RANGEFC: Interval-Aware Temporal Fact Checking for Knowledge Graph

Abdullah Qamar*, Umair Qudus*, Michael Röder, and Axel-Cyrille Ngonga Ngomo

Data Science Group (DICE), Heinz Nixdorf Institute, Department of Computer Science, Paderborn University
qamar@mail.uni-paderborn.de
{umair.qudus,michael.roeder,axel.ngonga}@uni-paderborn.de
https://dice-research.org

Abstract. Volatile assertions in knowledge graphs are often valid over intervals rather than a single point in time. Existing temporal fact-checking and validation methods predominantly treat temporal validity as a time-point prediction task, generating a single timestamp and failing to jointly model the start and end boundaries required for interval-based assertions. To address this gap, we propose RANGEFC, an interval-aware neural architecture designed to infer discrete validity intervals $[\hat{y}_s, \hat{y}_e]$ for knowledge graph assertions. RANGEFC leverages frozen pre-trained embeddings and employs a relation-aware Mixture-of-Experts (MoE) to adapt to diverse temporal regimes, such as short political tenures and long organizational lifespans. To ensure coherent predictions, we introduce a coupled interval prediction head and optimize using *overlap-aware objectives*, including a soft Intersection-over-Union (IoU) loss. We further construct and release a Wikidata-derived dataset of temporal assertions with year-range annotations to enable interval-based evaluation. Experimental results show that our model improves IoU by on average 3.7% over strong time-point baselines, particularly for relations with heterogeneous temporal patterns and long validity spans. These findings highlight the importance of interval modeling for temporal reasoning in knowledge graphs. Our implementation is publicly available at https://github.com/dice-group/RangeFC.

Keywords: Temporal Knowledge Graphs · Fact Checking · Interval Prediction · Mixture-of-Experts · Wikidata · Temporal Reasoning · Intersection-over-Union · Knowledge Graph Embeddings.

1 Introduction

Many real-world assertions are inherently time-dependent: they hold only within specific periods rather than at a single point in time [20]. For example, positions, affiliations, or organizational roles are valid only during well-defined intervals. However, knowledge graphs (KGs) such as Wikidata [15], which serve as the backbone for applications such as

* First two authors contributed equally to this work.

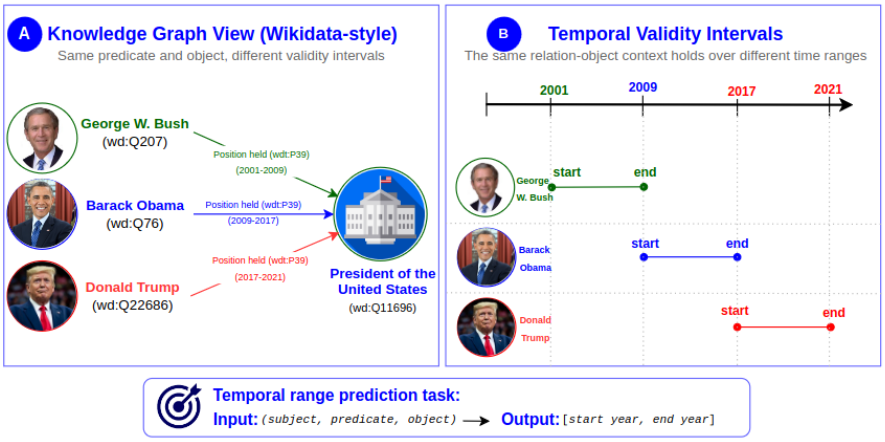

Fig. 1: Motivating example: The same predicate–object pair (e.g., *position held: President of the United States*) is associated with different validity intervals for different entities. Temporal validity is inherently interval-based rather than a single time point.

question answering and semantic search [17,21,1], often present such assertions without fully capturing their temporal extent. Ignoring temporal validity can therefore lead to misleading or incomplete conclusions [4,3,18].

Figure 1 illustrates a typical scenario in Wikidata: multiple entities share the same predicate–object pair but differ in their validity intervals. While such assertions are naturally represented as time spans, most existing approaches reduce temporal validity to a single timestamp [18,9,16]. This simplification ignores duration and overlap, and fails to capture the full semantics of temporal validity. We use this example throughout the paper. The assertion (`Barack Obama`, `position held`, `President of the United States`) is valid over the interval $[2009, 2017]$, not a single year. While time-point models predict one timestamp, interval prediction captures the full validity range and enables time-specific validation.

Recent advances in temporal fact checking have incorporated temporal signals from structured and unstructured sources [9,18]. In particular, embedding-based [16] and hybrid approaches such as TemporalFC [18] achieve strong performance by estimating the plausibility of assertions at specific time points. However, these approaches fundamentally formulate the problem as time-point prediction, either by regressing a single timestamp or ranking candidate years. As a result, they are inherently limited in modeling assertions whose validity spans extended and variable time intervals.

In practice, many assertions are associated with validity intervals, defined by a start and an end time. Modeling such intervals introduces several challenges. First, the start and end boundaries are interdependent, requiring coherent predictions that respect temporal order. Second, temporal patterns vary significantly across relations, ranging from short-lived events (e.g., political terms) to long-lasting facts (e.g., organizational lifespans). Third, standard evaluation metrics such as mean absolute error (MAE) do not

capture the semantic quality of predicted intervals, as they ignore the degree of overlap with the ground truth.

In this paper, we argue that temporal fact checking should be formulated as an interval prediction problem rather than a time-point estimation task. To this end, we propose RANGEFC, an interval-aware model that explicitly predicts validity intervals $[\hat{y}_s, \hat{y}_e]$ for a given assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$. Our approach builds on pre-trained embeddings and introduces a relation-aware Mixture-of-Experts (MoE) architecture to capture heterogeneous temporal patterns across relations. The model combines a factorized calendar head with a regression head to jointly capture global temporal structure and precise boundary localization. To ensure coherent predictions, we employ a coupled interval prediction head with consistency constraints that ensure $\hat{y}_e \geq \hat{y}_s$ during training and inference. Furthermore, we optimize the model using overlap-aware objectives, including a soft Intersection-over-Union (IoU) loss, to directly align training with interval quality.

To evaluate our approach, we construct and release a Wikidata-derived dataset of temporal assertions annotated with year ranges. We compare RANGEFC against strong time-point baselines and analyze performance across relations and interval characteristics. Our results show that explicitly modeling intervals leads to consistent improvements in overlap-based metrics, particularly for relations with diverse temporal regimes. This paper makes the following contributions:

- We formulate temporal fact checking for complete KG assertions as an interval prediction task rather than a time-point prediction task.
- We propose a relation-aware Mixture-of-Experts architecture for modeling heterogeneous temporal patterns across relations.
- We introduce overlap-aware training objectives for validity range prediction, including a soft IoU loss.
- We construct a Wikidata-derived benchmark with explicit start/end year annotations for evaluating interval-based temporal fact checking.
- We evaluate against TemporalFC-based time-point baselines and show improved interval quality, with IoU gains of up to 3.7%.

