scientific reports

Check for updates

OPEN

One-class edge classification through heterogeneous hypergraph for causal discovery

Marcos Paulo Silva Gôlo & Ricardo Marcondes Marcacini

Causal discovery from event pairs is essential for understanding complex real-world phenomena. Large language models (LLMs) have shown strong capabilities in capturing the semantics of events and inferring plausible cause-effect relations from text. However, they typically process each event pair in isolation and struggle to model the global event structure, which limits their ability to capture interdependencies among multiple events. Graph-based methods offer a structural alternative by explicitly modeling connections between events, but they often lack relational expressiveness, as relations are treated as implicit edges rather than as entities. Homogeneous hypergraphs address this by representing relations as nodes, enabling richer modeling of multi-event interactions and more expressive causal reasoning. Nevertheless, this strategy frequently leads to disconnected structures, hindering information aggregation through graph neural networks (GNNs). To address these challenges, we propose eCHOLGA (edge Classification through Heterogeneous One-cLass Graph Autoencoder), a novel method that leverages heterogeneous hypergraphs to model causal relationships more effectively. eCHOLGA integrates semantic features extracted from language models into the graph structure, enhancing the representation of events and their relations. By transforming relations into nodes and introducing additional node and edge types, it improves topological connectivity and enables GNNs to learn more informative edge representations. Furthermore, our method adopts a one-class learning strategy, requiring only positive (causal) examples for training, which reduces labeling effort. In addition to its effectiveness, eCHOLGA enhances interpretability and provides insights into the causal discovery process. Experimental results show that eCHOLGA outperforms state-of-the-art methods, establishing it as a promising approach for causal discovery in event pairs.

Keywords Event causal discovery, One-class learning, Heterogeneous graphs, Text pair causal discovery, Hypergraph for edge classification

Understanding the causal relationships between real-world events, known as causal discovery, is a fundamental yet challenging task with significant societal relevance. Identifying cause-effect links can support evidence-based decision-making in areas such as public policy, disaster management, public health, and economic planning[1]. In these contexts, understanding why an event occurs is often more valuable than merely knowing what occurred, as it enables proactive intervention, reduces risks, and improves resource allocation. While most research in causal discovery has focused on structured data and controlled settings, the widespread availability of unstructured textual event reporting in news articles, social media, and official documents has introduced new opportunities and methodological challenges. However, although textual descriptions offer rich contextual information, their inherent ambiguity and variability pose significant challenges for causal inference. As a result, there is a growing demand for computational methods capable of identifying causal relationships directly from natural language. In machine learning, this task is referred to as causal discovery in text or event pairs[2–4].

Recent advances in language models (LMs) have significantly improved the ability to identify causal relationships between pairs of textual events[2–4]. More recently, state-of-the-art approaches have employed large language models (LLMs), which can capture nuanced semantic dependencies between sentences and generalize across different domains[5]. These models are effective at inferring plausible causal relations based on textual patterns and world knowledge. However, most LLM-based methods process each event pair independently and do not explicitly capture the broader structure of causal relationships among multiple events. While longer context windows allow for more events to be considered simultaneously, this alone does not guarantee

Institute of Mathematical and Computer Sciences, University of Sao Paulo, São Carlos, São Paulo, Brazil. email: marcosgolo@usp.br; ricardo.marcacini@usp.br

a coherent understanding of how events are interrelated. Moreover, extended contexts may lead to diluted attention and reduced accuracy, especially when causal dependencies span long distances. As a result, LLMs may overlook important patterns that only emerge when events are analyzed collectively. To overcome these limitations, we can explore a combination of LLMs with graph-based representations, which explicitly model causal connections and support information aggregation across events. We argue that this approach combines the contextual strength of LLMs with the structural reasoning capabilities of graphs, enabling more effective and robust causal discovery.

In a graph-based representation, each node corresponds to an event, and edges denote causal or non-causal relationships[6]. This structure supports information aggregation and enables reasoning across multiple events, addressing some of the blind spots of LLMs when operating independently on event pairs. However, even with graphs, significant challenges remain. Most graph-based methods represent relationships implicitly as edges, which limits the model's ability to reason about the relationships themselves as entities. Furthermore, traditional graphs often lack the expressiveness needed to capture heterogeneous information associated with relationships, as they focus primarily on node-level attributes. Additionally, the prevalent use of Graph Neural Networks (GNNs) in these settings introduces architectural biases: GNNs are primarily designed to learn node embeddings and have limited capacity to represent rich relational semantics[7]. These factors constrain the effectiveness of graph-based causal discovery, especially when causal interactions are diverse and context-dependent.

