CognTKE: A Cognitive Temporal Knowledge Extrapolation Framework

Wei Chen[1,2], Yuting Wu[3], Shuhan Wu[1,2], Zhiyu Zhang[1,2], Mengqi Liao[1,2], Youfang Lin[1,2], Huaiyu Wan[1,2] *

[1]School of Computer Science & Technology, Beijing Jiaotong University, Beijing, China
[2]Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence, Beijing, China
[3]School of Software, Beijing Jiaotong University, Beijing, China
{w_chen, ytwu1, wushuhan, zyuzhang, mqliao, yflin, hywan}@bjtu.edu.cn

Abstract

Reasoning future unknowable facts on temporal knowledge graphs (TKGs) is a challenging task, holding significant academic and practical values for various fields. Existing studies exploring explainable reasoning concentrate on modeling comprehensible temporal paths relevant to the query. Yet, these path-based methods primarily focus on local temporal paths appearing in recent times, failing to capture the complex temporal paths in TKG and resulting in the loss of longer historical relations related to the query. Motivated by the Dual Process Theory in cognitive science, we propose a **Cogn**itive **T**emporal **K**nowledge **E**xtrapolation framework (CognTKE), which introduces a novel temporal cognitive relation directed graph (TCR-Digraph) and performs interpretable global shallow reasoning and local deep reasoning over the TCR-Digraph. Specifically, the proposed TCR-Digraph is constituted by retrieving significant local and global historical temporal relation paths associated with the query. In addition, CognTKE presents the global shallow reasoner and the local deep reasoner to perform global one-hop temporal relation reasoning (System 1) and local complex multi-hop path reasoning (System 2) over the TCR-Digraph, respectively. The experimental results on four benchmark datasets demonstrate that CognTKE achieves significant improvement in accuracy compared to the state-of-the-art baselines and delivers excellent zero-shot reasoning ability.

Introduction

Temporal knowledge graphs (TKGs) store the structured dynamic facts in the real world, consisting of a sequence of knowledge graph (KG) snapshots accompanied by corresponding timestamps. Each fact in TKGs can be represented as a quadruple in the form of (subject, relation, object, time), e.g., *(The World Cup, Be_held_at, Qatar, 2022)*. Reasoning over TKGs aims to infer new facts based on the historical KG snapshots, which can be divided into two settings: interpolation and extrapolation. According to the given history facts in time interval $[t_0, t_{|\mathcal{T}|}]$, the former attempts to complete missing facts that appear in the past time $t \in [t_0, t_{|\mathcal{T}|}]$, while the latter focuses on predicting facts (events) happening in the future time $t > t_{|\mathcal{T}|}$. Due to its ability to infer future new events, extrapolation in TKG has a wider range of application scenarios (Chen et al. 2022b; Guan et al. 2023), such as stock prediction, traffic prediction, and medical assistance. This paper primarily emphasizes the more challenging TKG extrapolation task that forecasts future unknown facts (or events).

*Corresponding authors.

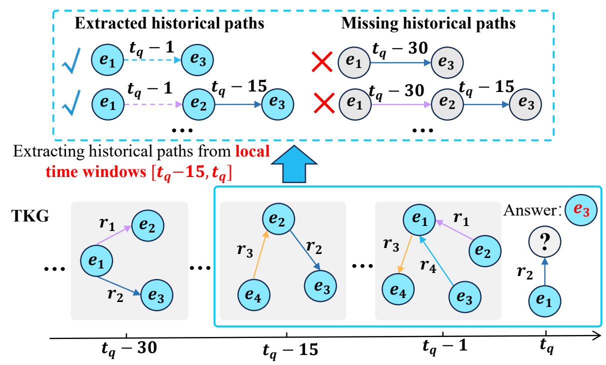

Figure 1: An example illustrates longer history temporal relation path loss. Relations are denoted using solid lines with different colors, and inverse relations are denoted using dashed lines. Black solid line denotes the identity relation.

Many arts for TKG extrapolation have been extensively explored and achieved excellent prediction performance. These studies mainly involve two categories (Mei et al. 2022; Liang et al. 2024a): embedding-based methods and path-based methods. Embedding-based methods learn the evolution embeddings of entity and relation by modeling the facts repeating and cyclic patterns, such as CyGNet (Zhu et al. 2021) and CENET (Xu et al. 2023), and the local adjacent facts evolution patterns, such as RE-GCN (Li et al. 2021b) and RETIA (Liu et al. 2023). However, due to the black-box characteristics of embedding-based methods, they lack explicit evidence to explain the reasoning results.

Naturally, some path-based methods such as TLogic (Liu et al. 2022b) and KartGPS (Xin and Chen 2024), are proposed to better capture the chain of evidence for TKG reasoning. Nevertheless, path-based methods focus on temporal relational paths modeling in the specific time window closest to the query time, resulting in the loss of longer historical information. As shown in Figure 1, given a local time window $[t_q - 15, t_q - 1]$, existing path-based methods only capture paths $(e_1, {r_4}^{-1}, e_3, t_q - 1)$[1] and $(e_1, {r_1}^{-1}, e_2, t_q - 1) \rightarrow (e_2, r_2, e_3, t_q - 15)$ based on the query $(e_1, r_2, ?, t_q)$, but ignoring distant historical paths information, such as $(e_1, r_1, e_2, t_q - 30) \rightarrow (e_2, r_2, e_3, t_q - 15)$. Although the time window of path modeling can extend to the entire history, it makes path retrieval expensive and may bring numerous earlier temporal relations irrelevant to the query. In addition, the expressiveness of path-based methods is still limited to individual paths, which leads to only a small subset of first-order logical formulas being represented in TKGs. Subgraphs can naturally be more informative than paths and are demonstrated to handle complex logical formulas for reasoning (Lin et al. 2023; Xin and Chen 2024). Unfortunately, efficiently retrieving valuable path information from the entire history to construct subgraphs and conducting reasoning in TKGs poses a significant challenge.

Deriving inspiration from the dual process theory (Evans and BT 1984; Evans 2003, 2008; Sloman and A 1996), a theory in cognitive science that explains the cognitive processes involving in thinking and reasoning. Dual process theory argues that the reasoning system of humans first retrieves much relevant information for quick decisions through a shallow and intuitive process called System 1. Another controllable and logical reasoning process called System 2, performs deeper sequential thinking and complex reasoning based on System 1. Although existing extrapolation methods (Li et al. 2021a; Liu et al. 2022a) explore the two-process theory to improve performance, they are still limited in terms of the interpretability of reasoning results.