The remainder of this paper is structured as follows. Section 2 introduces the necessary preliminaries. Section 3 reviews related work. Section 4 presents the proposed interval-aware model. Section 5 describes dataset construction, Section 6 presents the experimental setup, and Section 7 reports and discusses the results. Finally, Section 8 concludes the paper and outlines directions for future work.

2 Preliminaries

We formalize temporal fact checking as an interval prediction problem over a knowledge graph (KG). We first introduce the underlying KG and then define the prediction task.

Definition 1 (KG). *Let $\mathcal{E}$ be a set of entities and $\mathcal{R}$ a set of relations. A KG is a set of assertions*

$$\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}, \tag{1}$$

where each assertion is represented as $(s, p, o) \in \mathcal{G}$ with subject $s \in \mathcal{E}$, predicate $p \in \mathcal{R}$, and object $o \in \mathcal{E}$.

Definition 2 (Temporal extension of KG). *We consider a discrete set of calendar years* $\mathcal{T} = \{y_{\min}, \ldots, y_{\max}\}$. *A temporal assertion is associated with a validity interval*

$$[y_s, y_e] \subseteq \mathcal{T}, \quad \textit{with } y_s \leq y_e, \tag{2}$$

indicating that the assertion (s, p, o) *holds for all* $t \in [y_s, y_e]$. *A temporal assertion extends an assertion with a validity interval. We denote such an assertion by*

$$x = (s, p, o, [y_s, y_e]),$$

where $s, o \in \mathcal{E}$, $p \in \mathcal{R}$, *and* $y_s, y_e \in \mathcal{T}$ *are the start and end years of the interval, respectively. The set of temporal assertions is therefore defined as*

$$\mathcal{G}_I \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathcal{T} \times \mathcal{T}.$$

In contrast to time-point temporal fact checking, where the goal is to predict a single timestamp $t \in \mathcal{T}$, our setting requires predicting two dependent temporal boundaries $(\hat{y}_s, \hat{y}_e)$.

Definition 3 (Interval prediction). *Given an assertion* (s, p, o), *the task of interval prediction is to predict a validity interval* $[\hat{y}_s, \hat{y}_e]$ *to form a correct temporal fact* $(s, p, o, \hat{y}_s, \hat{y}_e)$, *where* $\hat{y}_s$ *and* $\hat{y}_e$ *denote the beginning and end of the validity period, respectively.*

Unlike time-point prediction, evaluating interval predictions requires measuring the overlap between predicted and ground-truth intervals.

Definition 4 (Interval Quality). *Let the predicted interval be* $\hat{I} = [\hat{y}_s, \hat{y}_e]$ *and the ground truth be* $I = [y_s, y_e]$. *We define the temporal Intersection-over-Union (IoU) as:*

$$\mathrm{IoU}(\hat{I}, I) = \frac{\max\big(0, \min(\hat{y}_e, y_e) - \max(\hat{y}_s, y_s) + 1\big)}{\max(\hat{y}_e, y_e) - \min(\hat{y}_s, y_s) + 1}. \tag{3}$$

IoU measures the degree of temporal overlap between predicted and ground-truth intervals, and ranges from 0 (no overlap) to 1 (perfect alignment).

3 Related Work

Temporal KGs extend KGs by associating assertions with temporal information, enabling reasoning over time-dependent assertions. Temporal KG Embedding (TKGE) models incorporate temporal signals into entity and relation representations for tasks such as link prediction and forecasting [6,13,8,2,23,14,12,22,5,16,7,24]. Early approaches such as HyTE [6], T-TRANSE [13], and TA-TRANSE [8] extend translation-based KGEs with temporal representations, while TDistMult [23], and TA-DistMult [14] adapt tensor factorization to temporal settings. RE-NET [12] models historical interactions via recurrent networks, while TeRo [22] and T-COMPLEX [5] improve temporal relational reasoning. Despite this progress, most TKGE approaches target relation prediction at specific timestamps rather than the full temporal validity interval of an assertion [18].

Temporal fact checking assesses whether an assertion holds at a given point in time. Existing approaches combine TKGEs with neural architectures and external evidence to

estimate assertion plausibility conditioned on time [16,9,18]. TemporalFC [18] predicts the most plausible timestamp for an assertion, while DYHE [16] uses dihedron algebra in a 4D hyper-complex embedding space for the same purpose. In all cases, temporal fact checking is formulated as time-point prediction via timestamp regression or candidate ranking, without explicitly modeling validity intervals or optimizing for temporal overlap.

Interval-aware modeling has been studied within temporal KG completion. TEMT [11] uses pre-trained language models for text-enhanced completion, including time-interval prediction, while TeRo [22] models interval assertions via separate rotation-based representations for start and end times; recent surveys [24] summarize such interval-aware embedding models. These works target KG completion or representation learning with ranking-oriented objectives, whereas we address temporal fact checking for a *complete* assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$, directly predicting its validity boundaries $[\hat{y}_s, \hat{y}_e]$ via coupled prediction heads and overlap-aware training. MoE architectures are well established for modeling heterogeneous data through conditional routing [19], but remain unexplored for temporal fact checking; relation-aware expert specialization for diverse temporal regimes has, to our knowledge, not been studied previously. In summary, unlike prior temporal fact-checking work, we cast factual validity as an interval prediction problem and optimize directly for interval overlap.

4 Methodology

We propose an interval-aware model for temporal fact checking that predicts validity intervals $[\hat{y}_s, \hat{y}_e]$ for a given assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$.

Given an assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$, the goal is to predict its validity interval:

$$f_\theta : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \to \mathcal{T} \times \mathcal{T}, \tag{4}$$

$$f_\theta(s, p, o) = (\hat{y}_s, \hat{y}_e), \quad \text{such that } \hat{y}_e \geq \hat{y}_s, \tag{5}$$

where θ denotes the model parameters. In our running example of Figure 1, the input is an assertion such as $(\texttt{Barack Obama}, \texttt{position held}, \texttt{President of the United States})$, and the output is the interval $[2009, 2017]$. We use this assertion as a running example throughout this section: a correct model should place $\hat{y}_s$ near 2009 and $\hat{y}_e$ near 2017, while enforcing $\hat{y}_e \geq \hat{y}_s$.

