HTSE: Hierarchical Time-Surface Model for Temporal Knowledge Graph Embedding

Langjunqing Jin
Huazhong University of Science and Technology
Feng Zhao (zhaof@hust.edu.cn)
Huazhong University of Science and Technology
Hai Jin
Huazhong University of Science and Technology

Research Article

Keywords: knowledge graph embedding, semantic hierarchy, time surface, temporal prediction

Posted Date: October 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2178549/v1

HTSE: Hierarchical Time-Surface Model for Temporal Knowledge Graph Embedding

Langjunqing Jin, Feng Zhao and Hai Jin

National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, Hubei, China.

*Corresponding author(s). E-mail(s): zhaof@hust.edu.cn; Contributing authors: langjunqingj@hust.edu.cn; hjin@hust.edu.cn;

Abstract

Representation learning based on temporal knowledge graphs (TKGs) has attracted widespread interest, and TKG embeddings express time entity and relation tokens and exhibit strong dynamics. Despite the significance of the dynamics and the persistent updates in TKGs, most studies have been devoted to static knowledge graphs. Moreover, previous temporal works ignored the semantic hierarchies observed in knowledge modelling cases, which are common in real-world applications. Inaccurate semantic expressions caused by incomplete projections might not capture complex topological structures very well. To solve this problem, a novel *hierarchical time-surface embedding* (HTSE) model is proposed for the representation learning of entities, relations and time. Specifically, a unified relation-oriented hierarchical space aims to distinguish relations at different semantic levels of a hierarchy, and entities can naturally reflect the corresponding hierarchy. Then, a time surface aims to enhance the temporal characteristics, and quadruples are learned through exponential mapping and tangent planes in the time surface. According to extensive experiments, *HTSE* can achieve remarkable performance on five benchmark datasets, outperforming baseline models for time scope prediction, temporal link prediction and hierarchical relation embedding tasks.Furthermore, the qualitative analysis is used to demonstrate the explainable strategy for hierarchical embeddings and their significance in TKGs.

Keywords: knowledge graph embedding, semantic hierarchy, time surface, temporal prediction

1 Introduction

Because obtaining rich semantic information from multirelational graph-structured data has become a challenge in artificial intelligence research, multiclass knowledge graphs (KGs) have been proposed and developed for various applications. These include large KGs, such as YAGO, Wikidata, Freebase, WordNet, and DBpedia[19], and they have been utilized in numerous mainstream applications, such as automatic question-and-answer systems, information retrieval systems, and recommendation systems. Traditional static KGs are known to take the form of triples (s,p,o), where s represents a subject, p represents a predicate and o is an object. For example, the triple (LeBron James, plays for, Cleveland) was true at a time node corresponding to 2018. However, some facts (entities or relationships) in a knowledge base may change over time, which can lead to conflicts between new facts and previous facts, thus making the representations of the corresponding KGs inaccurate. For example, the triple (LeBron James, plays for, Lakers) replaced the previous triple in 2022.

To accommodate such dynamics, temporal knowledge graphs (TKGs) have been proposed. In particular, for the commonly used Global Database of Events, Language, and Tone (GDELT)[13] and the International Crisis Early Warning System (ICEWS) database[15], TKGs can reflect remarkably important differences by incorporating time information into existing graph data in the corresponding static KGs. This approach can more accurately reflect the dynamics of facts and the timeliness of real KGs. However, the incompleteness has yet to be addressed in existing TKGs. Thus, related embedding tasks, such as temporal link prediction, have remained challenging.

TKG embedding is a representation learning method in which time information, entities and relations are expressed in a time-specific space in vector form to achieve the transformation of high-dimensional data into low-dimensional vectors. Several existing mainstream TKG embedding models, such as HyTE[6], TTransE[12] and ChronoR[19], ignore the algebraic representations of curvature vectors and inherent topological information. This results in the inaccurate projection of vectors in scenarios with rich interactions between temporal properties and multirelational features in TKGs. However, a surface model is suitable for this scenario because it can solve this issue by embedding high-dimensional curvature vectors. Therefore, it is necessary to explore a time surface model for TKG embedding. This is one of the major motivations for this work.

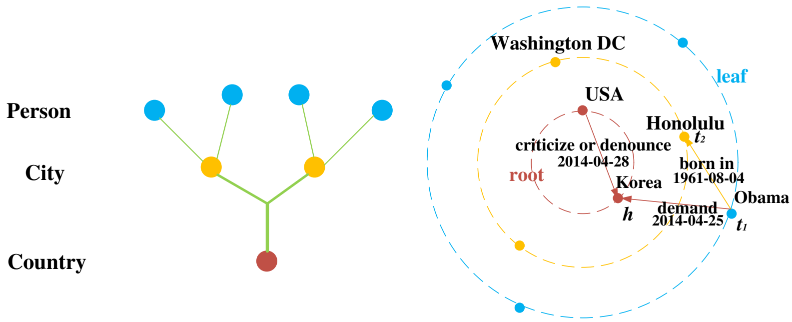


Fig. 1 Details of hierarchy modelling and projection on a time surface

Another major motivation is the ignorance of modelling semantic hierarchy in existing models. A semantic hierarchy is an indispensable TKG property. For instance, the quadruple (*Rio de Janeiro*, *capital of*, *Brazil*, 1764-1956) is true before relocation of the capital to Brasilia, where Brazil is at a higher level than *Rio de Janeiro*. In Fig. 1, the quadruple (Obama, born in, Honolulu, 1961-08-04) presents a person born in a city where the city is at a higher level than the person. More specifically, the ICEWS dataset contains quadruples (Obama, demand, Korea, [2014-04-25]) and (USA, criticize or denounce, Korea, [2014-04-28]), where a person 'Obama' is at a lower level than country 'USA' and 'Korea' in the hierarchy. Although some works have focused on the issue of static semantic hierarchs, they have been limited to static embedding and ignore temporal scenarios. This can lead to an inability to model complex relations and incomplete semantic expressions. Therefore, it is still challenging to research a strategy to represent the semantic hierarchy accurately and effectively in TKGs. Then, it is crucial to explore a time-surfaced model that can process entities and relations in each TKG hierarchy.

