HTCCN: Temporal Causal Convolutional Networks with Hawkes Process for Extrapolation Reasoning in Temporal Knowledge Graphs

Tingxuan Chen[1], Jun Long[2*], Liu Yang[1*], Zidong Wang[1], Yongheng Wang[3], and Xiongnan Jin[3]

[1]School of Computer Science and Engineering, Central South University, Hunan, China
[2]Big Data Institute, Central South University, Hunan, China
[3]Zhejiang lab, Zhejiang, China
{chentingxuan, junlong, yangliu, zdwang}@csu.edu.cn
{wangyh, xiongnan.jin}@zhejianglab.com

Abstract

Temporal knowledge graphs (TKGs) serve as powerful tools for storing and modeling dynamic facts, holding immense potential in anticipating future facts. Since future facts are inherently unknowable, effectively modeling the intricate temporal structure of historical facts becomes paramount for accurate prediction. However, current models often rely heavily on fact recurrence or periodicity, leading to information loss due to prolonged evolutionary processes. Notably, the occurrence of one fact always influences the likelihood of another. To this end, we propose HTCCN, a novel Hawkes process-based temporal causal convolutional network designed for temporal reasoning under extrapolation settings. HTCCN employs a temporal causal convolutional network to model the historical interdependence of facts and leverages Hawkes to model link formation processes inductively in TKGs. Importantly, HTCCN introduces dual-level dynamics to comprehensively capture the temporal evolution of facts. Rigorous experimentation on four real-world datasets underscores the superior performance of HTCCN.

1 Introduction

Knowledge Graphs (KGs) serve as graph-structured bases for storing human knowledge, promising numerous real-world scenarios in recommender systems (Fan et al., 2022), information retrieval (Gaur et al., 2022), and Q&A (Cui et al., 2023). Typically, KGs express and store knowledge in the form of $(subject, relation, object)$. However, facts undergo continuous changes over time, and facts exhibit interactions with one another. To effectively express this dynamic information, researchers construct Temporal Knowledge Graphs (TKGs) enabling the representation of each fact as a quadruple $(subject, relation, object, timestamp)$. Generally, a TKG is represented as a sequence of static KG snapshots, each associated with a timestamp indicating the occurrence of facts at a specific time.

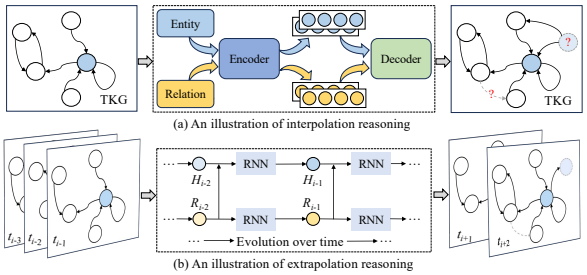

Figure 1: An illustration of interpolation reasoning and extrapolation reasoning.

The TKG reasoning task involves inferring new facts from known facts and is generally approached in two main settings, i.e. interpolation and extrapolation (Wang et al., 2023). In the interpolation setting, the goal is to recover historically missing facts. Conversely, in the extrapolation setting, historical facts are used to predict future facts (Liu et al., 2022a). This paper specifically focuses on the extrapolation setting, which poses significant challenges and remains an unsolved problem.

The prevailing interpolation reasoning method, as depicted in Fig. 1(a), typically extends static reasoning methods to TKGs. This method involves employing an encoder to derive embeddings of relations and entities within a particular time hyperplane. Subsequently, a decoder is used to predict incomplete facts at historical timestamps (Liu et al., 2023). However, relying solely on representations of relations and entities based on historical hyperplanes can lead to the model conflating future facts within the candidate set, thereby impeding its ability to accurately reason about future facts.

The key to the extrapolation reasoning task lies in fully understanding historical facts. As illus-

* Corresponding Author

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 4056–4066

trated in Fig. 1(b), many efforts center around extracting essential information from complex time-structured data to facilitate answering queries and making informed judgments (Bai et al., 2023). Representative methods such as RE-NET (Jin et al., 2020) and RE-GCN (Li et al., 2021b) utilize R-GCN (Schlichtkrull et al., 2018) to capture structural information from historical snapshots, while RNNs are employed to model the temporal information of facts. More recent approaches, including CluSTeR (Li et al., 2021a) and TiRGN (Li et al., 2022a), reason based on historically observed facts, enabling them to handle repetitive or periodic facts.

However, existing RNN-based methods model temporal characteristics of facts as equidistant, which does not accurately reflect real-world fact sequences that often have varying temporal intervals (Chen et al., 2020). Additionally, during the inference process, these methods tend to rank the probability scores of all candidate entities across the entire graph, disregarding the crucial aspect of factual evolution (Zhang et al., 2023b). We contend that fact evolution is highly essential when reasoning about future facts, while the fundamental concept of fact evolution lies in the conditional intensity function, which represents the probability of a fact occurring over time.

To address extrapolation reasoning challenges in TKGs, we propose the **H**awkes process-based **T**emporal **C**ausal **C**onvolutional **N**etwork, denoted as HTCCN. This innovative framework rests upon three fundamental pillars: (1) The adept utilization of a temporal causal convolutional network, empowered by inflated convolutions, which adeptly captures historical fact dependencies and effectively addresses the complex temporal dynamics over diverse time intervals. (2) The seamless integration of Hawkes processes, harnessed to inductively model the link formation processes within TKGs, concurrently establishing conditional intensity functions to capture structural and temporal information in TKGs. (3) The incorporation of dual-level dynamics, which comprehensively capture the evolution of facts.

In summary, the main contributions of this paper are as follows:

- We propose a temporal extrapolation reasoning model HTCCN, which adeptly takes into account the sequence of historical facts with variable temporal intervals, the interplay of causality among these facts, and the dynamic evolution of the facts themselves.
- We formulate a dual-level dynamic modeling mechanism to simultaneously capture both the collective features of nodes and the individual characteristics of facts. This is achieved by leveraging node-level dynamics and fact-level dynamics, offering a more comprehensive understanding of the temporal evolution in TKGs.
- Extensive experiments are conducted across four widely recognized TKG datasets, and the noteworthy enhancements observed across nearly all performance metrics underscore the effectiveness of HTCCN in the context of TKG extrapolation.

