CEGRL-TKGR: A Causal Enhanced Graph Representation Learning Framework for Temporal Knowledge Graph Reasoning

Jinze Sun[1], Yongpan Sheng[1,2](✉), Lirong He[3], Yongbin Qin[4], Ming Liu[5], Tao Jia[1]

[1]College of Computer and Information Science, Southwest University, Chongqing 400715, China
[2]School of Big Data & Software Engineering, Chongqing University, Chongqing 401331, China
[3]School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[4]College of Computer Science and Technology, The State Key Laboratory of Public Big Data, Guizhou University
[5]Deakin University, 3125, Victoria, Australia

ssunjinze@outlook.com   {shengyp2011,ronghe1217}@gmail.com   ybqin@gzu.edu.cn   m.liu@deakin.edu.au   tjia@swu.edu.cn

Abstract

Temporal knowledge graph reasoning (TKGR) is increasingly gaining attention for its ability to extrapolate new events from historical data, thereby enriching the inherently incomplete temporal knowledge graphs. Existing graph-based representation learning frameworks have made significant strides in developing evolving representations for both entities and relational embeddings. Despite these achievements, there's a notable tendency in these models to inadvertently learn biased data representations and mine spurious correlations, consequently failing to discern the causal relationships between events. This often leads to incorrect predictions based on these false correlations. To address this, we propose an innovative **C**ausal **E**nhanced **G**raph **R**epresentation **L**earning framework for TKGR (named CEGRL-TKGR). This framework introduces causal structures in graph-based representation learning to unveil the essential causal relationships between events, ultimately enhancing the performance of the TKGR task. Specifically, we first disentangle the evolutionary representations of entities and relations in a temporal knowledge graph sequence into two distinct components, namely causal representations and confounding representations. Then, drawing on causal intervention theory, we advocate the utilization of causal representations for predictions, aiming to mitigate the effects of erroneous correlations caused by confounding features, thus achieving more robust and accurate predictions. Finally, extensive experimental results on six benchmark datasets demonstrate the superior performance of our model in the link prediction task.

1 Introduction

Knowledge graphs (KGs) have gained significant promise in natural language processing or knowledge engineering perception tasks (Chen et al., 2022a). They model real-world factual knowledge using multi-relationship graph structures. However, factual knowledge in reality is constantly evolving, resulting in the form of event knowledge. This has led to the development and application of temporal knowledge graphs (TKGs). TKG encodes the relationship information of entities and events and their timing for capturing the dynamics of entities and their relationships over time (Gastinger et al., 2022). Thus, analyzing the TKG provides a comprehensive understanding of the evolving events, based on which various time-dependent applications have been developed, including time-sensitive semantic search (Barbosa et al., 2013), policy making (Deng et al., 2020), stock forecasting (Feng et al., 2019), and more (Chen et al., 2022a).

The reliability of applications depends on accurate predicting, which highly relies on data integrality. However, existing TKGs are inevitably incomplete due to the partial observation of real-world (Liang et al., 2022). To address this limitation and enhance the representation capability of the TKG, temporal knowledge graph reasoning (TKGR) models are proposed and aim to extrapolate new facts and relationships in the TKG according to their historical temporal information. Existing models explore different strategies to achieve satisfactory results on the TKGR task. GHNN (Han et al., 2020) and GHT (Sun et al., 2022) model historical facts as point-in-time processes. TKGR-RHETNE (Sun et al., 2023) jointly models the relevant historical event and temporal neighborhood event context of events in the TKG. RE-NET (Jin et al., 2020) and RE-GCN (Li et al., 2021) introduce graph neural networks (GNN) into sequence models to capture structural and temporal dependencies between entities. TKGR-GPRSCL (Xiong et al., 2024) captures complex structure-aware information by encoding paths across entities and obtaining temporal correlations in the complex plane. TLogic (Liu et al., 2022) and TITer (Sun et al., 2021) design

interpretable models based on logical rules and reinforcement learning, respectively. Despite the achievements of previous studies, they have overlooked the reality that there are numerous confounding factors in the TKG, such as shallow patterns and noisy links. However, these confounding factors commonly misguide the reasoning process in the TKG, resulting in the acquisition of incorrect dependencies and the generation of non-causal predictions (Sui et al., 2022).

To address the aforementioned issues, we advocate for the integration of causal theory into TKGR to guide learning of the essential causal relationships between events and mitigate the impact of confounding factors on the TKGR task. Specifically, we first construct a structural causal model (Zečević et al., 2021) to comprehensively analyze and model the TKGR task from a causal perspective. Then, based on the causal model, we propose a new framework, namely **C**ausal **E**nhanced **G**raph **R**epresentation **L**earning (CEGRL-TKGR), to disentangle confounding factors from the essential causal factors in the TKG. *To the best of our knowledge, this is the first study to incorporate causal intervention in a graph representation learning framework for learning the evolutionary representations of entities and relations in the TKG.* To conclude, our contributions in this paper are 3-folds:

- We propose a novel **C**ausal **E**nhanced **G**raph **R**epresentation **L**earning framework for **T**emporal **K**nowledge **G**raph **R**easoning, called CEGRL-TKGR, to uncover the essential causal relationships between events and mitigate the impact of confounding factors.

- The proposed CEGRL-TKGR framework disentangles the evolutionary representations of entities and relations into causal and confounding representations. Then, it applies causal interventions to perform backdoor adjustments of representations, prioritizing predicted causal features while minimizing the impact of spurious correlations introduced by confounding features.

- Comprehensive experimental results demonstrate that CEGRL-TKGR outperforms state-of-the-art baselines on six real-world datasets in the link prediction task. Further, comprehensive studies confirm the contribution of

the introduced causal structures and interventions[1].