To overcome the above issues and fill the research gaps, this work proposes a novel ***h****ierachical* ***t****ime-****s****urface* ***e****mbedding* (HTSE) model. First, the given entities and relations are embedded in a relation-oriented hierarchical space and can clearly reflect the different hierarchies of importance levels among the multihierarchical entities connected by each relation. Then, the resulting quadruples are projected onto a time surface and used to represent the rich interactions via exponential mapping, which can improve the semantic expression ability of *HTSE*. Our experiments demonstrate the superiority of the prediction process and the validity of *HTSE* hierarchy modelling. We summarize the significant contributions of this work below.

- To address the issue of ignoring hierarchical knowledge when modelling, a relation-oriented hierarchical modelling strategy is proposed to capture semantic hierarchies more completely with regard to the entities and relations in TKGs. First, three separate imaginary components α, β, and γ are used in the embeddings of the entities and relations within the hierarchical space. Then, relations at different semantic levels of the hierarchy can be

distinguished, and entities can naturally reflect the corresponding hierarchy. The processed entities and relations are taken as the quadruple bases following the introduction of the time attribute.
- A *HTSE* model is proposed to address the shortcoming of semantic expression and to capture the deep topology information resulting from uneven quadruple projections. Different from existing geometry-based methods, in the proposed approach, the semantic hierarchy space is transformed into a time surface, and various elements of a quadruple can be expressed on this time surface first. Next, the time surface is divided into several local manifolds by timestamps, and the quadruples are accurately and more intuitively embedded via an exponential mapping approach. In other words, the surfaced embedding model can improve the accuracy of representation learning with respect to semantic hierarchies and time dynamic characteristics.
- Extensive experiments are conducted on temporal link prediction and time information prediction. The experimental results show that the *HTSE* model can improve the accuracy of embeddings and the capacity for knowledge completion over that of the baseline models. The advantages of the proposed hierarchical knowledge modelling strategy can be illustrated in an analysis of hierarchical relation embeddings. Furthermore, the explainable strategy for link prediction on different hierarchies is shown in the qualitative analysis.

The remainder of this work is as follows. The necessary related work is introduced in Section 2. The main components and training process of the *HTSE* model are discussed in detail in Section 3. Extensive experiments are analysed and presented in Section 4. The paper concludes and provides an outlook for future development in Section 5.

2 Related Work

According to previous research, related embedding models can be classified into temporally unaware embedding learning models and temporal-aware learning models.

Temporal-unaware embedding models: A number of mature static KG embedding methods have recently become available. To address the issues of semantic loss and promote the accuracy of complex relation embeddings, various embedding space models have been proposed, such as the TransE[2], TransH[23], and TransR[14] models. These translation-based methods follow the principle of the closest distance between entity and relation vectors and have achieved success. To solve the issue of algebraic ill-posedness and the insufficient adaptivity of geometric forms, different types of geometric-based models have been proposed with excellent performance achieved through link prediction tasks. ManifoldE[24] reviews knowledge representations in a manifold space in regard to topology. Topology-aware associations are also effectively exploited between relations in TACT[4]. Furthermore, aiming to solve issues such as symmetry, antisymmetry and inversional relations using

linear function, RotatE[20] and LineaRE[18] proposed rotation/liner ideas to embed the relations of entity pairs into the corresponding space. Static hierarchical embedding models, such as HAKE[26], [3], HittER[5] and HBE[16], make use of different operations of hyperbolic reflection to the multiple hierarchical relation patterns of the model to achieve better results.
Temporal-aware embedding models: Recent studies on TKG embedding models, including extended models and entity dynamics models, have sought to enhance the performance of temporal prediction[10]. To make them more influential and extensible, these methods are actually extensions of previous static KG models. [12] and [1] focus on introducing relational embedding variables extended to quadruples to upgrade a static model to a model with a time attribute, e.g., TTransE, TA-DistMult and DE-SimplE[7]. RE-NET[11] models a sequence of events through a recurrent neural network event encoder and an adjacent aggregator. HyTE[6] is based on a hyperplane representation of the time space and expands the integration of the temporal information and element representations into TKGs. DyERNIE[8] introduces dynamic evolution in the form of a Riemannian manifold to capture the dynamic characteristics of TKGs using velocity vectors. TIMEPLEX[9] is an improved variant of ComplEx[22] that automatically utilizes the recursive properties of relations and temporal interactions. Know-Evolve[21] and EvoKG[17] use a deep evolution network structure based on a knowledge system for temporal reasoning. ChronoR[19] captures the rich interaction between temporal knowledge graphs and multirelation features with a high-dimensional rotation as its transformation operator.

Our proposed HTSE model is a type of dynamic geometric model. In particular, HTSE shares similarities with DyERNIE, studying dynamic relation embeddings based on Riemannian manifolds. However, there are two major differences:

- **The aims are different**.
  DyERNIE aims to learn multirelational data through dynamic relation embeddings. Although it can capture the geometric features in a KG, it ignores the explainable strategy for hierarchical embeddings. Instead, the *HTSE* aims to model hierarchical space and provide explicit evidence for temporal link prediction associated with different hierarchical levels when modelling hierarchical relations.
- **The methods to model temporal information are different**.
  DyERNIE aims to model temporal relation embeddings with Riemannian manifolds, which let the entity representations evolve based on a velocity vector defined at each timestamp. The different methods are limited to ignoring the weights of hierarchical structures. Instead, the *HTSE* utilizes surfaced projections to address the existing drawbacks by assigning separate weights to different layers on the local manifold, which significantly outperforms DyERNIE.

Another successful existing model is not considered to represent semantic hierarchy and is limited to nonsurfaced TKG projection. Therefore, a novel model *HTSE* is proposed and analysed comprehensively in this paper.

3 The Proposed HTSE

In this paper, a novel *HTSE* model is proposed and analysed. We emphasize the process of semantic hierarchy modelling and the derivation of a time surface embedding space in this section. *HTSE* contains four parts: a **relation-specific and hierarchy-aware space**, a **time surface and exponential mapping** module, a **score function** and a **training** process. Distinct gradations are present among these four parts.

3.1 Relation-Specific and Hierarchy-Aware Space

A TKG possesses dynamic characteristics among its entities and relations over time. *HTSE* transfers entities and relations to a relation-oriented space by integrating the embeddings of entities and relations to a hierarchy-aware space.