In this paper, we propose a novel **Cogn**itive **T**emporal **K**nowledge **E**xtrapolation framework (CognTKE), which combines the strengths of both embedding-based and path-based methods. CognTKE involves two processes: retrieving temporal relations from the TKG to build subgraphs and performing global shallow reasoning and local deep reasoning over subgraphs. Specifically, we present a temporal cognitive relation directed graph (TCR-Digraph) as a core subgraph in CognTKE, which expands the scope of temporal relation paths to subgraphs, preserving crucial temporal relation paths for TKG reasoning. In TCR-Digraph, the first layer consists of query-related global one-hop historical facts and the subsequent layers consist of query-related local multi-hop historical facts. To effectively learn the historical temporal relation paths at different layers, we propose two reasoners sequentially encode global one-hop temporal relation and local multi-hop complex paths over TCR-Digraph, respectively. The two distinct learning processes bear resemblance to the concepts of System 1 and System 2 in the dual process theory, which sequentially performs fast thinking for intuitive reasoning and slow thinking for complex reasoning, respectively. In summary, our contributions are as follows:

- We propose a novel CognTKE framework for TKG extrapolation based on human cognition, which efficiently and sequentially handles the global shallow one-hop temporal relation reasoning and local deep complex multi-hop temporal path reasoning over TCR-Digraph.
- We introduce a novel TCR-Digraph that captures crucial local and global temporal relation information based on TKG, providing more comprehensive evidence for the interpretation of TKG reasoning.
- Extensive experiments on four public datasets demonstrate that CognTKE achieves significant improvement in accuracy over the state-of-the-art baselines and has strong zero-shot reasoning ability.

[1] ${r_4}^{-1}$ denotes the inverse relation of the relation r_4

Related Work

Interpolation on TKGs Interpolation on TKGs aims to infer the historical missing facts. Most interpolation works are developed by extending static KG reasoning methods. For instance, TTransE (Leblay and Chekol 2018) is a translation-based method and introduces time constraints between facts that have common entities. RotateQVS (Chen et al. 2022a) models temporal evolutions as rotations in quaternion vector space, effectively capturing a variety of complex relational patterns. HGE-TNTComplEX (Pan et al. 2024) projects temporal facts onto a product space of heterogeneous geometric subspaces and employs temporal-geometric attention mechanisms for interpolation.

Extrapolation on TKGs Extrapolation on TKGs aims to predict future unseen facts and can be classified as embedding-based methods and path-based methods. Some embedding-based methods explore the cyclic and repetitive patterns of historical facts from a global perspective. For example, CyGNet (Zhu et al. 2021) employs a copy-generation mechanism to learn query-relevant entities that are important in repeated historical facts. Other approaches study the local dependency of historical facts based on the local KG snapshots. For example, RE-GCN (Li et al. 2021b) models the evolution of entities and relations at each timestamp by using the local historical dependency. CEN (Li et al. 2022) exploits a curriculum learning approach to model variable snapshot graph sequence lengths. Recent embedding-based methods such as TiRGN (Li, Sun, and Zhao 2022), integrate the above two historical patterns to achieve more accurate reasoning. LogCL (Chen et al. 2024) adopts a contrastive learning to enhance the representation of local and global representation of queries. Since the black box characteristics of embedding-based methods, it is difficult to provide explainable evidence for reasoning results. Path-based methods generate temporal rules by sampling local temporal relations, and extend temporal rules to predict future facts. For example, TLogic (Liu et al. 2022b) mines the temporal logic rules from the loop relation path in a given time window and designs a strategy for evaluating the confidence of candidate temporal logic rules. TempValid (Huang et al. 2024) designs a rule-adversarial and a time-aware negative sampling strategies to obtain competitive results.

Our Approach

Preliminaries

A TKG $\mathcal{G}$ is denoted as a sequence of KG snapshots, i.e., $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_{|\mathcal{T}|}\}$. The KG snapshot $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$ consists of a set of valid facts at $t \in \mathcal{T}$, where $\mathcal{E}, \mathcal{R}, \mathcal{T}$ denote the set of entities, relations, and timestamps, respec-

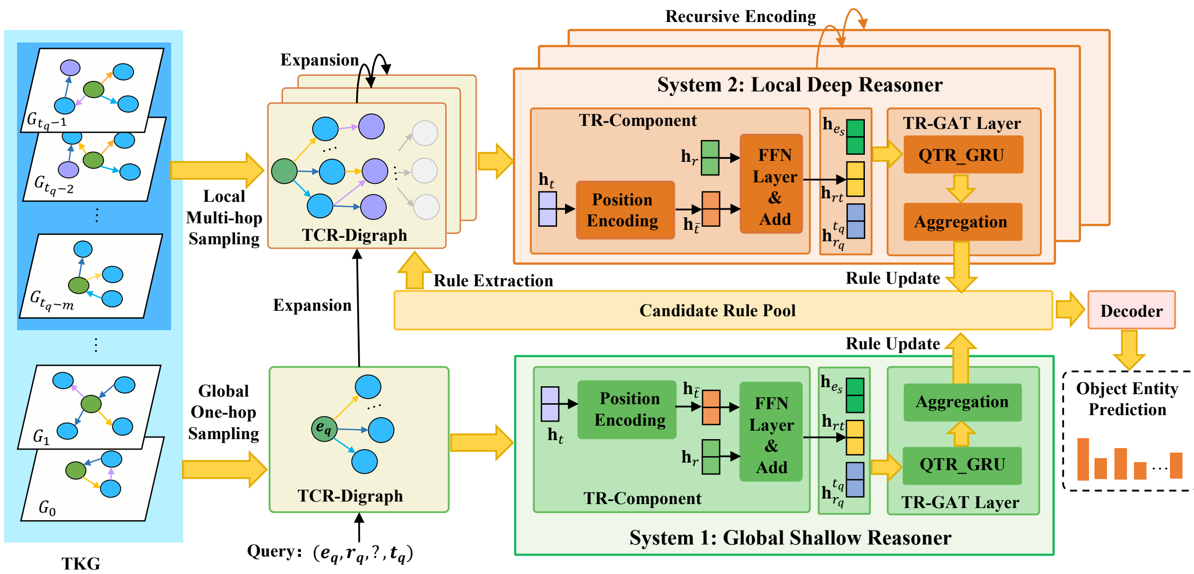

Figure 2: An illustrative diagram of the proposed CognTKE architecture.

tively; $\mathcal{F}_t$ denotes the set of facts at t. A fact (or an event) is represented in the form of a quadruple $(e_s, r, e_o, t) \in \mathcal{F}_t$, where an edge r serves as an association between the subject entity e_s and object entity e_o at t.

