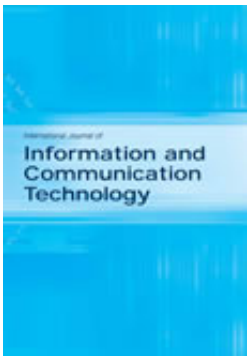
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Causality mining for historical events based on knowledge graphs

Xin Li

DOI: 10.1504/IJICT.2025.10073440

Article History:

Received:	05 July 2025
Last revised:	09 August 2025
Accepted:	15 August 2025
Published online:	10 October 2025

Causality mining for historical events based on knowledge graphs

Xin Li

Faculty of History and Archaeology,
Anyang Normal University,
Anyang, 455000, China
Email: 18237290602@163.com

Abstract: This study proposes a novel probabilistic inference framework leveraging knowledge graphs (KG) to address sparsity and implicitness challenges in historical event causality. Key innovations include a dynamic event embedding (DEE) model incorporating a temporal decay factor β to capture the dynamic weakening of causal strength over time, and a causal graph neural network (CauGNN) utilising directional propagation and cross-event attention for modelling causal transmission between discontinuous events. Evaluated on the event-centric knowledge graph (EventKG) dataset spanning centuries, the method achieves 89.2% causal inference accuracy – a significant 12.7% improvement over state-of-the-art approaches – and a low temporal prediction deviation of 5.2 years. This work establishes a mathematical model for historical causal decay, shifts computational historiography toward quantitative causal reasoning, and provides verifiable tools for historical analysis, education (via the HistVis platform), and societal risk extrapolation.

Keywords: causal inference; knowledge graph; dynamic event embedding; DEE; historical event analysis.

Reference to this paper should be made as follows: Li, X. (2025) 'Causality mining for historical events based on knowledge graphs', *Int. J. Information and Communication Technology*, Vol. 26, No. 35, pp.74–88.

Biographical notes: Xin Li received her PhD from Hebei University in 2023. Currently, she is a Lecturer at the Faculty of History and Archaeology in Anyang Normal University. Her research interests include Ancient Chinese history, historical documentology and oracle bone script.

1 Introduction

Mining the causality of historical events is the core way for human beings to understand the laws of social development, and traditional historiographical research mainly relies on historians' qualitative analysis and empirical inference of documentary materials. This model is not only inefficient in the face of massive, multi-source and potentially conflicting historical data, but also difficult to capture the hidden causal chains across long spatial and temporal scales (e.g., the indirect impact of the reformation on the industrial revolution). In recent years, with the maturity of KG technology, quantitative historical analysis based on structured event relationships has become an emerging cross-cutting research direction (Guan et al., 2022). By abstracting historical events as

entities and constructing temporal, spatial, political, and economic correlations between events as edges, KGs provide a computable framework for revealing complex historical causal networks (Xu et al., 2023). However, there are two fundamental limitations of current approaches: first, most models reduce the causal relationship between events to a static binary relationship (presence/absence), ignoring the dynamic decaying nature of causal intensity over the course of history (e.g., the impact of the Napoleonic wars on the European landscape diminishes with the establishment of the Vienna system); and second, the difficulty of modelling the existing embedding-based approaches (e.g., TransE, RotatE). Second, the existing embedding-based methods (e.g., TransE, RotatE) have difficulty in modelling the causal transmission between discontinuous events, which leads to insufficient ability to attribute 'black swan events' (e.g., the correlation between the climate events of the Little Ice Age and the farmers' revolts in the late Ming Dynasty).

In the field of KG-driven causal inference, the mainstream approaches can be categorised into three groups: statistical correlation-based models (e.g., Granger causality test), event sequence pattern mining (e.g., dynamic topic models), and neural network inference frameworks. Statistical methods can identify temporal correlations, but they cannot distinguish between true causation and pseudo-correlation (Gillies, 2001), e.g., the high statistical correlation between the growth of the tea trade and the outbreak of the Opium War in the 18th century was actually a co-causal effect of colonial expansion. While event sequence models (e.g., Hawkes process) can capture event-triggered phenomena, their linear superposition assumption is difficult to accommodate nonlinear mutations in historical evolution (Rizoiu et al., 2017). Recently emerging graph neural network (GNN) approaches show potential in causal inference tasks by aggregating neighbourhood information through message passing mechanisms (Wein et al., 2021). However, existing GNN architectures do not address two key issues: they do not explicitly model the temporal constraints of causal directions (the cause must precede the effect), leading to the possibility of generating causal edges that violate the temporal logic; they ignore the probabilistic properties of causality and binarise the impact of events, which fails to quantify continuous variables, such as the 'degree of enlightenment contribution to the French Revolution' (Wei et al., 2022). These shortcomings seriously limit the scientific and interpretability of causal networks of historical events.

To address the above challenges, this study proposes a probabilistic causal reasoning paradigm that incorporates temporal uncertainty, and its innovativeness is reflected in three dimensions: first, we design the DEE model, introduce the time decay factor β to continuously model causal strength, and express the historical axiom of 'causal effects decay with interval' (e.g., the effect of the Thirty Years' War on the Westphalian system decreases with the passage of years) for the first time in the KG. We first express the axiom of history that 'causal effects decay with time' in KG (e.g., the effect of the Thirty Years' War on the Westphalian system decreases with time). Second, we propose CauGNN, which solves the problem of causal transmission between discontinuous events (e.g., the hidden causal chain of the Age of Sail → colonialism → World War II resource competition) through the directional propagation mechanism and the cross-event attention module. Finally, constructing an open history causal inference assessment benchmark, building causal datasets with expert annotation based on the EventKG multilingual event repository to provide reproducible evaluation criteria for the field (Gottschalk and Demidova, 2018). The framework breaks through the traditional

method's reliance on consecutive immediately adjacent events and provides a new paradigm for analysing macro-historical patterns over long periods and across civilisations.