Given a graph containing sets of entities E and relations R, we represent the algebraic representation learning of entity pairs as $h, t \in E$. Then, we assume that relation vectors are $R_i = [r_{i1}, r_{i2}, \ldots, r_{im}] \in R$ at the i^{th} hierarchy level. As illustrated in Fig. 2, entities and relations are represented by different coloured circles and directed arrows, respectively, on an equipotential surface expressed by dashed lines. The three axes form the corresponding semantic space. In hierarchical graph G_r, the green/blue points at the same level represent the head/tail entity vectors h_i and t_i, respectively, and the corresponding circles represent the embedding modulus of these head/tail entities. Yellow dots and purple dots denote entities belonging to different hierarchies. The red arrows indicate the hierarchical relations R_i. This component completes the triple modelling and supports the transformation to quadruples with time stamps.

Specifically, given a subgragh $G_r = (h, r, t)$ in hierarchy, *HTSE* represents the embeddings of entities within the hierarchical space. The embeddings h' and t' of the head and tail entities are represented as:

$$\begin{aligned} h' &= h_r + \alpha \boldsymbol{h}_{\alpha,h} + \beta \boldsymbol{h}_{\beta,h} + \gamma \boldsymbol{h}_{\gamma,h} \\ t' &= t_r + \alpha \boldsymbol{t}_{\alpha,t} + \beta \boldsymbol{t}_{\beta,t} + \gamma \boldsymbol{t}_{\gamma,t} \end{aligned} \tag{1}$$

, where h' and t' form a new hierarchical entity pair, and the hierarchical relations $r \in R$ link these entities to each other, h_r and $t_r \in \mathbb{R}^n$. Three imaginary components α, β and γ represent the three hierarchical space vector decomposition units, and $\alpha\beta\gamma = \alpha^2 = \beta^2 = \gamma^2 = -1$, $\boldsymbol{h}_{\alpha,h}$, $\boldsymbol{h}_{\beta,h}$ and $\boldsymbol{h}_{\gamma,h} \in \mathbb{R}^n$ are the corresponding entity representations.

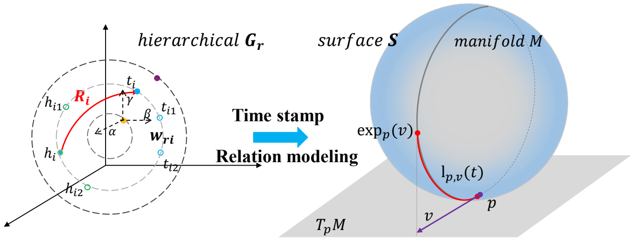


Fig. 2 Details of hierarchy modelling and projection on a time surface

3.2 Time Surface and Exponential Mapping

Structured knowledge should be valid only within a specific time range, and not considering this temporal information can result in dynamic deletions within fact expressions. Therefore, it is worth proposing a new time surface model for entities, relation reasoning, and time period prediction. This section introduces a time surface-aware model that is proposed as an alternative to existing methods. Specifically, the space processed as discussed in Section 3.1 is defined as G_r, and a time axis passing through G_r is transformed to incorporate time information. The transformational time surface is set as a combination of several local manifolds M and tangent planes T_pM, representing the correlations between entities and relations on the surface (as depicted in Fig. 2). Moreover, $\exp_p(v)$ represents the connection in the time surface between two time properties of quadruples; it is an exponential mapping from the starting to ending time stamp, and the temporal relation modelling is represented by the red real line. The red line $l_{p,v}$ represents the temporal relation after projection on the time surface.

Definition: If $\exp_p : T_pM \to M \Leftrightarrow \exp_p(v) = l_{p,v}(1)$ exists, $\exp_p$ is an **exponential mapping**.

Since the geodesic $l_{p,v}$ is defined locally, $\exp_p(v)$ can only be defined for an open subset on T_pM, for example, by setting $B_p(\delta) = \{v \mid v \in T_pM, \|v\| < \delta\}$, where δ is a constant associated with point p. The exponential mapping is: $\forall$ vector $v \in B_p(\delta)$, $\exists l_{p,v}(\tau)$ is the only geodesic, whose parameter value range is $[-1, 1]$; hence, $\exp_p(v) = l_{p,v}(1)$ holds. The geometric significance of the exponential mapping means that $\exp_p(v)$ is the point starting from p with the initial tangent vector v of length $\|v\|$. From the properties of geodesic $l_{p,v}(\tau)$, we know that $\left\|\frac{d(l_{p,v}(\tau))}{d\tau}\right\|$ is an invariable constant. Then, we can obtain that

$\frac{d(l_{p,v}(\tau))}{d\tau}|_{\tau=0} = v$ from the curvature, that is, the initial tangent vector v, $\left\|\frac{d(l_{p,v}(\tau))}{d\tau}\right\| \equiv \|v\|$. At the same time, $\exp_p(v) = l_{p,v}(1)$, $p = l_{p,v}(0)$, is known for deriving the arc length from point p to $\exp_p(v)$ as follows:

$$\int_0^1 \left\|\frac{d\left(l_{p,v}(\tau)\right)}{d\tau}\right\| d\tau = \int_0^1 \|v\| d\tau = \|v\| \tag{2}$$

. Therefore, the arc length at point p along the geodesic to $\exp_p(v)$ is exactly equal to $\|v\|$, as depicted by the red curve in Fig. 2. Finally, $\exp_p(\tau) = \frac{d\exp_p(\tau)}{d\tau}$ is obtained. This proves the feasibility of the exponential mapping.

By analysing Fig. 2 and combining it with the above derivation procedure, the temporal relation in the time interval $[\tau_i, \tau_i + \tau]$ can be defined as:

$$r_\tau = \exp_p\left(v_0\tau\right) + r_0 \tag{3}$$

where τ_i is the starting time stamp in the local manifold and $r_0 \in M$ is the initial relation vector indicating the initial embeddings without time fluctuations. $v_0 \in T_pM$ represents the temporal relation-specific curvature vector, which is defined in a tangent space that captures the dynamic features of the relations and entities over a time transformation. Graphic relations are based on derivations from the initial embeddings in combination with a tangent vector to easily represent the stationary semantics of relations/entities. Then, special attention is given to additional temporal information. The final three-component element is represented as (h_r, r_τ, t_r), and a time component can be added to form a new quadruple $(h_r, r_\tau, t_r, [\tau_1 : \tau_2])$.