This formulation differs fundamentally from time-point prediction, which estimates a single timestamp $t \in \mathcal{T}$. Interval prediction requires: 1. joint boundary estimation with consistency constraints to ensure $\hat{y}_e \geq \hat{y}_s$, and 2. overlap-aware evaluation, as boundary-wise errors do not capture interval quality. To address these challenges, we design an interval-aware architecture composed of four key components: input representation, a shared encoder, relation-aware specialization, and coupled interval prediction.

Figure 2 illustrates the overall architecture of the proposed model. Given an assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$, pre-trained entity and relation embeddings are first concatenated and passed through a shared MLP encoder to obtain a latent representation. This representation is then processed by a relation-aware MoE module, where a gating network conditioned on both the input and the relation embedding selects a subset of experts. The outputs of the selected experts are combined via a weighted aggregation to produce an expert-mixed representation.

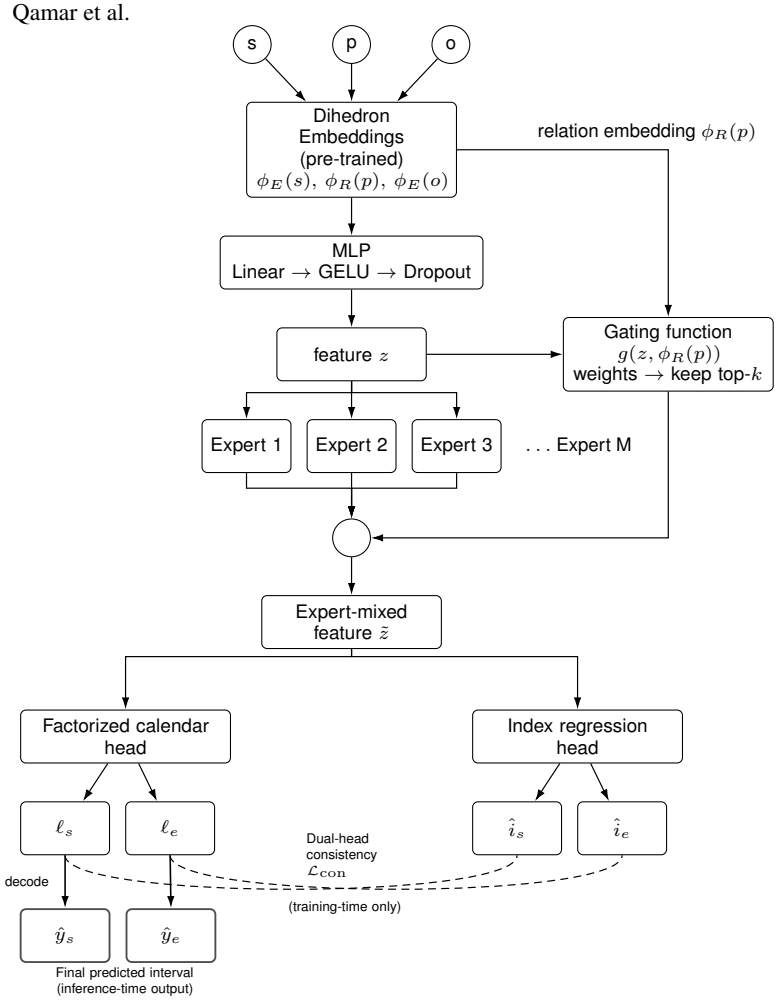

Fig. 2: RangeFC architecture: an assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$ is encoded, refined via relation-aware MoE, and passed to a calendar head (logits ℓ_s, ℓ_e) and a regression head (indices $\hat{i}_s, \hat{i}_e$), coupled via consistency loss $\mathcal{L}_{\text{con}}$ (dashed). Only the calendar head is decoded into the final interval $[\hat{y}_s, \hat{y}_e]$.

On top of this representation, two complementary prediction heads are applied: 1. a factorized calendar head that produces distributions over candidate start and end years, and 2. a regression head that predicts continuous boundary indices. Both heads are trained jointly and connected via a consistency constraint that ensures their predictions to agree, as detailed in Section 4.4.

4.1 Input Representation

As illustrated in Figure 2, we begin by constructing a feature representation for each assertion using pre-trained embeddings for entities and relations:

$$\phi_E : \mathcal{E} \to \mathbb{R}^{d_e}, \quad \phi_R : \mathcal{R} \to \mathbb{R}^{d_r}, \tag{6}$$

where d_e and d_r are the embedding dimensionalities for entities and relations, respectively. For an assertion $(\mathtt{s}, \mathtt{p}, \mathtt{o})$, the input feature vector is constructed as:

$$z_0 = \mathrm{concat}\big(\phi_E(s), \phi_R(p), \phi_E(o)\big). \tag{7}$$

For our running example, z_0 is the concatenation of the pre-trained embedding of *Barack Obama*, the embedding of the relation *position held*, and the embedding of *President of the United States*. These embeddings are kept *frozen* during training to isolate the effect of the interval prediction head.

4.2 Shared Encoder

The input vector z_0 is processed by a multi-layer perceptron (MLP) to obtain a shared representation:

$$z = \mathrm{MLP}(z_0), \tag{8}$$

where the MLP projects the input vector $z_0 \in \mathbb{R}^{2d_e+d_r}$ to a shared hidden space $\mathbb{R}^d$ (i.e., the hidden dimension of the MLP).

4.3 Relation-aware Mixture-of-Experts

Temporal patterns vary significantly across relations. To capture this heterogeneity, we employ a MoE module conditioned on the relation.

Let $M \in \mathbb{N}$ denote the number of experts, and let $\{E_m\}_{m=1}^{M}$ be a set of expert networks, where each $E_m : \mathbb{R}^d \to \mathbb{R}^d$ maps an input representation to a transformed feature space. Let $z \in \mathbb{R}^d$ denote the encoded representation of an assertion, and recall $\phi_R(p) \in \mathbb{R}^{d_r}$ is the embedding of relation p (Section 4.1). A gating function $g : \mathbb{R}^{d+d_r} \to \mathbb{R}^M$, implemented as a two-layer MLP, takes the concatenated vector $\mathrm{concat}(z, \phi_R(p))$ as input and produces a score vector over the M experts; we write this as $g(z, \phi_R(p))$ for brevity in the equations below. The scores are normalized using a softmax function:

$$\alpha = \mathrm{softmax}\big(g(z, \phi_R(p))\big), \quad \alpha \in \Delta^{M-1}, \tag{9}$$

where Δ^{M-1} denotes the $(M-1)$-dimensional probability simplex, ensuring that $\sum_{m=1}^{M} \alpha_m = 1$ and $\alpha_m \geq 0$.