2 Related Work

Reasoning on TKGs has drawn a lot of attention, mainly including neural network-based methods and query-based methods.

Neural network-based approaches utilize deep neural networks to learn the underlying features of timestamps for temporal reasoning. RE-NET (Jin et al., 2020) and TiRGN (Li et al., 2022a) combine recurrent neural networks and neighborhood aggregators to model fact sequences. RE-GCN (Li et al., 2021b) recursively models KG sequences using recursive evolutionary networks to learn an evolutionary representation of facts for each timestamp. TANGO (Han et al., 2021) employs neural ordinary differential equations to model the structural information of each candidate entity in a continuous time domain. CEN (Li et al., 2022b) develops an evolutionary model that represents all candidate entities by considering the history of a few recent timestamps. GHT (Sun et al., 2022) introduces two Transformer variants to harness structural and temporal information in TKGs, alongside developing a conditional intensity function for temporal prediction. TECHS (Lin et al., 2023) utilizes a graph convolutional network combined with temporal encoding and heterogeneous attention mechanisms to focus on attention-driven structural features. Meanwhile, RETIA (Liu et al., 2023) integrates a relational aggregation module with LSTM to model relations and track the temporal progression of hyper-relations. However, these models model the temporal characteristics of facts as isometric, which fail to accurately reflect the dynamic evolutionary patterns of facts in the real world.

Query-based approaches in temporal reasoning concentrate on modeling entities and relations related to queries. For instance, CyGNet (Zhu et al., 2021) reasons future facts by modeling query-related history facts through a copy-generation network. GHNN (Han et al., 2020b) introduces a point-in-time process to model temporal information and considers the 1-hop subgraph of query entities. TITer (Sun et al., 2021) utilizes a temporal-aware reinforcement learning strategy to continuously transfer query nodes to new nodes using relevant temporal facts and generate representation vectors of unseen entities using the IM module. CluSTeR (Li et al., 2021a) combines reinforcement learning and graph convolutional networks for reasoning on TKGs. Meanwhile, xERTE (Han et al., 2020a) addresses the TKG inference problem by using query subgraphs with time-aware neighbor sampling and attention propagation extensions, and additionally introduces a novel backward updating scheme to simulate human reasoning behavior. However, these approaches primarily focus on the history associated with a query, overlooking the complex dynamic interactions between facts.

3 Problem Definition

In this paper, we view a TKG as a sequence of timestamp ascension-based subgraphs, namely, $G_{(1,\tau)} = \{G_1, G_2, \cdots, G_\tau\}$, where $G_t = (V_t, E_t, F_t)$ corresponds to the subgraph at time $t(t \in 1,\ 2, \cdots, \tau)$ with an entity set V_t and a relation set E_t. Each fact in F_t is denoted by a quadruple (s, r, o, t), in which $s, o \in V_t$ and $r \in E_t$.

Based on the historical observed facts in $G_{(1,\tau)}$, this paper concentrates on reasoning under the extrapolation setting, i.e., predicting potential object entity o or subject entity s via answering queries query $(s, r, ?, t)$ or $(?, r, o, t)$.

4 Method

4.1 Overview of HTCCN

Leveraging the principles of the Hawkes process, we propose the Temporal Causal Convolutional Network (HTCCN) to effectively model the link formation processes within TKGs inductively. The overall framework of HTCCN is illustrated in Fig. 2. A key strength of our approach lies in its ability to seamlessly integrate both fact and node dynamics into the model. This dual-level integration enables us to capture not only the individual features of facts but also the collective features of nodes, resulting in a more holistic understanding of the temporal evolution within TKGs.

4.2 Hawkes Process-based TCCN

Recognizing the TKG as a multi-relational heterogeneous graph, previous studies (Jin et al., 2020; Li et al., 2022a) have employed R-GCN (Schlichtkrull et al., 2018) and RE-GCN (Li et al., 2021b) to aggregate multi-relation and multi-hop neighbor information at a single timestamp, storing the interplay between nodes in the form of hidden vectors.

However, within historical snapshots, multiple facts frequently relate to an entity concurrently, yet only a minority of the neighboring information contributes to query resolution. R-GCN falls short in addressing this issue as it treats all messages equally, overlooking the significance of some. As a response, we introduce HTCCN to realize the heightened importance of concurrent facts for reasoning about future facts.

4.2.1 Temporal Causal Convolution

At each timestamp, we aim to derive a temporal representation that possesses the ability to extend into the feature space for new nodes, while preserving the inherent fact characteristics constituted by the original entities and relations. To fulfill this objective, we let each node engage in recursive processes, encompassing the reception, aggregation, and mapping of multiple relations and multi-hop neighbor information across layers. Notably, our approach diverges from RE-GCN, which sums relational embeddings to entity embeddings within GCNs for local information aggregation. Instead, we opt to initially integrate entity embeddings and relational embeddings to ensure fact plausibility, subsequently enhancing their aggregation through the use of 1D convolutional operations. Formally, let $h^l_{i,t} \in \mathbb{R}^{d_l}$ denote the d_l-dimensional embedding of the entity i at t in the l-th layer, which is computed by

$$\mathrm{h}^{l+1}_{o,t} = \sigma\left(\sum_{(s,r,o,t)\in F_t} \frac{1}{c_o}\mathrm{W}^l_r\left(\Gamma_1(\mathrm{h}^l_{s,t} \parallel \mathrm{r}_t)\right) + \mathrm{W}^l_o \mathrm{h}^l_{o,t}\right) \tag{1}$$

where σ is an activation function (e.g., RReLu), W^l_o represents a learnable weight matrix that maps the embedding of entity o itself from the previous layer, W^l_r is a relation-specific learnable weight matrix to map the embeddings of subject s and relation r, c_o is a normalizing factor equal to the in-degree of o, Γ_1 denotes the 1D convolutional