2 Related Work

2.1 Temporal Knowledge Graph Reasoning

TKGR in extrapolation settings focuses on predicting new facts about the future based on historical events. Specifically, CyGNet (Zhu et al., 2021) uses a copy-generating mechanism to capture the global repetition rate of facts. GHNN (Han et al., 2020) and GHT (Sun et al., 2022) construct a temporal point process (TPP) to capture the temporal dynamics of successive events, predicting future facts by estimating the conditional probability of the TPP. In recent years, with the successful application of GNN in many dynamic scenarios (Zhang et al., 2022), they have also been introduced into structural-semantic dependency models in TKGR. RE-NET (Jin et al., 2020) used a neighborhood aggregator and cyclic event encoder to model historical facts as subgraph sequences. RE-GCN (Li et al., 2021) uses RGCN (Schlichtkrull et al., 2018) to learn evolutionary representations of entities and relationships at each timestamp. CEN (Li et al., 2022a) uses length-aware convolutional neural networks (CNNS) to process evolutionary patterns of different lengths. There are also some studies to solve the TKGR problem through path search. For example, TLogic (Liu et al., 2022) completes link prediction tasks based on temporal logic rules learned from temporal knowledge graphs. TITer (Sun et al., 2021) proposes a TKG prediction model based on reinforcement learning, which uses time-shaped rewards based on Dirichlet distribution to guide model training. All of the methods discussed above have limitations in modeling entity and relationship representations, in particular ignoring cause-and-effect relationships between different entities, which we believe is key to making correct predictions.

2.2 Causal Representation Learning

In graph causal representation learning, researchers have explored various methods to improve the explanatory power and generalization performance of GNNs. By applying the principles of causal reasoning to graph-structured data, the researchers

---

[1]To illustrate the evaluation of our CEGRL-TKGR framework and facilitate further research on this topic, we have made the experimental details and source code of the framework publicly available at https://github.com/shengyp/CEGRL-TKGR.

sought to address the challenges GNNs face when dealing with complex systems such as social networks, molecular maps, and syntax trees of program code. DIR (Wu et al., 2022) is proposed to reveal the intrinsic interpretability of GNNs by discovering invariant reasons, which involves splitting input graphs into causal and non-causal fruit graphs and training the two classifiers through invariant risk loss functions. GOOD (Chen et al., 2022b) improves the cross-domain generalization of graphs by distinguishing invariant subgraphs from other parts of graphs that are susceptible to domain transfer. CAL (Sui et al., 2022) introduces de-confounding training to distinguish the key and secondary parts of the graph and eliminate the confounding effect of the secondary parts on model prediction. CFLP (Zhao et al., 2022) points out that the causal relationship between graph structure and link presence is often ignored, and proposed to generate counterfactual links to enhance training data and reduce reliance on false associations. Zevcevic *et al.* (Zečević et al., 2021) theoretically analyze the relationship between GNNs and structural causal models (SCMs) and design a new class of neuro-causal models. However, none of the work has been done to combine causal learning with the TKGR task.

3 Preliminary

3.1 Notations and Task Formulation

A TKG $\mathcal{G}$ can be formalized as a sequence of knowledge graph slices $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$, where $\mathcal{G}_t = \{(e_s, r, e_o, t) \in \mathcal{G}\}$ denotes a knowledge graph slice that consists of facts that occurred at the timestamp t range from t_0 to t_n. Here, e_s and e_o represent the subject and object entities, respectively, and r denotes the predicate as a relation type. Besides, $\mathbf{e}_s$, $\mathbf{r}$, $\mathbf{e}_o$ written in bold represent their embeddings. The objective of TKGR task is to predict either the subject in a give query $(?, r, e_o, t)$ or the object in a given query $(e_s, r, ?, t)$ with $t > t_n$.

3.2 A Causal Perspective on the GNN-Based TKGR Task

3.2.1 GNN-based TKGR Paradigm

Inspired by previous GNN-based modeling in a casual look (Didelez and Pigeot, 2001; Sui et al., 2022), we abstract the GNN-based TKGR process through a structural causal figure, as shown in Fig. 1, encompassing five distinct variables. The connectivity from one variable to another epito-
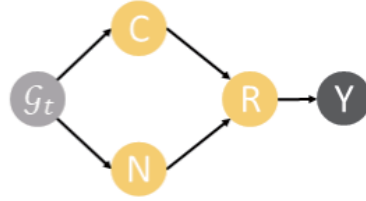

Figure 1: The GNN-based structural causal graph for the TKGR task.

mizes the causal relationship, delineated as the cause $\rightarrow$ effect. The variables are described as follows:

- Graph data $\mathcal{G}_t$: The knowledge graph at each timestamp t, manifests as a directed multi-relationship figure.

- Causal Feature C: These features epitomize the causal essence of the targeted entity, providing a fundamental understanding of its inherent dynamics.

- Confounding Feature N: These features, discerned from GNN, embody the confounding attributes, unveiling the potential biases or trivial patterns ingrained in graph-based learning methodologies.

- Representation R: These representations are the entity and relational representations of the output of the final GNN layer after learning for $\mathcal{G}_t$.

- Prediction Y: Denoted as TKGR as the link prediction, this aspect transitions through the decoder, rendering the ultimate reasoning based on the preceding representation.

The causal embedding encapsulates the causal features C, authentically mirroring the implicit knowledge inherent in the knowledge graph $\mathcal{G}_t$. Conversely, N symbolizes the confounding features, which might be spawned by data biases, data noise, or superficial patterns within graph-based learning methodologies. These confounding features forge a backdoor pathway between C and Y, fostering spurious correlations that don't contribute to accurate reasoning. Functionally, the structural operation denoted by $C \rightarrow R \leftarrow N$ portrays a GNN, wherein both the causal features C and the confounding features N, as discerned by the target entity from the graph data, exert a direct impact on the output R of the GNN. Subsequently,

the output R of GNN directly sways the model inference outcome, illustrated as $R \rightarrow Y$.

In the graph-based TKGR paradigm, causal and confounding features are not decoupled for each entity or relationship embedding. Using causal graphs, we aspire to explicitly separate causal embeddings and confounding embeddings from entity or relational representations, and aim to mitigate the effects of confounding features by performing causal interventions. This endeavor not only clarifies the inference process but also endeavors to refine the accuracy and reliability of the GNN-based TKGR mechanism.

3.3 Causal Intervention Strategies

Beyond fostering a novel comprehension of GNN-based TKGR, causal theory avails analytical instruments predicated on causal figures, such as causal intervention. Causal intervention facilitates a profound examination of the factors precipitating inference outcomes. As delineated by Fig. 1., confounding feature N and causal feature C can be discerned from the knowledge graph $\mathcal{G}_t$. These features are contemplated in the representation R of entities and relations, thereby establishing a backdoor pathway represented as $N \leftarrow \mathcal{G}_t \rightarrow C \rightarrow R \rightarrow Y$, with N serving as the quick bridge between C and Y.