3.3 Score Function

To develop an adaptive measurement to support the score function, we adopt the standardized Euclidean distance. The standardized Euclidean distance measurement method has been demonstrated to be the most accurate method for calculating the distance between vectors. This method can overcome the inaccuracy of traditional methods caused by the uneven distribution of data in each dimension, which can lead to low link prediction accuracy. For this purpose, each component is first standardized to an equal mean variance to balance the dimensional components as follows. We set the standard deviation of the weighted vector for entities/relations to S_k, and the score function of *HTSE(control)* is

$$f(h, r, t, \tau) = \left\|\sum_{k=1}^{n} \frac{h_r + r_\tau - t_r}{S_k}\right\|_2 - \mu \left\|r_\tau - r\right\|_2 \tag{4}$$

To model these relations better, we incorporate the idea of error control to extend *HTSE*. $\|r_\tau - r\|_2$ aims to ensure that the time-specific relation vector

r_τ is the best nearest neighbour for calculating the distance from the original vector r, and μ can control the constraint as a hyperparameter.

3.4 Training

We explore the set of negative samples. A negative sampling method for temporal perception is proposed to emphasize temporal information that considers all quadruples in a TKG. Similar to previous works, the ranking loss based on a marginal style is minimized as follows:

$$L = \sum_{\tau \in \mathrm{T}} \sum_{x \in Q_\tau^+} \sum_{y \in Q_\tau^-} \max\left(0, f_\tau(x) - f_\tau(y) + \gamma\right) \quad -\mu \left\| r_\tau - r \right\|_2 \tag{5}$$

where Q_τ^+ are valid quadruples, and Q_τ^- are negative samples, both of which contain time information.

$$\{Q_\tau^- = \{(h', r, t', \tau) \mid \{h' \in E\} \cup \{t' \in E\} \ \ , (h', r, t', \tau) \notin Q_\tau^+\} \tag{6}$$

where γ is defined as the margin between Q_τ^+ and Q_τ^-, $\mu \left\| r_\tau - r \right\|_2$ controls the extension error, the relation-specific weight vector satisfies $\left\| w_r \right\|_2 = 1$, $\left\| h_r \right\|_2, \left\| t_r \right\|_2, \left\| r_\tau \right\|_2 \leq 1$, and $\mu \in [0, 1)$, $\forall \tau \in T$.

Since the model possesses a geometric structure, we make use of Riemannian stochastic gradient descent (RSGD)[8], in which the Riemannian gradient $\Delta_R L$ is obtained by normalizing the Δ_E Euclidean gradient with respect to the inverse of the metric tensor of the time surface.

The characteristic *HTSE* steps are shown in Algorithm 1. The triples of the quadruples in the datasets are introduced as inputs in $\Delta = (h, r, t)$. Then, we process the relations and entities separately by calculating entropy values e_j to model the unified relation-specific and entity-weighted spaces. In addition, it is necessary to adopt exponential mapping on the time surface T by projecting time-aware space and providing accurate calculations. Finally, the projected vectors and time information on the surface can be obtained. It is obvious that the weighted process and the emphasis on time attributes are incisively and vividly reflected by displaying (1) and (3) in the whole process.

4 Experiment

Extensive experiments are performed on benchmark datasets, including YAGO11k, Wikidata12k[25], ICEWS14, ICEWS05-15[1] and GDELT[8]. The experimental results are compared with those of several representative models. Our goals are as follows.

- Evaluating our model and comparing it with static and dynamic models in terms of temporal link prediction.
- Illustrating the advantage of *HTSE* based on surfaced projection with respect to temporal scope prediction.

Algorithm 1 *HTSE Time Surface Estimation*

Input: Entity pairs h and t, relations r, transformed vectors $v_{h,r,\cdot}$ and $v_{t,r,\cdot}$;
Three hierarchical space vector decomposition units α, β and γ;
Initial relation r_0, initial curvature vector v_0
local manifold M, time surface T
segment K: $K_1, K_2, \ldots, K_m$, tangent plane T_pM

Output: Projected entity and relation vectors on the time surface: h', t' and r_τ

for $i = 1 : k$ **do**
  for $j = 1 : m$ **do**
    $r \sim U(N(j))$ and $r \sim U\left(K_{\mathrm{i}} \backslash \{m\}\right)$
    $\boldsymbol{v}_{h,r,\cdot} = \boldsymbol{v}_h \otimes (\alpha \boldsymbol{h}_{\alpha,h} + \beta \boldsymbol{h}_{\beta,h} + \gamma \boldsymbol{h}_{\gamma,h})$
    $\boldsymbol{v}_{t,r,\cdot} = \boldsymbol{v}_t \otimes (\alpha \boldsymbol{t}_{\alpha,t} + \beta \boldsymbol{t}_{\beta,t} + \gamma \boldsymbol{t}_{\gamma,t})$
  end for
  $r_\tau = \exp_p(v_0 t) + r_0$
end for
$h' = h_r + \alpha \boldsymbol{h}_{\alpha,h} + \beta \boldsymbol{h}_{\beta,h} + \gamma \boldsymbol{h}_{\gamma,h}$
$t' = t_r + \alpha \boldsymbol{t}_{\alpha,t} + \beta \boldsymbol{t}_{\beta,t} + \gamma \boldsymbol{t}_{\gamma,t}$
$r_\tau = \exp_p(v_0 \tau) + r_0$

Table 1 The sizes of the categories contained in the five benchmark datasets

Dataset	#Entities	#Relations	#Train	#Validation	#Test
YAGO11k	10622	10	16408	2051	2050
Wikidata12k	12554	24	32497	4062	4062
ICEWS14	7128	230	72826	8941	8943
ICEWS05-15	10488	251	368962	46275	46092
GDELT	7691	240	1734399	238765	305241

- Analysing the differences between the results of the state-of-the-art (SOTA) models and *HTSE* regarding hierarchical relation embedding.
- Presenting queries based on fact datasets to demonstrate the strategy validity of our model for hierarchical embeddings.