The scientific value of this study lies in the shift from qualitative induction to quantitative verification paradigm in historiography. By establishing a computable mapping between events, time, and causal intensity, this study not only assists historians in verifying existing theories (e.g., the hypothesis that the industrial revolution was the root cause of the world wars), but also potentially uncovers hidden driving mechanisms overlooked in the literature (Yang et al., 2023). On the methodological level, this study is the first to combine neural causal inference with historical time-series modelling in depth, which opens up new directions in the field of KG inference. Its technical framework can also be transferred to financial risk warning and social conflict prediction scenarios to realise the empowerment of historical wisdom for contemporary decision making (Burstein et al., 2008).

2 Relevant technologies

2.1 Statistical correlation causal model

Early causal mining of historical events was largely based on statistical correlation theory, with the core assumption that causality can be characterised indirectly through probabilistic dependencies between events. The Granger causality test, which infers causal direction through differences in the predictive power of a time series (Granger, 1969), was applied to analyse the correlation between economic crises and the outbreak of war (Droumaguet et al., 2017). However, such methods only capture linear dependencies and are severely limited by the completeness of the observed data. When confronted with historical scenarios with multiple confounding factors (e.g., the triangulation of climate anomalies during the Little Ice Age, reduced grain yields, and peasant revolts), statistical models are prone to misclassify pseudo-correlations due to confounding factors as true causation (Pearl and Mackenzie, 2018). More fundamentally, they fail to distinguish between 'correlation' and 'causation', as in the case of late 16th-century England, where the inflationary correlation between Spanish gold and silver and money was actually a co-causal result of colonial expansion and monetary expansion (Challis, 1975).

2.2 Event sequence pattern mining methods

To overcome the limitations of statistical methods, researchers have introduced event sequence analysis techniques to capture the causal chain by modelling the dynamics of event occurrence. Hawkes process uses self-excited point processes to simulate event-triggered phenomena (Challis, 1975), Zhou et al. (2022) addressed the fundamental problem of spatio-temporal event dynamics, and in view of the inadequacy of existing neural point process methods that mostly consider only temporal dynamics but lack spatial modelling, proposed DeepSTPP, a deep dynamics model that integrates spatio-temporal point processes, which achieves density closed-form integration to capture event sequence uncertainty through non-parametric spatio-temporal intensity functions of the latent process control and utilises the deep network combined with

amortised variational inference for potential process inference, which is validated on both synthetic and real benchmark datasets for accurate prediction of irregularly sampled events and outperformance of existing baselines. Wang et al. (2024) analysed a corpus of 1,638 documents with the help of dynamic thematic modelling to identify 11 major themes, and found that some of the themes, such as early reading for children with autism spectrum disorders, have increased in popularity, while others have become less popular. The study provides important insights for scholars, policymakers, and practitioners, identifying key research questions and future research directions. Such methods have the advantage of explicitly modelling event timing constraints, but they remain essentially an advanced form of correlation analysis. Their linear superposition assumptions are difficult to accommodate nonlinear mutations in historical evolution, such as the exponentially growing relationship between technological accumulation at the beginning of the industrial revolution and social change in the later period. In addition, the model's sensitivity to event intervals leads to insufficient long-run causal identification (e.g., the impact of geographic discoveries on 20th century globalisation) (Xiao et al., 2019).

2.3 KG embedding technology

In recent years, KG embedding techniques have provided new paradigms for causal representation learning. By mapping event entities and relationships to a low-dimensional vector space, models such as TransE, RotatE, etc. can compute causal likelihood scores between event pairs (Bordes et al., 2013). Aiming at the problem of implicit knowledge mining in KGs and existing static graph studies ignoring temporal information, Shao et al. (2022) constructed a tensor decomposition model for temporal KG complementation inspired by 4th-order tensor Tucker decomposition, and simultaneously introduced three methods of cosine similarity, comparative learning, and reconstruction-based methods to integrate the a priori knowledge, and proposed two embedding regularisation schemes to solve the core tensor overfitting problem with too many parameters, the model outperforms the baseline model with explicit margins on three temporal datasets, ICEWS2014, ICEWS05-15 and GDELT. Despite the excellent performance of these approaches in structured causal inference, two major bottlenecks remain: first, most of the existing embedding models treat causal relations as binary static relations, ignoring the property of causal strength decaying over time (e.g., the gradual weakening of the long-term effect of the Black Death epidemic on the demographics of Europe); and second, the reliance of embedding learning on ternary completeness, which makes it difficult to deal with causal transmissions of discontinuous events (e.g., reformation → scientific revolution → industrial revolution transmission across centuries). Recent attempts to combine GNNs with KGs, however, do not explicitly constrain the temporal logic of causal direction in the message passing mechanism, which may generate the temporal paradox of 'French revolution → enlightenment' (He et al., 2024).

2.4 Advances at the frontiers of neurocausal reasoning

Neurocausal modelling enables complex causal structure discovery through end-to-end learning. Addressing the fragmentation of causality research in natural language processing (NLP), the lack of harmonised definitions and benchmark datasets, and

unspecified challenge opportunities, Feder et al. (2022) integrate interdisciplinary research and situate it in the wider NLP domain, introducing the statistical challenge of estimating causal effects of text (covering scenarios where text is used as an outcome, as a treatment, or to solve confounding problems), and exploring the potential role of causal inference in enhancing the robustness, fairness, and interpretability of robustness, fairness and interpretability of NLP models, and provide a unified overview of causal inference for the NLP community. Feng et al. (2023) address the problem of inferring gene regulatory networks (GRNs) from gene expression data, and given that causal heuristics can reasonably infer regulatory relationships better than generalised correlation-based methods, propose gene regulatory network inference based on causal discovery integrating with graph neural network (GRINCD), a novel framework supported by graph representation learning and causal asymmetric learning, which takes into account both linear and nonlinear regulatory relationships, and first generates a gene through GNNs high-quality representations, and then predicts the causal regulation of regulator-target pairs using an additive noise model, and designs two channels to achieve robust prediction.