To orchestrate a causal prognosis hinging on the causal feature C, it necessitates the modeling of $P(Y \mid C)$. However, the backdoor path distorts the probability distribution $P(Y \mid C)$ through the confounding effect of N, thereby necessitating the disentanglement of the backdoor pathway from N to Y. It is imperative to stymie this backdoor pathway to mitigate the repercussions of the hybrid embedding, thereby enabling the model to reason robustly by leveraging the causal feature to the fullest. Causality theory is a potent toolkit to address this backdoor path dilemma.

We engage the do-calculus for executing causal interventions on variable C, intending to sever the backdoor path $N \leftarrow \mathcal{G}_t \rightarrow C \rightarrow R \rightarrow$ Y. Our objective is to estimate $P(Y \mid do(C))$, as opposed to muddling it with $P(Y \mid C)$. By using Bayes' theorem with the causal postulation, we can extrapolate the ensuing expression:

$$P(Y \mid do(C)) = \sum_{n \in N} P(Y \mid C, n) P(n). \quad (1)$$

The equation above illustrates that to gauge the causal influence of C on Y, it's requisite to take into account the inference outcomes of both causal and confounding features. This can be perceived as re-coupling the disentanglement feature embeddings, utilizing them for deductive reasoning at future timestamps. However, C and N are usually unobservable, and it is difficult to obtain them directly at the data level, which makes the calculation of the Eq. (1) very challenging. In the next section, we discuss ways to overcome this problem.

4 The Proposed CEGRL-TKGR Framework

4.1 The Overall Architecture of TKGR-GPRSCL

We detail the CEGRL-TKGR framework for learning representations of entities and relationships based on causal features and confounding features. CEGRL-TKGR consists of three parts: (1) The representation learning part that learns the structure dependence in each $\mathcal{G}_t$; (2) The decoupling learning part that learns the entity and relation representations; (3) The decoder part that is designed based on the time interval. The overall architecture of the framework is shown in Fig. 2.

4.2 Entity and Relation Evolution Representation

Within each $\mathcal{G}_t$, representation learning of entities and relationships involves the aggregation of multiple relationships, as well as information from multiple hop neighbors under a single timestamp. Between adjacent $\mathcal{G}_t$, we expect to accurately capture the order dependencies inherent in the subgraph with different timestamps. Drawing inspiration from the RE-GCN model (Li et al., 2021), we employ the ω-layer RGCN, which hinges on structure modeling and a recurrent mechanism to progressively update the representations of entities and relations. This approach allows for a more nuanced understanding and modeling of the dynamic interactions within the graph over time.

$$\mathbf{e}_{o,t}^{l+1} = \text{RReLu} \left(\sum_{(e_s, r, e_o) \in \mathcal{G}_t} \frac{1}{d_{e_o}} \mathbf{W}_1^l \left(\Phi \left(\mathbf{e}_{s,t}^l, \mathbf{r}_t \right) \right) \right.$$
$$\left. + \mathbf{W}_2^l \mathbf{e}_{o,t}^l \right), \quad (2)$$

$$\mathbf{E}_t = \text{GRU} \left(\mathbf{E}_{t-1}, \mathbf{E}_t' \right). \quad (3)$$

In the Eq. (2), we describe how the embedding $\mathbf{e}_{o,t}^{l+1}$ of entity e_o at time step t and layer $l+1$ is computed. We integrate the information of all entities

(a) Entity and Relation Evolution Representation (b) Temporal Gap Guided Decoder
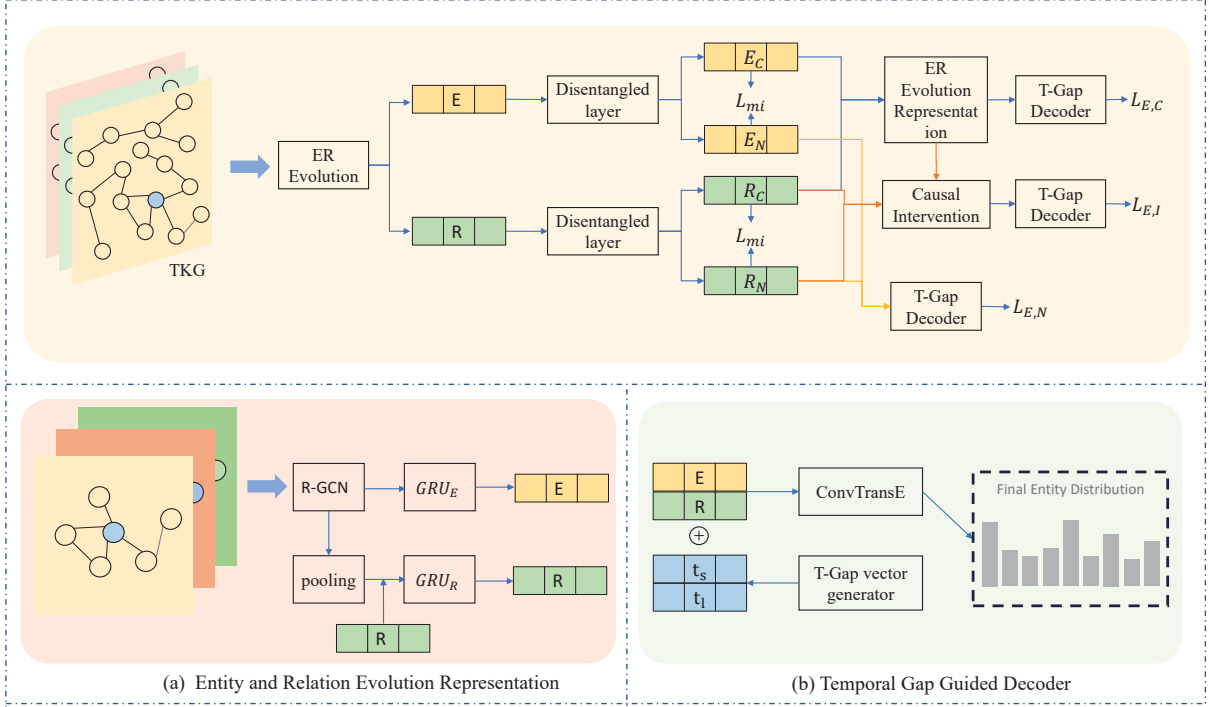
Figure 2: The overall architecture of our proposed CEGRL-TKGR framework.

and relations connected to entity e_o in the knowledge graph $\mathcal{G}_t$. $\mathbf{W}_1^l, \mathbf{W}_2^l$ is learnable weights and Φ has the option of addition or one-dimensional convolution. In the Eq. (3), we showcase how the entity embedding matrix $\mathbf{E}_t$ is updated via the GRU. Specifically, we take the entity embedding matrix $\mathbf{E}_{t-1}$ at the previous time step $t-1$ and the aggregated entity embedding matrix $\mathbf{E}_t'$ as inputs to obtain the entity embedding matrix $\mathbf{E}_t$ at the current time step t.