To encourage expert specialization, we apply top-k routing. Let $\mathcal{K}_k$ denote the set of indices corresponding to the k largest gate weights. We retain only these weights and set all remaining expert weights to zero. The selected weights are then renormalized:

$$\tilde{\alpha}_m = \begin{cases} \dfrac{\alpha_m}{\sum_{j \in \mathcal{K}_k} \alpha_j}, & \text{if } m \in \mathcal{K}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The final representation is computed as a weighted combination of the selected expert outputs:

$$\tilde{z} = \sum_{m \in \mathcal{K}_k} \tilde{\alpha}_m \, E_m(z). \tag{11}$$

The routing mechanism shown in Figure 2 selects the top-k experts for each assertion while still using relation and input information to guide the selection. As a result, the model can allocate capacity dynamically and learn specialized temporal patterns for different relations, such as short political tenures and longer organizational lifespans.

For our running example, the gating function conditions on both z and $\phi_R(\textit{position held})$ to select experts suited to short, sharply bounded intervals typical of political tenures (e.g., 2001–2009, 2009–2017, 2017–2021 for the U.S. presidency in Figure 1), as opposed to the longer, more loosely bounded intervals typical of organizational lifespans (relation `P571`).

4.4 Coupled Interval Prediction Head

To predict coherent validity intervals, the proposed model uses two complementary prediction heads on top of the expert-mixed representation $\tilde{z}$. The first head is an index regression head, which predicts continuous start and end indices in the discrete year-index space. The second head is a factorized calendar head, which produces start and end logits over the complete set of candidate years.

Let $E_T \in \mathbb{R}^{|\mathcal{T}| \times d_t}$ denote a trainable calendar embedding table, where each row represents one year in the discrete timeline and d_t is the dimensionality of the year embeddings. The factorized calendar head projects the representation $\tilde{z} \in \mathbb{R}^d$ into two query vectors using learned projections $W_s, W_e \in \mathbb{R}^{d_t \times d}$:

$$q_s = W_s \tilde{z}, \tag{12}$$

$$q_e = W_e \tilde{z}, \tag{13}$$

so that $q_s, q_e \in \mathbb{R}^{d_t}$ share the same dimensionality as the rows of the calendar embedding table E_T. The start and end logits are then computed by comparing these query vectors with the shared calendar table:

$$\ell_s = q_s E_T^\top, \tag{14}$$

$$\ell_e = q_e E_T^\top. \tag{15}$$

This formulation allows the model to score all candidate start and end years while sharing temporal structure through the same calendar representation.

In parallel, the index regression head directly predicts continuous boundary indices:

$$\hat{i}_s = (|\mathcal{T}| - 1)\, \sigma\big(f_s(\tilde{z})\big), \tag{16}$$

$$\hat{i}_e = \hat{i}_s + \big(|\mathcal{T}| - 1 - \hat{i}_s\big)\, \sigma\big(f_\Delta(\tilde{z})\big), \tag{17}$$

where $\sigma(\cdot)$ denotes the sigmoid function, mapping both terms into $[0, 1]$ so that $\hat{i}_s \in [0, |\mathcal{T}| - 1]$ and the offset term in Eq. (17) is non-negative by construction, guaranteeing $\hat{i}_e \geq \hat{i}_s$. Both heads are trained jointly: the calendar head captures global temporal structure over the entire timeline, while the regression head provides precise boundary localization in continuous index space. Applying $\mathrm{softargmax}(\cdot)$ to the calendar logits ℓ_s, ℓ_e yields a continuous index in the same space as $\hat{i}_s, \hat{i}_e$; to encourage agreement

between the two heads, we include a consistency term (Eq. (21)) between these two continuous index estimates.

For our running example, the calendar head should assign most of its probability mass to 2009 for the start year and 2017 for the end year, while the regression head should predict continuous indices $\hat{i}_s, \hat{i}_e$ corresponding to these same two years; a disagreement between the two (e.g., the regression head predicting 2010 while the calendar head favors 2009) is penalized by $\mathcal{L}_{\text{con}}$, introduced below in Equation 21.

The boundaries are predicted jointly from a shared representation and coupled through consistency and ordering constraints. At inference time, the final validity interval is decoded from the *calendar head* alone: we take $\hat{y}_s = \arg\max(\ell_s)$ and $\hat{y}_e = \arg\max(\ell_e)$, directly yielding calendar years without an intermediate index-to-year mapping, since the calendar table E_T is already indexed by year. The regression head is not used to produce the final decoded interval; instead, it supplies complementary training-time supervision (via $\mathcal{L}_{\text{reg}}$ and $\mathcal{L}_{\text{con}}$, introduced in Section 4.5) that shapes the shared representation $\tilde{z}$ and regularizes the calendar head toward precise, well-localized boundary predictions.

Unlike the regression head (Equation (17)), which guarantees $\hat{i}_e \geq \hat{i}_s$, the calendar head's independent logits may violate ordering. We apply a decoding correction $\hat{y}_e \leftarrow \max(\hat{y}_s, \hat{y}_e)$ to enforce coherence. During training, $\mathcal{L}_{\text{ord}}$ (Eq. (22)) penalizes such violations, while $\mathcal{L}_{\text{con}}$ indirectly reinforces ordering by coupling the calendar head with the regression head. This primarily prevents rare crossing cases (e.g., predicting an end year before the start year in our running example).

4.5 Learning Objectives

The model is trained with a combination of boundary-level, distributional, consistency, and overlap-aware objectives. This combination ensures both accurate boundary estimation and high-quality interval overlap.

Boundary-level Regression Loss. Boundary-level supervision is applied in the discrete year-index space. Let i_s and i_e denote the gold start and end indices, and let $\hat{i}_s$ and $\hat{i}_e$ denote the corresponding regression outputs. We use a robust Huber loss [10], which behaves as a squared-error (L_2) loss for small residuals and as an absolute-error (L_1) loss for large residuals, transitioning at a threshold δ:

$$\text{Huber}(x, y) = \begin{cases} \frac{1}{2}(x-y)^2, & |x-y| \leq \delta, \\ \delta\left(|x-y| - \frac{1}{2}\delta\right), & \text{otherwise}, \end{cases} \tag{18}$$

This makes training less sensitive to outlier-length intervals (e.g., organizational lifespans) that would otherwise dominate the gradient. Formally,

$$\mathcal{L}_{\text{reg}} = \text{Huber}(\hat{i}_s, i_s) + \omega_e \, \text{Huber}(\hat{i}_e, i_e), \tag{19}$$

where ω_e controls the relative weight of the end-boundary error. This weighting is useful because end years are often more variable and harder to predict than start years. For our running example, i_s and i_e correspond to the year indices for 2009 and 2017, respectively.