4.1 Fundamental Setup

The abovementioned datasets contain facts associated with time annotations. The dataset statistics are summarized in Table 1. The hyperparameters and optimization procedure are presented in the experimental implementation details.

Baselines and Evaluation Protocol To provide an overall presentation of the superiority of HTSE, we select several excellent representative learning models with static and temporal properties as our baselines. Specifically, we use TransE[2] and RotatE[20] as representative static models. As representative TKG embedding models, we choose the corresponding baselines, i.e., HyTE[6], DyERNIE[8], TIMEPLEX[9] and ChronoR[19]. We adopt the MRR and Hits@n (n=1,3,10) standard metric models to evaluate the link prediction performance. For each quadruple $q = (h, r_\tau, t, \tau)$ in the test set Q_τ, $|Q_\tau|$ is the size of Q_τ, MRR $= \frac{1}{|Q_\tau|}\sum_{\mathrm{i}=1}^{|Q_\tau|}\frac{1}{\mathrm{rank}i}$, where rank i denotes the ranking of the first correct answer in the i^{th} Q_τ, and Hits@n is defined in [2].

Furthermore, similar to the operations executed under the 'raw' and 'filtered' settings in TransE[2], inspired by [9], we report a filtered version of Hit@3. Specifically, we replace the head/tail entities with other entities when testing quadruples during evaluation. The resulting corrupted quadruples may be correct. The 'raw' and 'filtered' indices are treatments for the test set, where the 'filtered' is the correct quadruple filtered out of the corrupted quadruples.

Experimental Implementation Details To present an impartial comparison between *HTSE* and the baselines, we utilize the experimental setup for the classic HyTE and DyERNIE baselines as our experimental basis and select the optimal parameters for each model. Then, we use RSGD to train the baselines and optimize the hyperparameter setups in accordance with the MRRs obtained on the validation set. We set the maximum number of epochs to 5000 and fix the minibatch size to 1024. The remaining settings are as follows: the embedding dimensions $d = 100, 200, 300, 500, 1000$, the learning rate $l_\gamma = 0.05, 0.001, 0.005, 0.01, 0.1$, $\eta = 1, 3, 6, 12$, the margin $\gamma = 3, 6, 12, 36, 48, 120$, and the error control parameter μ is varied in the range $[0, 1]$.

4.2 Temporal Link Prediction

In the evaluation, we aim to illustrate the advantages of our model with respect to hierarchy and time surface modelling in comparison with other models. TransE[2] and RotatE[20] are experimentally chosen for comparison due to their common embedding space and lack of consideration for temporal dynamics and hierarchies. HyTE[6], TIMEPLEX[9], DyERNIE[8], and ChronoR[19] are chosen due to their use of hyperplanes, Euclidean distances, complex spaces and Riemann manifolds instead of surface models.

The significant task is to predict the missing entity for an incomplete quadruple. Unlike previous works involving static KGs, this task can predict missing entities for quadruples in a TKG. More formally, for negative samples derived from the gold-standard quadruple (h, r_τ, t, τ), we perform prediction on two categories: $(h, r_\tau, ?, \tau)$ and $(?, r_\tau, t, \tau)$. Following the same filtered DyERNIE[8] settings, we evaluate our model with the MRR and Hits@1, 3, 10 metrics mentioned in the 4.1 Evaluation Protocol. The results obtained with the above experimental settings are presented in Table 4.2.

Results and Observations From Table 4.2, on the YAGO11k and Wikidata12k datasets, *HTSE* produces excellent link prediction results compared with those of the promising HyTE, TIMEPLEX and DyERNIE models as well

Table 2 Temporal link prediction performance on the datasets

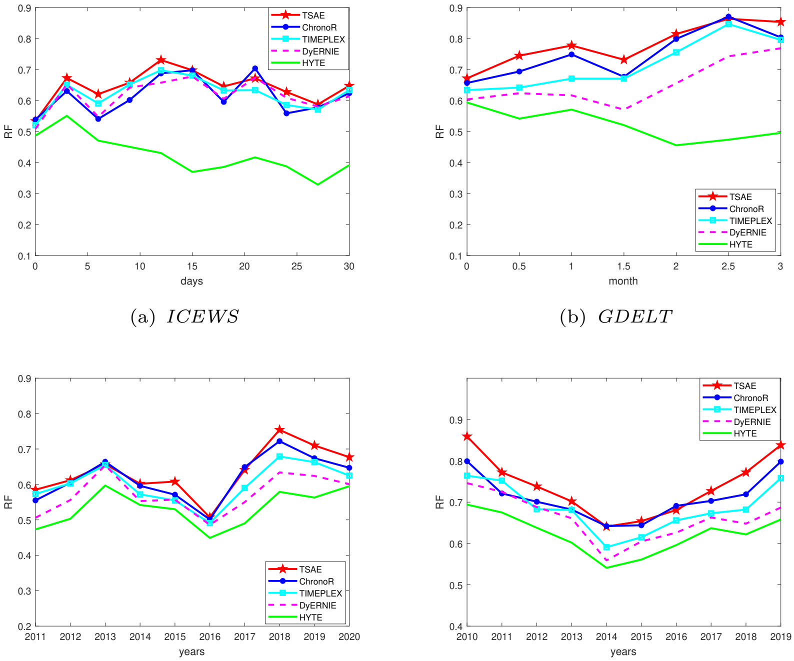
Model		Static model		Temporal model				
		TransE	RotatE	HyTE	ChronoR	DyERNIE	TIMEPLEX	**HTSE(C)**
YAGO11k	MRR	.103	.167	.136	**.248**	.214	.246	**.251±.004**
	H1	.060	.103	.132	.253	.243	.169	**.260±.003**
	H3	.183	.197	.213	.335	.334	.389	**.407±.001**
	H10	.244	.305	.301	.344	.491	.484	**.518±.003**
Wikidata12k	MRR	.178	.221	.253	.375	.402	.336	**.411±.001**
	H1	.121	.116	.147	.147	.275	.228	**.301±.004**
	H3	.192	.239	.197	.209	.417	**.477**	.469±.004
	H10	.344	.461	.483	.489	.589	.532	**.603±.004**
GDELT	MRR	.113	.203	.118	.230	.457	**.470**	**.503±.001**
	H1	.000	.204	.000	.141	**.387**	.312	.371±.003
	H3	.158	.277	.165	.248	.479	.427	**.502±.005**
	H10	.312	.389	.326	.403	.589	.599	**.621±.001**
ICEWS-14	MRR	.283	.400	.298	.526	.669	.604	**.674±.002**
	H1	.094	.388	.108	.418	.599	.601	**.620±.001**
	H3	.367	.480	.416	.592	.714	.515	**.726±.006**
	H10	.638	.724	.655	.725	.797	.771	**.819±.004**
ICEWS05-15	MRR	.294	.575	.316	.513	.739	.640	**.759±.005**
	H1	.090	.388	.116	.392	.679	.546	**.720±.009**
	H3	.354	.522	.445	.578	**.773**	**.787**	.766±.003
	H10	.663	.709	.681	.748	.855	.818	**.878±.004**