For relations, ensuring consistency with the entity embedding updates within the subgraph sequence is crucial. To achieve this consistency, a specialized GRU tailored for relations is employed for the update process. This mechanism facilitates a harmonized evolution of both entity and relation causal embeddings over the sequence of subgraphs:

$$\mathbf{r}_t' = \text{pooling}\left(\mathbf{E}_{t-1}, R_t\right) \oplus \mathbf{r}, \quad (4)$$
$$\mathbf{R}_t = \text{GRU}\left(\mathbf{R}_{t-1}, \mathbf{R}_t'\right), \quad (5)$$

where $\mathbf{r}_t'$ is an aggregation of all entities connected to relation r via a mean pooling operation, and $\mathbf{R}_t'$ is obtained by concatenating this result with the embeddings of all relations. Eventually, we update the relation embedding matrix $\mathbf{R}_t$ using a GRU.

4.3 Disentangled Causal and Confounding Features

In the previous subsection, the entity and relation representations are learned based on GNN-

contained causal and confounding factors, and we separate them at the presentation level, which provides a solution to the previously mentioned problem of not being able to separate these two features at the data level. To do this, we introduce a decoupling module to decouple causal and confounding features. Taking the entity embedding matrix as an example, it is represented as follows:

$$\mathbf{D}_{E,C}, \mathbf{D}_{E,N} = \text{softmax}(\text{MLP}(\mathbf{E})), \quad (6)$$
$$\mathbf{E}_C = \mathbf{E} \odot \mathbf{D}_{E,C}, \mathbf{E}_N = \mathbf{E} \odot \mathbf{D}_{E,N}. \quad (7)$$

We want the two embeddings learned from the decoupling module to be as independent as possible, which is essential to accurately separate causal and confounding features (Chen et al., 2023). Mutual information is a basic quantity to measure the nonlinear correlation of two random variables. Minimizing mutual information is a feasible scheme to decouple causal features from confounding features. Specifically, we implement this process with contrastive log-ratio upper-bound MI estimator (Cheng et al., 2020; Wu et al., 2021), which utilizes variational distributions q and a neural network to approximate the true distribution. We define the objective function as follows:

$$\mathcal{L}_{mi} = \mathbb{E}_{p(\mathbf{E}_C, \mathbf{E}_N)} \left[\log q_\theta(\mathbf{E}_N | \mathbf{E}_C)\right] \\ - \mathbb{E}_{p(\mathbf{E}_C)} \mathbb{E}_{p(\mathbf{E}_N)} \left[\log q_\theta(\mathbf{E}_N | \mathbf{E}_C')\right]. \quad (8)$$

We perform the same operation with relation

10

embedding decoupling, after which we obtain $\mathbf{R}_C$ and $\mathbf{R}_N$.

4.4 Temporal Gap Guided Decoder

After the causal and confounding embeddings of entities and relations in the derived data, we use a specially crafted decoder to determine the likelihood score of potential entities and relations. Events or facts in a data stream may span different periods. For example, major political events may occur in rapid succession over a short period, while certain rare natural phenomena may occur sporadically and at longer intervals. With this in mind, it is reasonable to consider the time intervals of events to get an accurate picture of their temporal relationship. The key to the design of our decoder is the time interval vector, which guides the decoding process in considering the event time interval while calculating the fraction. Formulaic as:

$$\mathbf{t}_s = \boldsymbol{\alpha}_s t + \boldsymbol{\beta}_s, \; \mathbf{t}_l = \boldsymbol{\alpha}_l t + \boldsymbol{\beta}_l. \tag{9}$$

Here, $\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s, \boldsymbol{\alpha}_l$, and $\boldsymbol{\beta}_l$ signify learnable parameters. Adopting ConvTransE as our decoder, we introduce four variables, which traverse a one-dimensional convolutional layer followed by a fully connected layer, culminating in the extraction of a probability vector encompassing all entities. This process is mathematically articulated as:

$$\mathbf{p}_C \left(e_o \mid e_s, r, t \right) = \mathrm{ReLU} \left(\mathrm{ConvTransE} \left(\mathbf{e}_{s,C,t}, \right. \right.$$
$$\left. \left. \mathbf{r}_{C,t}, \mathbf{t}_s, \mathbf{t}_l \right) \right) \mathbf{E}_{C,t}.$$
$$\tag{10}$$

We apply the same decoding process to the confounding features to get $\mathbf{p}_N \left(e_o \mid e_s, r, t \right)$.

4.5 Causal Intervention and Training Objective

Causal-based embedding learns the intrinsic causes that cause events to occur, so the reasoning results obtained from causal-based embedding are expected to yield reasonable input results. We define the supervised classification loss as follows:

$$\mathcal{L}_{E,C} = \sum_{(e_s, r, e_o, t) \in \mathcal{G}} \mathbf{y}_t \log \mathbf{p}_C \left(e_o \mid e_s, r, t \right), \tag{11}$$

where $\mathbf{y}_t$ is label vector. Conversely, confounding features are conceptualized to address conceivable biases or superficial patterns emanating from the training dataset. Given their inability to aid in inference, we proceed to compute their output average across all entity categories and encapsulate the loss as:

$$\mathcal{L}_{E,N} = \frac{1}{|\mathbf{E}_{N,t}|} \sum_{(e_s, r, e_o, t) \in \mathcal{G}} \mathrm{KL} \left(\mathbf{y}_u, \right. $$
$$\left. \log \mathbf{p}_N \left(e_o \mid e_s, r, t \right) \right), \tag{12}$$

where KL denotes the KL-Divergence, $\mathbf{y}_u$ represents the uniform distribution.