Definition 1 (Temporal Relation Path over TKG). *The temporal relation path with length l is a set of l quadruples $(e_1, r_1, e_2, t_1) \rightarrow \cdots \rightarrow (e_{l-1}, r_{l-1}, e_l, t_{l-1}) \rightarrow (e_l, r_l, e_{l+1}, t_l)$ with $t_1 \leq ... \leq t_{l-1} \leq t_l$, that are connected origin entity e_1 to destination entity e_{l+1} sequentially.*

Definition 2 (Problem Statement). *TKG extrapolation aims to forecast the future unknown object entity (or subject entity) given a query $(e_q, r_q, ?, t_q)$ (or $(?, r_q, e_q, t_q)$) according to previous historical KG snapshots $\{\mathcal{G}_0, \mathcal{G}_1, ..., \mathcal{G}_{t_q-1}\}$.*

Note that, the inverse relation quadruples (e_s, r^{-1}, e_o, t) and identity quadruples (e_s, r_{self}, e_s, t) are added to the TKG dataset, where $r_{self} \in \mathcal{R}$ is an identity relation. Without loss of generality, the TKG reasoning goal can be expressed as the prediction of object entities.

Architecture Overview

The whole framework of CognTKE is shown in Figure 2, involving the three processes: (1) Retrieving temporal relations from the TKG to build TCR-Digraph; (2) Performing global shallow encoding and local deep encoding over TCR-Digraph to generate crucial candidate rules to associate with the query; (3) Based on the candidate rule pool, a decoder is employed to calculate the scores of entities.

Temporal Cognitive Relation Digraph

Temporal relation paths have been proven to possess high transferability and interpretability in TKG (Liu et al. 2022b; Xin and Chen 2024). However, the limitation lies in their ability to model complex multi-hop reasoning due to a small subset of local temporal relation paths learned in TKGs. Inspired by (Zhang and Yao 2022; Wu et al. 2023; Liang et al. 2024b), we introduce a temporal cognitive relation directed graph (TCR-Digraph) that provides rule-like explanatory and inductive capabilities to explore the significant historical chain of evidence. Next, we will introduce some definitions related to TCR-Digraph.

Definition 3 (Layered Graph(Battista et al. 1999)) *The layered graph is a directed graph with exactly one source node and one sink node (destination node). All edges are directed, connecting nodes between consecutive layers and pointing from $L-1$-th layer to L-th layer.*

Definition 4 (Temporal Relation Directed Graph, TR-Digraph). *The TR-Digraph $\bar{\mathcal{G}}_{e_q^{t_q}, e_o^t | L}$ is a layered graph with the source entity (query entity) $e_q^{t_q}$ at query time t_q and the sink entity (target entity) e_o^t at history time t, where $t_q > t$ and L is the number of layers in the TR-Digraph. Entities on the same layer are different from each other. Each edge $(r_L t_L)^L$ formed by a combination of relation r and history time t is a direction, connecting an entity in the $L-1$-th layer to an entity in the L-th layer. $\bar{\mathcal{G}}_{e_q^{t_q}, e_o^t | L}$ is set as $\emptyset$ if there is no temporal relation path connecting $e_q^{t_q}$ and e_o^t with length L.*

Here, the quadruples with the reverse or identity relations are considered in the TR-Digraph. Through the above definition, any temporal relation path between query entity $e_q^{t_q}$ at query time t_q and the target entity e_o^t at specific history time t in the TR-Digraph $\bar{\mathcal{G}}_{e_q^{t_q}, e_o^t | L}$ is denoted as $e_q^{t_q} \rightarrow (r_1 t_1)^1 \rightarrow (r_2 t_2)^2 \rightarrow \cdots \rightarrow (r_L t_L)^L \rightarrow e_o^t$ with length L and $t_q \geq t_L$, where $(r_L t_L)^L$ connects an entity in the $L-1$-layer to the entity in L-layer. From a high-level perspective, the TR-DiGraph can be seen as a subgraph composed of multiple temporal paths that originate from a single source node. This implies that we can construct the query TR-Digraph by iteratively retrieving the historical tempo-

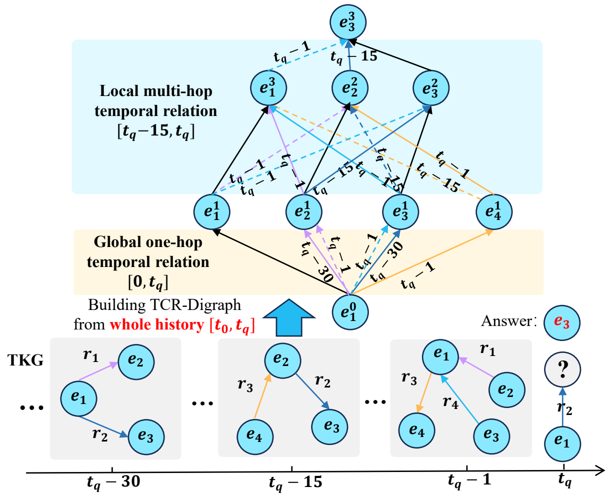

Figure 3: An illustrative of TCR-Digrph $\hat{\mathcal{G}}_{e_1,e_3|3}$ formed by the TKG.

ral relationship paths associated with the query entity $e_q^{t_q}$ in the TKG. Thus, all temporal relation paths between the query entity $e_q^{t_q}$ and the target entity e_o^t can be preserved in TR-Digraph. Unlike existing methods (Liu et al. 2022b; Sun et al. 2021) that impose temporal order constraints on consecutive temporal relations in each temporal relation path, we advocate for exploring additional features between temporal relations, such as causality, without limiting temporal order.

