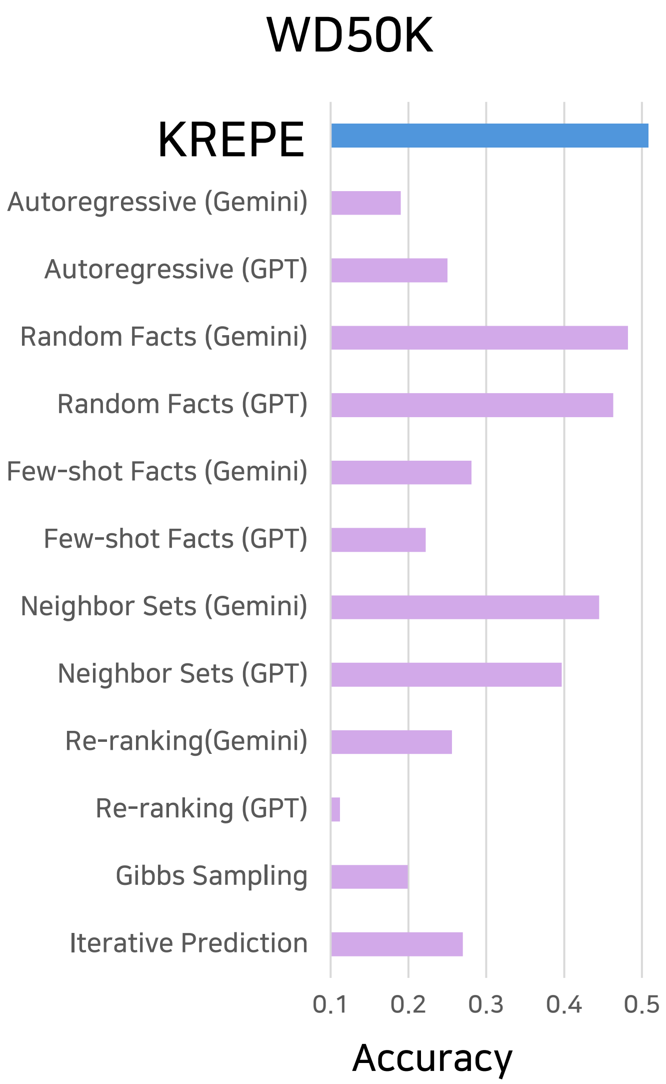
- Random Facts (based on an LLM)
  - An LLM is prompted to generate facts by utilizing 1,000 randomly sampled training facts

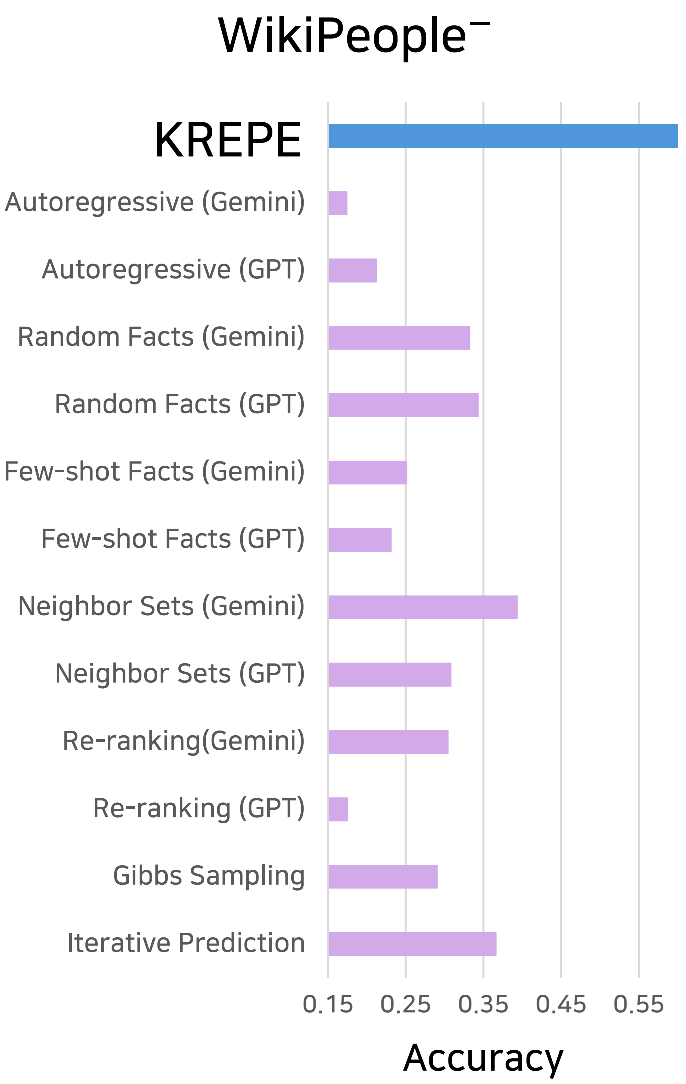
- Autoregressive (based on an LLM)
  - An LLM generates components step-by-step, prompted to utilize training facts containing the previously predicted component for generation