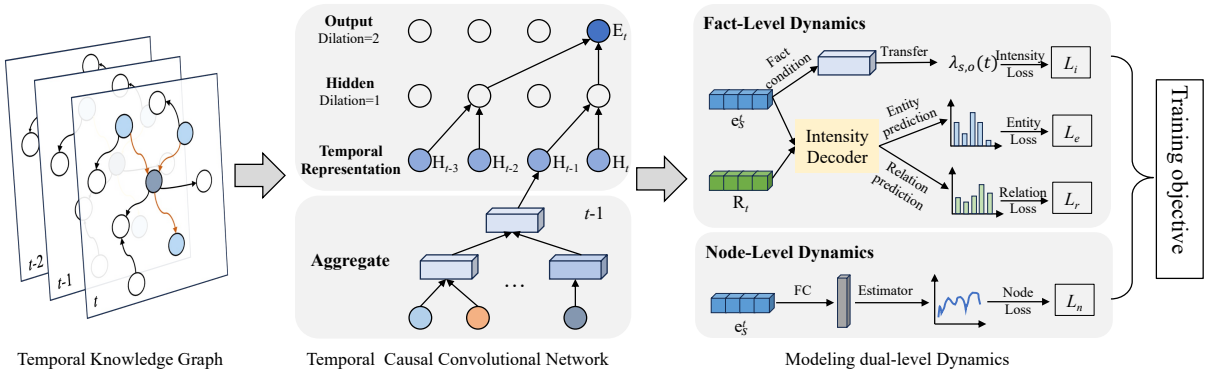

Figure 2: Illustration of the proposed HTCCN model. The temporal causal convolutional network (TCCN) encodes evolutionary representations of entities and relations. The dual-level dynamics capture individual features of facts and collective features of nodes.

operator, and $\|$ means concentrate operation.

In essence, the temporal representation of a node is computed through the reception and aggregation of messages from both itself and its historical neighbors from the preceding layer. Self-messages are crucial in capturing the base intensity, whereas messages from historical neighbors are instrumental in capturing the excitement induced by historical facts. Subsequently, these computations lead to the generation of a temporal representation sequence denoted as $\mathbf{H}_{1:t} = \{\mathrm{H}_1, \mathrm{H}_2, ..., \mathrm{H}_t\}$, where $\mathrm{H}_t = \left\{\mathrm{h}_{1,t}^{l+1}, \mathrm{h}_{2,t}^{l+1}, ..., \mathrm{h}_{n,t}^{l+1}\right\}$ denotes the temporal representation sequence of entities at time t.

As facts strictly adhere to a temporal sequence, we devise a temporal causal convolutional network in which convolutions are restricted to use inputs at time t or earlier, preserving the temporal causality. Additionally, we introduce a dilation factor, allowing filters to span regions larger than their size by skipping input values in specific steps. This strategic choice expands the model's receptive field, providing it with a longer effective memory. Let d denote the dilation factor, the temporal causal convolutions can be expressed as follows:

$$\mathbf{E}_t = (\mathbf{H}_{1:t} *_d \Gamma_2)(t) = \sum_{k=0}^{K-1} \Gamma_2(k)\mathbf{H}_{1:t}(t - d \cdot k) \quad (2)$$

where Γ_2 represents the convolution operator, and K is the size of the kernel. Importantly, the model can employ a stacked configuration of multiple temporal causal convolutional networks, thereby expanding its receptive field and efficiently capturing long-term temporal dependencies with a smaller number of layers. Given the temporal representation sequence $\mathbf{H}_{1:t}$, the output $\mathbf{E}_t$ is a sequence of the entity embeddings at time t.

After that, we employ a GRU (Gated Recurrent Unit) to update the relation embedding, while ensuring consistency with the update of entity embeddings within the subgraph sequence.

$$\mathbf{r}'_t = [\mathcal{F}_{mean}(\mathbf{E}_{t-1}, \mathcal{E}_{r,t}); \mathbf{r}] \quad (3)$$

$$\mathbf{R}_t = \mathrm{GRU}(\mathbf{R}_{t-1}, \mathbf{R}'_t) \quad (4)$$

where $\mathcal{F}_{mean}$ is a mean pooling operation, $\mathcal{E}_{r,t}$ includes all entities under relation $\mathbf{r}$ at timestamp t, $\mathbf{R}'_t$ is the set of all relations $\mathbf{r}'_t$. The final relation embedding matrix $\mathbf{R}_t$ is updated by $\mathbf{R}_{t-1}$ and $\mathbf{R}'_t$ by a single GRU layer.

4.2.2 Hawkes Process on TKG

In the sequences of TKGs, the Hawkes process proves adept at modeling the fact-link formation process. More precisely, our approach HTCCN quantifies whether nodes s and o form a temporal link at timestamp t based on the conditional intensity of the fact,

$$\lambda_{s,o}(t) = \mu_{s,o}(t) + \sum_{(s,o',t') \in \mathcal{H}_t^s} \gamma_{o'}(t') \exp(-\delta(t - t')) \quad (5)$$

Here, $\mu_{s,o}$ signifies the baseline rate of the fact where entities s and o form a temporal link at timestamp t, unaffected by historical facts on s or o. $\mathcal{H}_t^s = \left\{(s, o', t') \mid o' \in \mathcal{N}_{t'}^{s,r}, t' < t\right\}$ denotes the set of historical facts on s with respect to time t, designating o' as a historical neighbor of s. $\gamma_{o'}(t')$ quantifies the excitement instilled by a historical neighbor o' at time t' on the present fact. Additionally, we introduce δ, a scalar parameter that is amenable to learning, serving to regulate the rate of the time decay effect concerning timestamp t.