We believe that causal intervention is the manifestation of causal features under the influence of confounding features, but we cannot directly conduct causal intervention at the data level to mitigate confounding effects. Therefore, we obtain intervention features that combine causal features and confounding features at the representation level of entities and relationships. Specifically, according to the backdoor adjustment Eq. (1), we first introduce a random addition procedure to obtain the intervention feature, and for the intervention feature we expect the decoder to still output the correct result:

$$\mathbf{E}_{I,t} = \phi \left(\mathbf{E}_{C,t}, \mathbf{E}'_{N,t} \right), \tag{13}$$

$$\mathbf{p}_I \left(e_o \mid e_s, r, t \right) = \mathrm{ReLU} \left(\mathrm{ConvTransE} \left(\mathbf{e}_{s,I,t}, \mathbf{r}_{I,t}, \right. \right.$$
$$\left. \left. \mathbf{t}_s, \mathbf{t}_l \right) \right) \mathbf{E}_{I,t}, \tag{14}$$

where $\mathbf{E}'_{N,t}$ is the confounding feature of the entites randomly sampled from $\mathbf{E}_{N,t}$. Then we define the loss as follows:

$$\mathcal{L}_{E,I} = \sum_{(e_s, r, e_o, t) \in \mathcal{G}} \mathbf{y}_t \log \mathbf{p}_I \left(e_o \mid e_s, r, t \right). \tag{15}$$

Finally, the loss function of the model for the link prediction task is as follows:

$$\mathcal{L}_E = \mathcal{L}_{E,C} + \lambda_1 \mathcal{L}_{E,N} + \lambda_2 \mathcal{L}_{mi} + \lambda_3 \mathcal{L}_{E,I}, \tag{16}$$

where $\lambda_1, \lambda_2, \lambda_3$ are designated as hyperparameters, and the first two are used to determine the strength of decoupled learning of the model and the latter is used to determine the strength of causal intervention of the model.

5 Experiments and Analysis

5.1 Experimental Settings and Implementation Details

Datasets. We evaluate our model and baselines on six benchmark datasets, including ICEWS14 (Garcia-Duran et al., 2018),

11

ICEWS18 (Jin et al., 2019), ICEWS05-15 (Garcia-Duran et al., 2018), YAGO (Mahdisoltani et al., 2014), WIKI (Leblay and Chekol, 2018) and GDELT (Leetaru and Schrodt, 2013). Statistics of the datasets are summarized in Table 1.

Table 1: Statistics of datasets in the experiments.

Dataset	# Entity	# Predict	# Train	# Valid	# Test	Time interval
ICEWS14	7128	230	63685	13823	13222	24 hours
ICEWS18	23033	256	373018	45995	49545	24 hours
ICEWS05-15	10488	251	322958	69224	69147	24 hours
YAGO	10623	10	161540	19523	20026	1 year
WIKI	12554	24	539286	67583	63110	1 year
GDELT	7691	240	1734399	238765	305241	15 mins

Baselines. For the link prediction task, we compare CEGRL-TKGR model with two categories of KGR models: (1) *static KGR models*, including TransE (Bordes et al., 2013), DistMult (Yang et al., 2015), ComplEx (Trouillon et al., 2016) and R-GCN(Schlichtkrull et al., 2018). We apply these models in static KGs that ignore timestamp information. (2) *TKGR models*, including TTransE (Leblay and Chekol, 2018), TA-DistMult (Garcia-Duran et al., 2018), TNTComplEx (Lacroix et al., 2019), RE-GCN (Li et al., 2021), GHT (Sun et al., 2022), EvoKG (Park et al., 2022), TITer(Sun et al., 2021), xERTE (Han et al., 2021), TLogic(Liu et al., 2022) and CEN(Li et al., 2022b).

Evaluation Metrics. The mean reciprocal rank (MRR) and Hits@k are standard metrics for the TKG link prediction task. MRR is the average reciprocal of the correct query answer rank. Hits@k indicates the proportion of correct answers among the top k candidates. We use a more reasonable time-aware filter setting to report our experimental results[2].

Implementation Details. The whole of training hyper-parameters and model configurations are summarized in Appendix A.1.

5.2 Experimental Results and Discussion

Table 2 and Table 3 report the experimental results of the link prediction task on six benchmark datasets. Static KG embedding methods fell far behind CEGRL-TKGR due to their inability to capture temporal dynamics in the TKG. Our method is also superior to other TKGR models in predicting events. The improved performance shows that surface patterns and noise are widely present in several real-world TKG datasets. The previous models

---

[2]The time-aware filtering setting filters out only the four groups that occur at query time and can simulate extrapolated prediction tasks in the real world (Sun et al., 2021).

are generally inadequate in design. CEGRL-TKGR based on evolutionary representation will learn the inherent confounding features in the TKG when gathering neighborhood information and transmitting historical information, and the model based on rule-based inference will mine the false correlation in the data, all of which will lead to the model-making non-causal predictions in the reasoning stage. Our model incorporates causal theory into the TKGR task and visibly separates causal features from confounding features. This helps to protect the model from surface patterns and noise present in the dataset and to uncover the real reasons that affect the formation of links between entities. TiTer and EvoKG show excellent performance on YAGO datasets because the former's historical fact search strategy works well on smaller datasets, while the latter's modeling of event timing works well on datasets containing events at relatively regular time intervals. More model configurations and experimental results are summarized in the Appendix.

6 Conclusion

In this paper, we revisit the GNN-based TKGR model from the causality perspective, on this basis, we propose a novel CEGRL-TKGR framework. By synergistically integrating causal structures with graph representation learning of the TKG, our framework overcomes the problem of existing TKGR models' learning biased data representations and mining for false correlations unintentionally. Comprehensive experimental results and analysis have proved the effectiveness of CEGRL-TKGR.

Limitations and Future Work. The proposed CEGRL-TKGR is an innovational causal enhanced graph representation learning framework for optimizing feature representations directly using causal technology for the TGKR task. The limitations of CEGRL-TKGR are as follows:

- From the dataset's perspective, our research primarily focuses on TKG datasets, which may not verify the generalization ability of the CEGRL-TKGR framework to those time-interval insensitive graph datasets. Additionally, we aim to further conduct case studies to enhance the interpretability of the framework in the reasoning procedure as future work.

- From the model's perspective, our research evaluates the TKGR task alone. Theoreti-

Table 2: Experimental results of link prediction on ICEWS series dataset. The best result in each column is boldfaced.