Calendar Classification Loss. In addition to the regression head, the factorized calendar head produces logits over all candidate years for the start and end boundaries. To avoid treating neighboring years as completely unrelated classes, we use Gaussian-smoothed targets around the gold year indices, i.e., $\tilde{y}_t = \mathcal{N}(t \mid y, \tau^2)$, normalized over the timeline, with y being the gold year index and τ controlling the smoothing bandwidth. The corresponding classification loss is:

$$\mathcal{L}_{\text{ce}} = \text{CE}(\ell_s, \tilde{y}_s) + \text{CE}(\ell_e, \tilde{y}_e), \tag{20}$$

where ℓ_s and ℓ_e are the start and end logits, and $\tilde{y}_s$ and $\tilde{y}_e$ are Gaussian-smoothed target distributions. For our running example, this means that, rather than treating 2017 as the only correct label for Obama's end year, adjacent years (e.g., 2016 or 2018) receive partial target probability, while distant years (e.g., 1995) do not.

Consistency and Ordering Losses. To align the two prediction heads, we include a consistency loss between the soft-argmax of the calendar logits and the continuous outputs of the regression head:

$$\mathcal{L}_{\text{con}} = \left\| \text{softargmax}(\ell_s) - \hat{i}_s \right\|_1 + \left\| \text{softargmax}(\ell_e) - \hat{i}_e \right\|_1 . \tag{21}$$

While the regression head inherently enforces $\hat{i}_e \geq \hat{i}_s$ via its formulation, the calendar head's logits do not guarantee this ordering. To encourage the calendar head to produce coherent start/end distributions, we penalize cases where the soft-argmax of the calendar logits violates temporal order:

$$\mathcal{L}_{\text{ord}} = \max\big(0,\, \text{softargmax}(\ell_s) - \text{softargmax}(\ell_e)\big). \tag{22}$$

This loss aligns the calendar head with the regression head's inherent ordering and provides additional supervision during early training, when the regression head may not yet have stabilized.

Overlap-aware Loss. Boundary errors alone do not fully capture interval quality. To provide differentiable overlap supervision, we compute a soft Intersection-over-Union loss directly on the index intervals $[\hat{i}_s, \hat{i}_e]$ and $[i_s, i_e]$, using the same discrete-inclusive convention as the evaluation metric in Definition 4, but evaluated on the continuous, pre-rounding indices so that gradients can flow to the regression head:

$$\mathcal{L}_{\text{IoU}} = 1 - \text{IoU}\big([\hat{i}_s, \hat{i}_e],\, [i_s, i_e]\big), \tag{23}$$

where

$$\text{IoU}([a,b],[c,d]) = \frac{\max\big(0,\ \min(b,d) - \max(a,c) + 1\big)}{\max\big(0,\ \max(b,d) - \min(a,c) + 1\big) + \epsilon}, \tag{24}$$

with ϵ added to the denominator for numerical stability. This matches the functional form of Eq. (4) up to the stabilizing constant ϵ, so $\mathcal{L}_{\text{IoU}}$ directly optimizes a differentiable relaxation of the evaluation metric rather than a separate quantity, and the two coincide (up to ϵ) once $\hat{i}_s, \hat{i}_e$ are rounded to integer year indices at decoding time. For our running example, predicting $\hat{I} = [2010, 2016]$ against the gold interval $I = [2009, 2017]$ yields $\text{IoU} = \frac{2016-2010+1}{2017-2009+1} = 7/9 \approx 0.78$, providing a graded overlap signal that complements the boundary-wise losses.

Final Objective. The final training objective combines all components:

$$\mathcal{L} = \mathcal{L}_{\text{reg}} + \lambda_{\text{ce}}\mathcal{L}_{\text{ce}} + \lambda_{\text{IoU}}\mathcal{L}_{\text{IoU}} + \lambda_{\text{ord}}\mathcal{L}_{\text{ord}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}}. \tag{25}$$

The regression loss provides stable boundary-level supervision, the calendar loss encourages year-aware classification, the consistency term aligns the two heads, and the IoU term promotes semantically meaningful interval overlap. Together, these objectives allow the model to optimize both boundary accuracy and overall interval quality.

5 RANGEFC Dataset

Existing temporal resources are not directly aligned with the interval fact-checking setting studied here. TemporalFC [18] focuses on single timestamp prediction and therefore does not provide paired start/end boundaries. Temporal KG completion benchmarks such as YAGO11k and Wikidata12k [6], used in prior interval-aware TKGE work [11,22], contain temporal intervals but are primarily designed for entity or relation ranking with metrics such as MRR and Hits@k rather than assertion-level interval prediction. Constraint-based temporal validation datasets [20] serve a different purpose, focusing on validation through discovered temporal constraints and including many literal-valued cases. In contrast, our benchmark provides explicit (s, p, o, y_s, y_e) instances, relation-aware splits, and evaluation using interval-overlap metrics such as IoU.

5.1 Dataset Construction

To evaluate interval-based temporal fact checking, we construct a dataset from Wikidata consisting of assertions annotated with validity intervals. We focus on relations that provide explicit temporal qualifiers (e.g., start time and end time), enabling the extraction of year ranges. For each relation $p \in \mathcal{R}$, we identify corresponding temporal properties (e.g., `P580` for start time and `P582` for end time). We extract instances of the form $(s, p, o, [y_s, y_e])$, where y_s and y_e are parsed from the associated qualifiers.

All timestamps are mapped to calendar years, yielding a discrete timeline $\mathcal{T}$. Intervals are represented as closed year ranges. We exclude assertions for which either the start or end boundary is missing, since the task requires complete validity intervals for supervised training and evaluation. We apply consistency checks to ensure valid intervals: 1. removal of instances with missing or malformed timestamps, 2. enforcement of $y_s \leq y_e$, and 3. deduplication of identical assertions. Each resulting instance is represented as a quintuple:

$$(s, p, o, y_s, y_e). \tag{26}$$

We create train, validation, and test splits using a *per-relation stratified strategy*, ensuring that each relation is represented across all splits while preventing leakage of identical intervals.

5.2 Dataset Characteristics

The resulting dataset contains $134{,}037$ assertions spanning five relation types with interval semantics. The data is split into $107{,}229$ training, $13{,}403$ validation, and $13{,}405$ test instances.