Subsequently, based on the embeddings $\mathbf{e}_s^t$ and $\mathbf{e}_o^t$ of entities s and o at timestamp t respectively,

we employ a transfer function f to materialize the conditional intensity as outlined in Eq. (5), as follows:

$$\lambda_{s,o}(t) = f(\mathbf{e}_s^t, \mathbf{e}_o^t) \tag{6}$$

where the crux of materializing the conditional intensity lies in fitting a transfer function f on the top of temporal causal convolution layers. To accommodate the diversity of facts, we adopt the softplus function, ensuring that f is adeptly tailored to the conditional intensity.

$$f(x) = \beta \cdot log(1 + exp(\frac{x}{\beta})) \tag{7}$$

Here, the parameter β serves as the scaling factor for the transfer function f, while 1 represents a vector of ones, ensuring that the scaling factors are centered around one. Note that the input to f can take various forms, such as the concatenation of $\mathbf{e}_s^t$ and $\mathbf{e}_s^t$ or the element-wise square of the difference between them. In our formulation, we adopt the latter, as it tends to yield superior empirical performance. One possible explanation is that the differential representation proves to be a robust predictor of whether a fact occurs between the two nodes. Finally, f incorporates a sigmoid activation to guarantee that the output remains positive as it denotes the conditional intensity of the fact.

4.3 Intensity Guided Decoder

After acquiring the conditional intensities for facts, to facilitate entity prediction and relation prediction under the extrapolation setting, we present Intensity-ConvTransE and Intensity-ConvTransR, drawing inspiration from the ConvTransE (Shang et al., 2019) framework. In these models, the decoder conducts a one-dimensional convolution operation on the concatenation of entity embeddings and relation embeddings, followed by scoring the resulting representations. Formally, the computation of the convolution operator is:

$$\begin{aligned} m_n =& \Gamma_c\left(\mathbf{e}_s^t, \mathbf{r}_t, \lambda_s(t), n\right) \\ =& \sum_{\tau=0}^{K-1} \mathbf{w}_c(\tau)(\mathbf{e}_s^t \cdot \lambda_s(t) \parallel \mathbf{r}_t)(n+\tau) \end{aligned} \tag{8}$$

where k equals the number of convolution kernels, K is the size of the kernel, $n \in [0, d)$ is the index of output vector, and $\mathbf{w}_c$ are learnable kernel parameters. The convolution operation integrates the conditional intensities while preserving the translational characteristics of the embedding. Then, the resulting output vectors $\mathbf{\Gamma}_c\left(\mathbf{e}_s^t, \mathbf{r}_t, \lambda_s(t)\right) = [m_0, m_1, \ldots, m_{d-1}]$ from each convolution kernel are aggregated into a matrix $\mathbf{M}_c \in \mathbb{R}^{k \times d}$. Ultimately, the Intensity-ConvTransE outputs the scores of entities by

$$\psi\left(\mathbf{e}_s^t, \mathbf{r}_t, \lambda_s(t)\right) = \sigma\left(\mathbf{W}_\varphi \cdot \mathcal{F}_{vec}\left(\mathbf{M}_c\right) + b_c\right)\mathbf{E}_t^o \tag{9}$$

where σ is an activation function, $\mathcal{F}_{vec}$ is a feature map function, $\mathbf{W}_\varphi$ is a learnable weight parameter for linear transformation, b_c is the basis, and $\mathbf{E}_t^o$ is the set of object embeddings. Intensity-ConvTransR calculates the scores similarly, with the distinction that it substitutes $\mathbf{r}_t$ with $\mathbf{e}_o^t$.

4.4 Dual-level Dynamics

To enhance our modeling of the TKG's evolution, we have developed a sophisticated dual-level joint training optimization strategy, known as fact-level dynamics and node-level dynamics.

- Fact-level dynamics: This level of dynamics primarily concerns the conditional intensity of facts and extrapolation reasoning. It places paramount importance on capturing the nuanced individual characteristics of facts, encompassing factors such as conditional intensities, entity and relation predictions.
- Node-level dynamics: In contrast, the node-level dynamics pivot towards the collective features of nodes in TKGs. It predominantly takes into account the network's topology, especially the number of links formed by nodes. This perspective allows us to capture the inherent tendencies of facts to connect with one another, grounded in the collective behaviors of nodes.

4.4.1 Fact-level Dynamics

Given a fact $(s, r, o, t) \in F_t$, we anticipate a higher conditional intensity $\lambda_{s,o}(t)$. Conversely, for the fact $(s, r, o, t) \notin F_t$ that did not occur, a lower conditional intensity is expected. To achieve this, we define a loss function for facts using negative log-likelihood. This optimization strategy aims to align the conditional intensity with the actual occurrence or non-occurrence of a fact. Specifically, the loss for an observed fact is defined as follows:

$$L_i(s, o, t) = -\log(\lambda_{s,o}(t)) - Q \cdot \mathbb{E}_{k \sim P_n} \log(1 - \lambda_{s,k}(t)) \tag{10}$$

where k is a negative node we sampled from a distribution P_n, ensuring that the fact $(s, r, k, t) \notin F_t$ has never occurred before. Additionally, Q represents the number of negative samples for each

positive fact. Typically, P_n is defined based on the node's degree, i.e., $P_n(v) \propto \deg(v)^{\frac{3}{4}}$.

Besides, continuing in the vein of TiRGN (Li et al., 2022a), we adopt a similar approach by treating entity and relation predictions as multi-label learning problems. Formulaically, the entity prediction loss L_e and the relation prediction loss L_r are as follows:

$$L_e(s,r,o,t) = \sum_{(s,r,o,t)\in F_t} \mathfrak{m}_t^e \log \mathbf{p}\left(o \mid s,r,t\right) \tag{11}$$

$$L_r(s,r,o,t) = \sum_{(s,r,o,t)\in F_t} \mathfrak{m}_t^r \log \mathbf{p}\left(r \mid s,o,t\right) \tag{12}$$

where $\mathbf{p}\left(o \mid s,r,t\right)$ and $\mathbf{p}\left(r \mid s,o,t\right)$ are the final probabilistic scores of entity and relation predictions by softmax function with Eq. (9). $\mathfrak{m}_t^e$ and $\mathfrak{m}_t^r$ are the mask vectors, where the element is 1 if the fact $(s,r,o,t) \in F_t$, otherwise 0.