To address these limitations, researchers have explored hypergraphs, which promote relationships to entities by modeling them as nodes, enabling a more flexible representation of multi-way interactions. When we use homogeneous hypergraphs, the resulting structure can often lack connectivity since an event can be associated with only one relation, which would generate a hypergraph with thousands of three-node components (event-relation-event). This fact weakens the use of Graph Neural Networks (GNNs) to effectively propagate and aggregate information. Figure 1 illustrates the limitations presented above, including those of homogenous hypergraphs. On the other hand, incorporating additional node types into the hypergraph can alleviate this issue by enriching the topological structure and enabling more robust information flow[8]. The introduction of such heterogeneous elements transforms the hypergraph into a heterogeneous hypergraph[9], which offers a more expressive and flexible framework for modeling complex and multi-faceted causal relationships among events.

We introduce a novel method for causal discovery in heterogeneous hypergraphs based on one-class learning. This work presents eCHOLGA (**e**dge **C**lassification through **H**eterogeneous **O**ne-c**L**Ass **G**raph **A**utoencoder). Our approach leverages heterogeneous hypergraphs to enhance representation learning with GNNs[7]. Additionally, we model causal discovery in event pairs through One-Class Learning, where the model is trained using only one class (causal relations) but can predict two classes (causal or non-causal relationships). This eliminates the need to cover the entire spectrum of non-causal relationships and reduces labeling effort, as only causal instances need to be annotated[10–12]. We further propose a novel heterogeneous triple loss function, combining the state-of-the-art one-class loss function[13], the reconstruction loss from graph autoencoders[14], and the cross-entropy loss for different node types. Finally, eCHOLGA introduces explicit three-dimensional representations to naturally enhance interpretability. Our key contributions are:

1. **Heterogeneous hypergraph** modeling transforming edges into nodes, and enrichment to incorporate additional node types such as topics, involved entities, location, and event date to create a more connected causal discovery graph.
2. **A novel triple heterogeneous loss function** designed for one-class learning in heterogeneous hypergraphs.
3. **We embed language model features** into our heterogeneous hypergraph to enrich its semantic representation.
4. **One-Class Learning** for causal discovery, enabling a more efficient labeling process and a natural problem formulation.
5. **Interpretable** representation learning on hypergraphs through GNNs and OCL for causal discovery.

Related work

LM-Based Causal Discovery. Hassanzadeh et al.[2] introduced NLM-BERT, an unsupervised method that leverages cosine similarity between the top-k most similar BERT embeddings to predict causality. Kayesh et al.[3] proposed fine-tuning BERT and its variations to detect causality using a semi-supervised dataset of 100,000

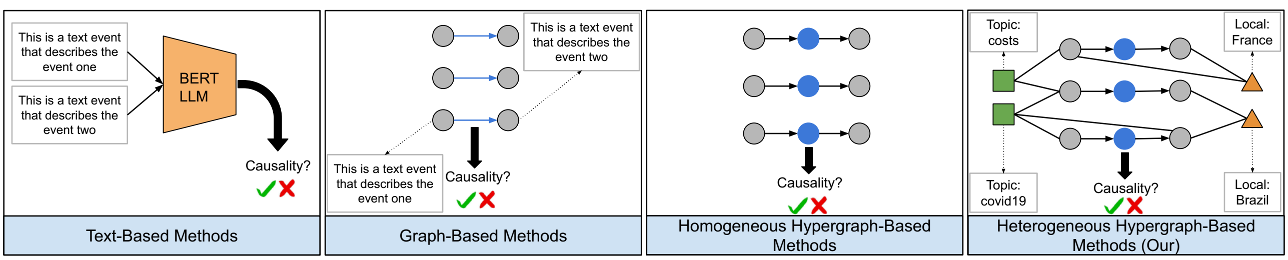

Fig. 1. Illustration of the styles of causality works. First, text-based, such as those exploring BERT and LLMs. Then, there are graph-based methods that classify edges as causal or not. Third, hypergraph-based works classify nodes that were edges in the causal graph. Finally, our strategy involves utilizing heterogeneous graphs to connect the causal hypergraph.

causal and non-causal sentence pairs. However, the fine-tuned models did not outperform NLM-BERT. Later, Kayesh et al.[4] extended this approach by incorporating an additional training dataset of 197,000 sentence pairs and experimenting with multiple methods and their combinations. The authors also explored a causality graph constructed from the training sets. Their CF-Context method integrates causal feature extraction, which involves embedding generation via graph-based representations, and contextual feature extraction.