The temporal range covers mainly years from 1428 to 2027, reflecting both historical and contemporary assertions. The dataset exhibits a highly imbalanced distribution across relations: organizational assertions dominate with over $100{,}000$ instances, while other relations such as office terms and military equipment contain a few thousand instances each.

The dataset captures diverse temporal regimes, including short-lived intervals (e.g., political tenures), medium-duration intervals (e.g., military service), and long-term spans (e.g., organizational lifecycles and biographical lifetimes). This variability makes the dataset well-suited for evaluating models that must adapt to heterogeneous temporal patterns. This construction yields a large-scale, real-world dataset that reflects the inherent temporal variability and imbalance present in Wikidata.

6 Experimental Setup

We use Dihedron-based pre-trained temporal embeddings for entities and relations [16]. The entity and relation embeddings are kept frozen during training, while only the interval prediction components are optimized for the task. The shared encoder is implemented as a multi-layer perceptron (MLP) with a hidden dimension of 1024 and GELU activation. The relation-aware MoE module consists of $M = 5$ experts with top-3 routing, where only the top-k experts (based on gating scores) contribute to the final representation. The factorized calendar head uses a trainable embedding dimension of 128. Models are trained using the objective defined in Section 4, combining boundary-level regression, calendar classification, consistency, ordering, and overlap-aware losses. We use the Adam optimizer with a learning rate of 1.9×10^{-3} and a batch size of 1024. Training runs for up to 120 epochs with early stopping based on validation IoU. The loss weights in Equation 25 are set to $\lambda_{\text{ce}} = 0.1$, $\lambda_{\text{IoU}} = 0.2$, $\lambda_{\text{ord}} = 0.07$, and $\lambda_{\text{con}} = 0.5$. The Huber loss threshold is $\beta = 0.74$, the end-boundary weight is $\omega_e = 1.0$ (no additional weighting), and the Gaussian smoothing bandwidth is $\tau = 1.35$. We use a dropout rate of 0.12, a fixed gating temperature of 1.43, and an auxiliary load-balancing weight of 0.02 for the MoE. To regularize the frozen pre-trained embeddings, we add Gaussian noise $\mathcal{N}(0, 0.01^2)$ during training.

We evaluate models on the test set and report average performance across all relations. In addition, we perform: 1. per-relation analysis, 2. evaluation across interval length buckets, and 3. ablation studies to assess the contribution of model components.

6.1 Baselines

To the best of our knowledge, most existing temporal fact-checking approaches focus on predicting a single timestamp rather than a validity interval, and dedicated interval-prediction baselines for this setting remain scarce. We therefore adapt strong time-point methods for comparison.

Table 1: Overall performance comparison. Higher IoU is better; lower MAE is better.

Model	IoU ↑	MAE_s ↓	MAE_e ↓
Time-point baseline	0.611	15.11	34.66
TemporalFC	0.624	**12.07**	28.74
RANGEFC	**0.661**	18.94	**21.37**

- **Time-point MLP:** a neural model with the same input representation as our approach, trained to predict a single timestamp $\hat{t}$.
- **TemporalFC model:** a state-of-the-art hybrid approach combining pre-trained embeddings with a neural predictor for time-point estimation [18].

To enable comparison with interval prediction, we extend these models to produce interval estimates. Specifically, we predict start and end years independently by applying the time-point predictor twice, once for $\hat{y}_s$ and once for $\hat{y}_e$.

This adaptation provides a controlled baseline, allowing us to evaluate the benefits of explicitly modeling temporal intervals over independent time-point estimation.

6.2 Evaluation Metrics

We evaluate interval quality using Intersection-over-Union (IoU), which measures the overlap between predicted and ground-truth intervals, defined in Equation 3. We also report mean absolute error (MAE) for start and end boundaries:

$$\text{MAE}_s = \mathbb{E}[|\hat{y}_s - y_s|], \quad \text{MAE}_e = \mathbb{E}[|\hat{y}_e - y_e|]. \tag{27}$$

Returning to the running example, predicting $[2010, 2016]$ for the Obama presidency would not be an exact match, but it would still have high temporal overlap with the gold interval $[2009, 2017]$. This is why we report IoU in addition to boundary-wise MAE.

7 Results

We evaluate the proposed interval-aware model against strong time-point baselines to assess its effectiveness in predicting temporal validity spans. Our analysis focuses on overall performance, relation-specific behavior, and the impact of interval characteristics.

7.1 Overall Performance

Table 1 reports the results on the test set. Our model outperforms time-point baselines in terms of Intersection-over-Union (IoU), highlighting the benefit of explicitly modeling validity intervals.

The results in Table 1 reveal a clear trade-off between boundary accuracy and interval quality. Our model achieves the highest IoU (0.661), outperforming the TemporalFC baseline (0.624) by $+3.7$ percentage points and the time-point MLP by $+5.0$ points. On this benchmark, this result suggests that explicitly modeling temporal intervals leads to substantially better overlap with ground-truth validity spans.

Table 2: Seed robustness of RANGEFC on the test split. Higher IoU is better; lower MAE is better.

Seed	**IoU** ↑	**MAE**$_s$ ↓	**MAE**$_e$ ↓
42 (reported)	0.661	18.94	21.37
0	0.652	19.39	21.17
1	0.653	19.12	21.16
7	0.655	18.85	20.73
123	0.655	18.98	21.00
Mean ± Std	**0.655 ± 0.003**	**19.06 ± 0.21**	**21.09 ± 0.24**

Table 3: Per-relation performance on the test split (lower MAE is better; higher IoU is better).

Relations	#Assertions	TemporalFC			RANGEFC		
		MAE$_s$	MAE$_e$	IoU	MAE$_s$	MAE$_e$	IoU
P571/P576 (Organizations)	10,635	**14.79**	31.26	0.595	22.89	**23.06**	**0.635**
P569/P570 (Biographical)	1,759	**1.42**	17.67	0.797	2.72	**13.82**	**0.822**
P1619/P3999 (Buildings)	549	**0.44**	22.08	0.738	2.85	**18.80**	**0.749**
P580/P582 (Political roles)	254	**4.51**	20.58	**0.417**	9.82	**12.86**	0.406
P729/P730 (Military)	208	**3.12**	21.10	0.618	7.93	**16.19**	**0.673**

Interestingly, TemporalFC achieves the lowest start boundary error (MAE$_s$ = 12.07), indicating that time-point models can localize individual timestamps accurately. However, this does not translate into high-quality interval predictions, as reflected by its lower IoU. In contrast, our model exhibits a higher MAE$_s$ (18.94), but significantly improves the end boundary prediction (MAE$_e$ = 21.37 vs. 28.74), suggesting that it better captures the temporal extent of assertions.