In practice, such TR-Digraph can bring much earlier historical information that is not directly relevant to the query, which not only makes the TR-Digraph construction and encoding expensive but also does not facilitate subsequent entity reasoning. Therefore, we propose an improved TCR-Digraph $\hat{\mathcal{G}}_{e_q^{t_q},e_o^t|L}$ based on TR-Digraph, which contains important global one-hop history quadruples and recent local multi-hop history quadruples relevant to the query. To be specific, the first layer of the TCR-Digraph is built by retrieving the global query-related historical facts in the whole history time windows $[0, t_q]$. Based on candidate entities obtained in the previous layer, the subsequent layers of the TCR-Digraph is built by further retrieving local historical facts associated with candidate entities in recent time windows $[t_q - m, t_q]$, where m is the length of local time windows. An example of TCR-Digraph $\hat{\mathcal{G}}_{e_1,e_3|3}$ in Figure 3, which is adopted to infer the new facts (e_1, r_2, e_3, t_q) based on layered paths in TCR-Digraph $\hat{\mathcal{G}}_{e_1,e_3|3}$. For this purpose, TCR-Digraph preserves the crucial temporal relation paths associated with the query, making it more expressive and providing significant reasoning evidence and lower computational cost from the entire history.

Recursive Encoding over TCR-Digraph

As discussed above, the TCR-Digraph provides the interpretable chain of evidence consisting of global one-hop temporal relation and local multi-hop temporal relation paths. However, how to model different temporal relation paths on TCR-Digraph for temporal knowledge reasoning poses another problem. Inspired by the dual process theory of cognitive science (Evans 2008), we present a global shallow reasoner and a local deep reasoner consisting of *temporal relation component* (TR-Component) and *temporal relation graph attention network* (TR-GAT). The global shallow reasoner as the first decision process aims to carry out intuitive global one-hop reasoning, which can be seen as fast thinking in human cognition. The local deep reasoner as the second decision process performs local complex multi-hop reasoning, which can be regarded as slow thinking in human cognition. Although both reasoners perform different functions, similar pipelines of iteratively encoding each layer of the TCR-Digraph are employed to update the candidate rule pool. Note that each layer construction and encoding of TCR-Digraph is performed alternately. The candidate rule pool is a latent representation of candidate entities in TCR-Digraph, which is considered as working memory (Baddeley 1992) used to store updated rules for subsequent entity prediction.

Temporal Relation Component The TR-Component aims to learn the representation of temporal relation in TCR-Digraph and serves as an input to TR-GAT. As historical facts associated with the query at different times have different effects on reasoning results, we adopt the positional encoding method (Vaswani et al. 2017) to map relative time into embedding representations. Formally, the relative time encoding formula is as follows:

$$\mathbf{TE}^{(pos,2i)} = \sin(pos/10000^{2i/d_{time}}), \tag{1}$$

$$\mathbf{TE}^{(pos,2i+1)} = \cos(pos/10000^{2i/d_{time}}), \tag{2}$$

where $\mathbf{TE} \in \mathbb{R}^{|\mathcal{T}| \times d_{time}}$ is the embedding matrix of relative time, pos denotes the location of the time value, i denotes the dimension of the embedding presentation, and d_{time} denotes the embedding size of relative time, $\sin(\cdot)$ and $\cos(\cdot)$ are sine and cosine functions, respectively.

To incorporate the relative time information into the relation while preserving the basic semantic information of the relation, we obtain the representation of the temporal relation rt by the Feed-forward network (FFN) layer (Vaswani et al. 2017) and add operation:

$$\mathbf{h}_{rt} = \sigma_1(\mathbf{W}_2(\sigma_1(\mathbf{W}_1[\mathbf{h}_r, \mathbf{v}_{\bar{t}} + \mathbf{b}_1]) + \mathbf{b}_2) + \mathbf{h}_r, \tag{3}$$

where $[,]$ denotes the concatenate operation of vectors, $\mathbf{h}_r \in \mathbf{H}$ denotes the embedding of relation, $\mathbf{H} \in \mathbb{R}^{|\mathcal{R}| \times d}$ is an initialized relation matrix, $\mathbf{v}_{\bar{t}} \in \mathbf{TE}$ is the embedding of relative time, $\bar{t} = t_q - t$ is the relative time between the historical fact appearing time and the query time, σ_1 is the $LeakyReLU$ activate function, $\mathbf{W}_1 \in \mathbb{R}^{d \times d}, \mathbf{W}_2 \in \mathbb{R}^{d \times (d + d_{time})}, \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^d$ are all learnable parameters, where d is the dimension size of the relation. Through the encoding of temporal relation, the quadruple (e_s, r, e_o, t) in TKGs is converted into the form of triple (e_s, rt, e_o).

Temporal Relation Graph Attention Layer The goal of the TR-GAT layer is to propagate temporal relation information in TCR-Digraph $\hat{\mathcal{G}}_{e_q^{t_q},e_o^t|L}$ centered on the query entity

$e_q^{t_q}$ and update the candidate rule pool that is a latent representation of candidate answer entities. To preserve the ordering of consecutive temporal relations in the TCR-Digraph during message propagation, we present a variant of GRU (QTR_GRU) in TR-GAT to perform message passing. The message passing function is as follows:

$$\mathbf{m}^l_{e_s,rt|r_q^{t_q}} = \mathrm{QTR\_GRU}(\mathbf{h}^{l-1}_{e_s}, \mathbf{h}^l_{rt}, \mathbf{h}^l_{r_q^{t_q}}), \tag{4}$$

where $\mathbf{h}^l_{r_q^{t_q}} \in \mathbb{R}^d$, $\mathbf{h}^l_{e_s} \in \mathbb{R}^d$ and $\mathbf{h}^l_{rt} \in \mathbb{R}^d$ are the representation of the query relation at query time t_q, the subject entity and temporal relation, respectively. Note that, all entity representations are initialized to the zero vector before performing message passing.