4.4.2 Node-level Dynamics

While facts can be individual, they do not occur in isolation. In particular, links are formed to connect nodes, indicating that their behavior is jointly influenced by the nodes they have in common. Therefore, we propose to manage the collective characteristics of nodes at the node level, capturing their "fact tendencies". In essence, different nodes exhibit varying tendencies to form new links with other nodes, and even the same node may display different tendencies at different times.

To be more precise, we can quantify the node dynamics of a node at time t by considering the number of new facts occurring on that node at time t, denoted as $N_s(t)$. Thus, we create a node dynamics estimator to predict the number of new facts on a specific node:

$$\hat{N}_s(t) = W_n \cdot \mathrm{e}_s^t + b_n \tag{13}$$

where W_n is a linear transformation matrix, b_n is the basis, e_s^t is the temporal representation of node s, and $\hat{N}_s(t)$ is the predicted number of new facts that occur on node s at time t. To ensure that the occurrence of facts aligns with the dynamics of continuously evolving nodes in TKGs, we devise the following node loss:

$$L_n(s,t) = \begin{cases} \vartheta \left|\hat{N}_s(t) - N_s(t)\right|^2, & |\hat{N}_s(t) - N_s(t)| < 1 \\ |\hat{N}_s(t) - N_s(t)| - \vartheta, & \text{otherwise} \end{cases} \tag{14}$$

where ϑ is the average degree of all nodes. The node loss aims to ensure that the estimator $\hat{N}_s(t)$ accurately fits the true dynamics $N_s(t)$ of all nodes and times.

Ultimately, we perform joint training of fact-level dynamics and node-level dynamics using the following comprehensive loss function:

$$\arg\min_{\Theta} \sum_{(s,r,o,t)\in F_t} \eta_1(L_i + L_e + L_r) + \eta_2 L_n \tag{15}$$

where Θ denotes the parameters of HTCCN. The hyperparameters η_1 and η_2 play a crucial role in controlling the relative contribution of fact-level dynamics and node-level dynamics, respectively.

5 Experiments

5.1 Experimental Setup

5.1.1 Datasets.

In our experiments, we utilized four extensively recognized real-world TKG datasets, i.e., ICEWS14 (Han et al., 2020a), ICEWS05-15 (Garcia-Duran et al., 2018), ICEWS18 (Ward et al., 2013), and GDELT (Zhang et al., 2022).

5.1.2 Evaluation Setting and Metrics.

To evaluate the performance of our proposed model and facilitate fair comparisons, we employ the well-established time-aware filtered settings (Jin et al., 2020; Zhu et al., 2021; He et al., 2021) and leverage two widely recognized evaluation metrics, i.e., Mean Reciprocal Rank (MRR) and Hits at N (H@N).

5.1.3 Baselines.

In our comprehensive evaluation, we systematically compare our newly proposed HTCCN model against a diverse array of state-of-the-art TKG models, including CyGNet (Zhu et al., 2021), TANGO (Han et al., 2021), RE-GCN (Li et al., 2021b), TITer (Sun et al., 2021), CEN (Li et al., 2022b), TiRGN (Li et al., 2022a), TLogic (Liu et al., 2022b), EvoKG (Gao et al., 2022), GHT (Sun et al., 2022), RPC (Liang et al., 2023), TiPNN (Dong et al., 2023b), DaeMon (Dong et al., 2023a), CENET (Xu et al., 2023), TECHS (Lin et al., 2023) and RETIA (Liu et al., 2023).

5.1.4 Model Configurations.

For all datasets, we maintain an embedding size of 200, aligning with the baseline method as established in (Zhu et al., 2021). We utilize 2 temporal causal convolutional layers with a dilation length of 2. Dropout is set to 0.2 to prevent overfitting. Regarding the intensity-guided decoder, we

Table 1: Performance (in percentage) for entity prediction task under the time-aware filtered setting.

Method	ICEWS14			ICEWS05-15			ICEWS18			GDELT		
	MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3
CyGNet	37.65	27.43	42.63	40.42	29.44	46.06	27.12	17.21	30.97	22.22	12.35	21.66
TANGO	36.81	27.32	40.86	42.86	32.72	48.14	28.97	19.51	32.61	19.66	12.50	20.93
RE-GCN	42.00	31.63	47.20	48.03	37.33	53.90	32.62	22.39	36.79	19.69	12.46	20.93
TITer	41.73	32.74	46.46	47.60	38.29	52.74	29.98	22.05	33.46	18.19	11.52	19.20
CEN	42.20	32.08	47.46	-	-	-	31.50	21.70	35.44	-	-	-
TiRGN	43.81	33.49	48.90	49.84	39.07	55.75	33.58	23.10	37.90	21.67	13.63	23.27
TLogic	43.04	33.56	48.27	46.97	36.21	53.13	29.82	20.54	33.95	19.80	12.20	21.70
EvoKG	27.18	-	30.84	-	-	-	29.28	-	33.94	19.28	-	20.55
GHT	37.40	27.77	41.66	41.50	30.79	46.85	27.40	18.08	30.76	20.04	12.68	21.37
RPC	44.55	34.87	49.80	51.14	39.47	57.11	34.91	24.34	38.74	22.41	14.42	24.36
TiPNN	-	-	-	-	-	-	32.17	22.74	36.24	21.17	14.03	22.98
DaeMon	-	-	-	-	-	-	31.85	22.67	35.92	20.73	13.65	22.53
CENET	41.30	32.58	-	47.13	37.25	-	29.65	19.98	-	19.73	12.04	-
TECHS	43.88	34.59	49.36	48.38	38.34	54.69	30.85	21.81	35.39	-	-	-
RETIA	42.77	32.29	47.78	47.27	36.65	52.91	32.43	22.24	36.48	19.73	12.54	21.01
HTCCN	**45.39**	**36.58**	**50.84**	**51.94**	**40.32**	**57.79**	**35.63**	**24.90**	**39.26**	**23.46**	**15.18**	**25.21**

