entropy

MDPI

*Article*

# IMF: Interpretable Multi-Hop Forecasting on Temporal Knowledge Graphs

**Zhenyu Du [†], Lingzhi Qu [†], Zongwei Liang [*], Keju Huang, Lin Cui and Zhiyang Gao**

College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China; dzy17@nudt.edu.cn (Z.D.)

* Correspondence: zongweiliang_nudt@163.com

† These authors contributed equally to this work.

**Abstract:** Temporal knowledge graphs (KGs) have recently attracted increasing attention. The temporal KG forecasting task, which plays a crucial role in such applications as event prediction, predicts future links based on historical facts. However, current studies pay scant attention to the following two aspects. First, the interpretability of current models is manifested in providing reasoning paths, which is an essential property of path-based models. However, the comparison of reasoning paths in these models is operated in a black-box fashion. Moreover, contemporary models utilize separate networks to evaluate paths at different hops. Although the network for each hop has the same architecture, each network achieves different parameters for better performance. Different parameters cause identical semantics to have different scores, so models cannot measure identical semantics at different hops equally. Inspired by the observation that reasoning based on multi-hop paths is akin to answering questions step by step, this paper designs an Interpretable Multi-Hop Reasoning (IMR) framework based on consistent basic models for temporal KG forecasting. IMR transforms reasoning based on path searching into stepwise question answering. In addition, IMR develops three indicators according to the characteristics of temporal KGs and reasoning paths: the question matching degree, answer completion level, and path confidence. IMR can uniformly integrate paths of different hops according to the same criteria; IMR can provide the reasoning paths similarly to other interpretable models and further explain the basis for path comparison. We instantiate the framework based on common embedding models such as TransE, RotatE, and ComplEx. While being more explainable, these instantiated models achieve state-of-the-art performance against previous models on four baseline datasets.

**Keywords:** temporal knowledge graphs; forecasting; interpretable reasoning

check for updates

**Citation:** Du, Z.; Qu, L.; Liang, Z.; Huang, K.; Cui, L.; Gao, Z. IMF: Interpretable Multi-Hop Forecasting on Temporal Knowledge Graphs. *Entropy* **2023**, *25*, 666. https://doi.org/10.3390/e25040666

Academic Editors: Zoran H. Perić, Vlado Delić, Vladimir Despotovic and Zoran Ognjanović

Received: 20 February 2023
Revised: 26 March 2023
Accepted: 8 April 2023
Published: 16 April 2023

## 1. Introduction

Knowledge graphs (KGs) are collections of triples, such as Freebase [1] and YAGO [2]. Temporal KGs introduce a new dimension into static knowledge graphs [3], i.e., a timestamp for each triple to form a quadruple. Although there are billions of triples in temporal KGs, they are still incomplete. These incomplete knowledge bases will lead to limitations in practical applications. Since temporal KGs involve the time dimension, the completion of temporal KGs can be divided into interpolation and forecasting. The former utilizes the facts of all timestamps to predict the triples at a particular moment; the latter employs historical facts to predict future triples. Due to the importance of temporal KG forecasting in event prediction, it has attracted growing attention recently. This paper mainly focuses on temporal KG forecasting.

Most current research on temporal KG completion focuses on interpolation [4–10]. Recently, there have been attempts to investigate temporal KG forecasting [3,4,7,11–13]. According to the interpretability, research on temporal KG forecasting can be divided into two categories. One type is the black-box model, which designs an unexplainable scoring function for quadruples' rationality. The other type is interpretable approaches. CyGNet [11] utilizes one-hop repetitive facts to realize prediction. Its performance is limited by the lack of direct repetitive knowledge of historical moments. xERTR [7], CluSTeR [3], and TITer [14] are all path-based temporal KG forecasting models. xERTR [7] adopts the inference subgraphs to aggregate local information around the question. CluSTeR [3] and TITer [14] manipulate reinforcement learning for the path search and improve the performance through temporal reasoning.

Thus far, however, there has been little discussion on the following two aspects. Firstly, uniformly measuring the paths of different hops requires handling the same semantics equivalently at different hops. Current models utilize separate networks to evaluate paths at different hops. Although each hop's network has the same architecture, each network acquires different parameters for better performance. Different parameters cause identical semantics to have different scores, so current models cannot truly compare multi-hop paths according to the same criteria. For example, xERTR [7] simply gathers the scores of different paths for comparison, which is mainly based on training datasets. Secondly, although current models can provide reasoning paths, the comparison of paths operates in a black-box fashion. The interpretability of the current models means providing the reasoning paths, which is an essential property of path-based models. These models lack an explanation of the preference for various paths, i.e., they cannot provide the basis for path comparison.

In practice, forecasting based on path searching aims to find the appropriate multi-hop paths, the combination of whose relations is equivalent to the question's relation. As we observe, reasoning based on multi-hop paths is akin to stepwise question answering. Inspired by stepwise question answering, this paper designs a new Interpretable Multi-Hop Reasoning (IMR) framework based on consistent basic models, which can uniformly integrate the paths of different hops and perform more interpretable reasoning.

The primary pathway of IMR can be as follows. IMR first transforms reasoning based on path searching into stepwise question answering based on basic KG embedding models [1,15–18] and IRN [19]. This framework calculates the unanswered parts of questions after each hop as the new question for the next hop during the stepwise question answering, which is named the remainder of questions in this paper. Moreover, IMR designs three indicators based on the unanswered parts of questions and the inferred tails: the query matching degree, answer completion level, and path confidence. The query matching degree, i.e., the matching degree between the reasoning tails and the original questions, measures the rationality of the new quadruples. The answer completion level, i.e., the matching degree between the relations of paths and that of the questions, measures the answer's completeness. Path confidence, i.e., the difference between the same entities with different timestamps, measures the reliability of the reasoning paths. IMR achieves the unified scoring of multi-hop paths and better explainable reasoning simultaneously with these indicators' combination.

The major contributions of this work are as follows. (1) A new Interpretable Multi-Hop Reasoning framework (IMR) is proposed in this paper, which provides a new framework for the specific design of forecasting models. Furthermore, IMR defines three indicators: the query matching degree, answer completion level, and path confidence. (2) Unlike other models that cannot measure the paths of different hops uniformly, IMR can measure the paths of different hops according to the same criteria and utilize multi-hop paths for inference. (3) IMR can provide reasoning paths similarly to other interpretable models and further explain the basis for path comparison. (4) Based on basic embedding models, IMR is instantiated as the specific model. Experiments on four benchmark datasets show that these instantiated models achieve state-of-the-art performance against previous models.

## 2. Related Work

**Static KG reasoning.** Knowledge graph reasoning based on representation learning has been widely investigated by scholars. These approaches to reasoning can be categorized into geometric models [1,17,20–22], tensor decomposition models [15,16,18,23], and deep learning models [24–26]. In recent years, some scholars have attempted to introduce GCN into knowledge graph reasoning [27], which can improve the performance of basic models. Some other scholars focus on multi-hop reasoning with symbolic inference rules learned from relation paths [28,29]. The above methods are all designed for static KGs, making it challenging to deal with temporal KG reasoning.

**Temporal KG reasoning.** Temporal KGs import the time dimension into static KGs, which makes the facts of a specific timestamp extremely sparse. The temporal KG reasoning task can be divided into two categories: reasoning about historical facts [4–8,30], i.e., interpolation on temporal KGs, and reasoning about future facts [3,4,7,11], i.e., forecasting on temporal KGs. The former predicts the missing facts of a specific historical moment based on the facts of all moments, and the latter predicts future events based only on the past facts. There are many studies on the task of temporal KG interpolation. However, these studies are all black-box models, which cannot explain predictions. Most of the proposed models for temporal KG forecasting are also black-box models. BoxTE [31] utilizes BoxEmbedding for temporal KG forecasting, which is expressive and possesses an inductive capacity. Recently, xERTR [7], CluSTer [3], and TITer [14] were shown to explain predictions to some extent. These models can provide the reasoning paths for the predictions. However, both models cannot truly handle multi-hop paths crossing the same criteria, which is more similar to the weighted combination. xERTR and TiTer combine the scores of paths with different hops by training weights. Experiments show that CluSTeR performs worse on paths with multiple hops than on paths with only one hop.

Most current temporal KG forecasting models are black-box models. Only some models can provide reasoning paths for prediction. Moreover, none of them can explain how path comparisons work and none of them can integrate paths of different hops uniformly.

## 3. Preliminaries

**The task of temporal KG forecasting.** Suppose that $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ represent the entity set, predicate set, and timestamp set, respectively. The temporal KG is a collection of quadruples, which can be expressed as

$$K = \{(e_s, r, e_o, t), e_s, e_o \in \mathcal{E}, r \in \mathcal{R}, t \in \mathcal{T}\} \tag{1}$$

$(e_s, r, e_o, t)$ denotes a quadruple; $e_s$ and $e_o$ represent the subject and object, respectively. $r$ represents the relation, and $t$ represents the time that the quadruple occurs. Suppose that facts happening before the selected time $t_k$ can be expressed as

$$G_{t_k} = \left\{ \left(e_i, r, e_j, t_i\right) \in K | t_i < t_k \right\} \tag{2}$$

Temporal KG forecasting predicts future links based on past facts. This means that its foundation is the process of predicting $e_o$ based on a question $\left(e_s, r_q, ?, t_q\right)$ and the previous facts $G_{t_q}$, where $r_q, t_q$ denote the relation and timestamp of the question. Temporal KG forecasting involves ranking all entities of the specific moment and obtaining the preference for prediction.