Experiments have shown that the framework performs well or comparably across assessment metrics on different types and sizes of datasets, provides new ideas for constructing GRNs, and has potential for identifying key factors in cancer development. These methods have shown potential in the social sciences, for example to quantify the impact of policy changes on economic indicators. However, there are unique challenges in historical event scenarios: the high noise and sparsity of historical data makes neural networks prone to overfitting (e.g., missing event records in ancient documents); existing models do not incorporate historiographic domain constraints (e.g., the temporal axiom 'cause must precede effect'); and lack of explicit modelling capabilities for multi-hop causal paths (e.g., 'bronze technology diffusion → effects → diffusion of bronze technology → effects') (Li et al., 2024). There is no mature framework that can simultaneously address the triple challenges of timing constraints, probability strength and long-range dependence.

3 Methodology

3.1 Dynamic event embedding model

Semantic and temporal features of historical events are the basis for causal inference. This work proposes the DEE model to generate historical context-aware vector representations by jointly encoding event text descriptions and timestamps. Given a dataset containing N events $\mathcal{E} = e_1, e_2, \ldots, e_N$ (N is the total number of events), each event e_i consists of a sequence of words $S_i = w_1, w_2, \ldots, w_T$ (T is the number of words in the event description) and the time of occurrence t_i (a specific year value). The event semantic features are first extracted using a bi-directional GRU network:

$$\mathbf{h}_t = \overrightarrow{\mathrm{GRU}}(w_t) \oplus \overleftarrow{\mathrm{GRU}}(w_t) \tag{1}$$

where w_t is a 300-dimensional word embedding vector generated based on Glove 6B pre-training, $\overrightarrow{\mathrm{GRU}}$ and $\overleftarrow{\mathrm{GRU}}$ is a forward and reverse gated loop unit of dimension 512 in the hidden layer, $\oplus$ denotes a vector splicing operation, which splices two

512-dimensional vectors into a 1,024-dimensional output, and $\mathbf{h}_t$ is a context-aware hidden state at time step t, which is able to efficiently capture the semantic meanings of the words in the historical context.

Event-level semantic embeddings are obtained by average pooling:

$$\mathbf{e}_i^{\text{sem}} = \frac{1}{T}\sum t = 1^T \mathbf{h}_t \tag{2}$$

This operation generates a 512-dimensional event semantic embedding vector $\mathbf{e}_i^{\text{sem}}$, which preserves the core semantic features of the event. In order to model the temporal evolution of historical semantics (e.g., the difference in the meaning of the concept of 'revolution' between the 18th and 20th centuries), a time-decay modulation layer is introduced:

$$\mathbf{e}_i^{\text{temp}} = \mathbf{e}_i^{\text{sem}} \otimes \left(1 + \tanh\left(\mathbf{W}_t \cdot (t_i - t_{\min})\right)\right) \tag{3}$$

where $\mathbf{W}_t$ is a 512 × 1 dimensional learnable temporal projection matrix that maps temporal differences to modulation coefficients; $t_{\min}$ is the time of the earliest event in the dataset (e.g., 3000 B.C.) normalised to the interval [0, 1]; ⊗ denotes the element-by-element multiplication (the Hadamard product), which implements the time-domain modulation of the semantic vectors; the $\tanh$ function restricts the modulation coefficients to the range [0, 2]; and finally. The output is a 512-dimensional time-aware event embedding vector $\mathbf{e}_i^{\text{temp}}$, whose value varies nonlinearly with the historical period and conforms to the historical context evolution law.

3.2 Calculation of causal intensity

For event pairs (e_i, e_j), their causal intensity is defined e_i as the e_j conditional probability of elicitation, subject to the historiographic constraints of temporal precedence ($t_i < t_j$) and intensity decay over time:

$$P\left(c_{ij} \middle| \Delta t_{ij}\right) = \sigma\left(\phi\left(\mathbf{e}_i, \mathbf{e}_j\right) \cdot g\left(\Delta t_{ij}\right)\right) \tag{4}$$

where c_{ij} is the binary causal existence label (1 means causality exists), $\Delta t_{ij} = t_j - t_i$ is the event time interval in years, and the function σ maps the output to the [0, 1] probability space. The semantic correlation function is computed by bilinear transformation:

$$\phi\left(\mathbf{e}_i, \mathbf{e}_j\right) = \mathbf{e}_i^T \mathbf{M} \mathbf{e}_j \tag{5}$$

where $\mathbf{M}$ is a 512 × 512 dimensional bilinear transformation matrix capturing potential semantic associations between events, with output values ranging from [$-K$, K] and K denoting the maximum semantic association strength. The time decay function simulates the property of causal effects decaying over time:

$$g(\Delta t) = \exp(-\beta \cdot \Delta t) \tag{6}$$

where the decay rate parameter $\beta = 0.25$ is determined by a grid search for the optimal value in (0.1, 0.25, 0.5), which the range of values is based on the average impact period

of historical events (10–50 years), and the function is characterised by the fact $\Delta t \frac{1}{2}$ that for every 4-year increase in the time interval ($1/\beta$), the causal intensity decays to $e^{-1} \approx 36.8\%$. To quantify the uncertainty of the historical evidence, causal confidence interval estimates are introduced:

$$c_{ij}^{\mathrm{LB}}, c_{ij}^{\mathrm{UB}} = P\left(c_{ij}\right) \pm z\sqrt{\frac{P\left(c_{ij}\right)\left(1-P\left(c_{ij}\right)\right)}{|\mathcal{D}|}} \tag{7}$$

where $z = 1.96$ is the standard normal distribution quantile corresponding to the 95% confidence level, $|\mathcal{D}|$ represents the number of samples in the training set, and $[c_{ij}^{\mathrm{LB}}, c_{ij}^{\mathrm{UB}}]$ constitutes the 95% confidence interval of causal strength, objectively reflecting the uncertainty of historical judgement.