This behavior highlights a key limitation of time-point approaches: optimizing for point-wise accuracy does not ensure coherent or well-aligned interval predictions. By jointly modeling start and end boundaries with dual prediction heads and overlap-aware objectives, our approach prioritizes global interval quality over local boundary precision, leading to superior performance under IoU. Table 1 reports results for seed 42; Table 2 confirms stability across five seeds (mean IoU = 0.655 ± 0.003).

7.2 Per-Relation Analysis

To better understand where the improvements originate, Table 3 reports performance across relation types.

Overall, our model achieves higher IoU on four out of five relations, with particularly strong gains for relations characterized by long or variable-duration intervals. For instance, on organizational assertions (`P571`), our model improves IoU from 0.595 to 0.635 (+4.0 points), and on military-related relations (`P729`), from 0.618 to 0.673 (+5.5 points). Similar improvements are observed for biographical (`P569`) and building-related (`P1619`) assertions, indicating consistent benefits of interval-aware modeling.

A key observation is that our model significantly reduces the end boundary error (MAE$_e$) across all relations. For example, on `P571`, MAE$_e$ decreases from 31.26 to 23.06, and on `P729` from 21.10 to 16.19. This suggests that the proposed coupled interval prediction mechanism is particularly effective at capturing the temporal extent of

Table 4: Performance across interval-length buckets (in years). Higher IoU is better; lower MAE is better. TempFC denotes TemporalFC, Ours denotes RANGEFC, and # Assertions denotes total number of Assertions.

		[0–1]	[2–5]	[6–10]	[11–25]	[26–50]	[51–100]	[101+]
Ours	IoU	0.36	**0.49**	**0.51**	**0.59**	**0.72**	**0.83**	0.65
	MAE_s	13.34	12.76	12.11	8.90	7.88	4.63	134.07
	MAE_e	**19.71**	**17.37**	**18.56**	**17.94**	**18.09**	**14.73**	**68.89**
TempFC	IoU	0.36	0.46	0.47	0.54	0.65	0.80	**0.68**
	MAE_s	**8.27**	**6.80**	**7.95**	**3.71**	**4.18**	**2.64**	**94.04**
	MAE_e	27.19	25.03	27.59	25.85	26.90	20.20	76.90
	# Assertions	817	1480	1192	2321	2327	4138	1130

assertions. As in Section 7.1, TemporalFC's lower MAEs does not translate into higher IoU.

An exception is the political roles relation (P580), where TemporalFC slightly outperforms our model in IoU (0.417 vs. 0.406). This relation is characterized by relatively short and less variable intervals. As a result, the benefit of explicit interval modeling is less pronounced.

Overall, these results show that interval-aware modeling is most beneficial for relations with diverse or long temporal spans, where independent time-point estimation fails to capture the full validity interval.

7.3 Effect of Interval Length and Overlap

To further analyze model behavior, we evaluate performance across different interval lengths. Table 4 reports results grouped by ground-truth interval duration.

For very short intervals ([0–1] years), both models achieve identical IoU of 0.36. However, as interval duration increases, the advantage of interval-aware modeling becomes more apparent. Our model outperforms TemporalFC across most medium and long-duration buckets. For example, in the [26–50] and [51–100] year ranges, our approach achieves IoU scores of 0.72 and 0.83, compared to 0.65 and 0.80 for TemporalFC.

A notable observation is that our model achieves lower end-boundary error (MAE_e), indicating improved estimation of interval duration. In contrast, TemporalFC generally achieves lower start-boundary error (MAE_s), reflecting its strength in predicting individual timestamps. Nevertheless, lower boundary error does not necessarily translate into better interval quality, as evidenced by the lower IoU scores in most duration buckets. For extremely long intervals (101+ years), TemporalFC slightly outperforms our model in IoU (0.68 vs. 0.65). This may be due to the higher sparsity and variance in this bucket, where predicting a central timestamp can still yield reasonable overlap.

Overall, the results highlight a key limitation of time-point approaches: approximating long validity spans with a single timestamp often leads to suboptimal overlap. By explicitly modeling interval boundaries and optimizing overlap-aware objectives, the proposed approach produces more accurate and semantically meaningful temporal

Table 5: Ablation study of RANGEFC, covering both the relation-aware MoE component and the interval-specific design choices.

Model variant	MAE_s	MAE_e	IoU	(ΔIoU)
Full model	**18.94**	**21.37**	**0.661**	–
No MoE	22.67	23.19	0.587	-0.074
Top-1 routing	22.22	22.73	0.597	-0.064
w/o consistency loss	19.19	21.87	0.628	-0.033
w/o soft-IoU objective	20.03	21.26	0.644	-0.017
w/o auxiliary ordering loss	19.91	21.59	0.646	-0.015
Regression-only	20.74	21.50	0.612	-0.049
Calendar-only	22.00	30.95	0.347	-0.314

predictions. These improvements are particularly relevant for downstream applications such as temporal question answering, and fact checking over temporal KGs.

7.4 Ablation Study

Table 5 reports component-level ablations of RANGEFC, covering both the MoE module and the interval-specific design choices. Removing MoE or restricting routing to top-1 costs 6–7 IoU points, confirming the benefit of relation-aware specialization. The dual-head design is similarly critical: regression-only and calendar-only variants drop IoU to 0.612 and 0.347, respectively, and removing the consistency loss lowers it to 0.628. The soft-IoU and ordering losses contribute smaller gains (0.015–0.017), as interval order is already enforced architecturally at decoding.

8 Conclusion

In this paper, we introduced an interval-aware formulation of temporal fact checking that models assertions' validity as a time range rather than a single timestamp. To address the challenges of interval prediction, we proposed a relation-aware MoE architecture with coupled interval prediction heads and overlap-aware learning objectives. Experimental results on a Wikidata-derived benchmark show that explicitly modeling validity intervals improves overlap quality, particularly for relations with heterogeneous temporal patterns and long-duration validity spans. As future work, we plan to explore uncertain and open-ended temporal intervals, incorporate textual evidence, and investigate downstream applications such as temporal question answering and temporal KG completion.

Acknowledgments

This work has been supported by the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW) within the project SAIL under the grant no NW21-059D, by the German Federal Ministry of Research, Technology and Space (BMFTR) within the project KI-Akademie OWL under the grant no 16IS24057B, and within the project "WHALE" (LFN 1-04) funded under the Lamarr Fellow Network programme by the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW).