**Temporal KG forecasting based on paths.** Knowledge graph embedding associates the entities $e \in \mathcal{E}$ and relations $r \in \mathcal{R}$ with vectors $\mathbf{e}, \mathbf{r}$. Different from static KGs, the entities in temporal KGs contain time information. The entity may contain different attributes at different moments. In order to better characterize the entity in temporal KGs, we associate each entity $e$ with a specific time label $t_i \in \mathcal{T}$, so the entity $e$ can be depicted as $e^{t_i}$ and its embedding can be denoted as $\mathbf{e^{t_i}}$. The set of quadruples directly associated with $e_s^{t_i}$, which can be defined as the 1-hop paths associated with $e_s^{t_i}$, can be expressed as $P_{(e_s,t_i)} = \left\{ \left(e_s, r, e_j, t_k\right) | \left(e_s, r, e_j, t_k\right) \in G_{t_i} \right\}$, where $e_s, e_j \in \mathcal{E}, r_p \in \mathcal{R}, t_k < t_i \in \mathcal{T}$. In this way, $P_{(e_s,t_i)}$ can represent all associated quadruples. The set of entities directly associated with $e_s^{t_q}$ in the path $P_{(e_s,t_q)}$, i.e., the 1-hop neighbors of $e_s^{t_q}$, can be denoted as $N_{(e_s,t_q)} =$

$\left\{e_i^{t_h} | (e_s, r, e_i, t_h) \in P_{(e_s, t_q)}\right\}$, where $e_s, e_i \in \mathcal{E}, r \in \mathcal{R}, t_h < t_q \in \mathcal{T}$. Given the question $\left(e_s, r_q, ?, t_q\right)$, the forecasting task can be depicted as requesting the entity $e_o$ based on path searching. For example, we search the path with $e_s$ as the starting point:

$$\left(e_s, r_{p_{(1)}}, e_1, t_1\right), \left(e_1, r_{p_{(2)}}, e_2, t_2\right), \ldots, \left(e_{i-1}, r_{p_{(i)}}, e_i, t_i\right) \tag{3}$$

where $r_{p_{(i)}}$ denotes the relations of the $i$th-hop. Thus, answers to the question may be $e_1, e_2, e_3, \ldots, e_i$, and the corresponding inference hop is $1, 2, 3, \ldots, i$, respectively. Moreover, $e_{s_{(i)}}, r_{q_{(i)}}$ denotes the remaining (or unanswered) subjects and relations of questions after the $i$th-hop paths, which will be explained in Section 4.3.2.

**Uniformly measuring paths of different hops.** Uniformly measuring paths of different hops requires models scoring paths of different hops according to the same criteria. For example, given question $\left(e_s, r_q, ?, t_q\right)$ and the searched 1-hop path $\left(e_s, r_p, e_1, t_1\right)$, the score obtained for the searched 1-hop path is $f$. If we find no path during the first hop, the original question is left to the second hop to solve. Thus, the remaining question (unanswered question) for the second hop is still $\left(e_s, r_q, ?, t_q\right)$. When the path searched at the second hop is also $\left(e_s, r_p, e_1, t_1\right)$, the score for the searched path at the second hop should also be $f$. As is shown in this example, we should score identical semantics equivalently even under different hops. Moreover, the equal comparison of paths provides the basis for the interpretability of path comparison. This attribute constrains models to have an identical scoring mechanism at each hop, i.e., each hop's separate networks for the models based on neural networks should have the same parameters. However, only IMR can meet the attribute.

**Fact matching based on TransE.** This paper is the first study of the design of interpretable evaluation indicators from the perspective of actual semantics. We instantiate IMR to better illustrate the design pathway and thus choose the basic embedding model TransE as the basis of IMR. In TransE, relations are represented as translations in the embedding space. If the triple $(e_s, r, e_o)$ holds in static KGs, TransE [1] assumes the following relationship.

$$|\mathbf{e}_s + \mathbf{r} - \mathbf{e}_o| = 0 \tag{4}$$

where $\mathbf{e}_s$, $\mathbf{r}$ and $\mathbf{e}_o \in \mathcal{R}^k$, and $k$ denotes the dimension of each vector.

For each quadruple $\left(e_s, r_q, e_o, t_q\right)$ in temporal KGs, the relation $r_q$ can also be taken as the translation from the subject $e_s$ to the object $e_o$, i.e., $\mathbf{e}_s^{t_q} + \mathbf{r}_q = \mathbf{e}_o^{t_q}$. We suppose that when the distance $d$ of quadruples is smaller, the quadruple will be better matched. The distance of the quadruple $\left(e_s, r_q, e_o, t_q\right)$ can be expressed as

$$d = \left|\mathbf{e}_s^{t_q} + \mathbf{r}_q - \mathbf{e}_o^{t_q}\right| \tag{5}$$

The relations in KG embedding models indicate the translations between entities, whose specific design determines the complexity of the indicators designed by IMR. The design route of IMR originates from the perspective of reasoning from actual semantics, which is not limited to specific basic models. The consistent basic model of IMR-TransE is TransE, i.e., all IMR-TransE's specific formulas are based on TransE, which will not be explained below. To limit the length of the paper, we move the details of IMR-TransE and IMR-ComplEx to Appendix A.2.

## 4. IMR: Interpretable Multi-Hop Reasoning

We introduce the Interpretable Multi-Hop Reasoning framework (IMR) in this section. We first provide an overview of IMR in Section 4.1. IMR comprises three modules: the path searching module, query updating module, and path scoring module. The path searching module searches related paths hop by hop from the subjects of questions, involving path sampling and entity clipping, whose motivation and design are presented in Section 4.2. The query updating module calculates the remaining questions hop-by-hop for each path, involving the update of the subject and relations, whose motivation and design are introduced in Section 4.3. The path scoring module designs three indicators: the question matching degree, answer completion level, and path confidence. This module combines three indicators to evaluate each path, whose motivation and design are presented in Section 4.4. We introduce training strategies and the regularizations on state continuity in Section 4.5. IMR conducts uniform path comparisons based on consistent basic models. To better illustrate this framework, we also include the corresponding instance model (IMR-TransE) in Sections 4.3–4.5. The detailed implementations of IMR-RotatE and IMR-ComplEx are included in Appendix A.2.

### *4.1. Framework Overview*

We notice that predicting unknown facts based on paths is akin to answering questions, i.e., the question can be answered directly via finding triples with an equal relation or gradually by utilizing the multi-hop equivalent paths. Inspired by this observation, we take the task of link prediction as stepwise question answering. IMR primarily consists of searching for paths hop by hop, updating the remaining questions for each path, and filtering the best answers based on three indicators: the question matching degree, answer completion level, and path confidence.

We show a toy example in Figure 1. Given a question $(e_s, r_q, ?, t_q)$ and the previous facts $G_{t_q}$, the task of forecasting is predicting the missing object $e_o$. The steps of IMR are as follows.

**Step 1**: Starting from the subject $e_s$, we first acquire the associated quadruples $P_{(e_s,t_q)}$, namely 1-hop paths. We temporally bias the neighborhood sampling using an exponential distribution for the neighbors [7]. The distribution negatively correlates with the time difference between node $e_s$ and its neighbor $N_{(e_s,t_q)}$. Then, we calculate the remaining questions (the remaining subject $e_{s_{(1)}}$ and the remaining relation $r_{q_{(1)}}$) for each sampled path. Finally, IMR scores 1-hop paths based on three indicators, which is discussed in Section 4.4.

**Step 2**: To prevent the path searching from exploding, the model samples the tails of 1-hop paths for the 2-hop path searching. As shown by the pink arrow in Figure 1, the tails of 1-hop paths are clipped according to the scores of 1-hop paths. For the 2-hop paths searched from the clipped tails, IMR samples the paths negatively correlated with time distances. Then, IMR calculates the remaining questions for each 2-hop path (the remaining subject $e_{s_{(2)}}$ and the remaining relation $r_{q_{(2)}}$) and scores the 2-hop paths based on three indicators.

**Step 3**: Rank the scores of 1-hop and 2-hop paths to obtain the preference answer.

### *4.2. Path Searching Module*

Inspired by the observation that reasoning based on multi-hop paths is akin to stepwise question answering, this module searches related paths hop by hop from the subjects of questions.

**Path sampling.** For the path searching from the starting subject $e_s^{t_q}$, the number of triples in $P_{(e_s,t_q)}$ may be very large. To prevent the path searching from exploding, we sample a subset of the paths. In fact, the attributes of entities in temporal KGs may change over time. Consider the observation that when $t_1$ is closer to $t_q$, the attributes of $e_s^{t_1}$ should be more similar to those of $e_s^{t_q}$. We also verify the correlation between attributes and the time distance in Appendix A.6. Therefore, we are more prone to sample nodes whose time is closer to $t_q$. In this paper, we employ time-aware exponentially weighted sampling in xERTR [7]. xERTR temporally biases the neighborhood sampling using an exponential distribution of temporal distance.

**Entity pruning.** The search for next-hop paths is based on the tails of previous-hop paths, so the number of paths is increased by the exponent of dimensions. To avoid the explosion of next-hop path searching, this paper proposes to select the top-K entities for the next-hop search based on the sorted scores of the previous hops.

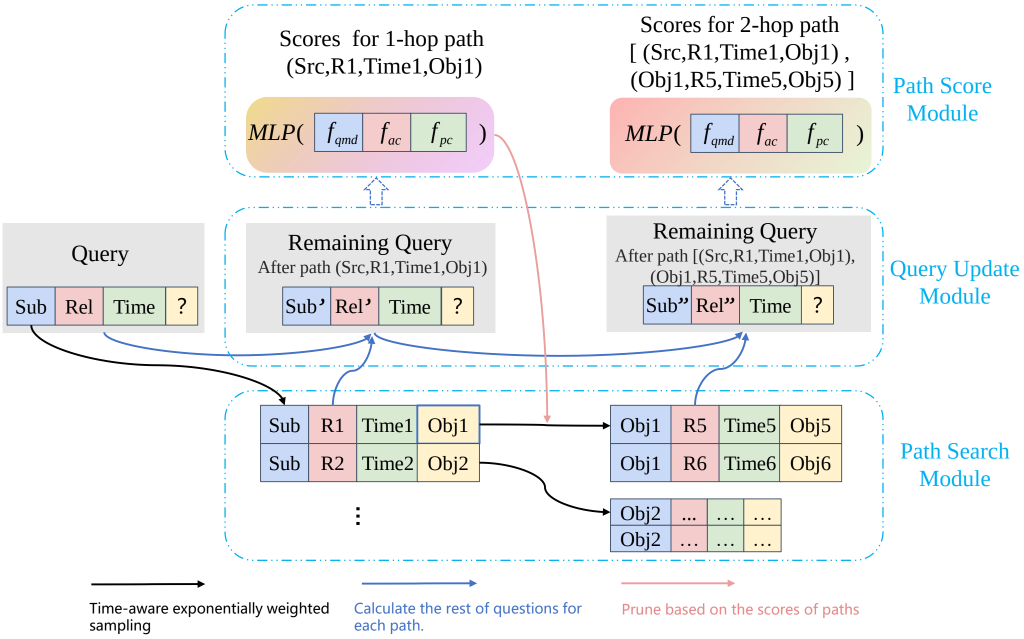


**Figure 1.** The architecture of IMR. We take the 2-hop path search as an example. The black and red arrows denote time-aware exponentially weighted sampling and pruning based on the scores of paths, respectively (Section 4.2). The blue arrows denote the calculation of the rest of the questions for each path (Section 4.3). (Sub, Rel, ?, Time) is regarded as the original question, which can be denoted as $\left(e_s, r_q, ?, t_q\right)$. The searched two paths are [(Sub,R1,Obj1,Time1)] and [(Sub,R1,Obj1,Time1),(Obj1,R5,Obj5,Time5)], which can be denoted as $\left[\left(e_s, r_{p_{(1)}}, e_1, t_1\right)\right]$ and $\left[\left(e_s, r_{p_{(1)}}, e_1, t_1\right), \left(e_1, r_{p_{(2)}}, e_2, t_2\right)\right]$, respectively. (Sub′, Rel′, ?, Time) and (Sub″, Rel″, ?, Time) denote the remaining questions after the 1-hop and 2-hop path, which can be taken as $\left(e_{s_{(1)}}, r_{q_{(1)}},, ?, t_q\right)$, $\left(e_{s_{(2)}}, r_{q_{(2)}}, ?, t_q\right)$, respectively.