use 50 channels and a kernel size of 4*3. Model parameters are optimized using the Adam optimizer (Kingma and Ba, 2015) with a learning rate of 0.001. The training epoch is set to 20, a duration deemed adequate for achieving convergence in most scenarios. All experimentation is conducted on a GeForce GTX 3080 Ti. The baseline results are sourced from previous papers (Li et al., 2022a; Liang et al., 2023; Zhang et al., 2023a). The size of the learnable parameters can be found in Appendix A.

5.2 Performance Comparison

The reasoning performance of HTCCN with baselines is presented in Table 1. Across the four benchmark datasets, HTCCN consistently demonstrates superior performance compared to all 15 state-of-the-art baselines spanning from 2021 to 2023.

Specifically, compared to the second-best performance, HTCCN achieves an average improvement of 2.45% in MRR, 2.39% in Hits@1, and 2.23% in Hits@3. Notably, HTCCN exhibits particularly significant improvements in Hits@1, with a remarkable increase of approximately 5.27% and 2.64% on the GDELT and ICEWS14 datasets, respectively. These results underscore the ability of HTCCN to reason future facts under the extrapolation setting. The training and testing time overhead of the model can be found in Appendix B.

Most baselines like RPC, TiRGN, RGCRN, RENET, TANGO, xERTE, and RE-GCN, employ RNN to consider adjacent timestamp facts and perform well in the experiments. However, HTCCN leverages a temporal dilated causal convolutional network and a dual-level dynamic modeling mechanism to capture a broader range of structural fact features and collective features of fact nodes. Consequently, HTCCN surpasses these approaches and effectively captures fact evolution.

Compared to GHT gets the answer entity directly comparing intensity values, HTCCN computes entity scores via Intensity-ConvTransE, which integrates the conditional intensities while preserving the translational characteristics of the embedding, making it more suitable for extrapolated reasoning. More comparative analysis with GHT can be found in Appendix C.

However, as faced with datasets containing numerous timestamps, the reasoning performance of TITer declines due to challenges in finding suitable entities within a large search space, as observed on the GDELT dataset. In contrast, HTCCN excels in reasoning about a set of candidate entities with high conditional intensity, effectively avoiding concerns regarding time and space complexities. Consequently, HTCCN demonstrates superiority across datasets with varying numbers of timestamps.

5.3 Ablation study on different modules

The ablation experiments aimed at investigating model effectiveness are conducted on all four TKG datasets, employing the MRR evaluation metric. The primary focus is to dissect the efficacy of three pivotal design modules, i.e., temporal causal convolutional networks (TCCN), Hawkes processes, and intensity-guided decoder. The experiment results are presented in Fig. 3, showcasing a comparative assessment across five sub-models, i.e.,

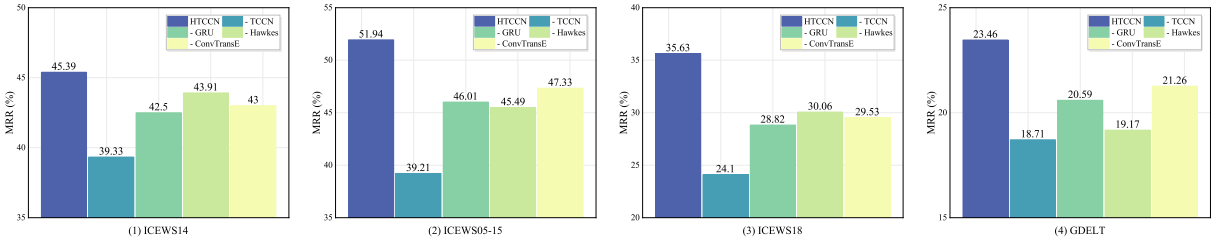

Figure 3: Effect of main modules. By systematically removing the three primary modules, namely, TCCN, Hawkes, and the intensity-guided decoder, the reasoning performance of HTCCN exhibits varying degrees of deterioration. This outcome underscores the efficacy of the three meticulously designed modules in effectively modeling the evolution of facts.

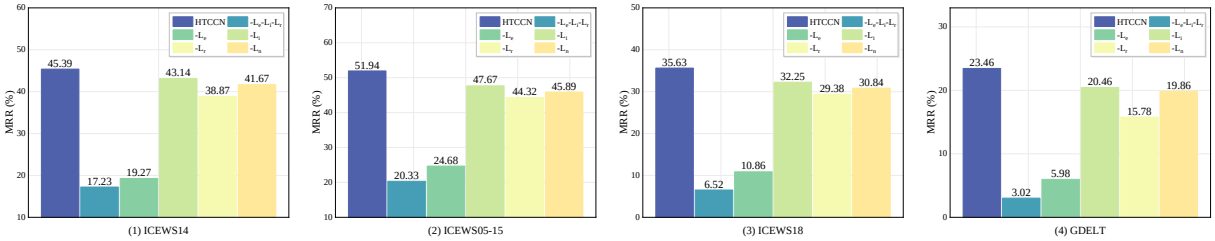

Figure 4: Effect of dual-level dynamics. The fact-level dynamics, encompassing fact-conditional intensity, entity prediction loss, and relation prediction loss, are instrumental in capturing the individual attributes of facts, while the node-level dynamics focus on the collective features of nodes. A noteworthy observation emerges when we examine the impact of removing fact-level dynamics and node-level dynamics independently.