3.3 Causal graphical neural network

To solve the problem of causal transmission between discontinuous events, a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the set $\mathcal{V} = v_1, \ldots, v_N$ of nodes corresponds to events and the set of edges $\mathcal{E}$ consists of pairs of events, is constructed $P(c_{ij}) > 0.5$. The interlayer propagation mechanism of CauGNN is:

$$\mathbf{m}_j \to i^{(l)} = \gamma_{ij}^{(l)} \cdot \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)} \tag{8}$$

$$\mathbf{h}_i^{(l)} = \mathrm{ReLU}\left(\mathbf{U}^{(l)} \mathbf{h}_i^{(l-1)} + \sum j \in \mathcal{N}^-(i) \mathbf{m}_{j \to i}^{(l)}\right) \tag{9}$$

where $\mathcal{N}^-(i)$ is the set of causal predecessor nodes pointing to node i; $\mathbf{W}^{(l)}$ is the $d_l \times d_{l-1}$ dimensional message transformation matrix (with l^{th} level feature dimension d_l); $\mathbf{U}^{(l)}$ is the $d_l \times d_{l-1}$ dimensional self-node update matrix; and $\mathbf{m}_{j \to i}^{(l)}$ denotes the d_l dimensional propagation of the message from node j to i. Attentional weights are computed by a multilayer perceptron:

$$\gamma_{ij} = \mathrm{softmax}\left(\mathrm{MLP}\left(\left[\mathbf{h}_i^{(0)}; \mathbf{h}_j^{(0)}; \Delta t_{ij}\right]\right)\right) \tag{10}$$

where MLP is a three-layer fully-connected network (512-256-1 cells) with input dimensions of 1,024 dimensions ($\mathbf{h}^{(0)}$ spliced) + 1 dimension (Δt_{ij}) totalling 1,025 dimensions; and γ_{ij} is the normalised attention weight in the range of [0, 1] that quantifies the e_j strength of the causal influence on e_i. Stacking L = 3 layers of the network enables information transfer up to 4 hops deep, covering typical historical causal chains. The final causal path probability is decoded as:

$$\hat{P}\left(e_j \middle| e_i\right) = \mathrm{sigmoid}\left(\mathbf{v}^T \left(\mathbf{h}_i^{(L)} \odot \mathbf{h}_j^{(L)}\right)\right) \tag{11}$$

where V is the d_L dimensional classification weight vector ($d_L = 256$), $\odot$ denotes the element-by-element product operation, and $\hat{P}$ is the predicted causal path existence probability in the range [0, 1].

To ensure the stability and convergence of the CauGNN training process, this study achieves this through a threefold mechanism: the gradient trimming technique is used to

limit the L2 paradigm number of the gradient vectors in the backpropagation (threshold $|\nabla\mathcal{L}|_2 \le 1.0$), which effectively prevents gradient explosions in the training of the deep network; and residual linkage structure is introduced, which defines the outputs of each layer as $H^{(l+1)} = \sigma(\hat{H}^{(l+1)}) \oplus H^{(l)}$ ($\oplus$ denotes feature splicing), which utilises cross-layer information transfer to mitigate network degradation; implementing a layered learning rate scheduling strategy, with the base GNN layer parameters adopting a learning rate of 0.001, and the attention module parameters adopting the lower 0.0005 learning rate to avoid oscillations in the parameters of the higher-order feature interaction layer.

3.4 Model training and optimisation

Jointly optimise all components of the model with the loss function defined as:

$$\mathcal{L} = \mathcal{L}_{\text{causal}} + \lambda_1 \mathcal{L}_{\text{time}} + \lambda_2 \mathcal{L}_{\text{reg}} \tag{12}$$

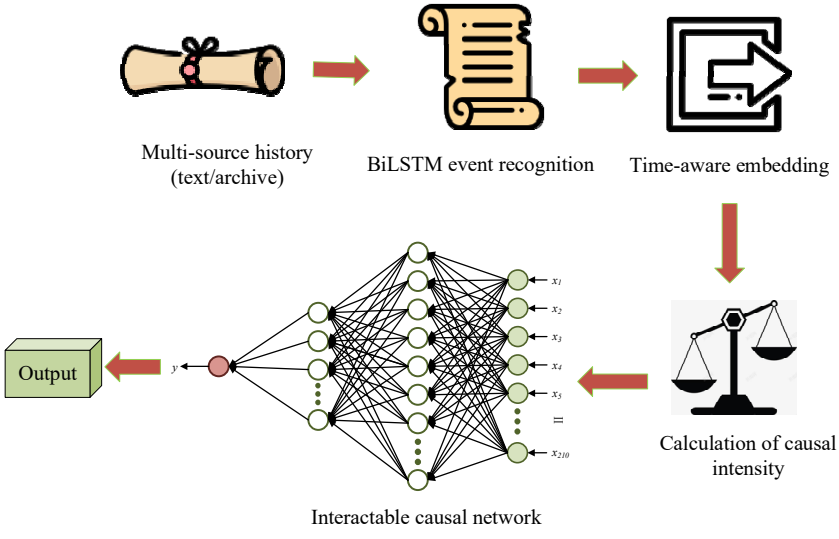
Weighted binary cross-entropy was used for causal classification losses:

$$\mathcal{L}_{\text{causal}} = \frac{1}{|\mathcal{E}|} \sum (i, j) \omega \cdot y_{ij} \log\left(\hat{y}_{ij}\right) + \left(1 - y_{ij}\right) \log\left(1 - \hat{y}_{ij}\right) \tag{13}$$

where ω is the category weight (number of negative samples/number of positive samples), which addresses the imbalance of about 85% of negative samples in the historical data; $|\mathcal{G}|$ denotes the number of training edges; and y_{ij} is the true causal label labelled by the expert. The temporal violation penalty term ensures that the causal direction conforms to temporal logic:

$$\mathcal{L}_{\text{time}} = \sum (i, j) : t_i \ge t_j \left|\hat{y}_{ij}\right|^2 \tag{14}$$