### 4.3. Query Updating Module

Given a question $\left(e_s, r_q, ?, t_q\right)$, there may be a few relations directly equivalent to $r_q$ in the temporal KGs for the task of link prediction. More questions need to go through multi-hop paths to infer the outcome. In question answering, a complex question can be decomposed into multiple sub-questions, with one sub-question answered at each step. Thus, inference based on the multi-hop path is equivalent to answering complex questions step by step. Moreover, we need to remove the part resolved to focus on the remaining questions. IMR proposes to update the question according to the last hop and focus on finding the unsolved parts. The query updating module mainly calculates the remaining questions, i.e., the unanswered questions.

The embedding of entities is first introduced in this subsection, followed by the query updating module of IMR-TransE.

#### 4.3.1. Entity Representation

The attributes contained in the entities may change over time. This paper divides the entity embeddings of each timestamp into a static representation and dynamic representation.

$$\mathbf{e} = act\left(MLP([\,\mathbf{e_{sta}} \,||\, \mathbf{e_{dy}}\,])\right) \tag{6}$$

Here, the vector $\mathbf{e_{sta}}$ denotes the static embedding, which captures time-invariant features and global dependencies over the temporal KGs. The vector $\mathbf{e_{dy}}$ represents the dynamic embedding for each entity that changes over time. $||$ denotes the operation of concatenation and $MLP(\cdot)$ denotes the multilayer perceptron (MLP). $act(\cdot)$ denotes the activation function. We provide more details about $\mathbf{e_{sta}}$ and $\mathbf{e_{dy}}$ in Appendix A.3.

### 4.3.2. Question Updating

Each path contains a different set of relations. After each hop, the question needs to discard the processed semantic, i.e., to obtain the remaining subject and relation of the question.

**Question updating for IMR-TransE.** As shown in Figure 1, the subject and relation of the question after the *i*-th hop path are updated based on Equation (5) as follows.

$$\mathbf{e}_{s_{(i)}} = \mathbf{e}_{s_{(i-1)}} + \mathbf{r}_{p_{(i)}} \tag{7}$$

$$\mathbf{r}_{q_{(i)}} = \mathbf{r}_{q_{(i-1)}} - \mathbf{r}_{p_{(i)}} \tag{8}$$

where the embedding $\mathbf{e}_{s_{(i)}}$ and $\mathbf{r}_{q_{(i)}}$ represent the remaining subject and relation of the question after the *i*-hop path, respectively. Moreover, $\mathbf{e}_{s_{(0)}} = \mathbf{e}_s$, $\mathbf{r}_{q_{(0)}} = \mathbf{e}_q$ and $\mathbf{r}_{p_{(i)}}$ denotes the relation of *i*-th hop path and *i* is the number of hops for each path.

## *4.4. Path Scoring Module*

For the question (Sub, Rel, ?, Tq), we search the 2-hop path (Sub, R1, Obj1, T1),(Obj1, R2, Obj2, T2). The pink box indicates that the original question and the tail of the path are combined as a quadruple to measure the rationality of searched tails, i.e., the question matching degree $f_{qmd}$. The purple box represents the comparison between the question's relation and the path relations to measure the semantic equivalence between the question and the path, i.e., the answer completion level $f_{ac}$. These green boxes compare the attributes of the same entities with different timestamps to measure the reliability of the search path, i.e., the path confidence $f_{pc}$.

We evaluate the path searching from three perspectives. First, the searched tails should match the original question, which means that the correct tails searched by paths and the question should satisfy the consistent basic embedding model. Secondly, the ideal path should be the search for equivalent semantics for relations, not merely the search for the correct tails. It is necessary to ensure the correctness of semantic equivalence, i.e., the path is semantically equivalent to the relation of the question. Finally, considering the particularity of the temporal KGs, the attributes of the same entity may change over time. The current sampling strategy for path searching is to sample adjacent timestamp triples of the same entity. When the attribute value of the entity changes significantly over time, it is inappropriate to perform this sampling strategy for the next hop. We need to ensure that the same entity with different timestamps has similar properties in the same path. In this way, three indicators have been developed by IMR to measure the rationality of the reasoning path, respectively: the question matching degree, answer completion level, and path confidence. Although the current methods, such as models based on reinforcement learning, can have complicated designs, the score functions simply belong to a type of question matching degree. We provide a detailed analysis of the correlation between IMR and reinforcement-learning-based models in Appendix A.5.

### 4.4.1. Question Matching Degree

For the tails found by path searching, we need to measure the matching degree between the tails and the question, the question matching degree. In fact, the scoring function applied by some traditional reinforcement learning methods is a type of question matching degree. As shown in the yellow box in Figure 2, for the entity $e_{p_{(i)}}^{t_i}$ searched by the paths with $i$ hops, we combine the entity $e_i^{t_i}$ and the question $\left(e_s, r_q, ?, t_q\right)$ into a new quadruple $\left(e_s, r_q, e_{p_{(i)}}^{t_i}, t_q\right)$.

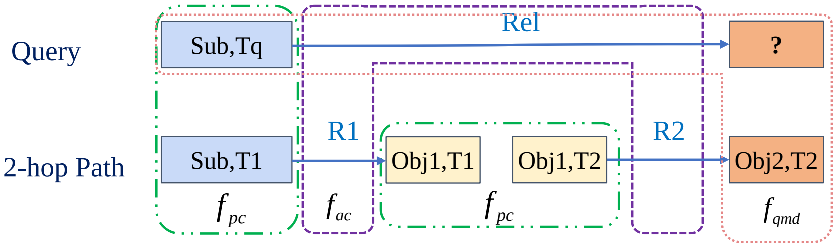


**Figure 2.** A brief illustration of the path scoring module.

**Question matching degree for IMR-TransE.** Question matching degree $f_{qmd}$ in IMR-TransE calculates the distance of the constructed quadruple based on TransE [1]. The better the entity matches the question, the smaller the distance of quadruples will be. The calculation of $f_{qmd}$ for $i$th-hop path is as follows.

$$f^{i}_{qmd} = \left\| \mathbf{e}^{\mathbf{t_q}}_{\mathbf{s}} + \mathbf{r_q} - \mathbf{e}^{\mathbf{t_i}}_{\mathbf{p_{(i)}}} \right\|_{p} \tag{9}$$

where the $p$-norm of a complex vector $V$ is defined as $\|V\|_p = \sqrt[p]{|V_i|^p}$. We use the L1-norm for all indicators in the following.

### 4.4.2. Answer Completion Level

Among the paths to the right tails, some paths are not related to the semantics of the question. Although these paths can infer the tail, these paths are invalid due to being unrelated to the question in semantics. Therefore, IMR designs an index to measure the semantic relevance between the path and the question. Answer completion level $f_{ac}$ indicates whether the combination of path relations can reflect the relation of the question in semantics. IMR takes the remaining relations of the question as the answer completion level, which is calculated based on the distance between the relations of paths $r_{p_{(1)}}, r_{p_{(2)}}, \ldots$ and the relation $r_q$. The fewer the relations of a question that remain, the more complete the answer given by the combination of path relations.

**Answer completion level for IMR-TransE.** The calculation of $f_{ac}$ for $i$th-hop path in IMR-TransE is as follows.

$$\begin{aligned} f^{i}_{ac} &= \left\| \mathbf{r}_q - \mathbf{r}_{p_{(1)}} - \mathbf{r}_{p_{(2)}} - \mathbf{r}_{p_{(3)}} - \ldots - \mathbf{r}_{p_{(i)}} \right\|_p \\ &= \left\| \mathbf{r}_{q_{(1)}} - \mathbf{r}_{p_{(2)}} - \mathbf{r}_{p_{(3)}} - \ldots - \mathbf{r}_{p_{(i)}} \right\|_p \\ &= \left\| \mathbf{r}_{q_{(i)}} \right\|_p \end{aligned} \tag{10}$$

### 4.4.3. Path Confidence

Path searching is the process of searching for the next-hop paths based on the tail of the previous hop. When searching for a path, the current sampling strategy is to sample adjacent timestamp triples of the same entity. There are deviations between the same entities with different timestamps in temporal KGs. The premise of this sampling strategy is that only when entities have similar attributes under different timestamps, the path searching is valid. When the entity's attributes change significantly over time, performing an effective next path search is inappropriate. The reasoning path is more reliable when the deviations between entities are smaller. IMR designs path confidence $f_{pc}$, i.e., the error between the subject of the updated question $e_{s_{(i)}}$ and the tails $e^{t_i}_{p_{(i)}}$ of the path with $i$ hops.

**Path confidence for IMR-TransE.** The calculation of $f_{pc}$ for $i$th-hop path in IMR-TransE is as follows.

$$f^{i}_{pc} = \left\| \mathbf{e}_{s_{(i)}} - \mathbf{e}^{t_i}_{p_{(i)}} \right\|_p \tag{11}$$

where $\mathbf{e}_{q_{(i)}}$ represents the remaining subject of the question updated by paths of the length $i$, and $\mathbf{e}^{t_i}_{p_{(i)}}$ represents the tail reasoned by the $i$-hop paths.