Since the construction of the TCR-Digraph $\hat{\mathcal{G}}_{e_q^{t_q}, e_o^t|L}$ is dynamic and highly correlated with query entities and query time, while remaining independent of the query relation. Different queries with the same query entity at the same query time share the same TCR-Digraph, e.g., *(The World Cup, Be_held_at, Qatar, 2022)* and *(The World Cup, Be_won_by, Argentina, 2022)*, leading to the same evidence being utilized for reasoning. Thus, we adopt the graph attention (GAT) (Velickovic et al. 2018) mechanism to encode the query information into the attention weight to distinguish the importance of different edges in the TCR-Digraph. The attention score $\mathbf{a}^l_{e_s,rt|r_q^{t_q}}$ for each triple (e_s, rt, e_o) in the TCR-Digraph is calculated as follows:

$$\mathbf{c}^l_{e_s,rt|r_q^{t_q}} = \sigma_2(\mathbf{W}^l_6 \sigma_1(\mathbf{W}^l_3 \mathbf{h}^l_{e_s} + \mathbf{W}^l_4 \mathbf{h}^l_{rt} + \mathbf{W}^l_5 \mathbf{h}^l_{r_q^{t_q}})), \tag{5}$$

$$\mathbf{a}^l_{e_s,rt|r_q^{t_q}} = \frac{\exp(\mathbf{c}^l_{e_s,rt|r_q^{t_q}})}{\sum_{(\tilde{e_s},\tilde{rt})\in\mathcal{N}^l_{e_o}} \exp(\mathbf{c}^l_{\tilde{e_s},\tilde{rt}|r_q^{t_q}})}, \tag{6}$$

where $\mathbf{W}^l_3, \mathbf{W}^l_4, \mathbf{W}^l_5, \mathbf{W}^l_6$ represent trainable weight matrices, $\mathbf{c}^l_{e_s,rt|r_q^{t_q}}$ denotes the attention weight, and $\mathcal{N}^l_{e_o}$ is the in-degree neighbors of the object entity e_o in TCR-Digraph. Thus, the candidate rule representations of the aggregated entities are calculated as follows:

$$\tilde{\mathbf{h}}^l_{e_o} = \mathbf{W}^l_7 \sum_{(e_s,rt)\in\mathcal{N}^l_{e_o}} \mathbf{a}^l_{e_s,rt|r_q^{t_q}} \mathbf{m}^l_{e_s,rt|r_q^{t_q}}, \tag{7}$$

$$\hat{\mathbf{h}}^l_{e_o} = \frac{\tilde{\mathbf{h}}^l_{e_o}}{\sqrt{indegree(e_o)}}. \tag{8}$$

Among them, $\mathbf{W}^l_7$ is the weight matrix. $indegree(e_o)$ denotes the number of in-degree object entities and serves as a scaling factor to adjust the entity representation. To further retain the feature of existing candidate entities, we utilize GRU (Cho et al. 2014) to control the updating of representations of entities. Thus, the final representation of the candidate entity e_o in l-th layer of TR-GAT can be calculated as follows:

$$\mathbf{h}^l_{e_o} = GRU(\mathbf{h}^{l-1}_{e_o}, \hat{\mathbf{h}}^l_{e_o}). \tag{9}$$

Interpretable Reasoning and Analysis

Prediction After obtaining the final candidate entity representation that includes the appropriate temporal relation rule information, we adopt a simple MLP (Multilayer perceptron) as the decoder to calculate the prediction score of all entities as follows:

$$score(e_o) = f(e_q, r_q, e_o, t_q) = \mathbf{W}_o \mathbf{h}^l_{e_o}, \tag{10}$$

where $\mathbf{W}_o \in \mathbb{R}^{1\times d}$ is the weight matrix. If an entity is not included in the TCR-Digraph corresponding to the query, it will be assigned a score of 0.

We consider entity prediction as a multi-label classification task, and utilize the multi-class log-loss to optimize the parameters Θ of CognTKE, which has been proven to be effective(Zhang and Yao 2022; Lacroix, Usunier, and Obozinski 2018),

$$\begin{aligned}\mathcal{L} = \sum_{(e_q,r_q,e_o,t_q)\in\mathcal{F}_{train}} &(-f(e_q,r_q,e_o,t_q) + \\ &log(\sum_{o'\in\mathcal{E}} \exp(f(e_q,r_q,e_{o'},t_q))),\end{aligned} \tag{11}$$

where (e_q, r_q, e_o, t_q) denotes positive fact in training set $\mathcal{F}_{train}$, and $(e_q, r_q, e_{o'}, t_q)$ denotes other fact with the same query $(e_q, r_q, ?, t_q)$.

Interpretable Analysis Although CognTKE acquires the graph structure of TCR-Digraph through the GAT mechanism, it still possesses the capacity to encode path-based logical rules of the same nature as those utilized in rule induction models, such as TLogic (Liu et al. 2022b) and TECHS(Lin et al. 2023).

Theorem 1. *Given a quadruple (e_q, r_q, e_o, t_q), let p be temporal relation rt, $\mathcal{C}$ be a set of temporal relation paths $e_q^{t_q} \to p_k^1 \to p_k^2 \to \cdots \to p_k^L \to e_o^t$($t_q > t$) that are formed by a set of rules between $e_q^{t_q}$ and e_o^t with the form:*

$$r_q(X,Y) \leftarrow p_k^1(X,Z_1) \wedge p_k^2(Z_1,Z_2) \wedge \cdots \wedge p_k^L(Z_{L-1},Y),$$

where $X, Y, Z_1, \ldots, Z_{L-1}$ are free variables that are bounded by unique entities. Assuming there exists a directed graph $\hat{\mathcal{G}}_{\mathcal{C}}$ that is built by $\mathcal{C}$, a parameter setting Θ, and a threshold $\lambda \in (0,1)$ for CognTKE. $\hat{\mathcal{G}}_{\mathcal{C}}$ can equal to the TCR-digraph if CognTKE's edges have attention weights $\mathbf{c}^\ell_{e_s,rt|r_q^{t_q}} > \theta$ in $\hat{\mathcal{G}}_{e_q^{t_q}, e_o^t|L}$, where θ is a learned decision boundary parameter.

According to Theorem 1, CognTKE is capable of encoding any temporal logical rule that corresponds to a path in the TCR-Digraph. This implies that if a set of temporal relation paths is highly correlated with the query quadruple, CognTKE can identify them through the attention weights, making them interpretable.

Experiments

Datasets To evaluate CognTKE on entity prediction task, we adopt four benchmark datasets that are widely used for TKG extrapolation, including ICE14, ICE18, ICE05-15 (Jin et al. 2020), and WIKI (Li et al. 2021b). Following the preprocessing strategy (Li et al. 2021b; Han et al. 2021; Li et al. 2022), we split all datasets into training, validation, and test sets with the proportions of 80%/10%/10% based on chronological order.