LLM-Based Causal Discovery. Large Language Models (LLMs) have been applied to causal discovery in textual event data[5,15]. LLMs are pre-trained on vast corpora containing trillions of words and can generate text from a given input. This enables querying LLMs in natural language to determine whether one sentence causes another, effectively functioning as causal discovery models. The PyWhy-LLM library was developed specifically for causal detection tasks[5,15]. However, BERT- and LLM-based methods do not explicitly model the relationships between events, as they analyze each pair of texts in isolation rather than leveraging structured graph representations. Even studies such as Kayesh et al.[4], which incorporate graphs, do not utilize them as the core of the method, resulting in inferior results compared to purely text-based approaches.

GNN-Based Causal Discovery. Minghim et al.[16] introduced a method that extracts word embeddings to construct a causal graph, which is then processed using a gated GNN. A fully connected NN was employed for causal discovery. Similarly, Sakaji and Izumi[17] explored a hybrid approach using BERT and GNNs, specifically Graph Attention Networks (GATs), highlighting improvements over Graph Convolutional Networks (GCNs). The model employed two fully connected neural networks to classify sentences separately as causes or effects. Sasaki et al.[8] took a different approach, performing edge classification on a causal graph where nodes represent sentences and edges represent causal relationships. A classifier was applied to the final layer of the GNN. However, a key limitation of GNN-based methods is that they primarily focus on message passing between nodes (events), rather than learning edge representations, which are crucial for causal discovery. Additionally, most GNN-based approaches rely on binary supervised learning, requiring labeled examples of both causal and non-causal relationships. This poses a significant challenge since defining non-causal relationships is complex and costly due to their broad scope.

Hypergraphs. Hypergraphs have been extensively explored in the literature to address different tasks using Graph Neural Networks[18]. Zhao et al.[19] investigate the use of hypergraphs for traffic flow prediction. Ju et al.[20] employ hypergraphs for semi-supervised graph classification. Yi et al.[21] propose a method for learning to generate hypergraphs to perform semi-supervised node classification. Gao et al.[22] develop a novel GNN framework for hypergraphs designed to capture correlations in multi-modal and multi-type data. Wu et al.[23] introduce a contrastive learning approach for hypergraphs in the context of node classification. Finally, Wu et al.[24] propose a reconstruction error–based method to enhance node classification performance in hypergraphs. These studies model hypergraphs using the concept of hyperedges, where each hyperedge connects multiple nodes, and the set of nodes in the hypergraph coincides with that of the original graph. These works consistently report performance gains when employing hypergraph-based modeling, indicating the promising potential of hypergraphs. It is important to note, however, that the hypergraph formulation adopted in these studies differs from the one proposed in this work. Here, we define a hypergraph where each edge in the original graph is transformed into a new node, resulting in a hypergraph with a larger number of nodes and edges. Gôlo and Marcacini[12] proposed eCOLGAT, a method that leverages homogeneous hypergraphs to enhance representation learning for edges. eCOLGAT enables GNNs to learn more effective edge representations. Moreover, the authors modeled causal discovery using one-class learning (OCL), which offers two key advantages: (i) It eliminates the need to define the full scope of non-causal relationships; and (ii) It reduces the annotation effort, as only causal relationships need to be labeled. However, eCOLGAT generates hypergraphs from homogeneous graphs, which can lead to disconnected hypergraph structures in cases where relationships exist between events that share other relations. We highlight that this issue is particularly common in event-based causal discovery scenarios, as observed in the datasets explored by this work.

Table 1 summarizes the key characteristics of related studies. The **(L)LMs** row indicates (Large) LM studies. The **GNNs** row identifies works that leverage GNNs. The **OCL** row specifies whether the approach is based on the OCL. The **3D** row highlights studies that propose interpretability. The **HyperG.** row denotes whether hypergraphs are explored. We adopt the definition of hypergraphs from Jo et al.[7], which defines a hypergraph as a graph where edges have been transformed into nodes. Consequently, we do not include causality-related hypergraph studies that define hypergraphs as graphs with hyperedges (groups of nodes). The **Het.** row identifies studies that leverage heterogeneous hypergraphs.

Methods	NLM-BERT++	CF-Context	LLMs	GNNs	HGNNs	eCOLGAT	eCHOLGA
Reference	2,3	4	5	8,16,17	19–24	12	–
(L)LMs	✓	✓	✓	✓		✓	✓
GNNs				✓	✓	✓	✓
OCL						✓	✓
3D						✓	✓
HyperG.					✓	✓	✓
Het.							✓

Table 1. Synthesis of related work of eCHOLGA. Gaps in interpretable one-class methods based on heterogeneous hypergraphs are filled by eCHOLGA.