#### 4.4.4. Combination of Scores

IMR merges indicators with positive weights to obtain the final score of each path, i.e., $f = w_{qmd} * f_{qmd} + w_{ac} * f_{ac} + w_{pc} * f_{pc}$, where $w_{qmd}, w_{ac}, w_{pc} \in \mathbf{R}^+$.

**Entity aggregation for IMR.** Considering that the searched paths may lead to entities with different timestamps, IMR adopts specific aggregation for searched entities. First, entities with the same timestamp may be inferred by different paths, so IMR needs to combine the scores of entities with unique timestamps. Considering that only one path matches the question best, IMR employs max aggregation to various paths reaching the same entities with the same timestamp. Moreover, specific paths may infer the same entity with different timestamps. IMR performs average aggregation on the scores of entities with different timestamps. Finally, IMR obtains the score of each entity at the question timestamp.

### *4.5. Learning*

We utilize binary cross-entropy as the loss function, which is defined as

$$\begin{aligned} L = &-\frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\left|\varepsilon^p_q\right|}\sum_{e_i\in\varepsilon^p_q}\left(y_{e_i,q}\log\left(\frac{f_{e_i,q}}{\sum_{e_i\in\varepsilon^p_q} f_{e_i,q}}\right)\right) \\ &+\frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\left|\varepsilon^p_q\right|}\sum_{e_i\in\varepsilon^p_q}\left((1-y_{e_i,q})\log\left(1-\frac{f_{e_i,q}}{\sum_{e_i\in\varepsilon^p_q} f_{e_i,q}}\right)\right) \end{aligned} \tag{12}$$

where $\varepsilon^p_q$ represents the set of entities reasoned by selected paths. $y_{e_i,q}$ represents the binary label that indicates whether it is the answer for $q$ and $Q$ represents the training set. $f_{e_i,q}$ denotes the score obtained by Section 4.4.4 for each path. We jointly learn the embeddings and other model parameters by end-to-end training.

**Regularization.** For the same entity with different timestamps, the closer its time distance is, the closer its dynamic embedding is [32]. IMR proposes the regularization on continuity for the dynamic vectors of entities.

The specific regularization for IMR is as follows.

$$reg = \left\|\mathbf{e}^{t_j}_k - \mathbf{e}^{t_{j-1}}_k\right\|_p + \left\|\mathbf{e}^{t_j}_k - \mathbf{e}^{t_{j+1}}_k\right\|_p \tag{13}$$

where $\mathbf{e}^{t_j}_k$ denotes the dynamic embedding of the $k$-th entity at the $j$-th timestamp. $\mathbf{e}^{t_{j-1}}_k, \mathbf{e}^{t_{j+1}}_k$ denotes the dynamic embedding of the previous and later timestamp against $e^{t_j}_k$, respectively. $\|\cdot\|_p$ denotes the $p$ norm of the vectors and we take $p$ as 1 in this paper.

## 5. Experiments

### *5.1. Datasets and Baselines*

To evaluate the proposed module, we consider two standard temporal KG datasets Integrated Crisis Early Warning System (ICEWS) [33], WIKI [34], and YAGO [35]. The ICEWS dataset contains information about political events with time annotations. We select two subsets of the ICEWS dataset, i.e., ICEWS14 and ICEWS18, containing event facts in 2014 and 2018. WIKI and YAGO is a temporal KG that fuses information from Wikipedia with WordNet [36]. Following the experimental settings of HyTE [37], we deal with year-level granularity by dropping the month and date information. We compare IMR and baseline methods by performing the temporal KG forecasting task on ICEWS14, ICEWS18, WIKI, and YAGO. Details of these datasets are listed in Table 1. We adopt the same dataset split strategy as in [38].

**Table 1.** Statistics of three benchmark datasets.

| Dataset | ICEWS14 | ICEWS18 | WIKI | YAGO |
|---|---|---|---|---|
| entity | 7128 | 23,033 | 12,554 | 10,623 |
| relation | 230 | 256 | 24 | 10 |
| timestamp | 365 | 304 | 232 | 189 |
| training | 63,685 | 373,018 | 539,286 | 161,540 |
| validation | 13,823 | 45,995 | 67,538 | 19,523 |
| test | 13,222 | 49,545 | 63,110 | 20,026 |

We compare the performance of IMR-TransE against the temporal KG reasoning models, including TTransE [34], TA-DistMult/TA-TransE [30], DE-SimplE [39], TNTComplEx [32], CyGNet [11], RE-Net [38], TANGO [40], TITer [14], and xERTR [7].

In the experiments, the widely used Mean Reciprocal Rank (MRR) and Hits@1,3,10 are employed as the metrics. The filtered setting for static KGs is not suitable for the reasoning task under the exploration setting, as mentioned in xERTR [7]. This paper adopts the time-aware filtering scheme, which only filters out genuine triples at the question time.

### *5.2. Experimental Results*

**Main results.** Tables 2 and 3 show the comparison between IMR-TransE, IMR-RotatE, IMR-ComplEx, and other baseline models on ICEWS, WIKI, and YAGO. Overall, the instantiated models of IMR outperform the baseline models in all metrics while being more interpretable, which convincingly verifies its effectiveness. Due to the limited paper length, a detailed analysis of the interpretability is provided in Appendix A.1. Compared to the best baseline (TiTer), IMR-TransE obtains a relative improvement of 3.3% and 2.5% in MRR and Hits@1, averaged on ICEWS, WIKI, and YAGO. Moreover, different IMR models achieve the best performance across unique datasets due to basic models.

**Table 2.** Results comparison on ICEWS14 and ICEWS18. Compared metrics are time-aware filtered MRR (%) and Hits@1/3/10 (%), which are multiplied by 100. The best results among all models are in bold.

| | ICEWS14 | | | | ICEWS18 | | | |
|---|---|---|---|---|---|---|---|---|
| | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** |
| TTransE | 13.43 | 3.11 | 17.32 | 34.55 | 8.31 | 1.92 | 8.56 | 21.89 |
| TA-DistMult | 26.47 | 17.09 | 30.22 | 45.41 | 16.75 | 8.61 | 18.41 | 33.59 |
| DE-SimplE | 32.67 | 24.43 | 35.69 | 49.11 | 19.30 | 11.53 | 21.86 | 34.80 |
| TNTComplEx | 32.12 | 23.35 | 36.03 | 49.13 | 27.54 | 19.52 | 30.80 | 42.86 9 |
| CyGNet | 32.73 | 23.69 | 36.31 | 50.67 | 24.93 | 15.90 | 28.28 | 42.61 |
| RE-NET | 38.28 | 28.68 | 41.34 | 54.52 | 28.81 | 19.05 | 32.44 | 47.51 |
| xERTE | 40.79 | 32.70 | 45.67 | 57.30 | 29.31 | 21.03 | 33.51 | 46.48 8 |
| TANGO-Tucker | – | – | – | – | 28.68 | 19.35 | 32.17 | 47.04 |
| TANGO-DistMult | – | – | – | – | 26.75 | 17.92 | 30.08 | 44.09 |
| TITer | 41.73 | 32.74 | 46.46 | 58.44 | 29.98 | 22.05 | 33.46 | |
| IMR-TransE | **44.76** | **35.64** | **49.49** | **62.30** | 32.45 | 22.97 | 36.05 | 49.36 |
| IMR-RotatE | 44.21 | 35.13 | 48.72 | 62.04 | 32.67 | 23.53 | 36.76 | 50.67 |
| IMR-ComplEx | 44.03 | 34.55 | 49.21 | 62.11 | **33.33** | **24.07** | **37.65** | **51.51** |

**Comparison of multi-hop paths.** Figure 3 shows the performance of IMR-TransE on ICEWS, WIKI, and YAGO as the maximum length of paths increases. The performance basically continues rising with the increase in the paths' length. However, as the maximum length of paths increases, the performance on ICEWS18 hardly improves. Further analysis on ICEWS18 in [3] explains that there are no strong dependencies between the relations of the question and the multi-hop paths. Thus, longer paths provide little gain for inference.

Moreover, as the maximum length of paths increases, the number of inference paths increases exponentially and most of the invalid paths will suppress the performance of IMR-TransE. In order to ensure that the performance of the model does not decrease, we propose to control the sampling number of next-hop paths to limit the total number of multi-step paths and suppress the impact of noisy samples. This paper set the number of next-hop samplings to 5. In summary, experiments show that unified indicators designed by IMR based on consistent basic models can uniformly measure the paths of different hops, allowing better reasoning based on paths with different hops, which verifies the claim in Section 4.4. We present an extra ablation study on three indicators in IMR-TransE in Appendix A.4.

**Table 3.** Results comparison on WIKI and YAGO. Compared metrics are time-aware filtered MRR (%) and Hits@1/3/10 (%), which are multiplied by 100. The best results among all models are in bold.

| | WIKI | | | | YAGO | | | |
|---|---|---|---|---|---|---|---|---|
| | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** |
| TTransE | 29.27 | 21.67 | 34.43 | 42.39 | 31.19 | 18.12 | 40.91 | 51.21 |
| TA-DistMult | 44.53 | 39.92 | 48.73 | 51.71 | 54.92 | 48.15 | 59.61 | 66.71 |
| DE-SimplE | 45.43 | 42.6 | 47.71 | 49.55 | 54.91 | 51.64 | 57.30 | 60.17 |
| TNTComplEx | 45.03 | 40.04 | 49.31 | 52.03 | 57.98 | 52.92 | 61.33 | 66.69 |
| CyGNet | 33.89 | 29.06 | 36.10 | 41.86 | 52.07 | 45.36 | 56.12 | 63.77 |
| RE-NET | 49.66 | 46.88 | 51.19 | 53.48 | 58.02 | 53.06 | 61.08 | 66.29 |
| xERTE | 71.14 | 68.05 | 76.11 | 79.01 | 84.19 | 80.09 | 88.02 | 89.78 |
| TANGO-Tucker | 50.43 | 48.52 | 51.47 | 53.58 | 57.83 | 53.05 | 60.78 | 65.85 |
| TANGO-DistMult | 51.15 | 49.66 | 52.16 | 53.35 | 62.70 | 59.18 | 60.31 | 67.90 |
| TITer | 75.50 | 72.96 | 77.49 | 79.02 | 87.47 | 84.89 | 89.96 | 90.27 |
| IMR-TransE | 80.41 | 76.04 | 84.91 | 85.95 | 90.24 | 87.91 | 92.65 | 92.77 |
| IMR-RotatE | 79.43 | 74.36 | 84.59 | 85.79 | **90.34** | **88.10** | 92.69 | **92.78** |
| IMR-ComplEx | **80.54** | **76.12** | **84.98** | **85.97** | 90.19 | 87.80 | **92.71** | 92.78 |