Supplemental Material Statement

The source code of RANGEFC, the scripts to recreate the full experimental setup, dataset, and the required libraries can be found on our GitHub page.[1]

Declaration of Use of Generative AI

During the preparation of this manuscript, the authors used DeepSeek open source LLM solely for language polishing, structural refinement, and consistency checking. All content was carefully reviewed and edited by the authors. The authors verified all technical content, references, tables, and experimental results, and take full responsibility for the final manuscript.

References

1. Athreya, R.G., Ngonga Ngomo, A.C., Usbeck, R.: Enhancing community interactions with data-driven chatbots–the dbpedia chatbot. In: Companion Proceedings of World Wide Web. p. 143–146. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018), https://doi.org/10.1145/3184558.3186964
2. Balazevic, I., Allen, C., Hospedales, T.: TuckER: Tensor factorization for knowledge graph completion. In: EMNLP-IJCNLP. pp. 5185–5194. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1522, https://aclanthology.org/D19-1522
3. Budhdeo, S., Zhang, J., Abdulle, Y., Agapow, P.M., McKechnie, D., Archer, M., Shah, V., Forte, E., Noori, A., Zitnik, M., Ashrafian, H., Sharma, N.: Scoping review of knowledge graph applications in biomedical and healthcare sciences. Wellcome Open Research **10**, 66 (2025). https://doi.org/10.12688/wellcomeopenres.23599.1, https://doi.org/10.12688/wellcomeopenres.23599.1, [version 1; peer review: 2 approved with reservations]
4. Cai, L., Mao, X., Zhou, Y., Long, Z., Wu, C., Lan, M.: A survey on temporal knowledge graph: Representation learning and applications. ArXiv **abs/2403.04782** (2024), https://api.semanticscholar.org/CorpusID:268296889
5. Chekol, M.W.: Tensor decomposition for link prediction in temporal knowledge graphs. In: Proceedings of the 11th on Knowledge Capture Conference. p. 253–256. ACM, New York, NY, USA (2021). https://doi.org/10.1145/3460210.3493558, https://doi.org/10.1145/3460210.3493558
6. Dasgupta, S.S., Ray, S.N., Talukdar, P.: HyTE: Hyperplane-based temporally aware knowledge graph embedding. In: EMNLP. pp. 2001–2011. Association for Computational Linguistics, Brussels, Belgium (2018). https://doi.org/10.18653/v1/D18-1225, https://aclanthology.org/D18-1225
7. Demir, C., Moussallem, D., Heindorf, S., Ngomo, A.C.N.: Convolutional hypercomplex embeddings for link prediction. In: Asian Conference on Machine Learning. pp. 656–671. PMLR (2021)
8. García-Durán, A., Dumančić, S., Niepert, M.: Learning sequence encoders for temporal knowledge graph completion. In: EMNLP. pp. 4816–4821. Association for Computational Linguistics, Brussels, Belgium (2018). https://doi.org/10.18653/v1/D18-1516, https://aclanthology.org/D18-1516

[1] Source code: https://github.com/dice-group/RangeFC

9. Gerber, D., Esteves, D., Lehmann, J., Bühmann, L., Usbeck, R., Ngonga Ngomo, A.C., Speck, R.: Defacto-temporal and multilingual deep fact validation. Web Semant. **35**(P2), 85–101 (Dec 2015). https://doi.org/10.1016/j.websem.2015.08.001, https://doi.org/10.1016/j.websem.2015.08.001
10. Huber, P.J.: Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics **35**(1), 73 – 101 (1964). https://doi.org/10.1214/aoms/1177703732, https://doi.org/10.1214/aoms/1177703732
11. Islakoglu, D.S., Chekol, M., Velegrakis, Y.: Leveraging pre-trained language models for time interval prediction in text-enhanced temporal knowledge graphs (2023), https://arxiv.org/abs/2309.16357
12. Jin, W., Zhang, C., Szekely, P.A., Ren, X.: Recurrent event network for reasoning over temporal knowledge graphs. CoRR **abs/1904.05530** (2019), http://arxiv.org/abs/1904.05530
13. Leblay, J., Chekol, M.W.: Deriving validity time in knowledge graph. In: Companion Proceedings of the The Web Conference 2018. p. 1771–1776. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018). https://doi.org/10.1145/3184558.3191639, https://doi.org/10.1145/3184558.3191639
14. Ma, Y., Tresp, V., Daxberger, E.A.: Embedding models for episodic knowledge graphs. Journal of Web Semantics **59**, 100490 (2019). https://doi.org/10.1016/j.websem.2018.12.008, https://www.sciencedirect.com/science/article/pii/S1570826818300702
15. Malyshev, S., Krötzsch, M., González, L., Gonsior, J., Bielefeldt, A.: Getting the most out of wikidata: Semantic technology usage in wikipedia's knowledge graph. In: International Semantic Web Conference. pp. 376–394. Springer International Publishing, Cham (2018)
16. 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: ESWC. pp. 253–269. Springer International Publishing, Cham (2022)
17. Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web **8**(3), 489–508 (2017)
18. Qudus, U., Röder, M., Kirrane, S., Ngomo, A.C.N.: Temporalfc: A temporal fact checking approach over knowledge graphs. In: The Semantic Web – ISWC 2023: 22nd International Semantic Web Conference, Athens, Greece, November 6–10, 2023, Proceedings, Part I. p. 465–483. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-47240-4_25, https://doi.org/10.1007/978-3-031-47240-4_25
19. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q.V., Hinton, G.E., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net (2017), https://openreview.net/forum?id=B1ckMDqlg
20. Soulard, T., Saïs, F., Raad, J.: Explainable temporal fact validation through constraints discovery in knowledge graphs. In: Curry, E., Acosta, M., Poveda-Villalón, M., van Erp, M., Ojo, A., Hose, K., Shimizu, C., Lisena, P. (eds.) The Semantic Web. pp. 227–244. Springer Nature Switzerland, Cham (2025)
21. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (Sep 2014). https://doi.org/10.1145/2629489, https://doi.org/10.1145/2629489
22. Xu, C., Nayyeri, M., Alkhoury, F., Shariat Yazdi, H., Lehmann, J.: TeRo: A time-aware knowledge graph embedding via temporal rotation. In: CICLing. pp. 1583–1593. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020). https://doi.org/10.18653/v1/2020.coling-main.139, https://aclanthology.org/2020.coling-main.139

23. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: ICLR (2015), http://arxiv.org/abs/1412.6575
24. Zhang, Y., Kong, X., Shen, Z., Li, J., Yi, Q., Shen, G., Dong, B.: A survey on temporal knowledge graph embedding: Models and applications. Knowledge-Based Systems **304**, 112454 (2024). https://doi.org/10.1016/j.knosys.2024.112454, https://www.sciencedirect.com/science/article/pii/S0950705124010888