as the traditional static models by virtue of considering a projected time surface instead of a traditional space, which can allow high-dimensional data to be more accurately captured. Moreover, the experimental results show improvements greater than 4% in terms of both the MRR and Hits@10 metrics. These findings suggest that the hierarchy modelling in our model is more adaptable than other methods for link prediction.

From Table 4.2, on the GDELT dataset, *HTSE* also achieves somewhat better performance except in terms of the Hits@1, as this metric may be affected by the time stamps of the validation facts. According to the other results, the temporal models are superior to the static models, thus demonstrating the significance of capturing time information. On the ICEWS14 and ICEWS05-15 datasets, the results of *HTSE* are also superior to those of the existing excellent models on average due to the more accurate expressivity in our model.

Overall, *HTSE* outperforms the other models in terms of link prediction due to its advantages of hierarchy modelling and more accurate projections to preserve temporal interactions. In addition, error control is beneficial for improving the results because it can ensure that the time-specific relation vector r_τ for calculating the distance from the original vector r is the best nearest neighbour.

Link Prediction Performance over Time To clearly and comprehensively illustrate the superior performance of *HTSE* in terms of future temporal link prediction performance over time, we take the time information as an index and present corresponding comparisons on the ICEWS, GDELT, Wikidata and YAGO datasets, as depicted in the line chart in Fig. 3. Specifically, the vertical axis *filtered Hit@3* represents the Evaluation Protocol in Section 4.1, and the horizontal axis includes different time stamps (day, month and year). Here, we present the performance of *HTSE* under the 'filtered' settings to show that it exhibits superior temporal link prediction capabilities over time after the removal of corrupted validations or test triples. Furthermore,



(a) *ICEWS* (b) *GDELT*

(c) *Wikidata* (d) *YAGO*

Fig. 3 Results of temporal link prediction for future time stamps

we choose HyTE[6], DyERNIE[8], ChronoR[19] and TIMEPLEX[9] for comparison because of their poor projection abilities and the fact that they ignore hierarchies.

Fig. 3 shows that the performance of these models fluctuates at different time stamps. We notice that *HTSE*, corresponding to the **red line**, consistently outperforms the other models for different time intervals, and its performance varies irregularly over time. Furthermore, this finding suggests that hierarchy modelling and surface projection achieve enhanced link prediction performance over time compared to that of other models. Inspired by these results, we consider that future performance can be predicted by simulating and fitting numerical curves following an existing time sequence. This is beneficial for addressing the issue of predicting future time periods.

4.3 Temporal Scope Prediction Results

In this part, we focus on illustrating the advantage of *HTSE* based on surface projection for temporal scope prediction. Inspired by [9], the *PG* score function is proposed to evaluate the accuracy of time stamp completion and the temporal information prediction results obtained in TKG embeddings. In addition, we test the results on the ICEWS14, ICEWS05-15 and GDELT datasets. HyTE [6], TIMEPLEX [9], DyERNIE [8], and ChronoR [19] are chosen for

Table 3 *PG* scores obtained for temporal information prediction

Model	HyTE	DyERNIE	TIMEPLEX	ChronoR	**HTSE(Our)**
*ICEWS14	44.3%	**70.0%**	53.9%	**70.2%**	69.5%
*ICEWS05-15	39.1%	70.4%	63.0%	63.9%	**71.3%**
*GDELT	36.6%	61.6%	55.6%	62.8%	**64.2%**

comparison because these models have certain abilities to predict temporal scopes, and none of them focus on hierarchical knowledge representations and surface projections.

Considering the scarcity and completeness of facts in a TKG, the ability to predict time information is indispensable. We wish to predict the time instance and time interval that target a given test quadruple $(h, r_\tau, t, ?)$. According to the established time surface, the relations and entities are projected onto this surface to check the plausibility of the test triple.

To correctly predict the temporal interval, we should make use of the optimal nearest fact. Specifically, for $(h, r_\tau, t, ?)$, the gold-standard time interval is $T_{gold} = [t_g^s, t_g^e]$ (which consists of the starting and ending time stamps of a true fact), and this interval should be compared to the predicted interval $T_{pre} = [t_p^s, t_p^e]$ to determine the similarity of the prediction to the true fact.

In terms of the chosen evaluation metrics, the metric in TKBC [9] is not entirely applicable to this task because it is designed to address the problem of large differences in the time proportion for the TAC metric [9]. For instance, two groups of intervals exist, golden interval $[2014, 2017]$ compared with predicted interval $[2010, 2013]$ and golden interval $[7, 10]$ compared with predicted interval [11,14]. Although the two groups share the same TAC score $\left[\frac{1}{1+|t_s^1-t_s^2|} + \frac{1}{1+|t_e^1-t_e^2|}\right]$, the former time proportion is obviously less than the latter. That is, a 4-year difference in 10 years may usually be considered more serious than in 2010. However, the time proportion in datasets cannot be large for this task. In response, we set the *PG* score function as a metric inspired by improving the TAC metric with a two-parametric calculation to enhance the accuracy of the score results.