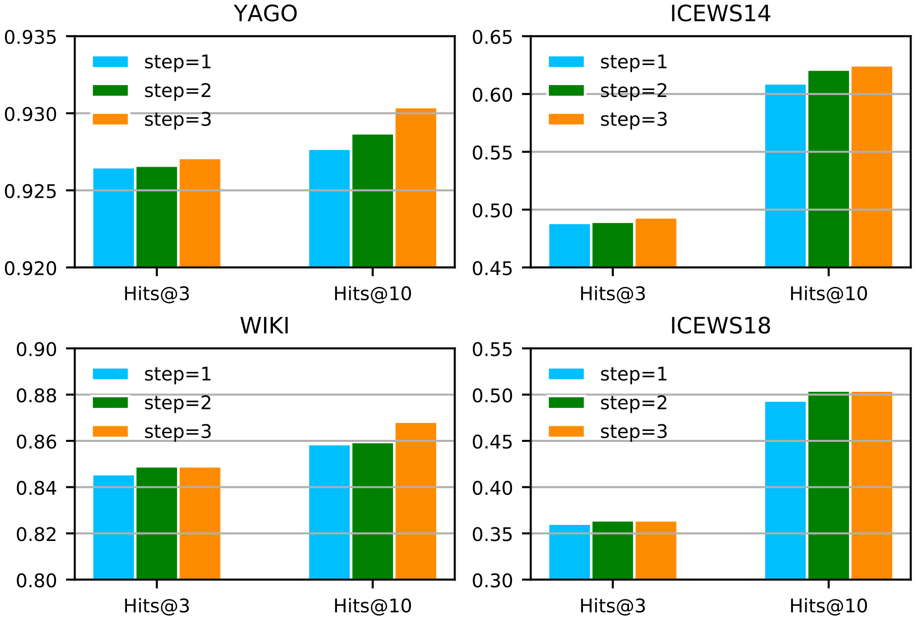


**Figure 3.** Comparison of the performance of paths with different maximum hops on four datasets. We average the output of four experiments with different random seeds and fixed hyperparameters.

## 6. Conclusions

We propose an Interpretable Multi-Hop Reasoning framework for temporal KG forecasting tasks. IMR transforms reasoning based on path searching into stepwise question answering based on consistent basic models. Moreover, IMR develops three indicators to measure the answer and reasoning paths, and this is the first study to develop interpretable evaluation indicators from the perspective of actual semantics for the temporal KG forecasting task. IMR can measure the paths of different hops according to the same criteria and be more explainable. Extensive experiments on four benchmark datasets demonstrate the effectiveness of our method. In the future, we plan to enhance the prediction by integrating different paths reaching the same tail, which will be more effective and interpretable. We will also continue to explore the models based on GAT [3] for temporal KG forecasting tasks.

**Author Contributions:** Supervision, Z.L. and K.H.; writing—review and editing, Z.D. and L.Q.; validation, L.C. and Z.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

### *Appendix A.1. Case Studies and Interpretability*

For the question (John Kerry, Make a visit, ?, 2014-11-11), we extract some of the paths for the case study in Table A1. The lower the scores or indicators in Table A1, the better the performance of the path. We compare the paths based on the total score, analyze various aspects of the paths based on detailed indicators, and verify the interpretation of the model with actual semantics.

The first block of Table A1 selects reasoning paths with the same objects to analyze the answer completion level. First, we compare path 1-1 and path 1-2. The score of path 1-1 is lower than that of path 1-2. As we analyze the three indicators further, we find that the answer completion level of path 1-1 is smaller than that of path 1-2. The comparison of the answer completion level indicates that the relation of path 1-1 should be closer to the relations of the question. Practically, path 1-1 has the same relation as the question, which is closer to the relation of question than path 1-2. Thus, actual semantics verify the interpretation of the model. Comparing path 1-4 and path 1-5, we find that the total score of path 1-4 is lower than that of path 1-5, and the answer completion level of path 1-5 is higher than that of path 1-4. IMR shows that the combination of reasoning relations of path 1-4 is better than that of path 1-5. In fact, these two paths for inference do not seem to be particularly appropriate to the question. Nevertheless, the combination of relations **[Meet at a 'third' location + Make a visit]** is actually closer to the relation of the question **[Make a visit]** than the combination of relations **[Consult + Consult]**. To summarize, the first set of experiments shows that the answer completion level can effectively indicate how well the combination of path relations equals the question's relation, verifying the statement in Section 4.4.

**Table A1.** Reasoning paths searched for the question **(John Kerry, Make a visit, ?, 2014-11-11)** and their scores, respectively.

| **Question:** | **John Kerry** | **Make a Visit** | **Oman** | **2014-11-11** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Path-ID** | **Reasoning Path** | | | | | | | | **Score** | | | |
| | $e_s$ | $r_2$ | $e_2$ | $t_2$ | $e_2$ | $r_3$ | $e_3$ | $t_3$ | $f_{ac}$ | $f_{qmd}$ | $f_{pc}$ | **Combined Score** |
| path 1-1 | John Kerry | Make a visit | Oman | 2014-11-09 | - | - | - | - | 0 | 74 | 74 | 137 |
| path 1-2 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-09 | - | - | - | - | 26 | 74 | 69 | 169 |
| path 1-3 | John Kerry | (Reversed) Host a visit | Oman | 2014-11-09 | - | - | - | - | 27 | 74 | 76 | 178 |
| path 1-4 | John Kerry | Meet at a 'third' location | Catherine Ashton | 2014-11-10 | Catherine Ashton | Make a visit | Oman | 2014-11-09 | 38 | 74 | 90 | 206 |
| path 1-5 | John Kerry | Consult | Mohammad Javad Zarif | 2014-11-10 | Mohammad Javad Zarif | Consult | Oman | 2014-11-09 | 73 | 74 | 107 | 254 |
| path 2-1 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-10 | - | - | - | - | 26 | 47 | 41 | 119 |
| path 2-2 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-09 | - | - | - | - | 26 | 74 | 69 | 170 |
| path 2-3 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-05 | - | - | - | - | 26 | 89 | 83 | 196 |
| path 2-4 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-02 | - | - | - | - | 26 | 90 | 82 | 197 |
| path 2-5 | John Kerry | Reversed Meet at a 'third' location | Catherine Ashton | 2014-11-10 | Catherine Ashton | Express intent to meet or negotiate | Oman | 2014-11-03 | 49 | 91 | 101 | 246 |
| path 2-6 | John Kerry | Reversed Meet at a 'third' location | Catherine Ashton | 2014-11-10 | Catherine Ashton | Express intent to meet or negotiate | Oman | 2014-11-05 | 49 | 89 | 100 | 242 |
| path 2-7 | John Kerry | Make a visit | China | 2014-11-05 | - | - | - | - | 0 | 88 | 88 | 162 |
| path 2-8 | John Kerry | Make a visit | North Atlantic Treaty Organization | 2014-06-25 | - | - | - | - | 0 | 87 | 87 | 160 |
| path 2-9 | John Kerry | Make a visit | Canada | 2014-10-27 | - | - | - | - | 0 | 85 | 85 | 157 |
| path 3-1 | John Kerry | Reversed Meet at a 'third' location | Catherine Ashton | 2014-11-10 | - | - | - | - | 53 | 46 | 40 | 155 |
| path 3-2 | John Kerry | Express intent to meet or negotiate | Oman | 2014-11-09 | - | - | - | - | 26 | 74 | 69 | 169 |
| path 3-3 | John Kerry | Make a visit | Afghanistan | 2014-07-21 | - | - | - | - | 0 | 88 | 88 | 162 |
| path 3-4 | John Kerry | Make a visit | Afghanistan | 2014-07-21 | Afghanistan | Reversed Make statement | Barack Obama | 2014-07-18 | 34 | 94 | 104 | 241 |
| path 3-5 | John Kerry | Make a visit | Angola | 2014-08-05 | Angola | (Reversed) Make statement | Anthony Foxx | 2014-08-04 | 35 | 93 | 105 | 241 |
| path 3-6 | John Kerry | (Reversed) Make a visit | Catherine Ashton | 2014-11-10 | Catherine Ashton | Make a visit | Oman | 2014-11-09 | 33 | 74 | 85 | 197 |

**Table A2.** Reasoning paths searched for the query **(Citizen (Nigeria), Use unconventional violence, ?, 8016)** and their scores, respectively.

| **Query:** | **Citizen (Nigeria)** | **Use Unconventional Violence** | **Secretariat (Nigeria)** | **8016** | | | | |
|---|---|---|---|---|---|---|---|---|
| **Path-ID** | **Reasoning Path** | | | | **Score** | | | |
| | $e_s$ | $r_2$ | $e_2$ | $t_2$ | $f_{ac}$ | $f_{qmd}$ | $f_{pc}$ | **Combined Score** |
| path 4-1 | Citizen (Nigeria) | Use unconventional violence | Militant (Nigeria) | 7968 | 0 | 162 | 162 | 215 |
| path 4-1 | Citizen (Nigeria) | Use unconventional violence | Militant (Nigeria) | 7728 | 0 | 185 | 185 | 245 |
| path 5-2 | Citizen (Nigeria) | Reversed Use unconventional violence | Terrorist (Boko Haram) | 7824 | 72 | 204 | 199 | 359 |
| path 5-3 | Citizen (Nigeria) | Reversed Use unconventional violence | Terrorist (Boko Haram) | 7776 | 72 | 174 | 168 | 319 |
| path 5-4 | Citizen (Nigeria) | Reversed Use unconventional violence | Militant (Boko Haram) | 7872 | 72 | 206 | 202 | 363 |
| path 5-5 | Citizen (Nigeria) | Reversed Use unconventional violence | Militant (Boko Haram) | 7776 | 72 | 173 | 166 | 317 |
| path 5-6 | Citizen (Nigeria) | Reversed Use unconventional violence | Militant (Boko Haram) | 7752 | 72 | 175 | 167 | 319 |
| Path 6-1 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7992 | 73 | 95 | 95 | 220 |
| path 6-2 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7872 | 73 | 174 | 168 | 321 |
| path 6-3 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7848 | 73 | 177 | 171 | 324 |
| path 6-4 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7824 | 73 | 178 | 171 | 325 |
| path 6-5 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7680 | 73 | 180 | 173 | 328 |
| path 7-1 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7848 | 73 | 177 | 171 | 324 |
| path 7-2 | Citizen (Nigeria) | Make an appeal or request | Government (Nigeria) | 7848 | 78 | 167 | 158 | 315 |
| path 7-3 | Citizen (Nigeria) | Reversed fight with small arms and light weapons | Boko Haram | 7848 | 73 | 177 | 171 | 324 |
| path 7-4 | Citizen (Nigeria) | Reversed Make an appeal or request | Tony Momoh | 7848 | 80 | 207 | 205 | 377 |
| path 7-5 | Citizen (Nigeria) | Reversed Express intent to meet or negotiate | South Africa | 7848 | 85 | 169 | 165 | 330 |
| path 7-6 | Citizen (Nigeria) | Reversed Bring lawsuit against | Fessehaye Yohannes | 7848 | 80 | 210 | 206 | 379 |