Model	ICE14				ICE18				ICE05-15				WIKI			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
CyGNet	35.05	25.73	39.01	53.55	24.93	15.90	28.28	42.61	36.81	26.61	41.63	56.22	33.89	29.06	36.10	41.86
TANGO	36.80	27.43	40.90	54.94	28.68	19.35	32.17	47.04	42.86	32.72	48.14	62.34	50.43	48.52	51.47	53.58
RE-GCN	40.39	30.66	44.96	59.21	30.58	21.01	34.34	48.75	48.03	37.33	53.85	68.27	77.55	73.75	80.38	83.68
CEN	42.20	32.08	47.46	61.31	31.50	21.70	35.44	50.59	46.84	36.38	52.45	67.01	78.05	75.05	81.90	84.90
TITer	40.87	32.28	45.45	57.10	29.98	22.05	33.46	44.83	47.69	37.95	52.92	65.81	75.50	72.96	77.49	79.02
TLogic	43.04	33.56	48.27	61.23	29.82	20.54	33.95	48.53	46.97	36.21	53.13	67.43	-	-	-	-
TiRGN	44.04	33.83	48.95	63.84	33.66	23.19	37.09	54.22	50.04	39.25	56.13	70.71	81.65	77.77	85.12	87.08
CENET	39.02	29.62	43.23	57.49	27.85	18.15	31.63	46.98	41.95	32.17	46.93	60.43	40.52	32.91	45.11	53.08
RETIA	42.76	32.28	47.77	62.75	32.43	22.23	36.48	52.94	47.26	36.64	52.90	67.76	78.58	74.79	81.45	84.60
TECHS	43.88	34.59	49.36	61.95	30.85	21.81	35.39	49.82	48.38	38.34	54.69	68.92	75.98	-	-	82.89
TempValid	45.78	35.50	**51.34**	**65.06**	33.50	23.91	37.89	52.33	50.31	39.46	56.71	70.55	83.19	74.64	**90.12**	**97.54**
CognTKE	**46.06**	**36.49**	51.11	64.49	**35.24**	**25.21**	**39.93**	**54.71**	**53.13**	**42.62**	**59.42**	**72.70**	**83.21**	**80.01**	86.07	87.34

Table 1: The prediction results of MRR and Hits@1/3/10 on all datasets.

Baselines To demonstrate the effectiveness of CognTKE for TKG reasoning, CognTKE is compared with up-to-date TKG extrapolation methods including CyGNet (Zhu et al. 2021), TANGO (Han et al. 2021), TITer (Sun et al. 2021), RE-GCN (Li et al. 2021b), CEN (Li et al. 2022), TiRGN (Li, Sun, and Zhao 2022), TLogic (Liu et al. 2022b), CENET (Xu et al. 2023), RETIA (Liu et al. 2023), TECHS (Lin et al. 2023), and TempValid(Huang et al. 2024).

Evaluation Metrics We utilize two evaluation metrics: mean reciprocal rank (MRR) and Hits (k=1, 3, 10), which are commonly employed to assess the effectiveness of TKG reasoning methods. We follow works (Han et al. 2021; Sun et al. 2021) to report the experimental results with the time-aware filtered setting, which only filters out the quadruples occurring at the query time. Note that, all experimental results in this paper regarding MRR and Hits@1/3/10 are reported as percentages.

Overall Results

The overall experimental results of CognTKE and baselines on four benchmark datasets are displayed in Table 1. The best results are marked in bold, and the second-best results are reported using underlining.

It can be seen that CognTKE consistently outperforms the best path-based baseline TempValid on the ICE18 and ICE05-15 datasets. On ICE14 and WIKI datasets, CognTKE achieves better results compared to TempValid on MRR and Hits@1 metrics, and worse on Hits@3/10 metrics. Since TempValid's cyclic temporal rules can effectively model simple temporal rules on the ICE14 and WIKI datasets, it struggles to mine more complex temporal rules on the more challenging ICE18 and ICE05-15 datasets. This demonstrates that our CognTKE is capable of modeling complex temporal relation rules. Compared with other path-based methods TLogic, TlTer and TECHS, CognTKE achieves significant performance improvement. This is mainly because the path evidence obtained by the TLogic, TlTer and TECHS is restricted, while the TCR-Digraph of CognTKE can enjoy more temporal relation rule semantics.

The embedding method TiRGN outperforms other embedding methods because TiRGN integrates local and global historical information, but the interpretability of reasoning results is poor. By integrating the global one-hop temporal relations and the local multi-hop paths, CognTKE not only outperforms the performance of TiRGN, but also has strong interpretability.

Model	ICE14		ICE18		ICE05-15	
	MRR	Hits@3	MRR	Hits@3	MRR	Hits@3
CognTKE	**46.06**	**51.11**	**35.24**	**39.93**	**53.13**	**59.42**
-w/o System1	36.12	40.22	29.01	32.58	40.61	45.08
-w/o System2	44.88	50.08	33.45	38.01	44.49	50.16
-w/o Ti	45.44	50.53	34.01	38.47	51.49	57.85
-w/o QTR	44.72	49.57	34.55	39.12	49.82	56.03

Table 2: The results of ablation study on ICE14, ICE18 and ICE05-15 datasets.

Ablation Study

To further better understand each component of CognTKE that contributes to the prediction results, we conduct ablation studies on ICE14, ICE18, and ICE05-15 datasets. As reported in Table 2. *-w/o System1* denotes a variant of CognTKE that only considers the local deep reasoner; *-w/o System2* denotes a variant of CognTKE that only uses the global shallow reasoner; *-w/o Ti* denotes a variant of CognTKE that doesn't consider the TR-Component; *-w/o QTR* denotes a variant of CognTKE that doesn't use the QTR_GRU for message aggregation, but instead uses addition.