All studies incorporate (L)LMs, as processing event texts is fundamental. However, some rely solely on these representations, while others integrate them with GNNs. OCL for event-based causal discovery remains largely unexplored, marking the first research gap. Additionally, interpretable representation learning is a relatively new yet valuable direction for causal detection, identifying the second gap. Finally, the combination of heterogeneous hypergraphs with one-class learning has not been fully explored, presenting the third research gap. In the next section, we introduce our proposed method, which integrates heterogeneous hypergraphs and one-class learning to advance causal discovery in event pairs.

eCHOLGA: edge classification through heterogeneous one-cLAss graph autoencoder

We propose a novel method called edge Classification through Heterogeneous One-cLAss Graph Autoencoder (eCHOLGA) for causal detection in event pairs. Our method integrates heterogeneous hypergraphs, one-class learning (OCL), and graph neural networks (GNNs). eCHOLGA introduces several innovations for causal discovery in event pairs. First, it is the second one-class method developed for causal discovery and the first to use heterogeneous hypergraphs. Second, using heterogeneous hypergraphs with GNNs for causal discovery in event pairs is a novel approach, as most hypergraph-based causal studies focus on hypergraphs that do not transform edges into nodes. Third, our method leverages three-dimensional representation learning to provide interpretability in causal discovery, which enhances the understanding of event pair relationships.

Heterogeneous hypergraphs address two critical gaps in causal discovery. First, they solve the edge-representation gap encountered in GNNs, and second, they enhance graph connectivity when new events emerge or when the homogeneous causal graph becomes naturally disconnected[7,25], such as presented in Fig. 1. eCHOLGA leverages a state-of-the-art one-class loss function to cluster causal relationships near the center of a latent sphere[13]. Our method learns a three-dimensional latent space to provide interpretable representations, where causal relations are positioned inside the sphere and non-causal relations are placed outside. This space is learned via a graph autoencoder, using reconstruction loss as a constraint to the sphere loss function, which improves edge representation learning. Additionally, eCHOLGA introduces an innovative triple loss function by combining the one-class loss with a heterogeneous loss based on node types. This third loss not only improves representation learning but also serves as a secondary constraint for the one-class loss.

An illustration of eCHOLGA is presented in Fig. 2. Our workflow can be described in four main steps:

1. In this step, we obtain the causality graph. This graph is directed, since if one event causes another, the reverse is not necessarily true. Therefore, we model the relationships as a directed graph. Causality is explored through a one-class learning approach. Consequently, we have causal edges, where one event causes another, and unlabeled edges, which will later be classified as causal or non-causal;
2. After obtaining the causality graph, we transform it into a homogeneous hypergraph. This strategy consists of converting the edges to be classified into nodes. In this way, for each edge in the original graph, we create a new node and two additional edges connecting the three nodes (the two original nodes and the new node representing the original edge);
3. Once the homogeneous hypergraph is built, we enrich it to obtain a heterogeneous hypergraph. The main motivation is to add more information to the causality scenario and make the hypergraph more connected. To achieve this, we extract components from the events and create additional node types, as events can share these components. Furthermore, we add edges between the node-edge types, adding some relation between these nodes to enrich our heterogenous hypergraph;
4. After generating the heterogeneous hypergraph, we apply graph neural networks to obtain interpretable representations in a one-class learning setting. We explore three loss functions that jointly guide the learning process, enabling the model to learn meaningful representations and classify unlabeled edges as causal or non-causal using only causal examples. At this point, it is important to note that we are classifying a node in the heterogeneous hypergraph that represents an edge from the original graph.

Causal discovery between event pairs can be defined as a binary classification task, where the input consists of two events and the output is a label indicating whether a causal or non-causal relationship exists. Let $e_1, e_2, e_3, ..., e_m \in E$ represent a set of m natural language sentences, each corresponding to an event, and let $causal, non-causal \in C$ denote the set of possible labels. The classification function can be defined as $f : E \to C$, which maps pairs of events $(e_i, e_j) \in E$ to either the causal or non-causal label. In this context, $\boldsymbol{E} \in \mathbb{R}^d$ represents the feature space for each event's sentence, where $d = 384$[26]. One-class learning (OCL) for causal discovery in event pairs is defined as the function f^*, which learns from a training set consisting only