Model	ICEWS14				ICEWS18				ICEWS05-15			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	22.48	13.36	25.63	41.23	12.24	5.84	12.81	25.10	22.55	13.05	25.61	42.05
Distmult	27.67	18.16	31.15	46.96	10.17	4.52	10.33	21.25	28.73	19.33	32.19	47.54
ComplEx	30.84	21.51	34.48	49.59	21.01	11.87	23.47	39.97	31.69	21.44	35.74	52.04
R-GCN	28.03	19.42	31.95	44.833	15.05	8.31	16.49	29.00	27.13	18.83	30.41	43.16
TTransE	13.43	3.11	17.32	34.55	8.31	1.92	8.56	21.89	15.71	5.00	19.72	38.02
TA-DistMult	26.47	17.09	30.22	45.41	16.75	8.61	18.41	33.59	24.31	14.58	27.92	44.21
TNTComplEx	32.12	23.35	36.03	49.13	21.23	13.28	24.02	36.91	27.54	19.52	30.80	42.86
Evo-KG	26.90	16.69	30.57	47.39	25.46	16.25	29.15	43.21	26.32	15.82	31.96	50.80
xERTE	40.79	32.70	45.67	57.30	29.31	21.03	33.51	46.48	46.62	37.84	52.31	63.92
TITer	40.59	31.41	45.47	57.62	29.55	21.37	33.10	44.87	46.62	36.46	52.29	65.23
TLogic	41.80	31.93	47.23	60.53	28.41	18.74	32.71	47.97	45.99	34.49	52.89	67.39
RE-GCN	42.00	31.63	47.20	61.65	32.62	22.39	36.79	52.68	48.03	37.33	53.90	**68.51**
CEN	41.93	31.71	46.86	61.36	29.41	19.60	33.91	49.97	47.04	36.58	52.60	67.18
GHT	38.28	28.43	42.85	57.47	28.38	18.78	32.01	47.27	42.90	31.76	46.77	64.64
CEGRL-TKGR	**42.74**	**32.32**	**48.05**	**62.68**	**32.90**	**22.70**	**36.91**	**52.95**	**48.35**	**37.63**	**54.22**	68.47

Table 3: Experimental results of link prediction on YAGO, WIKI, and GDELT datasets. The best result in each column is boldfaced.

Model	YAGO				WIKI				GDELT			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	38.97	26.87	42.45	56.05	23.46	16.53	28.45	35.71	-	-	-	-
Distmult	44.05	39.19	49.70	59.94	27.96	18.84	32.45	39.51	8.61	3.91	8.27	17.04
ComplEx	44.09	39.33	49.57	59.64	27.69	18.67	31.99	38.61	9.84	5.17	9.58	18.23
R-GCN	20.25	11.25	24.01	37.30	13.96	7.21	15.75	22.05	12.17	8.64	12.37	20.63
TTransE	31.19	18.12	40.91	51.21	29.27	21.67	34.43	42.39	-	-	-	-
TA-DistMult	54.92	48.15	59.61	66.71	44.53	39.92	48.73	51.71	10.34	6.25	10.44	21.63
TNTComplEx	57.98	52.92	61.33	66.69	45.03	40.04	49.31	52.03	19.53	12.41	20.75	33.42
Evo-KG	68.81	54.49	81.40	**92.41**	67.44	54.63	79.36	**85.98**	18.94	11.31	20.08	34.01
GHT	57.22	51.64	60.68	67.17	48.50	45.08	50.87	53.69	20.04	12.68	21.37	34.42
xERTE	84.19	80.09	88.02	89.78	73.60	69.05	78.03	79.73	19.45	11.92	20.84	34.18
TITer	87.47	80.09	**89.96**	90.27	73.91	71.70	75.41	76.96	18.19	11.52	19.20	31.00
RE-GCN	82.30	78.83	84.27	88.58	78.53	74.50	81.59	84.70	19.69	12.46	20.93	33.81
CEN	83.49	79.66	86.10	90.04	78.52	74.65	81.44	84.59	19.96	11.39	20.97	33.77
CEGRL-TKGR	**86.25**	**82.92**	88.72	91.70	**79.66**	**75.73**	**82.83**	85.59	**20.11**	**12.73**	**21.46**	**34.51**

cally, the GNN-based reasoning paradigm incorporated in the causal structure can be applied to any other graph representation learning tasks, such as triple classification (Jaradeh et al., 2021), triple set prediction (Zhang et al., 2024), and graph classification (Liu et al., 2023). In future work, we desire to explore powerful disentanglement methods and more advanced causal intervention strategies to improve the CEGRL-TKGR's performance for more rich graph representation learning-based tasks. Besides, the increased complexity of causal reasoning in the TKG is untouched.

- From the adaptation's perspective, to adapt the CEGRL-TKGR framework to more models, there are some hyper-parameters to control causal intervention and training. These hyper-parameters are sensitive to different models and datasets, hence it needs to take sufficient time to experiment to find the optimal values and combinations among them. Therefore, how to reduce the consumption in the above adaptation procedure upon the framework is

worthy of consideration.

References

Denilson Barbosa, Haixun Wang, and Cong Yu. 2013. Shallow information extraction for the knowledge web. In *Proceedings of IEEE 29th International Conference on Data Engineering*, pages 1264–1267.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, pages 2787–2795.

Guoxin Chen, Yongqing Wang, Fangda Guo, Qinglang Guo, Jiangli Shao, Huawei Shen, and Xueqi Cheng. 2023. Causality and independence enhancement for biased node classification. In *Proceedings of the 32th ACM International Conference on Information and Knowledge Management*, pages 203–212.

Yonghong Chen, Hao Li, Han Li, Wenhao Liu, Yirui Wu, Qian Huang, and Shaohua Wan. 2022a. An overview of knowledge graph reasoning: key technologies and applications. *Journal of Sensor and Actuator Networks*, 11(4):78.

Yongqiang Chen, Yonggang Zhang, Yatao Bian, Han Yang, MA KAILI, Binghui Xie, Tongliang Liu, Bo Han, and James Cheng. 2022b. Invariance principle meets out-of-distribution generalization on graphs. In *ICML 2022: Workshop on Spurious Correlations, Invariance and Stability*.

Pengyu Cheng, Weituo Hao, Shuyang Dai, Jiachang Liu, Zhe Gan, and Lawrence Carin. 2020. Club: A contrastive log-ratio upper bound of mutual information. In *International conference on machine learning*, pages 1779–1788. PMLR.

Songgaojun Deng, Huzefa Rangwala, and Yue Ning. 2020. Dynamic knowledge graph based multi-event forecasting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1585–1595.