(1) the original HTCCN model, (2) HTCCN with TCCN removed, denoted as "- TCCN", (3) HTCCN with TCCN replaced by GRU, denoted as "- GRU", (4) HTCCN with the Hawkes process removed, denoted as "- Hawkes", (5) HTCCN with the intensity-ConvTransE replaced by ConvTransE, denoted as "-ConvTransE". More module ablation experiments can be found in Appendix D.

On average, comparing the original HTCCN to the ablated models, the MRR metric experiences significant decreases of 3.2%, 7.43%, 7.50%, and 3.53% on the four TKGs, respectively. In particular, comparing "- TCCN" and "HTCCN", we reveal a significant enhancement attributed to TCCN, substantiating its effectiveness in TKG reasoning. Furthermore, the outcomes of "HTCCN" versus "-GRU" emphasize that TCCN captures more fact evolution features compared to traditional recurrent neural networks. The insight gained from the "- Hawkes" comparison underscores that the Hawkes process effectively models the fact evolution process on TKGs, enabling more accurate reasoning about potential future facts by capturing the impact (i.e., conditional intensity) of historical facts on future facts. Moreover, the analysis of "-ConvTransE" highlights that the intensity-aware decoder yields a notable enhancement in extrapolation reasoning compared to the commonly used ConvTransE.

In light of these promising results and the preceding analysis, we can assert that our proposed TCCN, Hawkes process, and intensity-aware decoder efficiently capture causal and evolutionary fact features, significantly enhancing the reasoning performance in TKGs.

5.4 Ablation study on dual-level dynamics

To gain deeper insights into the efficacy of our dual-level dynamics in capturing features at distinct granularities, we conduct a series of ablation studies to scrutinize the model's reliability. Fig. 4 succinctly portrays the results of these ablations. Notably, the fact-level dynamics (-L_e-L_i-L_r) exhibit the most substantial impact on performance, underscoring the pivotal role of individual fact features in temporal reasoning. Simultaneously, the node-level dynamics (-L_n) display a noticeable influence on reasoning, affirming the importance of capturing fact trends. Furthermore, the removal of fact-conditional intensity (-L_i) leads to a performance decline, reaffirming the potency of leveraging Hawkes to process temporal graphs effectively. In the context of entity prediction, the entity prediction loss (-L_e) exerts a more pronounced influence on performance compared to the relation prediction loss (-L_r). This can be attributed to the con-

siderably larger number of entities in all datasets compared to the number of relations. Thus, these findings provide compelling evidence that both fact-level dynamics and node-level dynamics significantly contribute to effective temporal reasoning.

6 Conclusion

In this paper, we propose HTCCN, a reasoning model tailored for TKGs under the extrapolation setting. HTCCN leverages Hawkes processes to effectively model the link formation process inductively within TKGs. Specifically, we introduce Temporal Causal Convolutional Networks based on Hawkes processes, enabling HTCCN to discern the significance of concurrent facts in reasoning about future facts. Additionally, we integrate node-level dynamics and fact-level dynamics, aiming to capture both collective features of nodes and individual features of facts, thereby achieving a more comprehensive understanding of the temporal evolution within TKGs. Experimental results on four benchmark datasets unequivocally demonstrate the notable advantages and effectiveness of HTCCN in temporal entity and relationship prediction. Furthermore, ablation experiments underscore the active roles played by fact-level dynamics and node-level dynamics in TKG inference.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. U2003208, No. 62172451, and No. 62306287), and the Zhejiang Provincial Natural Science Foundation of China (No. LY23F020012).

References

Luyi Bai, Mingcheng Zhang, Han Zhang, and Heng Zhang. 2023. Ftmf: Few-shot temporal knowledge graph completion based on meta-optimization and fault-tolerant mechanism. *World Wide Web*, 26(3):1243–1270.

Yitian Chen, Yanfei Kang, Yixiong Chen, and Zizhuo Wang. 2020. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399:491–501.

Hai Cui, Tao Peng, Feng Xiao, Jiayu Han, Ridong Han, and Lu Liu. 2023. Incorporating anticipation embedding into reinforcement learning framework for multi-hop knowledge graph question answering. *Information Sciences*, 619:745–761.

Hao Dong, Zhiyuan Ning, Pengyang Wang, Ziyue Qiao, Pengfei Wang, Yuanchun Zhou, and Yanjie Fu. 2023a. Adaptive path-memory network for temporal knowledge graph reasoning. *arXiv preprint arXiv:2304.12604*.

Hao Dong, Pengyang Wang, Meng Xiao, Zhiyuan Ning, Pengfei Wang, and Yuanchun Zhou. 2023b. Temporal inductive path neural network for temporal knowledge graph reasoning. *arXiv preprint arXiv:2309.03251*.

Huilian Fan, Yuanchang Zhong, Guangpu Zeng, and Chenhao Ge. 2022. Improving recommender system via knowledge graph based exploring user preference. *Applied Intelligence*, pages 1–13.

Yifu Gao, Linhui Feng, Zhigang Kan, Yi Han, Linbo Qiao, and Dongsheng Li. 2022. Modeling precursors for temporal knowledge graph reasoning via auto-encoder structure. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2044–2051.

Alberto Garcia-Duran, Sebastijan Dumančić, and Mathias Niepert. 2018. Learning sequence encoders for temporal knowledge graph completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4816–4821.

Manas Gaur, Kalpa Gunaratna, Vijay Srinivasan, and Hongxia Jin. 2022. Iseeq: Information seeking question generation using dynamic meta-information retrieval and knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10672–10680.

Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2020a. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In *International Conference on Learning Representations*.

Zhen Han, Zifeng Ding, Yunpu Ma, Yujia Gu, and Volker Tresp. 2021. Learning neural ordinary equations for forecasting future links on temporal knowledge graphs. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 8352–8364.

Zhen Han, Yunpu Ma, Yuyi Wang, Stephan Günnemann, and Volker Tresp. 2020b. Graph hawkes neural network for forecasting on temporal knowledge graphs. *Automated Knowledge Base Construction*, page 20.