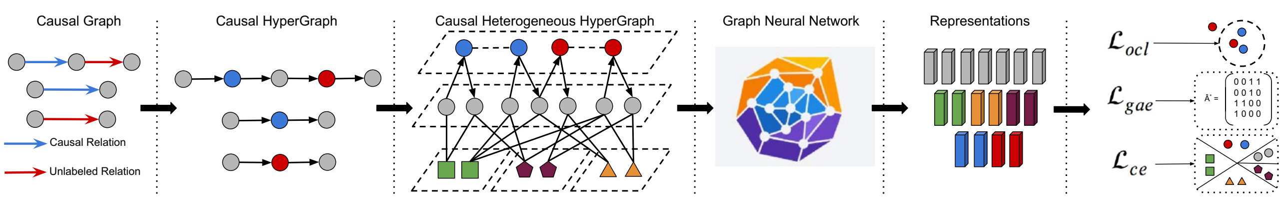

Fig. 2. Our proposed method eCHOLGA. We show all the steps from eCHOLGA: heterogeneous hypergraph generation, representation learning through GNNs, one-class sphere loss ($\mathscr{L}_{ocl}$), GAE loss ($\mathscr{L}_{gae}$), and cross-entropy loss for the node types ($\mathscr{L}_{ce}$).

of causal labels $((e_1, e_2); causal), ((e_5, e_3); causal), ..., ((e_i, e_j); causal)$, and approximates the unknown function *f*.

We model causality between event pairs using a directed graph. A directed graph is formally defined as $G = (V, \boldsymbol{A})$, where *V* is the set of nodes and $\boldsymbol{A}$ is the adjacency matrix representing the relationships between nodes. In this case, $V \equiv E$, meaning that the events themselves serve as the nodes, while $\boldsymbol{A}$ encodes the causal and non-causal relationships between them[27]. The directionality of edges is a natural property of causal relationships, as if one event causes another, the reverse is not necessarily true. Therefore, we preserve this directionality in the propagation process to avoid misleading the model during message passing. Although, in principle, directionality could be relaxed or treated as a parameter, since the causal graph can be transformed into a non-directional hypergraph, we deliberately maintain edge direction to ensure that information flows according to the causal order. This prevents the GNN from propagating information backward (from effects to causes), which could degrade the quality of learned representations. Our approach also incorporates Graph Neural Networks (GNNs), which are state-of-the-art models used for various tasks, including node classification[28]. However, due to GNNs' inherent limitations in learning edge representations and classifications, we turn to hypergraphs as an alternative representation for modeling causal relationships[7].

To overcome these limitations, we transform each edge in the directed graph into a node, thus converting the graph into a directed homogeneous hypergraph. Specifically, if node v_i is connected to node v_j, we introduce a new node v_o (referred to as a node-edge) and create connections between v_i and v_o, and between v_o and v_j [7]. We define a homogeneous hypergraph for causal discovery as $HG = (V^{hg}, \boldsymbol{A}^{hg})$, where V^{hg} represents all events represented by *V* and all edges of the original *G*, and $\boldsymbol{A}^{hg}$ represents all new edges between the events and node-edges nodes[7]. In this sense, the number of nodes in the hypergraph is the number of nodes and edges in the original graph[7]. The number of edges in the hypergraph is double the number of edges in the original graph, since for every two nodes with a relation, this is transformed into three nodes and two edges.

Through Hypergraphs, GNNs can now operate without the typical message-passing restrictions for edge representations, as the edges have effectively been transformed into nodes. Even after transforming the causal graph into a hypergraph, a problem still remains. Specifically, the resulting hypergraph may become disconnected. This issue arises when the dataset contains pairs of events that either never participate or only minimally participate in multiple relations. As a result, the hypergraph can become fragmented, consisting of several disconnected communities, each comprising three nodes: two representing the events and one representing the relation. This lack of connectivity between different parts of the graph can hinder the model's effectiveness in capturing global causal relationships.

One way to enhance the connectivity of the hypergraph is through enrichment, specifically by generating a heterogeneous hypergraph[25]. To enrich our hypergraph, we extracted several components from the event texts based on the 5W1H framework, Who, What, When, Where, Why, and How, which has been widely used in the literature to represent and enhance events in graph-based structures[29,30]. In our case, we selected the what, who, when, and where components, corresponding to the topic of the event, the entities involved, the temporal aspect, and the event location, respectively. We excluded the why and how components because they are highly abstract and open-ended, offering limited contribution to graph connectivity and often introducing noise during enrichment. To extract the topic component, we employed BERTopic[31], associating each event with a specific topic, except when the topic was labeled "−1" (a generic category for unclassified texts). For the who, when, and where components, we used large language models, specifically Google's Gemma2 (27B parameters). The prompts used for component extraction are available in the public code repository (see the Additional Information section). Each newly generated node was then represented by an embedding computed using BERT, based on its textual description. Finally, to ensure consistency among nodes, we applied a standardization process. For smaller datasets, this step was performed manually by the authors. For larger datasets, we adopted an automatic merging strategy based on embedding similarity, unifying nodes that referred to the same entity with minor textual variations.