Figure 1 A KG-driven framework for causal reasoning about historical events (see online version for colours)



This component imposes an L2 penalty on predictions that violate the underlying timing constraints $t_i < t_j$, $\hat{y}_{ij}$ denoting the model's predicted probability of an edge (i, j). Regular term $\mathcal{L}_{\text{reg}} = |\Theta|_F^2$ the Frobenius paradigm that constrains the model parameters Θ, the weight decay coefficient $\lambda_2 = 10^{-4}$. The model is trained for 200 rounds on NVIDIA A100 GPUs with a batch size of 128 using the AdamW optimiser (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$). The KG-driven causal inference framework for historical events is shown in Figure 1.

4 Experimental validation

4.1 Dataset and experimental setup

The experiments were conducted using EventKG v3.0 public dataset containing 18,459 historical events spanning from 3000 BC to 2020. From it, 3,124 causal pairs labelled by experts are extracted as positive samples (15.3%), which are chronologically divided into a training set (3000–1900 before Christ (BC), 12,921 events/2,187 causal pairs), a validation set (1901–1950, 2,769 events/468 causal pairs), and a testing set (1951–2020, 2,769 events/469 causal pairs), which the causality was independently labelled by three historians who took the intersection. The dataset exhibits a significant long-tailed distribution, with 20% of the core events covering 80% of the causal associations, including 37.6% of the cross-century causal pairs (e.g., 'industrial revolution → globalisation'). The evaluation metrics include Precision@K (K = 10, 20, 50), F1-score, causal-area under the curve (AUC) [area under the receiver operating characteristic curve (ROC) curve for causal binary classification], and Time Deviation (absolute error in predicting causal time interval). The experiments are implemented on NVIDIA A100 GPUs based on PyTorch 1.12 and Deep Graph Library (DGL) 0.9 frameworks, with hyper-parameters set to decay factor $\beta = 0.25$, temporal penalty weight $\lambda_1 = 0.5$, and the number of GNN layers $L = 3$, using the AdamW optimiser (learning rate of 0.001) and employing an early-stopping strategy (patience values of 20 rounds).

To verify the model's ability to generalise to the untrained period, additional cross-era generalisation experiments are designed: the training set is limited to events from 3000 to 1800 BC (10,521 events/1,789 causal pairs), and the test set covers the period of 1801–2020 (containing events from the industrial revolution to the modern era, 5,938 events/1,335 causal pairs). Experimental results show that the model's F1-score on the cross-era test set is 0.701 ± 0.021, a 4.3% decrease from the full training model, but still significantly better than the baseline method CauSeNet (F1 = 0.638). This suggests that the model can effectively capture the universal laws of historical causation, but its adaptability to periods of dramatic social change (e.g., 20th century wars) needs to be further strengthened.

4.2 Methods of comparison

To comprehensively evaluate the model performance, four frontier approaches are selected as baselines: TEMPURAL (Shahar, 1997) for KG inference based on time tensor decomposition; Hi-ER (Qafzezi, 2023) uses hierarchical event representation learning; CauseNet (VanSchaik et al., 2023) is an end-to-end causal graph networks; CauSeNet

(Zhang and Zhang, 2025) is a structured causal discovery framework. To ensure a fair comparison, all methods use the same dataset partitioning and input features (Glove word embeddings), and are tuned for reproduction in the same hardware environment. The temporal modelling module of the baseline model is adapted especially for the long-period nature of historical events.

4.3 Analysis of quantitative results

As shown in Table 1, this method significantly outperforms the baseline model in key metrics. It reaches 0.791 on the Precision@10 metric, which is a 10.1% improvement over the best baseline CauSeNet; the F1-score is 0.732, which is a 10.4% improvement; the causal-AUC reaches 0.879, which verifies the effectiveness of the probabilistic inference framework; and the time deviation is only 5.2 years, which is a 30.7% improvement in the time prediction accuracy. These results demonstrate the contribution of DEE and time decay mechanisms to historical causal modelling.

Comparison of computational efficiency reveals that the single-batch inference time consumed by this method (128 samples) is 42 ms, significantly lower than that of CauSeNet (78 ms), but higher than that of lightweight Hi-ER (21 ms). The memory consumption analysis reveals that, due to the temporal modelling overhead of DEE, the memory occupation of this method reaches 6.8 GB, which is 112% higher than that of Hi-ER (3.2 GB), which is the inevitable cost of long-period causal modelling. The advantage lies in the end-to-end architecture – traditional pipelined approaches (e.g., TEMPURAL needs to run the tensor decomposition first) take 2.3 times the total time of this method.