Yongquan He, Peng Zhang, Luchen Liu, Qi Liang, Wenyuan Zhang, and Chuang Zhang. 2021. Hip network: Historical information passing network for extrapolation reasoning on temporal knowledge graph. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, (IJCAI)*, pages 1915–1921.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2020. Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In *Proceedings of the 2020 Conference on Empirical*

Methods in Natural Language Processing (EMNLP), pages 6669–6683.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (Poster)*, San Diego, California, USA. Opneview.

Yujia Li, Shiliang Sun, and Jing Zhao. 2022a. Tirgn: time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 2152–2158. ijcai. org.

Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022b. Complex evolutional pattern learning for temporal knowledge graph reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 290–296.

Zixuan Li, Xiaolong Jin, Saiping Guan, Wei Li, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2021a. Search from history and reason for future: Two-stage reasoning on temporal knowledge graphs. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4732–4743.

Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021b. Temporal knowledge graph reasoning based on evolutional representation learning. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 408–417.

Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, and Xinwang Liu. 2023. Learn from relational correlations and periodic events for temporal knowledge graph reasoning. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1559–1568.

Qika Lin, Jun Liu, Rui Mao, Fangzhi Xu, and Erik Cambria. 2023. Techs: Temporal logical graph networks for explainable extrapolation reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1281–1293.

Kangzheng Liu, Feng Zhao, Hongxu Chen, Yicong Li, Guandong Xu, and Hai Jin. 2022a. Da-net: Distributed attention network for temporal knowledge graph reasoning. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1289–1298.

Kangzheng Liu, Feng Zhao, Guandong Xu, Xianzhi Wang, and Hai Jin. 2023. Retia: relation-entity twin-interact aggregation for temporal knowledge graph extrapolation. In *IEEE International Conference on Data Engineering*. IEEE.

Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022b. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4120–4127.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer.

Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3060–3067.

Haohai Sun, Shangyi Geng, Jialun Zhong, Han Hu, and Kun He. 2022. Graph hawkes transformer for extrapolated reasoning on temporal knowledge graphs. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7481–7493.

Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. 2021. Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8306–8319.

Jiapu Wang, Boyue Wang, Meikang Qiu, Shirui Pan, Bo Xiong, Heng Liu, Linhao Luo, Tengfei Liu, Yongli Hu, Baocai Yin, et al. 2023. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects. *arXiv preprint arXiv:2308.02457*.

Michael D Ward, Andreas Beger, Josh Cutler, Matthew Dickenson, Cassy Dorff, and Ben Radford. 2013. Comparing gdelt and icews event data. *Analysis*, 21(1):267–297.

Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4765–4773.

Fuwei Zhang, Zhao Zhang, Xiang Ao, Fuzhen Zhuang, Yongjun Xu, and Qing He. 2022. Along the time: Timeline-traced embedding for temporal knowledge graph completion. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2529–2538.

Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023a. Learning latent relations for temporal knowledge graph reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12617–12631.

Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023b. Learning long-and short-term representations for temporal knowledge graph reasoning. In *Proceedings of the ACM Web Conference 2023*, pages 2412–2422.

Cunchao Zhu, Muhao Chen, Changjun Fan, Guangquan Cheng, and Yan Zhang. 2021. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4732–4740.

A Model size

To provide a comprehensive evaluation, we conduct comparative analyses of HTCCN against the latest methods, including GHT, TiRGN, TECHS, and RETIA. These analyses encompass a range of aspects such as the number of parameters, parameter size, training time, and testing time, as shown in Table 2.

Table 2: Comparison of learnable parameters on ICEWS14.

Model	**param_sum**	**param_size (MB)**
GHT	4744424	18.099
TiRGN	13372441	51.012
TECHS	915206	3.491
RETIA	8242772	31.444
HTCCN	25108392	72.075

Although HTCCN has a relatively higher number of learnable parameters, we find that the overall parameter size is within acceptable limits considering modern computer storage capabilities.

B Model training and testing overhead

We compare the overheads of HTCCN, GHT, RETIA, and TECHS in terms of training and testing time, as shown in Table 3.

Table 3: Time overheads on ICEWS14.

Model	Train (min)	Test (min)
HTCCN	5:56	1:03
GHT	25:02	5:18
RETIA	45:52	4:35
TECHS	16:40	5:04

C Comparison with GHT

we conduct additional experiments with GHT using two different embedding dimensions, i.e., 100 and 200. These experiments are detailed in Table 4, labeled as GHT-100 and GHT-200, respectively.

Table 4: Comparison of GHT under 100 and 200 embedding dimensions on ICEWS14.

Model	MRR	H@1	H@3	H@10
HTCCN	45.39	36.58	50.84	66.07
GHT-200	38.90	28.26	44.06	59.72
GHT-100	37.40	27.77	41.66	56.19

Our findings indicate that while increasing the embedding dimension in GHT (to 200) does enrich the model with more semantic information, it does not lead to a substantial improvement in the model's reasoning performance. Contrastingly, our HTCCN model, with an embedding size of 200, demonstrates significantly superior performance compared to both versions of GHT.

D Module ablations

Furthermore, we investigate the compounded effects of simultaneously removing multiple components, the findings of which are presented in Table 5.

Table 5: Ablations with different components on ICEWS14.

Method	**MRR**	**H@1**	**H@3**
HTCCN	45.39	36.58	50.84
- TCCN - Hawkes	36.17	25.83	42.69
- TCCN - ConvTransE	35.84	24.15	41.02
- GRU - Hawkes	40.60	29.52	46.71
- GRU - ConvTransE	39.32	28.93	45.48
- Hawkes - ConvTransE	41.04	30.79	47.66

This multi-component ablation study reveals that the collective removal of components leads to a more pronounced decrease in evaluation metrics than the removal of individual components. This underscores each component's unique and indispensable role in HTCCN, thereby confirming the synergy between them.