Vanessa Didelez and Iris Pigeot. 2001. Causality: models, reasoning, and inference.

Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)*, 37(2):1–30.

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.

Julia Gastinger, Timo Sztyler, Lokesh Sharma, and Anett Schuelke. 2022. On the evaluation of methods for temporal knowledge graph forecasting. In *NeurIPS 2022 Temporal Graph Learning Workshop*.

Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2021. xerte: Explainable reasoning on temporal knowledge graphs for forecasting future links. In *Proceedings of the 9th International Conference on Learning Representations*.

Zhen Han, Yunpu Ma, Yuyi Wang, Stephan Günnemann, and Volker Tresp. 2020. Graph hawkes neural network for forecasting on temporal knowledge graphs. In *Proceedings of the 2nd Conference on Automated Knowledge Base Construction*.

Mohamad Yaser Jaradeh, Kuldeep Singh, Markus Stocker, and Sören Auer. 2021. Triple classification for scholarly knowledge graph completion. In *Proceedings of the 11th Knowledge Capture Conference*, pages 225–232.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2019. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *https://arxiv.org/abs/1904.05530*.

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*, pages 6669–6683.

Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2019. Tensor decompositions for temporal knowledge base completion. In *International Conference on Learning Representations*.

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *The Web Conference*, pages 1771–1776.

Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer.

Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022a. 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, 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*, pages 290–296.

Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021. Temporal knowledge graph reasoning based on evolutional representation learning. In *Proceedings of the 44th Annual 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, Xinwang Liu, and Fuchun Sun. 2022. Reasoning over different types of knowledge graphs: Static, temporal and multi-modal. *https://arxiv.org/abs/2212.05767*.

Xingyu Liu, Juan Chen, and Quan Wen. 2023. A survey on graph classification and link prediction based on gnn. *https://arxiv.org/abs/2307.00865*.

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

Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. 2014. Yago3: A knowledge base from multilingual wikipedias. In *7th biennial conference on innovative data systems research*. CIDR Conference.

Namyong Park, Fuchen Liu, Purvanshi Mehta, Dana Cristofor, Christos Faloutsos, and Yuxiao Dong. 2022. Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, pages 794–803.

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*, pages 593–607. Springer.

Yongduo Sui, Xiang Wang, Jiancan Wu, Min Lin, Xiangnan He, and Tat-Seng Chua. 2022. Causal attention for interpretable and generalizable graph classification. In *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1696–1705.

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.

Jinze Sun, Yongpan Sheng, Ling Zhan, and Lirong He. 2023. Tkgr-rhetne: A new temporal knowledge graph reasoning model via jointly modeling relevant historical event and temporal neighborhood event context. In *International Conference on Neural Information Processing*, pages 331–343. Springer.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 2071–2080.

Junkang Wu, Wentao Shi, Xuezhi Cao, Jiawei Chen, Wenqiang Lei, Fuzheng Zhang, Wei Wu, and Xiangnan He. 2021. Disenkgat: knowledge graph embedding with disentangled graph attention network. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pages 2140–2149.

Yingxin Wu, Xiang Wang, An Zhang, Xiangnan He, and Tat-Seng Chua. 2022. Discovering invariant rationales for graph neural networks. In *Proceedings of the 10th International Conference on Learning Representations*.

Lizhu Xiong, Yongpan Sheng, and Lirong He. 2024. Tkgr-gprscl: Enhance temporal knowledge graph reasoning with graph structure-aware path representation and supervised contrastive learning. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 200–212. Springer.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 3rd International Conference on Learning Representations*.

Matej Zečević, Devendra Singh Dhami, Petar Veličković, and Kristian Kersting. 2021. Relating graph neural networks to structural causal models. *https://arxiv.org/abs/2109.04173*.

Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. 2022. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4741–4753.

Wen Zhang, Yajing Xu, Peng Ye, Zhiwei Huang, Zezhong Xu, Jiaoyan Chen, Jeff Z Pan, and Huajun Chen. 2024. Start from zero: Triple set prediction for automatic knowledge graph completion. *IEEE Transactions on Knowledge and Data Engineering*.

Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, and Meng Jiang. 2022. Learning from counterfactual links for link prediction. In *Proceedings of the 39th International Conference on Machine Learning*, pages 26911–26926. PMLR.

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 35th AAAI Conference on Artificial Intelligence*, volume 35, pages 4732–4740.

A Appendix

A.1 Implementation Details

We set the dimension of all embeddings and hidden states to 200. The number of layers of the R-GCN is set to 1 for YAGO and 2 for the other datasets. The optimal number of historical snapshots is set to 8, 10, 10, 1, 2, and 6 for ICEWS14, ICEWS18, ICEWS05-15, YAGO, WIKI, and GDELT, respectively. To fair comparison, static graph constraints are added for ICEWS14, ICEWS18, and ICEWS05-15. The channel number for decoding is set to 50, and the kernel size is set to 4×3. We try several different values for λ_1, λ_2, and λ_3, and finally chose 0.5, 0.5, 0.3. We use Adam to optimize the parameters, with a learning rate of 0.001. All of the experiments are processed on a Linux server with CPU Xeon Gold 6142, RAM 64G, and Nvidia 4090 GPU.

A.2 Ablation Study

We investigate the effectiveness of causally enhanced and time-interval guided decoders for the link prediction task. Specifically, CEGRL-TKGR w/o TD means that no time interval vector is used to guide the decoder to work, and CEGRL-TKGR w/o CE means that the model removes causal decoupling and causal intervention parts. Table 4 shows the results of ablation experiments, which indicate the effectiveness of these two components. As can be seen from the results in the table, for datasets such as YAGO and WIKI that contain relatively regular time intervals, a temporal gap-guided decoder can capture this time interval pattern well enough to make accurate predictions. At the same time, it does not degrade performance even for time-interval insensitive datasets. Our causal enhancement module, under the independent constraint of emphasizing causal features and confounding features, eliminates the influence of the fast bridge through causal intervention, forcing the model to learn the intrinsic causes of the events. It is worth noting that our causal enhancement module can be seen as a flexible component that can be easily used in several GNN-based reasoning frameworks.