Specifically, the difference *PG* between two time intervals/scopes can be calculated as:

$$PG = \frac{1}{1 + t_p^s - t_g^s} + \frac{1}{1 + t_p^e - t_g^e} \tag{7}$$

The score function *PG* serves as a criterion for evaluating the accuracy of time predictions. As $t_p^s - t_g^s, t_p^e - t_g^e \to 0$, PG approaches its maximum value, i.e., $PG \to \sqrt{2}$. Thus, the higher the *PG* score (which is subject to the condition $0 < PG < \sqrt{2}$), the closer the predicted time interval is to the gold-standard value for a given fact. The results of TTransE, TIMEPLEX, DyERNIE, HyTE and *HTSE* are compared in Table 3. Note: the scores are converted to percentages with a maximum score

As indicated in this table, the results of *HTSE*, TIMEPLEX and DyERNIE are vastly superior to those of HyTE and TTransE. These findings show that a more adaptive projection space can result in better performance in terms of *PG*. Moreover, the results of 71.3% and 64.2% achieved by *HTSE* on the ICEWS05-15 and on GDELT datasets, respectively, are both the highest values. On the ICEWS14 dataset, our model is slightly inferior to DyERNIE and TIMEPLEX because of a few redundant entities in the datasets. *HTSE* outperforms DyERNIE and TIMEPLEX due to its innovative time surface, which offers higher representative power and thus improves the accuracy of the predicted temporal intervals. It is worth mentioning that DyERNIE and ChronoR are second only to *HTSE*, which further illustrates the benefits of utilizing specific geometry-based methods to enhance the performance of temporal scope inference by addressing the issue of incomplete projections in the embedding process. In other words, the surface projection benefits the presentation of temporal semantic information based on a given triple. This would be an exciting conclusion of this research.

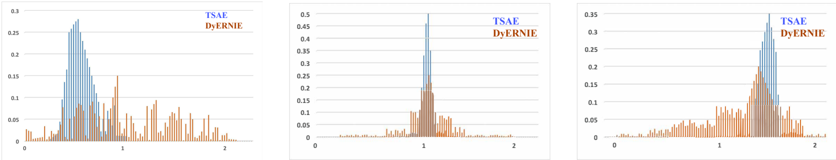
4.4 Analysis of Hierarchical Relation Embeddings

In this part, we aim to demonstrate that the *HTSE* model can effectively model hierarchical entities and relations at different levels by introducing a modulus in relation embeddings inspired by [26]. In addition, it is shown that *HTSE* is more accurate than the similar DyERNIE[8] model for entity matching at three types of hierarchical levels because it comprehensively analyses the embeddings of the hierarchical relations.

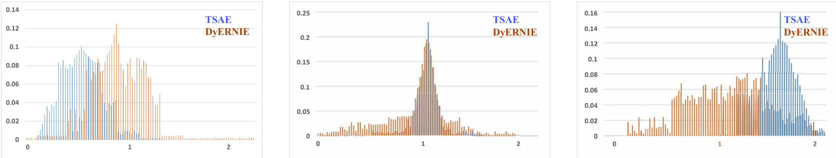
In Fig. 4, the distribution histograms concerning three types of relations are presented with the corresponding hierarchies. These relations are chosen from the ICEWS, GDELT, Wikidata and YAGO datasets. Specifically, the comparison involves three groups of examples concerning three types of hierarchical relations.

1. As shown in (a) and (d) in Fig. 4, the relations "$is\_affiliated\_to$" in the YAGO dataset and "$city\_of$" in the GDELT dataset indicate that the head entities are at lower semantic hierarchies than the corresponding tail entities.
2. As (b) and (e) in Fig. 4, the entities linking the relations "*make_a_visit*" in the ICEWS dataset and "*is_married_to*" in the YAGO dataset are at the same semantic hierarchy.
3. As shown in (c) and (f) in Fig. 4, the relations "*has_part*" and "*has_cause*" in the Wikidata dataset mean that this situation is exactly the opposite of (a) and (d) in the first item.

Similarly, we set h_m and t_m as each entry of head and tail entities, that is, the corresponding moduli are h_m and t_m. Then, following [26], we formulate corresponding relations as follows: $r_m = h_m^{-1} \circ t_m$. Therefore, we can obtain the modulus of relations as r_m. Experientially, the moduli of $r_m = 1$ indicate the same hierarchy. Notably, a smaller modulus (the horizontal axis) indicates a



(a) $is\_affiliated\_to$ (b) $make\_a\_visit$ (c) has_part



(d) $city\_of$ (e) $is\_married\_to$ (f) has_cause

Fig. 4 Distribution histograms of hierarchical relation embeddings for several temporal datasets

Table 4 Case study for a query based on a different hierarchy

Query1: $(Malaysia, host\_a\_visit, ?, 2014-02-14)$ **Candidate condition:** C0: Malaysia C1: Barack Obama C2: United States **Requirement:** $\|C_0\| > \|C_1\|$ and Lower hierarchy **Confirmed answer:** C1
Query2: $(?, Demand, Citizen, 2014-08-01)$ **Candidate condition:** C0: Citizen C1: United Kingdom C2:Lawyer/Attorney **Requirement:** $\|C_0\| = \|C_2\|$ and equal hierarchy **Confirmed answer:** C2
Query3: $(?, owns, Buffalo\_Stadium, 1961-1963)$ **Candidate condition:** C0: Malaysia C1: Houston_Astros C2: JimmyWynn **Requirement:** $\|C_0\| < \|C_1\|$ and higher hierarchy **Confirmed answer:** C1

lower semantic hierarchy. On this basis, we hope that a smaller model variance and a tighter distribution lead to clearer hierarchy modelling effects. Specifically, $moduli < 1$, and most entries are located on the left side of $moduli = 1$ in cases (a) and (d), illustrating the lower hierarchies of the head entities. Cases (c) and (f) are contrary to cases (a) and (d). Moreover, as indicated in Fig. 4, our *HTSE* model (the blue one) corresponds to a much tighter distribution than DyERNIE (the orange one), which represents a smaller variance and precisely proves our expectation. In conclusion, this experiment demonstrates that our model can model hierarchical relations and distinguish corresponding entities better than a similar model.

4.5 Qualitative Analysis

To demonstrate the explainable strategy for hierarchical embeddings in our model, we present some queries, corresponding candidates and requirements to evaluate the confirmed answer. Table 4 shows queries based on different hierarchies. For example, the query 1 means that 'Who is host the visit in

Malaysia? ', we may compare the candidate 'Barack Obama' with 'United States', then choose 'Barack Obama' as the confirmed answer because of the requirement 'Lower hierarchy and smaller moduli'. For the query 2 and query 3, we confirm corresponding answers based on their respective requirements. These cases fit true facts in real datasets, which can prove that hierarchy is very meaningful for temporal knowledge graph completion.