We propose an additional enhancement to strengthen the GNN's message passing by connecting relation nodes with similar nodes. Specifically, this strategy aims to improve message passing between nodes representing causal relationships. To achieve this, we employ a strategy based on Large Language Models (LLMs) and the heterogeneous hypergraph. The strategy involves exploring the event texts, along with related heterogeneous nodes (such as *topic*, *who*, *where*, and *when*), into the LLM to query whether the relationship between the events is potentially causal. If the LLM responds affirmatively, we connect the relationship between the events to the three nearest causal training relationships, based on the cosine similarity of their BERT embeddings. Conversely, if the LLM responds negatively, we connect the relationship to the three nearest predicted negatively relation nodes. We define the set of new heterogeneous nodes *topic*, *who*, *where*, and *when* as V^{he}. Thus, we define our heterogeneous hypergraph as $HHG = (V^{hhg}, \boldsymbol{A}^{hhg})$, where $V^{hhg} = V^{hg} \cup V^{he}$, and $\boldsymbol{A}^{hhg}$ represents the edges $\boldsymbol{A}^{hg}$ plus the edges between the events and the heterogeneous nodes, and the enriched edges between the node-edges. We illustrate our definitions of graph, homogenous hypergraph, and heterogeneous hypergraph in Fig. 3.

We leverage Graph Neural Networks (GNNs) to learn representations in our causal hypergraph. The GNNs take as input the structured representation of each node, $\boldsymbol{v}_i \in \boldsymbol{V}^{hhg}$, and the adjacency matrix $\boldsymbol{A}^{hhg}$, both of which are critical for the representation learning process. In this paper, we explore the Graph Convolutional Network (GCN)[32] defined by Equation 1 and its aggregate and combine steps (Equation 2):

$$\boldsymbol{H}^{(l+1)} = \sigma(\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}}^{hhg} \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{H}^{(l)} \boldsymbol{W}^{(l)}), \tag{1}$$

$$\boldsymbol{h}^{l}_{v_i} = \sigma(\boldsymbol{W}^{(l)} \cdot MEAN\{\boldsymbol{h}^{l-1}_{v_j} : v_j \in \{N_{v_i} \cup v_i\}\}, \tag{2}$$

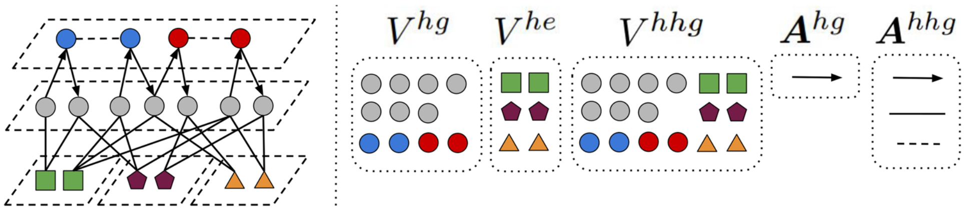

Fig. 3. Illustration of our homogeneous and heterogeneous hypergraph notations.

where, where $\boldsymbol{H}^{(l)}$ is the input to the l-th GNN layer, and $\boldsymbol{H}^{l+1}$ is the output of this layer, $\tilde{\boldsymbol{A}}^{hhg} = \boldsymbol{A}^{hhg} + \boldsymbol{I}$ is the adjacency matrix with self relations, $\boldsymbol{I}$ is the identity matrix, $\tilde{\boldsymbol{D}}$ is a diagonal matrix with $\tilde{\boldsymbol{D}} = \sum_j \tilde{\boldsymbol{A}}_{ij}^{hhg}$, σ is an activation function, $\boldsymbol{V}^{hhg} \equiv \boldsymbol{H}^{(0)}$, $\boldsymbol{W}^{(l)}$ represents trainable weights across the GNN layers $\mathrm{W} = \{\boldsymbol{W}^{(1)}, \cdots, \boldsymbol{W}^{(L)}\}$, and N_{v_i} are the v_i neighbors.