The second block of Table A1 selects the paths of the same reasoning relations to verify the path confidence and the question matching degree. Comparing paths 2-1, 2-2, 2-3, and 2-4, we observe that the scores of the paths are increasing. Additionally, the path confidence of these three paths is also growing. In fact, the time distance between the paths and the question is gradually increasing, which means that the reliability of the paths gradually decreases. The reliability indicated by path confidence is consistent with the actual reliability. Similarly, we find that the path confidence of path 2-5 is higher than that of path 2-6, indicating that path 2-5 is less reliable. The actual situation is that the timestamp of path 2-5 (2014-11-03 < 2014-11-05) is farther from the timestamp of the question, which is consistent with the explanation. Comparing path 2-9 with paths 2-7 and 2-8, respectively, the model further infers that the path confidence and question matching degree of path 2-9 are better than those of the other two paths. The actual situation is that the timestamp error with the question satisfies path 2-7 $>$ path 2-9 $>$ path 2-8. This is because the question matching degree covers the path confidence. Because the path confidence contains the error of the triple in the training dataset, the triple error covers the error caused by different timestamps, which makes path 2-9 more reliable than path 2-7. In general, the second set of experiments illustrates that the path confidence can effectively indicate the validity of each path.

In the third block of Table A1, we randomly select the paths, explain the paths based on these indicators, and verify them with the actual situation. We first sort three paths according to the answer completion level : path 3-1 $<$ path 3-2 $<$ path 3-3. Therefore, the semantic similarity of relations between the three paths and the question should satisfy path 3-3 $>$ path 3-2 $>$ path 3-1. The actual semantic similarity between the relations of paths and that of the question satisfies **Make a visit** $>$ **Express intent to meet or negotiate** $>$ **Meet at a 'third' location**, which is consistent with the interpretation of IMR. Sort three paths by path confidence: path 3-1 $<$ path 3-2 $<$ path 3-3. The reliability of the three inference paths should satisfy path 3-1 $<$ path 3-2 $<$ path 3-3. We observe that the time distance between the three paths and the question is gradually increasing, which verifies the explanation by path confidence. The analysis of paths 3-4 to 3-6 is similar to the analysis of former paths. Case studies show that IMR can provide reasoning paths and offer a valid basis for path comparison.

### *Appendix A.2. Details on IMR-RotatE and IMR-ComplEx*

#### Appendix A.2.1. IMR-RotatE

**RotatE.** RotatE [17] defines each relation as a rotation from head entities to tail entities in a complex vector space. Given a triplet $(\mathbf{h}, \mathbf{t}, \mathbf{r})$, we expect that $\mathbf{t} = \mathbf{h} \circ \mathbf{r}$, where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbf{C}^k$ are the embeddings, the modulus for each dimension of relations satisfies $|\mathbf{r}_i| = 1$, and $\circ$ denotes the Hadamard product. The score function for $(e_s, r_q, e_o, t_q)$ is

$$f_{r_q}\left(\mathbf{e_s^{t_q}}, \mathbf{e_o^{t_q}}\right) = -\left\|\mathbf{e_s^{t_q}} \circ \mathbf{r_q} - \mathbf{e_o^{t_q}}\right\|_2 \tag{A1}$$

where $\mathbf{e_s^{t_q}}, \mathbf{r_q}, \mathbf{e_o^{t_q}} \in \mathbf{C}^k$, $|\mathbf{r_q}_i| = 1$.

**Question updating for IMR-RotatE.**

$$\mathbf{e}_{s_{(i)}} = \mathbf{e}_{s_{(i-1)}} \circ \mathbf{r}_{p_{(i)}} \tag{A2}$$

$$\mathbf{r}_{q_{(i)}} = \mathbf{r}_{q_{(i-1)}} - \mathbf{r}_{p_{(i)}} \tag{A3}$$

**Question matching degree for IMR-RotatE.** Question matching degree $f_{qmd}$ in IMR-RotatE calculates the distance of the constructed quadruple based on RotatE [17]. The better the entity matches the question, the smaller the distance of quadruples will be. The calculation of $f_{qmd}$ for $i$th-hop path is as follows.

$$f_{qmd}^{i} = \left\| \mathbf{e_s^{t_q}} \circ \mathbf{r_q} - \mathbf{e_{p_{(i)}}^{t_i}} \right\|_p \tag{A4}$$

where the $p$-norm of a complex vector $V$ is defined as $\|V\|_p = \sqrt[p]{|V_i|^p}$. We use the L1-norm for all indicators in the following.

**Answer completion level for IMR-RotatE.** The calculation of $f_{ac}$ for $i$th-hop path in IMR-RotatE is as follows.

$$\begin{aligned} f_{ac}^{i} &= \left\| \mathbf{r}_q - \mathbf{r}_{p_{(1)}} - \mathbf{r}_{p_{(2)}} - \mathbf{r}_{p_{(3)}} - \ldots - \mathbf{r}_{p_{(i)}} \right\|_p \\ &= \left\| \mathbf{r}_{q_{(1)}} - \mathbf{r}_{p_{(2)}} - \mathbf{r}_{p_{(3)}} - \ldots - \mathbf{r}_{p_{(i)}} \right\|_p \\ &= \left\| \mathbf{r}_{q_{(2)}} - \mathbf{r}_{p_{(3)}} - \ldots - \mathbf{r}_{p_{(i)}} \right\|_p \\ &= \left\| \mathbf{r}_{q_{(i)}} \right\|_p \end{aligned} \tag{A5}$$

**Path confidence for IMR-RotatE.** The calculation of $f_{pc}$ for $i$th-hop path in IMR-RotatE is as follows.

$$f_{pc}^{i} = \left\| \mathbf{e}_{s_{(i)}} - \mathbf{e}_{p_{(i)}}^{t_i} \right\|_p \tag{A6}$$

where $\mathbf{e}_{q_{(i)}}$ represents the remaining subject of the question updated by paths of the length $i$, and $\mathbf{e}_{p_{(i)}}^{t_i}$ represents the tail reasoned by the $i$-hop paths.

### Appendix A.2.2. IMR-ComplEx

**ComplEx.** ComplEx [15] extends the real space to the complex space and constrains the embeddings for relations to be diagonal matrices. The bilinear product becomes a Hermitian product in the complex space. The score function for $(e_s, r_q, e_o, t_q)$ can be expressed as

$$f_{r_q}\left(\mathbf{e_s^{t_q}}, \mathbf{e_o^{t_q}}\right) = \mathbf{Re}\left(\left(\mathbf{e_s^{t_q}}\right)^{\mathbf{T}} \operatorname{diag}(\mathbf{r_q}) \mathbf{e_o^{t_q}}\right) \tag{A7}$$

where $\mathbf{e_s^{t_q}}, \mathbf{r_q}, \mathbf{e_o^{t_q}} \in \mathbf{C}^k$.

**Question updating for IMR-ComplEx.** Considering that such tensor decomposition models are difficult to interpret geometrically, the metrics of IMR-ComplEx are not computed stepwise. The index of each path is calculated independently, which will lead to a certain increase in the amount of calculation.

**Question matching degree for IMR-ComplEx.** Question matching degree $f_{qmd}$ in IMR-RotatE calculates the distance of the constructed quadruple based on RotatE [17]. The better the entity matches the question, the smaller the distance of quadruples will be. The calculation of $f_{qmd}$ for $i$th-hop path is as follows.

$$f_{qmd}^{i} = \mathbf{Re}\left(\left(\mathbf{e_s^{t_q}}\right)^{\mathbf{T}} \operatorname{diag}(\mathbf{r_q}) \mathbf{e_{p_{(i)}}^{t_i}}\right) \tag{A8}$$

where the $p$-norm of a complex vector $V$ is defined as $\|V\|_p = \sqrt[p]{|V_i|^p}$. We use the L1-norm for all indicators in the following.

**Answer completion level for IMR-ComplEx.** The calculation of $f_{ac}$ for $i$th-hop path in IMR-ComplEx is as follows.

$$f_{ac}^{i} = \left\| \mathbf{r}_q - \mathbf{r}_{p_{(1)}} \times \mathbf{r}_{p_{(2)}} \times \mathbf{r}_{p_{(3)}} \times \ldots \times \mathbf{r}_{p_{(i)}} \right\|_p \tag{A9}$$

**Path confidence for IMR-ComplEx.** The calculation of $f_{pc}$ for $i$th-hop path in IMR-ComplEx is as follows.

$$f_{pc}^{i} = \left\| \mathbf{e}_{s_{(i)}} - \mathbf{e}_{p_{(i)}}^{t_i} \right\|_p \tag{A10}$$

where $\mathbf{e}_{q_{(i)}}$ represents the remaining subject of the question updated by paths of the length $i$, and $\mathbf{e}_{p_{(i)}}^{t_i}$ represents the tail reasoned by the $i$-hop paths.