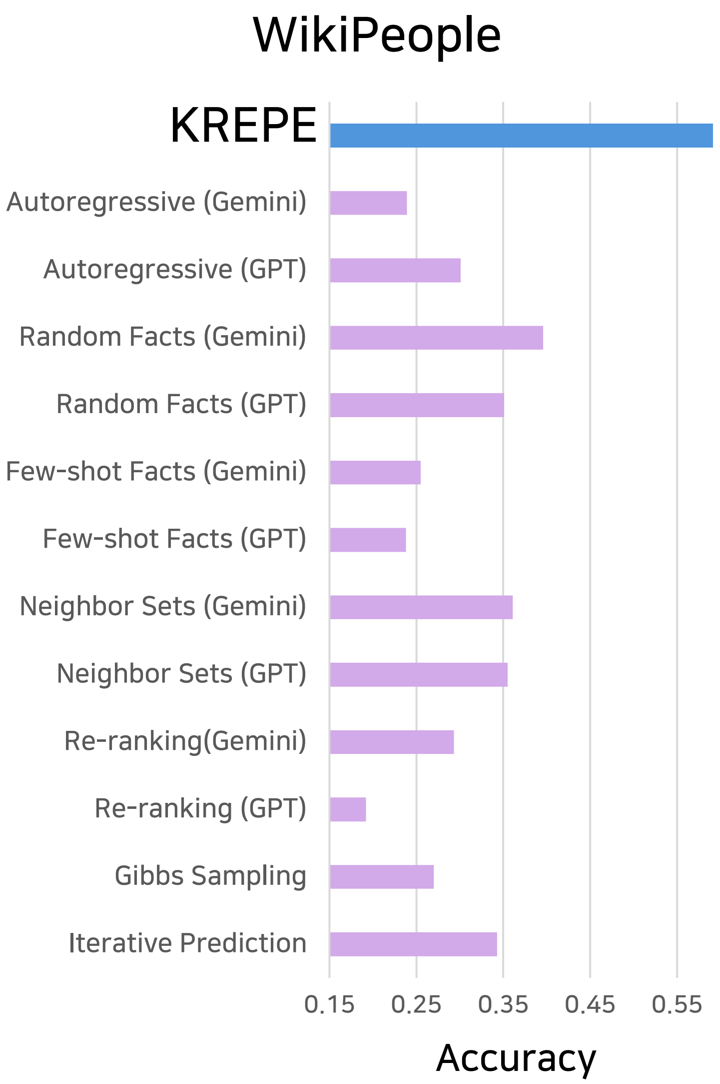
04 Fact Generation Results - Scratch



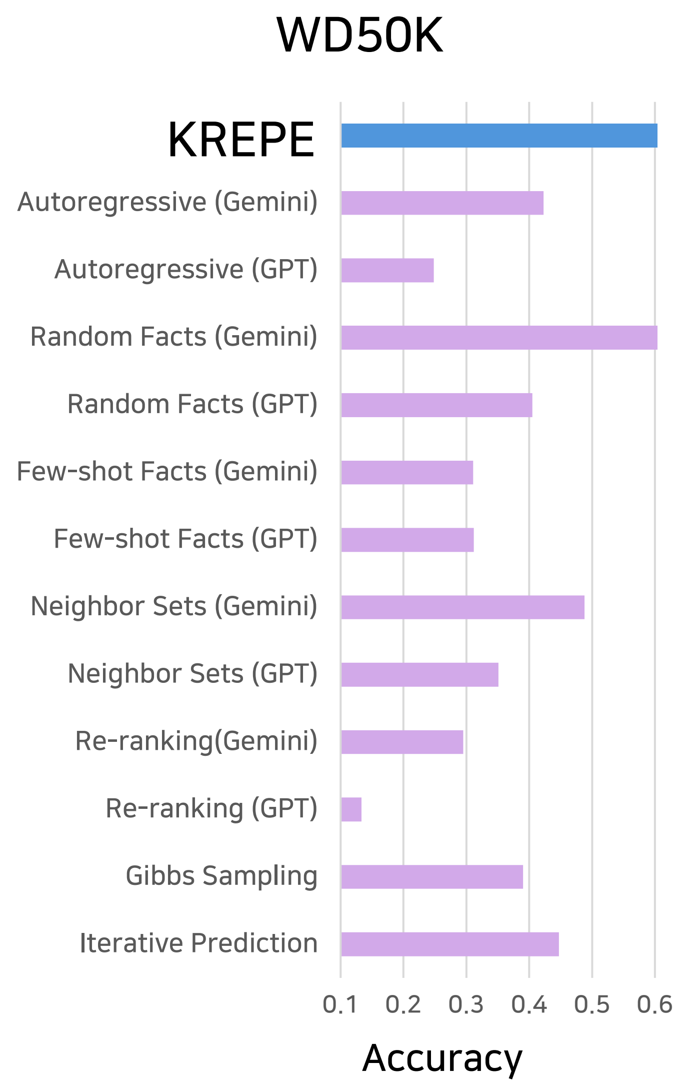
04 Fact Generation Results - Targeted

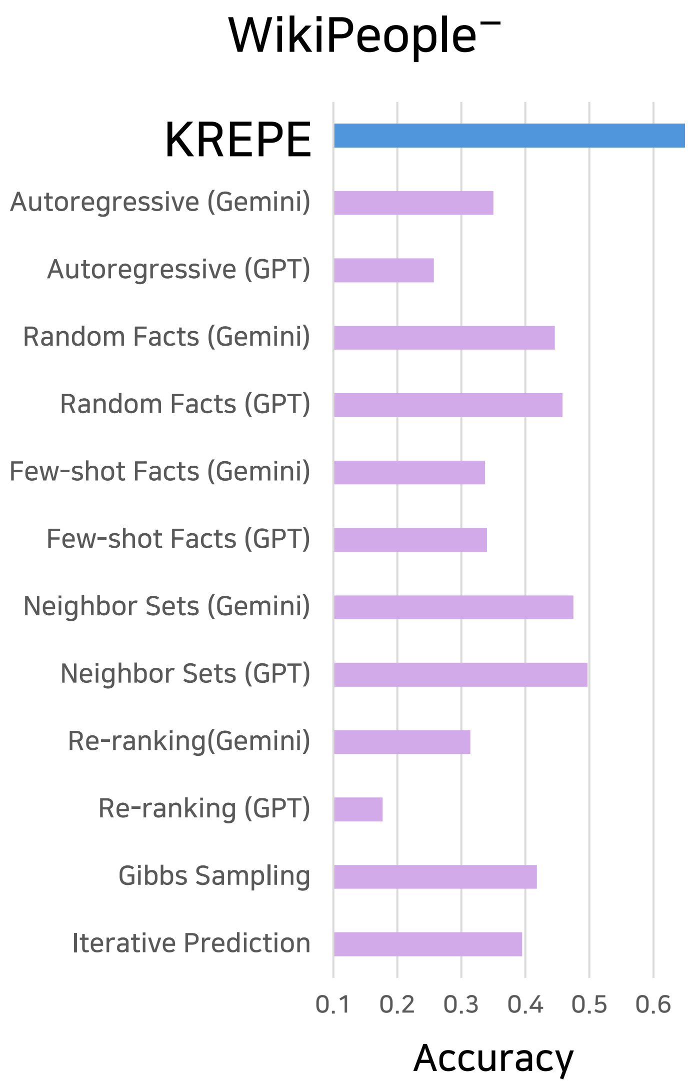


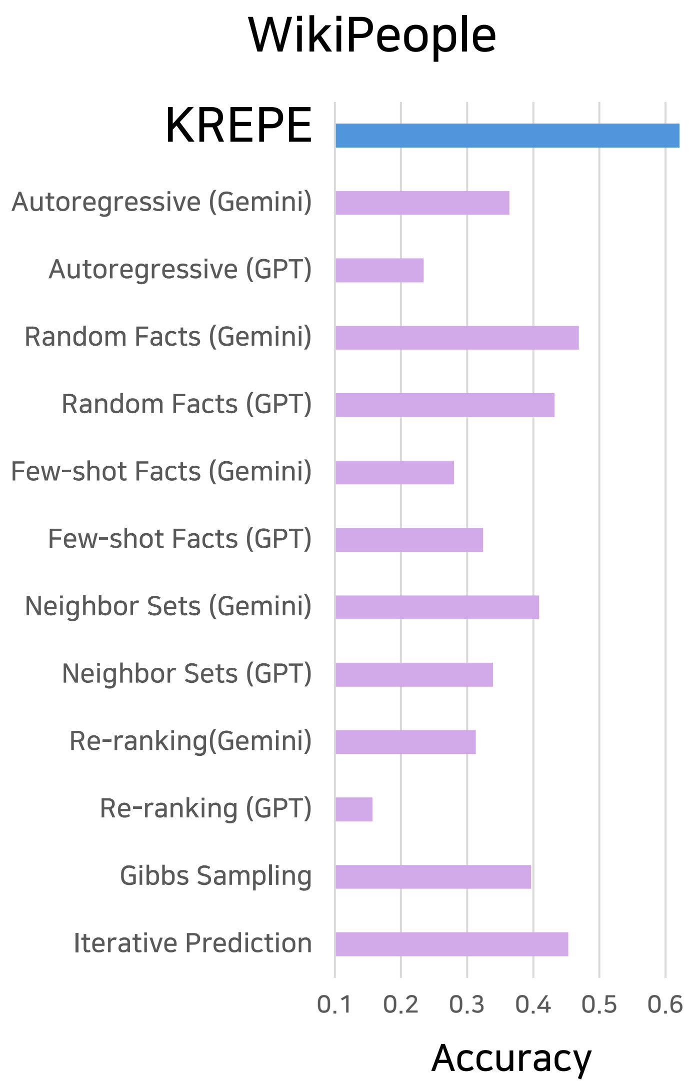




04 Fact Generation Results - Arbitrary







04 Conclusion

- Introduce **Fact Generation**, a novel task that aligns closely with real-world scenarios where **multiple components of a fact may be missing**
    - Advances beyond the traditional fill-in-the-blank link prediction in the HKG domain

- Propose **KREPE**, the first **HKG representation learning framework equipped with generative capabilities** via a masked discrete diffusion objective

- KREPE not only establishes a new state-of-the-art link prediction performance but also significantly outperforms LLM-based baselines in fact generation

05 References

- Some slides are made based on the following references.

  - P. Rosso et al., “Beyond Triplets: Hyper-Relational Knowledge Graph Embedding for Link Prediction”, TheWebConf, 2020.
  - M. Galkin et al., “Message Passing for Hyper-Relational Knowledge Graphs”, EMNLP, 2020.
  - Q. Wang et al., “Link Prediction on N-ary Relational Facts: A Graph-based Approach”, ACL Findings, 2020.
  - H. Luo et al., “HAHE: Hierarchical Attention for Hyper-Relational Knowledge Graphs in Global and Local Level”, ACL, 2023.
  - B. Xiong et al., “Shrinking Embeddings for Hyper-Relational Knowledge Graphs”, ACL, 2023.
  - C. Chung et al., “Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers”, KDD, 2023.
  - Z. Hu et al., “HyperFormer: Enhancing Entity and Relation Interaction for Hyper-Relational Knowledge Graph Completion”, CIKM, 2023.
  - J. Lee and J. J. Whang, “Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs”, ICML, 2025.
  - J. Lee, S. Kim, and J. J. Whang, “Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion”, ICML 2026.