GNNs with sphere loss functions are state-of-the-art for one-class graph neural networks[13,33]. These methods learn $\boldsymbol{h}_{v_i}^{L}$ encapsulating the nodes of interest. To detect causality through our hypergraph, we explore this strategy[13]. We use the sphere loss function $\mathscr{L}_{ocl}$ defined in Equation 3.

$$\mathscr{L}_{ocl}(\boldsymbol{W}) = \frac{1}{|V^{\text{in}}|} \sum_{i=1}^{|V^{\text{in}}|} \begin{cases} o_i + 1, & \text{if } o_i > 0 \\ \exp(o_i), & \text{otherwise} \end{cases}, \tag{3}$$

$$o_i = \|\boldsymbol{h}_{v_i}^{(L)} - \boldsymbol{c}\|^2 - r^2, \tag{4}$$

where Eq. 4 represents the value indicating whether the interest instance v_i is within the hypersphere with radius r and center $\boldsymbol{c}$ and V^{in} are the set of interest nodes-edges in our hypergraph (with the causal class).

By using only the $\mathscr{L}_{ocl}$ loss, all nodes would converge to the center, leading to a collapse of the representations. To address this issue, following the approach in Gôlo et al.[13], we incorporate our Graph Convolutional Network (GCN) layer into a Graph Autoencoder (GAE). GAEs have an unsupervised loss function that acts as a constraint to prevent this collapse, thus ensuring the nodes do not all converge to the center. Therefore, we combine the sphere loss function with the loss function of GAEs[13], which has demonstrated superior and state-of-the-art results compared to other methods[33]. The GAE utilizes GNN layers as an encoder and the inner product of the latent representations as a decoder to learn effective node embeddings. Equation 5 provides the formal definition of a GAE[34]:

$$GAE = \begin{cases} Encoder: \boldsymbol{H}^{(L)}, \hat{\boldsymbol{T}} = g(\mathrm{V}^r, \boldsymbol{A}^r; \mathrm{W}) \\ Decoder: \hat{\boldsymbol{A}}^r = \sigma(\boldsymbol{H}^{(L)} \cdot \boldsymbol{H}^{(L)\intercal}) \end{cases}, \tag{5}$$

$$\mathscr{L}_{gae}(\boldsymbol{W}) = -\frac{1}{|V|} \sum_{i=0}^{|V|} \sum_{j=0}^{|V|} (\boldsymbol{A}_{ij}^r \cdot log\hat{\boldsymbol{A}^r}_{ij} + (1 - \boldsymbol{A}_{ij}^r) \cdot log(1 - \hat{\boldsymbol{A}^r}_{ij})), \tag{6}$$

where $\hat{\boldsymbol{T}}$ is the predicted vector of node types by our model (will be used in our third loss in Equation 7), V^r and $\boldsymbol{A}^r$ are the node-edges (relations) and the adjacency matrix for these node-edges, i.e., we apply the reconstruction loss only in the relation nodes. $\sigma(.)$ is a logistic sigmoid function. The GAE loss function $\mathscr{L}_{gae}$ is defined in Equation 6 (binary cross entropy loss applied in the adjacency matrix).

To construct our triple loss and leverage the benefits of building a heterogeneous hypergraph, we incorporate a cross-entropy loss for node types, contributing to our heterogeneous loss. By adding heterogeneity components, we influence both the sphere and reconstruction loss functions. In addition to applying the sphere loss function to the primary relation nodes in the heterogeneous hypergraph, we also introduce a cross-entropy loss function for the node types (event, relation, topic, who, where, and when). This strategy aims to enhance the model's ability to correctly predict the node type based on its learned representation. Our motivation is that by learning representations that clearly distinguish between node types, the model can better separate interest nodes from non-interest nodes through the sphere loss. This loss function is defined as follows:

$$\mathscr{L}_{ce}(\boldsymbol{W}) = -\frac{1}{|V|} \sum_{i=0}^{|V|} (\boldsymbol{T}_i \cdot log\hat{\boldsymbol{T}}_i + (1 - \boldsymbol{T}_i) \cdot log(1 - \hat{\boldsymbol{T}}_i)). \tag{7}$$

where $\boldsymbol{T}$ is the vector with the node types and $\hat{\boldsymbol{T}}$ is the node types predicted vector.