5 Conclusion

In this work, a novel time surface-aware model for embedding is proposed to learn significant representations from TKGs. Our model employs a relation-oriented hierarchy modelling strategy to address the issue of ignoring semantic hierarchies. Another issue concerns the inaccurate semantic expressions caused by the limitations of incomplete projections. To address this issue, we use a time surface model with exponential mapping to enhance the representations of temporal characteristics. Experimental results indicate that *HTSE* achieves promising results and outperforms its geometric counterpart and other SOTA models. It demonstrates the advantages of utilizing surface-based spaces and hierarchical modelling for inference and prediction tasks in TKG embeddings. Furthermore, we believe that enhancing the interpretability of temporal causal embeddings will be the focus of future research.

Declarations

- Ethical Approval
  (applicable for both human and/ or animal studies. Ethical committees, Internal Review Boards and guidelines followed must be named. When applicable, additional headings with statements on consent to participate and consent to publish are also required)
  This declaration is not applicable.
- Competing interests
  (always applicable and includes interests of a financial or personal nature)
  This declaration is not applicable.
- Authors' contributions
  (applicable for submissions with multiple authors)
  Langjunqing Jin wrote the main mauscript text, including Introduction, Related work, Model and Experiments, etc. FengZhao revised the abstract, Introduction and Model. HaiJin revised the Reference. All authors reviewed the manuscipt.
- Funding
  (details of any funding received)
  This declaration is not applicable.
- Availability of data and materials
  (a statement on how any datasets used can be accessed)
  If the article is published later, these related materials will be posted to github.

References

[1] Alberto García-Durán, Sebastijan Dumančić, M.N.: Learning sequence encoders for temporal knowledge graph completion. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing(EMNLP). pp. 4816–4821 (2018)

[2] Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS). pp. 2787–2795 (2013)

[3] Chami, I., Wolf, A., Juan, D.C., Sala, F., Ravi, S., Ré, C.: Low-dimensional hyperbolic knowledge graph embeddings. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics(ACL). pp. 6901–6914 (2020)

[4] Chen, J., He, H., Wu, F., Wang, J.: Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In: Preoceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence(AAAI). pp. 6271–6278 (2021)

[5] Chen, S., Liu, X., Gao, J., Jiao, J., Zhang, R., Ji, Y.: Hitter: Hierarchical transformers for knowledge graph embeddings. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing(EMNLP). pp. 10395–10407 (2021)

[6] Dasgupta, S.S., Ray, S.N., Talukdar, P.P.: Hyte: Hyperplane-based temporally aware knowledge graph embedding. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing(EMNLP). pp. 2001–2011 (2018)

[7] Goel, R., Kazemi, S.M., Brubaker, M., Poupart, P.: Diachronic embedding for temporal knowledge graph completion. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence(AAAI). pp. 3988–3995 (2020)

[8] Han, Z., Chen, P., Ma, Y., Tresp, V.: Dyernie: Dynamic evolution of riemannian manifold embeddings for temporal knowledge graph completion. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 7301–7316 (2020)

[9] Jain, P., Rathi, S., Chakrabarti, S.: Temporal knowledge base completion: New algorithms and evaluation protocols. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 3733–3747 (2020)

[10] Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: Representation, acquisition and applications. IEEE Transactions on Neural Networks **33**, 495–514 (2022)

[11] Jin, W., Qu, M., Jin, X., Ren, X.: Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). pp. 6669–6683 (2020)

[12] Leblay, J., Chekol, M.W.: Deriving validity time in knowledge graph. In: Proceedings of the 27th World Wide Web Conference(WWW). pp.

1771–1776 (2018)

[13] Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location and tone, 1979-2012. ISA annual convention **2**, 1–49 (2013)

[14] Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI). pp. 2181–2187 (2015)

[15] Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., Scaramuzza, D.: The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam:. The International Journal of Robotics Research **36**(2), 142–149 (2017)

[16] Pan, Z., Wang, P.: Hyperbolic hierarchy-aware knowledge graph embedding for link prediction. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing(EMNLP). pp. 2941–2948 (2021)

[17] Park, N., Liu, F., Mehta, P., Cristofor, D., Faloutsos, C., Dong, Y.: Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In: Proceedings of Fifteenth ACM International Conference on Web Search and Data Mining(WSDM). pp. 794–803 (2022)

[18] Peng, Y., Zhang, J.: Lineare: Simple but powerful knowledge graph embedding for link prediction. In: Proceedings of the 20th IEEE International Conference on Data Mining (ICDM). pp. 422–431 (2020)

[19] Sadeghian, A., Armandpour, M., Colas, A., Wang, D.Z.: Chronor: Rotation based temporal knowledge graph embedding. In: Preoceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence(AAAI). pp. 6471–6479 (2021)

[20] Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J.: Rotate: Knowledge graph embedding by relational rotation in complex space. In: Proceedings of the 7th International Conference on Learning Representations (ICLR). pp. 1–18 (2019)

[21] Trivedi, R., Dai, H., Wang, Y., Song, L.: Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In: Proceedings of the International Conference on Machine Learning(ICML). pp. 3462–3471 (2017)

[22] Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: Proceedings of the 33rd International Conference on Machine Learning (ICML). pp. 2071–2080 (2016)

[23] Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI). pp. 1112–1119 (2014)

[24] Xiao, H., Huang, M., Zhu, X.: From one point to a manifold: knowledge graph embedding for precise link prediction. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence(IJCAI). pp. 1315–1321 (2016)

[25] Xu, C., Nayyeri, M., Alkhoury, F., Yazdi, H.S., Lehmann, J.: Temporal

knowledge graph completion based on time series gaussian embedding. In: Proceedings of the International Semantic Web Conference(ISWC). pp. 654–671 (2020)

[26] Zhang, Z., Cai, J., Zhang, Y., Wang, J.: Learning hierarchy-aware knowledge graph embeddings for link prediction. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI). pp. 3065–3072 (2020)