A.3 Parameter Sensitivity Analysis

In the CEGRL-TKGR, λ_1 and λ_2 jointly affect the disentanglement intensity of causal and confounding features, and λ_3 controls the intensity of causal intervention. We study the sensitivity of parameters in different benchmark datasets, as depicted in Fig. 3. Specifically, one parameter is fixed at 0.5 and the other parameter varies in [0,1] with a step size of 0.1. The model is relatively stable in most parameter selection cases, but on noisy datasets, the model has higher requirements for hyper-parameters, and extreme values will degrade the performance of the model. The best range for λ_1, λ_2 is about 0.5 to 0.7. λ_3 should be a relatively small value, ranging from 0.3 to 0.6.

A.4 Performance on Noisy Temporal Knowledge Graphs

To explore whether the proposed CEGRL-TKGR can alleviate noise and surface patterns, we randomly replace a certain percentage of positive triples in the training set of each TKG dataset in form of noisy TKGs. Taking YAGO and WIKI datasets as examples, we test the performance of CEGRL-TKGR and CEGRL-TKGR w/o CE under different noise deviations, respectively. The experimental results are shown in Fig. 4.

From the experimental results, we can draw the following conclusion: when the noise in the dataset increases, the performance of models lacking the recognition of causal features and confounding features will deteriorate sharply, and the performance of MRR and Hits@1 will decrease, which indicates that the CEGRL-TKGR w/o CE is easy to capture data bias and make wrong predictions based on it. In contrast, CEGRL-TKGR uses the causal enhancement module to effectively reduce the impact of confounding features and shows more stable performance on these two noisy TKG datasets. The performance degradations on MRR and Hits@1 are significantly smaller than those without the causal module.

Table 4: The ablation study of our model on the six benchmark datasets. "w/o" means "without".

| Model	ICEWS14		ICEWS18		ICEWS05-15		YAGO		WIKI		GDELT	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
CEGRL-TKGR w/o TD	42.21	62.43	32.67	52.68	48.13	68.33	84.71	90.56	78.54	84.37	19.93	34.50
CEGRL-TKGR w/o CE	41.89	61.65	32.62	52.54	48.03	68.20	81.93	88.39	79.04	84.79	19.66	33.71
CEGRL-TKGR	**42.74**	**62.68**	**32.90**	**52.95**	**48.35**	**68.47**	**86.25**	**91.70**	**79.66**	**85.59**	**20.11**	**34.51**
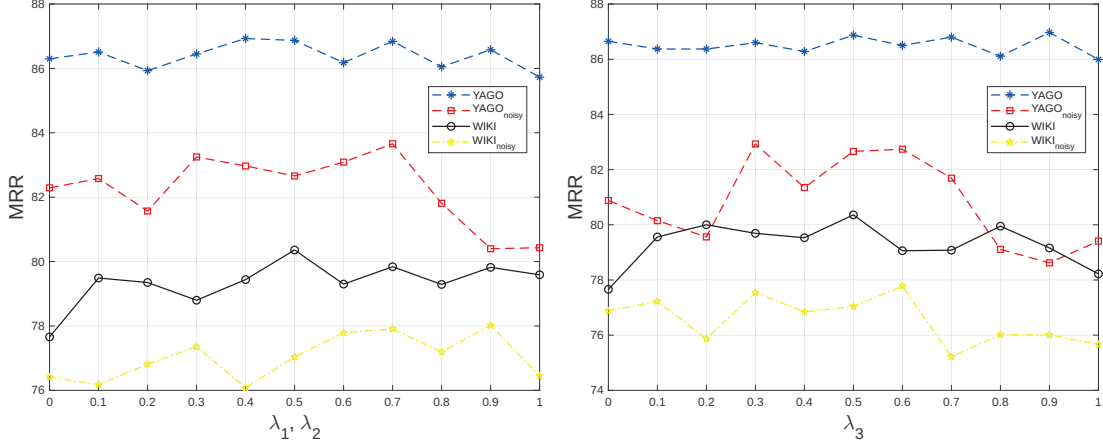
Figure 3: The parameters sensitivity analysis of loss coefficients λ_1, λ_2 and λ_3.
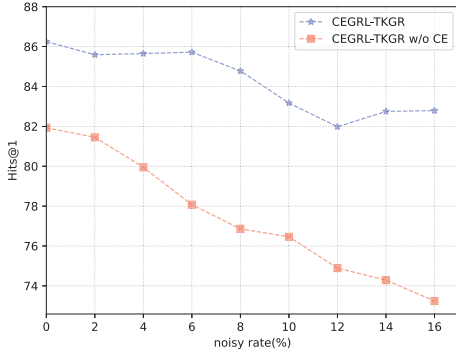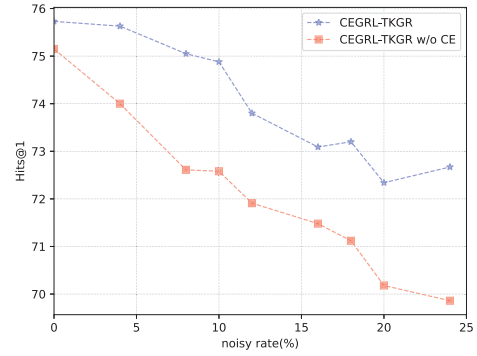

(a) MRR results on the YAGO dataset.
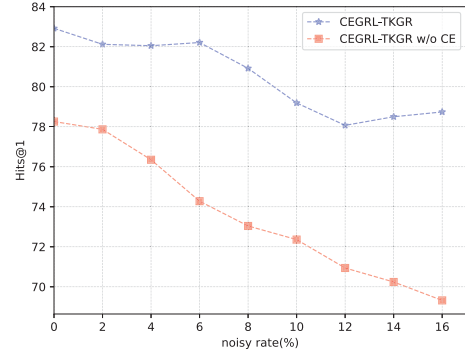
(b) Hits@1 results on the YAGO dataset.
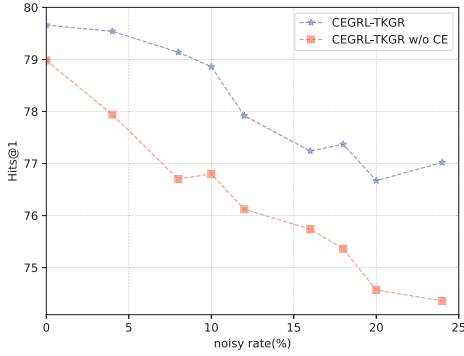
(c) MRR results on the WIKI dataset.

(d) Hits@1 results on the WIKI dataset.

Figure 4: The performance of CEGRL-TKGR and CEGRL-TKGR w/o CE on the noisy YAGO and WIKI datasets, respectively.