### *Appendix A.3. Entity Representation*

We denote the static embedding of the entity $\mathbf{e}_k$ with $\mathbf{e}_{\mathbf{sta}-k} \in \mathbf{R}^{\mathbf{d}}$, which is a vector independent of time. IMR-TransE adopts the static embedding in xERTR [41]. xERTR [41] proposes a generic time encoding to generate the time-variant part of entity representations, which can be denoted as $\Phi(t)$.

$$\Phi(t) = \sqrt{\frac{1}{d}}[\cos(\omega_1 t + \phi_1), \ldots, \cos(\omega_d t + \phi_d)], \Phi(t) \in \mathbf{R}^{\mathbf{d}} \tag{A11}$$

where $\omega_i, \phi_i, i = 1, 2, \ldots, d$ denote the frequencies and phase shift of time encoding, respectively. Employing this time encoding, quadruples with the same subject, predicate, and object can have different attention scores. Specifically, quadruples that occurred recently tend to have higher attention scores. This makes the embedding more interpretable and effective.

In fact, the attribute deviation caused by the time deviation is the only assumption obtained after statistics. It is the semantic attributes of entities that determine the reasoning. In order to avoid being only affected by time factors, we propose a new time-specific entity representation $\Psi_k(t) \in \mathbf{R}^{\mathbf{d}}$, i.e., each entity has a different representation at different timestamps. If each entity applies different representations at every moment, it will consume enormous resources. As most of the entities are only observed at limited timestamps, this paper characterizes the entities whose timestamps only appear in the training dataset. IMR utilizes the embedding of the separate entity when it last occurred in the training dataset to represent the embedding at the timestamps missing from the training dataset. Moreover, we apply regularizations on time continuity to avoid over-fitting caused by too many parameters. This regularization believes that the temporally continuous entities should have closer embeddings, which is described in Section 4.5. Finally, we combine $\Phi(t)$ and $\Psi_k(t)$ to construct $\mathbf{e}_{\mathbf{dy}-k}^{t} \in \mathbf{R}^{\mathbf{2d}}$.

$$\mathbf{e}_{\mathbf{dy}-k}^{t} = [\ \Phi(t) \ || \ \Psi_k(t)\ ] \tag{A12}$$

In summary, the embedding for each entity $e_k^t$ can be represented as follows:

$$\mathbf{e}_k^t = act\Big(MLP([\ \mathbf{e}_{\mathbf{sta}-k} \ || \ \mathbf{e}_{\mathbf{dy}-k}^{t}\ ])\Big) \tag{A13}$$

The entities' timestamps in actual datasets are sparse, e.g., ICEWS114 and YAGO have only 11 and 21 timestamps per entity on average. In view of the huge memory usage, we reduce the parameters by basis vectors in actual implementations. The entities' dynamic embeddings are linearly combined by 50 shared vectors. Table A3 shows the memory usage in the ablation experiments on entity-time-specific embeddings.

**Table A3.** The memory usage of the ablation experiments on entity-time-specific embeddings.

| Dataset | Ent-Time-Specific | Non-Ent-Time-Specific | Memory Increment |
|---|---|---|---|
| ICEWS14 | 45.21 G | 39.84 G | 5.37 G |
| ICEWS18 | 61.45 G | 46.00 G | 15.37 G |
| WIKI | 54.39 G | 21.36 G | 33.03 G |
| YAGO | 38.40 G | 26.60 G | 11.80 G |

We can find that using entities' dynamic embeddings brings an extra 5-15 G in memory usage, which is under the affordable range.

### *Appendix A.4. Combination of Indicators*

The three indicators measure different aspects of the path: the matching degree between answers and the question, the completeness of relational equivalence, and the reliability of the reasoning paths. We verify the performance of each metric through ablation experiments. As shown in Tables A4 and A5, the first block displays the performance with only one indicator, the second block presents the performance with a combination of two parameters, and the last is a combination of three indicators. The bottom line shows the error between the combination of the three parameters and the best result. Since the distribution varies across two datasets, there are certain differences in performance when employing a single indicator to rank paths. The model's performance has significantly improved after incorporating the three indicators in pairs, but a few differences remain. IMR-TransE can obtain the best inference performance in most datasets by combining three indicators. In summary, the experiment illustrates that the combination of three indicators designed by IMR-TransE can effectively measure the reasoning paths.

**Table A4.** The comparison of the three indicators in different combinations between YAGO and ICEWS14 datasets. We average the output of ten experiments with different random seeds and fixed hyperparameters. All metrics are multiplied by 100.

| **Dataset** | **YAGO** | | | | **ICEWS14** | | | |
|---|---|---|---|---|---|---|---|---|
| **Indicator** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** |
| $f_{qmd}$ | 87.32 | 92.53 | 92.76 | 89.87 | 22.61 | 39.20 | 55.32 | 33.48 |
| $f_{ac}$ | 87.79 | 92.67 | 92.78 | 90.18 | 31.67 | 46.02 | 59.21 | 41.05 |
| $f_{pc}$ | 87.74 | 92.67 | 92.77 | 90.15 | 25.65 | 43.03 | 58.25 | 36.63 |
| $f_{ac}, f_{qmd}$ | 87.95 | **92.67** | **92.77** | 90.26 | 34.91 | 49.26 | 61.12 | 43.82 |
| $f_{pc}, f_{qmd}$ | 87.74 | **92.67** | 92.75 | 90.15 | 25.64 | 43.16 | 58.30 | 36.63 |
| $f_{ac}, f_{pc}$ | 87.91 | 92.65 | **92.77** | 90.24 | 34.81 | 49.02 | **61.15** | 43.74 |
| $f_{ac}, f_{pc}, f_{qmd}$ | **88.31** | 92.66 | **92.77** | **90.48** | **34.96** | **49.27** | 61.09 | **43.89** |
| Distance to the best | 0 | 0.01 | 0 | 0 | 0 | 0 | 0.06 | 0 |

**Table A5.** The comparison of the three indicators in different combinations between WIKI and ICEWS18 datasets. We average the output of ten experiments with different random seeds and fixed hyperparameters. All metrics are multiplied by 100.

| **Dataset** | **WIKI** | | | | **ICEWS18** | | | |
|---|---|---|---|---|---|---|---|---|
| **Indicator** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** | **Hit@1** | **Hit@3** | **Hit@10** | **MRR** |
| $f_{qmd}$ | 70.75 | 83.39 | 85.87 | 77.12 | 12.76 | 26.47 | 43.75 | 22.66 |
| $f_{ac}$ | - | - | - | - | 20.41 | 33.50 | 47.48 | 29.45 |
| $f_{pc}$ | 70.72 | 83.35 | 85.31 | 77.00 | 14.92 | 29.00 | 45.58 | 24.82 |
| $f_{ac}, f_{qmd}$ | **76.12** | 84.90 | 85.94 | **80.46** | 23.05 | **36.20** | 49.47 | 31.84 |
| $f_{pc}, f_{qmd}$ | 73.85 | 84.12 | 85.65 | 78.99 | 13.10 | 26.38 | 43.27 | 22.75 |
| $f_{ac}, f_{pc}$ | 76.04 | 84.91 | 85.95 | 80.41 | 23.04 | 36.10 | 49.46 | 31.83 |
| $f_{ac}, f_{pc}, f_{qmd}$ | 76.09 | **84.92** | **85.96** | 80.44 | **23.15** | 36.12 | **49.52** | **31.89** |
| Distance to the best | 0.03 | 0 | 0 | 0.02 | 0 | 0.08 | 0 | 0 |

From the above experiments, we can only use two indicators in IMR-TransE. However, IMR can be instantiated based on other models. For example, the performance of IMR-RotatE with any two indicators is quite different. Thus, we should reserve all indicators for the best performance.

### *Appendix A.5. Correlation between IMR and Other Models*

**Correlation between IMR and PTransE.** Both IMR and PTransE consider measuring the semantic equivalence between relations. PtransE resembles the ensemble, which combines the scores of relations and triples in different models. IMR indicators are based on unified theoretical models (such as TransE or RotatE), which can effectively combine different paths. IMR can truly measure the paths of different hops under the same criteria. Moreover, IMR further designs path confidence for time attributes.

**Correlation between IMR and reinforcement-learning-based models.** First, the reinforcement learning models are black-box models, which cannot explain the basis of judgment. Moreover, reinforcement learning utilizes rewards, which is essentially a measure of the matching degree between tails and the question. This end-to-end design is essentially that of the question matching degree in IMR, which is unexplainable and complicated.

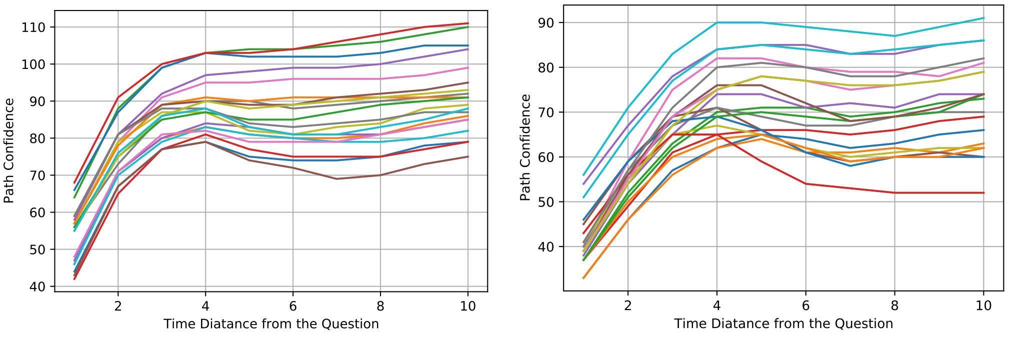
Moreover, IMR is the first to design indicators from the perspective of actual semantics, so we select the basic embedding models as the basis for IMR to better illustrate the pathway. The modeling of triples in TransE is elementary, so the formulas of indicators are simple. Compared with the complex greedy algorithm, it is natural to take the design of IMR as too simple. Although the design of IMR-TransE is simple, it achieves better performance than reinforcement learning models, such as Titer. The indicators of instantiated IMR models can be more complex and their performance will be better.

Finally, we should design other indicators of IMR based on consistent basic models (such as RotatE). Current reinforcement learning models are commonly based on multi-layer networks. We cannot further design the other two indicators.

### *Appendix A.6. Correlation between Path Confidence and Time Distance in IMR-TransE*

The current sampling strategy believes that the greater the time distance of the same entity, the greater the deviation of its semantic properties. Therefore, IMR adopts a time-negative sampling strategy to search for more effective paths. Path reliability is affected by semantic similarity, and negative time-aware correlation is a general situation or statistical result. IMR proposes path reliability to better measure the reliability of the searched path. Here, we utilize the path confidence of the same path with different timestamps to analyze the changes in semantic similarity over time. For the same problem, we find the same path with various timestamps. We randomly select 20 questions for the path search, and each question selects the same path containing ten different timestamps to calculate the path confidence. Figure A1 shows how the path confidence of each path changes with time and distance.