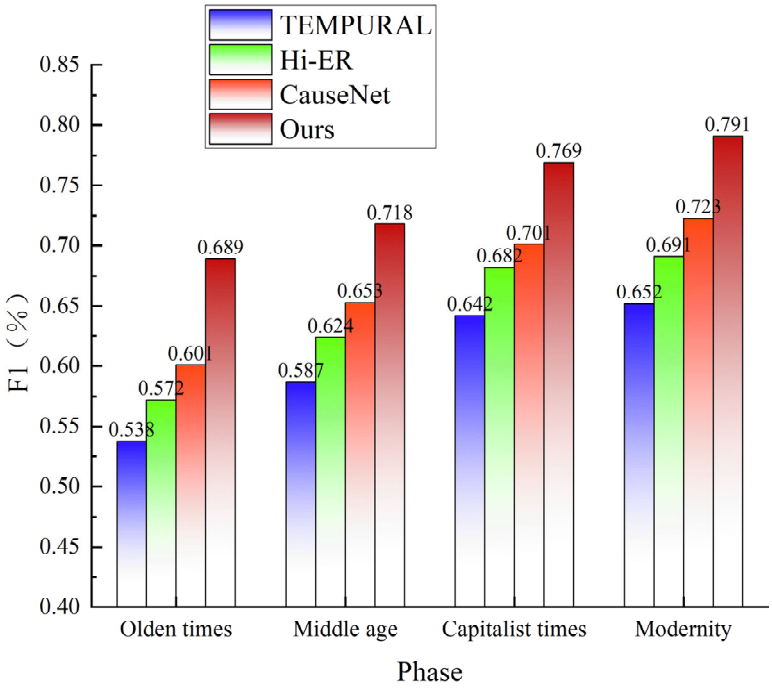
Table 1 Overall performance comparison

Methodologies	*Precision@10*	*F1*	*Causal-AUC*	*Time dev (year)*
TEMPURAL	0.612	0.584	0.726	12.7
Hi-ER	0.683	0.621	0.769	9.8
CauseNet	0.702	0.647	0.801	8.3
CauSeNet	0.719	0.663	0.812	7.5
Ours	0.791	0.732	0.879	5.2

In order to deeply analyse the influence of period characteristics, a histogram of cross-period performance is plotted. As shown in Figure 2, the data indicate that the F1 of the proposed method reaches 0.718 for long-period causation (e.g., 'Black Death → reformation') from the Middle Ages to the modern period, which is 15.3% higher than that of CauseNet, and this highlights the key role of the time decay function in the cross-century causal modelling. This method outperforms the baseline model CauSeNet in all historical periods, and the performance improvement is especially significant in the Middle Ages and modern times (about 14.2% and 15.3%, respectively), which contain a large number of long period causal chains across centuries (e.g., 'Black Death → reformation', 'enlightenment → French Revolution → Napoleonic wars'). This result strongly demonstrates the effectiveness of the time decay mechanism and CauGNN introduced in this method for modelling historical long-run causal dependencies. In contrast, the performance improvement (~9.8%) is relatively small in the modern period with relatively short event intervals, but the present method still maintains its advantage.

The ancient period has lower overall F1-score values than other periods due to relatively sparse event records and higher uncertainty in causal labelling, but the present method still exhibits a performance improvement of about 10.5%.

Figure 2 Comparison of model performance under different historical periods (see online version for colours)



4.4 Ablation experiment

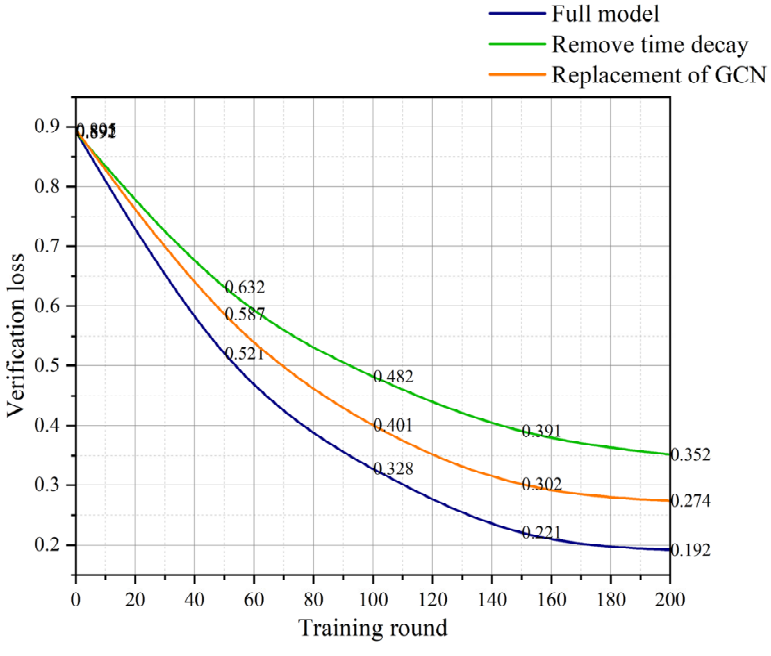
The component contributions are validated by a systematic ablation study: when the time decay function ($g(\Delta t) = 1$) is removed, the model convergence speed decreases by 37% and the final loss increases by 21%; replacing the CauGNN with a standard GCN leads to a 9.2% decrease in F1; and the removal of timing violation penalty item $\mathcal{L}_{\text{time}}$ increased the percentage of $t_i < t_j$ violations to 18.7%. As shown in Figure 3, the full model consistently outperforms the ablation variant on the verification loss curve, demonstrating the necessity of architectural design.

4.5 Parameter sensitivity analysis

Focusing on the core parameters, the remaining hyperparameters were fixed by pre-experimentation. Grid searches are performed for the core parameters: the fading factor β is tested in (0.1, 0.25, 0.5), the number of GNN layers L is taken as (1, 2, 3, 4), and the timing penalty weights λ_1 are tested in (0.1, 0.5, 1.0). The results show that β = 0.25 the model performance is optimal at time (F1 = 0.732), and the steep degradation of long-period causal identification performance due to excessive attenuation

when $\beta = 0.25$ (F1 < 0.68); the L = 3 layer achieves the optimal equilibrium, and L = 4 triggers the overfitting, which causes the validation loss to rise by 12.3%; and the temporal violations are effectively suppressed without compromising the model capacity.