From the results in Table 2, it can be observed that *-w/o System1* and *-w/o System2* perform consistently worse than CognTKE on ICE14, ICE18, and ICE05-15 datasets. This demonstrates that the removal of either the global shallow reasoner or the local deep reasoner results in a decline in performance. Another interesting observation is that the performance of *-w/o System1* is inferior to *-w/o System2*. The reason for this phenomenon is that *-w/o System2* can directly answer the majority of query questions by the global shallow reasoner, since most query questions are associated with global one-hop historical facts that appeared in history. On the other hand, *-w/o System1* primarily focuses on local complex multi-hop reasoning, and many query problems are unrelated to local historical facts. So *-w/o System1* is difficult

Model	CognTKE			TempValid	CognTKE			TempValid	CognTKE			TempValid
Test	ICE14				ICE18				ICE05-15			
Train	ICE14	ICE18	ICE05-15	ICE14	ICE14	ICE18	ICE05-15	ICE18	ICE14	ICE18	ICE05-15	ICE05-15
MRR	46.06	46.01	**46.71**	45.78	33.07	**35.24**	34.21	33.50	49.99	49.86	**53.13**	50.31
Hits@1	36.49	36.31	**37.11**	35.50	23.3	**25.21**	24.27	23.91	39.86	39.52	**42.62**	39.46
Hits@3	51.11	51.41	**51.77**	51.34	37.38	**39.93**	38.75	37.89	56.03	55.96	**59.42**	56.71

Table 3: The zero-shot reasoning results on ICE14, ICE18 and ICE05-15 datasets.

to infer the query problems associated with longer historical facts. Hence, combining the global shallow reasoner and the local deep reasoner leads to more accurate prediction results.

The results of *-w/o Ti* denote that the TR-Component is useful, since the TR-Component can help CognTKE distinguish time intervals between temporal relations and queries. The performance of *-w/o QTR* denotes the QTR_GRU is beneficial for the CognTKE to aggregate and update messages in TCR-Digraph. The main reason is that QTR_GRU considers the order between adjacent temporal relations, and such an order will guide CognTKE to infer more accurate results.

Zero-Shot Reasoning

The proposed CognTKE is an inductive approach that allows for transferability to new datasets sharing the same relations as the training dataset, achieving unknown entity prediction through zero-shot reasoning. To evaluate the effectiveness of zero-shot reasoning in CognTKE, we conduct experiments on ICEWS series datasets: ICE14, ICE18, and ICE05-15 datasets. ICEWS series datasets share a majority of the same relations. The results are shown in Table 3, we also bold the best results. It can be seen that CognTKE still performs better than the best path-based baseline TempValid, when ICE14 and ICE18 are used as test sets. This demonstrates that CognTKE exhibits strong capabilities for handling new data migrations. When ICE05-15 is considered as the test set, the prediction results trained on ICE14 and ICE18 datasets are close to TempValid, but worse than that trained on ICE05-15 dataset. The main reason is that the time span of ICE14 and ICE18 datasets is smaller than the span of ICE05-15, which cannot provide enough historical facts to generate corresponding temporal relation rules.

Case Study

We visualize two cases of TCR-digraphs learned by CognTKE on ICE14 dataset. We eliminate edges in the TCR-Digraph whose attention weights are below 0.4. Two queries with the learned structure consisting of temporal relations are shown in Figure 4. For the first query, we can observe that CognTKE can directly obtain the answer *Citizen_(India)* from the global one-hop temporal relation, without further local multi-hop judgment. For the second query, although the quadruple *(Ashraf_Ghani_Ahmadzai, Express_intent_to_meet_or_negotiate, North_Atlantic_Treaty_Organization, 2014-11-30)* directly contains the answer entity, it is still necessary to traverse the identity quadruple to reach the answer entity. The reason for this is mainly attributed to the lower degree of correlation between the two

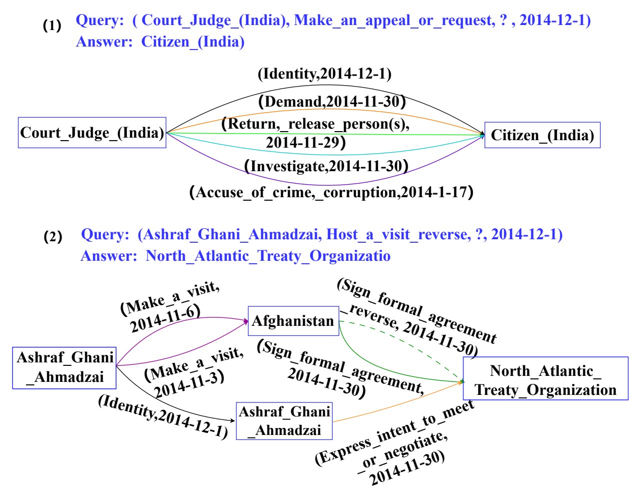

Figure 4: Visualization of the learned structures. Dashed lines mean inverse relations. entities are indicated by the blue rectangles.

relations *Host_a_visit_reverse* and *Express_intent_to_meet_or_negotiate*. The history relation *Make_a_visit* is semantically similar to the query relation, so it gets more weight. However, as the answer entity is not found in one-hop quadruples, exploring local two-hop quadruples becomes essential to infer a more accurate answer. Two cases in Figure 4 further demonstrate that CognTKE has a strong interpretive capability.

Conclusion

In this paper, we propose a novel cognitive temporal knowledge reasoning framework (CognTKE) according to dual process theory. In CognTKE, we introduce a novel TCR-Digraph that consists of significant local and global historical temporal relation paths associated with the query. A global shallow reasoner and a local deep reasoner are designed to perform inductive reasoning over the TCR-Digraph. TR-Component and TR-GAT in both reasoners are employed to distinguish the temporal relation paths in TCR-Digraph most relevant to the query. Extensive experiments on four datasets demonstrate that CognTKE outperforms state-of-the-art baselines and exhibits excellent zero-shot reasoning capabilities and strong interpretability.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62406022.

References

Baddeley, A. 1992. Working memory. *Science*, 255(5044): 556–559.

Battista, G. D.; Eades, P.; Tamassia, R.; and Tollis, I. G. 1999. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice-Hall. ISBN 0-13-301615-3.

Chen, K.; Wang, Y.; Li, Y.; and Li, A. 2022a. RotateQVS: Representing Temporal Information as Rotations in Quaternion Vector Space for Temporal Knowledge Graph Completion. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, 5843–5857. Association for Computational Linguistics.

Chen, W.; Wan, H.; Guo, S.; Huang, H.; Zheng, S.; Li, J.; Lin, S.; and Lin, Y. 2022b. Building and exploiting spatial-temporal knowledge graph for next POI recommendation. *Knowl. Based Syst.*, 258: 109951.

Chen, W.; Wan, H.; Wu, Y.; Zhao, S.; Cheng, J.; Li, Y.; and Lin, Y. 2024. Local-Global History-Aware Contrastive Learning for Temporal Knowledge Graph Reasoning. In *40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024*, 733–746. IEEE.

Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 1724–1734. ACL.

Evans; and BT, J. S. 1984. Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75(4): 451–468.

Evans, J. S. B. 2003. In Two Minds: Dual-process Accounts of Reasoning. *Trends in cognitive sciences*, 7(10): 454–459.

Evans, J. S. B. 2008. Dual-processing Accounts of Reasoning, Judgment, and Social Cognition. *Annu. Rev. Psychol.*, 59: 255–278.

Guan, S.; Cheng, X.; Bai, L.; Zhang, F.; Li, Z.; Zeng, Y.; Jin, X.; and Guo, J. 2023. What is Event Knowledge Graph: A Survey. *IEEE Trans. Knowl. Data Eng.*, 35(7): 7569–7589.

Han, Z.; Ding, Z.; Ma, Y.; Gu, Y.; and Tresp, V. 2021. Learning Neural Ordinary Equations for Forecasting Future Links on Temporal Knowledge Graphs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 8352–8364. Association for Computational Linguistics.

Huang, R.; Wei, W.; Qu, X.; Zhang, S.; Chen, D.; and Cheng, Y. 2024. Confidence is not Timeless: Modeling Temporal Validity for Rule-based Temporal Knowledge Graph Forecasting. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, 10783–10794. Association for Computational Linguistics.

Jin, W.; Qu, M.; Jin, X.; and Ren, X. 2020. Recurrent Event Network: Autoregressive Structure Inferenceover Temporal Knowledge Graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 6669–6683. Association for Computational Linguistics.

Lacroix, T.; Usunier, N.; and Obozinski, G. 2018. Canonical Tensor Decomposition for Knowledge Base Completion. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, 2869–2878. PMLR.

Leblay, J.; and Chekol, M. W. 2018. Deriving Validity Time in Knowledge Graph. In Champin, P.; Gandon, F.; Lalmas, M.; and Ipeirotis, P. G., eds., *The-Web-Conf*, 1771–1776. ACM.

Li, Y.; Sun, S.; and Zhao, J. 2022. Tirgn: time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *IJCAI*, 2152–2158.

Li, Z.; Guan, S.; Jin, X.; Peng, W.; Lyu, Y.; Zhu, Y.; Bai, L.; Li, W.; Guo, J.; and Cheng, X. 2022. Complex Evolutional Pattern Learning for Temporal Knowledge Graph Reasoning. In *ACL*, 290–296. Association for Computational Linguistics.

Li, Z.; Jin, X.; Guan, S.; Li, W.; Guo, J.; Wang, Y.; and Cheng, X. 2021a. Search from History and Reason for Future: Two-stage Reasoning on Temporal Knowledge Graphs. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, 4732–4743. Association for Computational Linguistics.

Li, Z.; Jin, X.; Li, W.; Guan, S.; Guo, J.; Shen, H.; Wang, Y.; and Cheng, X. 2021b. Temporal Knowledge Graph Reasoning Based on Evolutional Representation Learning. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 408–417. Association for Computing Machinery.

Liang, K.; Meng, L.; Liu, M.; Liu, Y.; Tu, W.; Wang, S.; Zhou, S.; Liu, X.; Sun, F.; and He, K. 2024a. A Survey of Knowledge Graph Reasoning on Graph Types: Static, Dynamic, and Multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liang, K.; Meng, L.; Zhou, S.; Tu, W.; Wang, S.; Liu, Y.; Liu, M.; Zhao, L.; Dong, X.; and Liu, X. 2024b. MINES: Message Intercommunication for Inductive Relation Reasoning over Neighbor-Enhanced Subgraphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 10645–10653.

Lin, Q.; Liu, J.; Mao, R.; Xu, F.; and Cambria, E. 2023. TECHS: Temporal logical graph networks for explainable extrapolation reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1281–1293.

Liu, K.; Zhao, F.; Chen, H.; Li, Y.; Xu, G.; and Jin, H. 2022a. DA-Net: Distributed Attention Network for Temporal Knowledge Graph Reasoning. In *Proceedings of the 31st*

ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022, 1289–1298. ACM.

Liu, K.; Zhao, F.; Xu, G.; Wang, X.; and Jin, H. 2023. RETIA: Relation-Entity Twin-Interact Aggregation for Temporal Knowledge Graph Extrapolation. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*, 1761–1774. IEEE.

Liu, Y.; Ma, Y.; Hildebrandt, M.; Joblin, M.; and Tresp, V. 2022b. TLogic: Temporal Logical Rules for Explainable Link Forecasting on Temporal Knowledge Graphs. In *Proceedings of the AAAI conference on artificial intelligence*, 4120–4127. AAAI Press.

Mei, X.; Yang, L.; Cai, X.; and Jiang, Z. 2022. An Adaptive Logical Rule Embedding Model for Inductive Reasoning over Temporal Knowledge Graphs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 7304–7316. Association for Computational Linguistics.

Pan, J.; Nayyeri, M.; Li, Y.; and Staab, S. 2024. HGE: Embedding Temporal Knowledge Graphs in a Product Space of Heterogeneous Geometric Subspaces. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, February 20-27, 2024, Vancouver, Canada*, 8913–8920. AAAI Press.

Sloman; and A, S. 1996. The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1): 3.

Sun, H.; Zhong, J.; Ma, Y.; Han, Z.; and He, K. 2021. TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 8306–8319. Association for Computational Linguistics.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *NeurIPS*, 5998–6008.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations*. OpenReview.net.

Wu, S.; Wan, H.; Chen, W.; Wu, Y.; Shen, J.; and Lin, Y. 2023. Towards Enhancing Relational Rules for Knowledge Graph Link Prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, 10082–10097. Association for Computational Linguistics.

Xin, H.; and Chen, L. 2024. KartGPS: Knowledge Base Update with Temporal Graph Pattern-based Semantic Rules. In *40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024*, 5075–5087. IEEE.

Xu, Y.; Ou, J.; Xu, H.; and Fu, L. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 4765–4773.

Zhang, Y.; and Yao, Q. 2022. Knowledge Graph Reasoning with Relational Digraph. In *The-Web-Conf*, 912–924. ACM.

Zhu, C.; Chen, M.; Fan, C.; Cheng, G.; and Zhang, Y. 2021. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks. In *Proceedings of the AAAI conference on artificial intelligence*, 4732–4740. AAAI Press.