Figure A1 shows that as the time distance between the paths and questions increases, the score of path confidence gradually increases, indicating that its confidence is gradually decreasing. Experiments show that the semantic deviation of the same entity increases as the time distance increases, which verifies the rationality of time-aware negative exponential sampling.



(**a**) YAGO (**b**) WIKI

**Figure A1.** The relation between path confidence and time distance. The questions and paths corresponding to each polyline are shown in (**a**,**b**).

### *Appendix A.7. The Offsetting Property in Question Updating*

In order to infer the correct tails, the query updating module should satisfy that the question still matches the same tail entity even after updating. As shown in Equation (A14), we take IMR-TransE to analyze the offsetting property.

$$\begin{aligned} \mathbf{e_{q\_i}} + \mathbf{r_{q\_i}} &= \mathbf{e_{q\_i-1}} + \mathbf{r_{pi}} + \mathbf{r_{q\_i-1}} - \mathbf{r_{pi}} \\ &= \mathbf{e_{q\_i-1}} + \mathbf{r_{q\_i-1}} \\ &= \mathbf{e_{q\_0}} + \mathbf{r_{q\_0}} \\ &= \mathbf{e_q} + \mathbf{r_q} \\ &= \mathbf{e_o} \end{aligned} \tag{A14}$$

This cancellation of the relation guarantees that the answers to questions will not change along with the paths. In addition, the offset will not appear in the calculation of the indicator. Only the subject of the question is applied in the calculation of the path confidence, and only the relation in the question is used in the calculation of the answer completion level.

## References

1. Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the NIPS 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013.
2. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A large ontology from wikipedia and wordnet. *J. Web Semant.* **2008**, *6*, 203–217. [CrossRef]
3. Li, Z.; Jin, X.; Guan, S.; Li, W.; Guo, J.; Wang, Y.; Cheng, X. Search from History and Reason for Future: Two-stage Reasoning on Temporal Knowledge Graphs. In Proceedings of the ACL/IJCNLP (1), Virtual Event, 1–6 August 2021; pp. 4732–4743.
4. Jin, W.; Zhang, C.; Szekely, P.A.; Ren, X. Recurrent Event Network for Reasoning over Temporal Knowledge Graphs. *arXiv* **2019**, arXiv:1904.05530.
5. Xu, C.; Nayyeri, M.; Alkhoury, F.; Yazdi, H.S.; Lehmann, J. Temporal Knowledge Graph Completion Based on Time Series Gaussian Embedding. In Proceedings of the ISWC (1), Athens, Greece, 2–6 November 2020; Lecture Notes in Computer Science. Springer: Berlin/Heidelberg, Germany, 2020; Volume 12506, pp. 654–671.
6. Jung, J.; Jung, J.; Kang, U. Learning to Walk across Time for Interpretable Temporal Knowledge Graph Completion. In Proceedings of the KDD, Singapore, 14–18 August 2021; ACM: New York, NY, USA, 2021; pp. 786–795.
7. Han, Z.; Chen, P.; Ma, Y.; Tresp, V. Explainable Subgraph Reasoning for Forecasting on Temporal Knowledge Graphs. In Proceedings of the ICLR, OpenReview.net, Vienna, Austria, 4 May 2021.
8. Wu, J.; Cao, M.; Cheung, J.C.K.; Hamilton, W.L. TeMP: Temporal Message Passing for Temporal Knowledge Graph Completion. In Proceedings of the EMNLP (1), Online, 16–20 November 2020; pp. 5730–5746.
9. Pavlović, A.; Sallinger, E. ExpressivE: A Spatio-Functional Embedding For Knowledge Graph Completion. *arXiv* **2022**, arXiv:2206.04192.
10. Wang, X.; Chen, J.; Wu, F.; Wang, J. Exploiting Global Semantic Similarities in Knowledge Graphs by Relational Prototype Entities. *arXiv* **2022**, arXiv:2206.08021.
11. Zhu, C.; Chen, M.; Fan, C.; Cheng, G.; Zhan, Y. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks. *arXiv* **2020**, arXiv:2012.08492.
12. Nayyeri, M.; Vahdati, S.; Khan, M.T.; Alam, M.M.; Wenige, L.; Behrend, A.; Lehmann, J. Dihedron Algebraic Embeddings for Spatio-Temporal Knowledge Graph Completion. In Proceedings of the Semantic Web—19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, 29 May–2 June 2022; Lecture Notes in Computer Science; Groth, P., Vidal, M., Suchanek, F.M., Szekely, P.A., Kapanipathi, P., Pesquita, C., Skaf-Molli, H., Tamper, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2022; Volume 13261, pp. 253–269. [CrossRef]
13. Chen, K.; Wang, Y.; Li, Y.; Li, A. Rotateqvs: Representing temporal information as rotations in quaternion vector space for temporal knowledge graph completion. *arXiv* **2022**, arXiv:2203.07993.
14. Sun, H.; Zhong, J.; Ma, Y.; Han, Z.; He, K. TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting. In Proceedings of the EMNLP, Virtual, 7–9 November 2021.
15. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex Embeddings for Simple Link Prediction. In Proceedings of the ICML, New York, NY, USA, 19–24 June 2016.
16. Yang, B.; Yih, W.T.; He, X.; Gao, J.; Deng, L. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *arXiv* **2015**, arXiv:1412.6575.

17. Sun, Z.; Deng, Z.; Nie, J.Y.; Tang, J. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *arXiv* **2019**, arXiv:1902.10197.
18. Nickel, M.; Tresp, V.; Kriegel, H. A Three-Way Model for Collective Learning on Multi-Relational Data. In Proceedings of the ICML, Bellevue, WA, USA, 28 June–2 July 2011.
19. Zhou, M.; Huang, M.; Zhu, X. An Interpretable Reasoning Network for Multi-Relation Question Answering. In Proceedings of the COLING, Santa Fe, NM, USA, 20–26 August 2018; pp. 2010–2022.
20. Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge Graph Embedding by Translating on Hyperplanes. In Proceedings of the AAAI, Québec City, QC, Canada, 27–31 July 2014.
21. Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In Proceedings of the AAAI, Austin, TX, USA, 14–18 November 2015.
22. Ji, G.; He, S.; Xu, L.; Liu, K.; Zhao, J. Knowledge Graph Embedding via Dynamic Mapping Matrix. In Proceedings of the ACL, Beijing, China, 26–31 July 2015.
23. Balazevic, I.; Allen, C.; Hospedales, T.M. TuckER: Tensor Factorization for Knowledge Graph Completion. *arXiv* **2019**, arXiv:1901.09590.
24. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2D Knowledge Graph Embeddings. In Proceedings of the AAAI, New Orleans, LA, USA, 2–7 February 2018.
25. Nguyen, D.Q.; Nguyen, T.; Nguyen, D.Q.; Phung, D.Q. A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network. *arXiv* **2018**, arXiv:1712.02121.
26. Nguyen, D.Q.; Vu, T.; Nguyen, T.; Nguyen, D.Q.; Phung, D.Q. A Capsule Network-based Embedding Model for Knowledge Graph Completion and Search Personalization. *arXiv* **2019**, arXiv:1808.04122.
27. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P. Composition-based Multi-Relational Graph Convolutional Networks. *arXiv* **2020**, arXiv:1911.03082.
28. Li, R.; Cheng, X. DIVINE: A Generative Adversarial Imitation Learning Framework for Knowledge Graph Reasoning. In Proceedings of the EMNLP/IJCNLP (1), Hong Kong, China, 3–7 November 2019; pp. 2642–2651.
29. Wang, H.; Li, S.; Pan, R.; Mao, M. Incorporating Graph Attention Mechanism into Knowledge Graph Reasoning Based on Deep Reinforcement Learning. In Proceedings of the EMNLP/IJCNLP (1), Hong Kong, China, 3–7 November 2019; pp. 2623–2631.
30. García-Durán, A.; Dumancic, S.; Niepert, M. Learning Sequence Encoders for Temporal Knowledge Graph Completion. In Proceedings of the EMNLP, Brussels, Belgium, 31 October–4 November 2018.
31. Messner, J.; Abboud, R.; Ceylan, İ.İ. Temporal Knowledge Graph Completion Using Box Embeddings. In Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, the Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022, Virtual Event, 22 February–1 March 2022; pp. 7779–7787.
32. Lacroix, T.; Obozinski, G.; Usunier, N. Tensor Decompositions for temporal knowledge base completion. *arXiv* **2020**, arXiv:2004.04926.
33. Boschee, E.; Lautenschlager, J.; O'Brien, S.; Shellman, S.; Starz, J.; Ward, M. *Icews Coded Event Data*; Harvard Dataverse: Cambridge, MA, USA, 2015; Volume 12.
34. Leblay, J.; Chekol, M. Deriving Validity Time in Knowledge Graph. In Proceedings of the Web Conference 2018, Lyon, France, 23–27 April 2018.
35. Mahdisoltani, F.; Biega, J.; Suchanek, F.M. YAGO3: A Knowledge Base from Multilingual Wikipedias. In Proceedings of the CIDR, Asilomar, CA, USA, 4–7 January 2015.
36. Miller, G.A. WordNet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
37. Dasgupta, S.S.; Ray, S.N.; Talukdar, P.P. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding. In Proceedings of the EMNLP, Brussels, Belgium, 31 October–4 November, 2018; pp. 2001–2011.
38. Jin, W.; Qu, M.; Jin, X.; Ren, X. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs. In Proceedings of the EMNLP, Online, 16–20 November 2020.
39. Goel, R.; Kazemi, S.M.; Brubaker, M.A.; Poupart, P. Diachronic Embedding for Temporal Knowledge Graph Completion. *arXiv* **2020**, arXiv:1907.03143.
40. Ding, Z.; Han, Z.; Ma, Y.; Tresp, V. Temporal Knowledge Graph Forecasting with Neural ODE. *arXiv* **2021**, arXiv:2101.05151.
41. Xu, D.; Ruan, C.; Körpeoglu, E.; Kumar, S.; Achan, K. Inductive Representation Learning on Temporal Graphs. *arXiv* **2020**, arXiv:2002.07962.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.