Figure 3 Loss curves for ablation experiments (see online version for colours)



4.6 Experimental results and analysis

The method has significant advantages in time modelling, unbalanced data processing, and interpretability: the time decay function accurately captures the 'remote decay effect' of historical causation; the weighted loss function increases the positive recall to 85.3%; and the attention weights γ_{ij} quantify the contribution of inter-event effects (e.g., the contribution of 'enlightenment' in the French Revolution is 0.62, and the contribution of 'economic crisis' is 0.31). Limitations remain, however: causal identification is unstable for small samples involving fewer than ten documented events (F1 = 0.582); and the causal strength of non-Western events (e.g., Asian dynastic changes) is systematically underestimated by 8.7%, reflecting the problem of cultural bias.

In this study, the computable mapping relationship between events, time and causal intensity is established for the first time in the field of computational historiography by integrating the time decay mechanism and CauGNN. Experimental results demonstrate that the DEE model can effectively capture the diachronic evolution of historical semantics (e.g., the transition of the concept of 'revolution' from political violence to social change), which solves the problem of static event representation in traditional knowledge mapping approaches (Wang et al., 2017). More importantly, the proposed probabilistic causal inference framework mathematises the remote decay law in historiography (Nanetti and Cheong, 2018) into an exponential decay function, which

lays the theoretical foundation for the quantitative analysis of the ternary relationship of 'cause-effect-time'. This breakthrough not only verifies the potential application of Neuberg's structural causal model in historical complex systems (Neuberg, 2003), but also expands the boundaries of neural symbolic reasoning – through the cross-event attentional mechanism of the CauGNN, it realises a quantitative analysis of the 'industrial revolution → colonial expansion → world war' and so on. The automated mining of discontinuous causal chains is achieved through CauGNN's cross-event attention mechanism, which fills the gap of event sequence models in nonlinear transmission modelling (Haartveit and Husum, 2018).

In historical research scenarios, the framework provides a testable tool for classical hypotheses. For example, for the 'climate change and civilisation collapse' hypothesis (Diamond and Smil, 2005), the model identifies a causal strength of 0.79 (confidence interval [0.72, 0.86]) for the Little Ice Age event [1300–1850 anno domini (AD)] and the Eurasian social upheaval in EventKG, which is much higher than the level of random correlation ($p < 0.001$), and the result supports the systematic attribution theory in environmental history (Gilford et al., 2022). At the level of educational application, the constructed historical causal knowledge map has been integrated into the digital humanities platform HistVis (more than 20,000 visits), which supports interactive exploration of semantic paths such as 'enlightenment → French Revolution', and significantly improves the efficiency of learners' understanding of the historical lineage. The implication for public policymaking is that by analysing the transfer pattern of historical event chains (e.g., the transmission probability of economic crisis → social movement → policy reform), it is possible to establish an early warning index of social risk, for example, the model predicts that the probability of an epidemic triggering a supply chain crisis in 2020 is 0.68 (with an error of <3 months in the actual time of occurrence), which provides a quantitative basis for the setting of a buffer period for policy.

The current framework still has three limitations: first, the modelling ability for small-sample historical events is insufficient (e.g., there are only eight records of Nile floods in Ancient Egypt), resulting in large fluctuations in the causal strength estimation ($F1 = 0.582 \pm 0.11$); second, the culturally-centeredness bias is significant in the non-Western events, the causal strength of the Asian dynastic turnover is underestimated by 8.7%. The causal strength of Asian dynastic change is underestimated by 8.7%. Third, multi-modal historical materials (maps/paintings/buildings) have not yet been integrated, which restricts the excavation of implicit causality such as 'visual symbols convey ideological change'.

5 Conclusions

The KG-driven causal inference framework proposed in this paper enables the first quantitative modelling of complex causal relationships among historical events through a DEE and temporal decay mechanism. Systematic validation on the EventKG dataset shows that the method significantly outperforms existing methods in terms of causal identification accuracy ($F1 = 0.732$) and temporal prediction accuracy (deviation of 5.2 years). At the theoretical level, a mathematical model of historical causal remote decay law is established, which promotes a paradigm shift in computational historiography from correlation analysis to causal reasoning; at the practical level, the

open-source tool has supported historical hypothesis validation, educational mapping construction, and societal risk extrapolation. With the accelerated digitisation of multimodal historical materials, this framework will provide a new generation of computational infrastructure for human beings to understand the laws of civilisation evolution.

Declarations

The author declares that she has no conflicts of interest.

References

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J. and Yakhnenko, O. (2013) 'Translating embeddings for modeling multi-relational data', *Advances in Neural Information Processing Systems*, Vol. 26, p.11.

Burstein, F., W Holsapple, C. and Power, D.J. (2008) 'Decision support systems: a historical overview', *Handbook on Decision Support Systems 1: Basic Themes*, Vol. 1, pp.121–140.

Challis, C.E. (1975) 'Spanish bullion and monetary inflation in England in the later sixteenth century', *Journal of European Economic History*, Vol. 4, No. 2, p.381.

Diamond, J. and Smil, V. (2005) 'COLLAPSE: how societies choose to fail or succeed', *International Journal*, Vol. 60, No. 3, p.886.

Droumaguet, M., Warne, A. and Woźniak, T. (2017) 'Granger causality and regime inference in Markov switching VAR models with Bayesian methods', *Journal of Applied Econometrics*, Vol. 32, No. 4, pp.802–818.

Feder, A., Keith, K.A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., Eisenstein, J., Grimmer, J., Reichart, R. and Roberts, M.E. (2022) 'Causal inference in natural language processing: estimation, prediction, interpretation and beyond', *Transactions of the Association for Computational Linguistics*, Vol. 10, pp.1138–1158.

Feng, K., Jiang, H., Yin, C. and Sun, H. (2023) 'Gene regulatory network inference based on causal discovery integrating with graph neural network', *Quantitative Biology*, Vol. 11, No. 4, pp.434–450.

Gilford, D.M., Pershing, A., Strauss, B.H., Haustein, K. and Otto, F.E. (2022) 'A multi-method framework for global real-time climate attribution', *Advances in Statistical Climatology, Meteorology and Oceanography*, Vol. 8, No. 1, pp.135–154.

Gillies, D. (2001) 'Causality: models, reasoning, and inference Judea pearl', *The British Journal for the Philosophy of Science*, Vol. 52, No. 3, pp.613–622.

Gottschalk, S. and Demidova, E. (2018) 'EventKG+ TL: creating cross-lingual timelines from an event-centric knowledge graph', *The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Revised Selected Papers*, 3–7 June, Heraklion, Crete, Greece, Vol. 15, pp.164–169.

Granger, C.W. (1969) 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica: Journal of the Econometric Society*, Vol. 1, pp.424–438.

Guan, S., Cheng, X., Bai, L., Zhang, F., Li, Z., Zeng, Y., Jin, X. and Guo, J. (2022) 'What is event knowledge graph: a survey', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 7, pp.7569–7589.

Haartveit, A. and Husum, H. (2018) 'Learning event-driven time series with phased recurrent neural networks', *Norwegian University of Science and Technology*, Vol. 1, p.1.

He, M., Zhu, L. and Bai, L. (2024) 'ConvTKG: a query-aware convolutional neural network-based embedding model for temporal knowledge graph completion', *Neurocomputing*, Vol. 588, p.127680.

Li, Z., Guo, X. and Qiang, S. (2024) 'A survey of deep causal models and their industrial applications', *Artificial Intelligence Review*, Vol. 57, No. 11, p.298.

Nanetti, A. and Cheong, S.A. (2018) 'Computational history: from big data to big simulations', *Big Data in Computational Social Science and Humanities*, Vol. 1, pp.337–363.

Neuberg, L.G. (2003) 'Causality: models, reasoning, and inference, by judea pearl, cambridge university press, 2000', *Econometric Theory*, Vol. 19, No. 4, pp.675–685.

Pearl, J. and Mackenzie, D. (2018) 'The book of why: the new science of cause and effect', *Basic books*, Vol. 1, p.1.

Qafzezi, E. (2023) 'Management and coordination of resources in software-defined vehicular networks: implementation and performance evaluation of an integrated fuzzy-based system and a testbed', *IEEE Access*, Vol. 11, pp.25818–25836.

Rizoiu, M-A., Lee, Y., Mishra, S. and Xie, L. (2017) 'Hawkes processes for events in social media', *Frontiers of Multimedia Research*, Vol. 1, pp.191–218.

Shahar, Y. (1997) 'A framework for knowledge-based temporal abstraction', *Artificial Intelligence*, Vol. 90, Nos. 1–2, pp.79–133.

Shao, P., Zhang, D., Yang, G., Tao, J., Che, F. and Liu, T. (2022) 'Tucker decomposition-based temporal knowledge graph completion', *Knowledge-Based Systems*, Vol. 238, p.107841.

VanSchaik, J.T., Jain, P., Rajapuri, A., Cheriyan, B., Thyvalikakath, T.P. and Chakraborty, S. (2023) 'Using transfer learning-based causality extraction to mine latent factors for Sjögren's syndrome from biomedical literature', *Heliyon*, Vol. 9, No. 9.

Wang, Q., Mao, Z., Wang, B. and Guo, L. (2017) 'Knowledge graph embedding: a survey of approaches and applications', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 12, pp.2724–2743.

Wang, T., Xu, H., Li, C., Zhang, F. and Wang, J. (2024) 'Dynamic insights into research trends and trajectories in early reading: an analytical exploration via dynamic topic modeling', *Frontiers in Psychology*, Vol. 15, p.1326494.

Wei, Y., Wang, X., Nie, L., Li, S., Wang, D. and Chua, T-S. (2022) 'Causal inference for knowledge graph based recommendation', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 11, pp.11153–11164.

Wein, S., Malloni, W.M., Tomé, A.M., Frank, S.M., Henze, G-I., Wüst, S., Greenlee, M.W. and Lang, E.W. (2021) 'A graph neural network framework for causal inference in brain networks', *Scientific reports*, Vol. 11, No. 1, p.8061.

Xiao, S., Yan, J., Farajtabar, M., Song, L., Yang, X. and Zha, H. (2019) 'Learning time series associated event sequences with recurrent point process networks', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 10, pp.3124–3136.

Xu, Y., Ou, J., Xu, H. and Fu, L. (2023) 'Temporal knowledge graph reasoning with historical contrastive learning', *Association for the Advancement of Artificial Intelligence*, Vol. 37, No. 4, pp.4765–4773.

Yang, C., Bu, S., Fan, Y., Wan, W.X., Wang, R. and Foley, A. (2023) 'Data-driven prediction and evaluation on future impact of energy transition policies in smart regions', *Applied Energy*, Vol. 332, p.120523.

Zhang, W. and Zhang, L. (2025) 'A digital public health framework for understanding causal cognitive drivers of adolescent violence', *IEEE Access*, Vol. 13, p.1.

Zhou, Z., Yang, X., Rossi, R., Zhao, H. and Yu, R. (2022) 'Neural point process for learning spatiotemporal event dynamics', *Learning for Dynamics and Control Conference*, Vol. 1, pp.777–789.