When combining multiple loss functions, a common challenge is balancing their scales, as one loss function could dominate the learning process and cause other tasks to be neglected[35]. To address this issue, we introduce three impact factors, α, β, and δ, to balance the contributions of each loss function. Additionally, graph representation learning may require adjustments in these impact factors at different stages of the learning process to optimize both node representation learning and causal discovery performance. For example, during the early stages of training, we assign more weight to the cross-entropy loss ($\mathscr{L}_{ce}$) to help the model distinguish between node types. As the training progresses, we gradually shift the focus toward $\mathscr{L}_{ocl}$ and $\mathscr{L}_{rec}$, ensuring that the representation of relation nodes aligns with the sphere, with causal relations encapsulated within the sphere and non-causal relations positioned outside. To facilitate this, we propose a strategy for dynamically adjusting the impact factors during training. Our final loss function is therefore defined as: $\mathscr{L} = (\mathscr{L}_{ocl} * \alpha) + (\mathscr{L}_{gae} * \beta) + (\mathscr{L}_{ce} * \delta)$, where α, β, and δ are impact factors for the losses.

We present the causal graph, causal hypergraph, causal heterogeneous hypergraph, the GCN step for representation learning in the heterogeneous hypergraph, as well as the sphere loss, GAE loss, and node type loss. In summary, our proposal not only introduces a novel approach for causal event representation through heterogeneous hypergraphs but also emphasizes interpretability as a central design principle. Interpretability has become increasingly important due to the growing demand for transparent and explainable models across different research domains. To this end, we propose that eCHOLGA inherently learns three-dimensional representations, thereby providing interpretability and enabling us to better understand the real-time representation learning process. This choice allows direct interpretation through the visualization of the latent space without the need for additional dimensionality reduction. Moreover, this 3D structure enables a clear understanding of the model decision since instances inside the sphere correspond to causal node–edges, while those outside represent non-causal ones. Overall, the three-dimensional design of eCHOLGA goes beyond improving performance. It transforms the model into an interpretable and interactive system, bridging the gap between representation learning and interpretability. This design choice provides researchers and practitioners with a concrete tool to observe, analyze, and trust the causal reasoning process, reinforcing interpretability as a core contribution of our proposal.

Experimental evaluation

This section presents the experimental evaluation. We present the four used datasets, experimental settings, results, and discussion. Our research goal is to demonstrate that our eCHOLGA proposal outperforms other SOTA methods for causal discovery in event pairs, through the new heterogeneous hypergraph proposed. Another goal is to demonstrate that our method learns low-dimensional representations, providing interpretability, even considering the heterogeneous scenario.

Datasets

We explore four causality datasets, each consisting of event pairs, where each event is represented by a text description. Each event pair is labeled as either *causal* or *non_causal*. The first dataset is the Risk Models dataset[2], which was introduced by[2] to analyze models created by expert analysts for configuring decision support systems[36]. These models are structured as graphs, where the nodes represent textual descriptions of conditions or events, and the edges denote causal relationships. The models are based on enterprise risk management, expert knowledge, literature reviews, and reports. To create the cause-effect pairs dataset, the authors transformed each edge in the graph into a pair of texts with a causal label.

The second dataset is FinCausal, which focuses on detecting causality associated with a quantified fact. In this dataset, an event refers to the emergence of a new object or context relative to a previous situation. Therefore, the dataset emphasizes detecting causality related to the transformation of financial objects embedded in quantified facts. The third dataset is Headlines[37], which focuses on collecting implicit causal relationships between sentences. This dataset consists of pairs of English and Russian news headlines, each labeled with a causality tag obtained through crowdsourcing. The authors intentionally excluded texts from the full news articles, as the headlines themselves were considered representative of events. The labels were assigned based on the rule that the first headline causes the second if the second headline would be impossible without the first. In other words, if the first event did not occur, the second event could not happen either.

The fourth dataset is the Twitter dataset[38], which contains tweets labeled as either *causal* or *non-causal*. The tweets in this dataset are related to the Commonwealth Games held on the Gold Coast in 2018. The dataset was manually annotated. Table 2 presents the details of the datasets. The Risk dataset comprises 402 causal pairs with 223 unique events. For the non-interest class, 402 pairs were randomly selected to be labeled as non-causal. The FinCausal dataset includes 536 causal pairs with 860 unique events. For the non-interest class, 540 non-causal pairs were randomly chosen. The Headlines dataset has 909 causal pairs, 1603 non-causal pairs, and 4463 unique events. The Twitter dataset comprises 459 causal pairs, 457 non-causal pairs, and 1772 events. Table 2 also provides the node information after hypergraph enrichment. Thus, Risk and FinCausal generate connected hypergraphs, while Headlines and Twitter generate disconnected hypergraphs, since they present a scenario closer to the real world with naturally obtained counterexamples.

Name	Causal	Non_causal	Events	Relation	Topic	Who	When	Where
Risk	402	402	223	769	2	27	0	3
FinCausal	536	540	860	986	25	66	16	23
Headlines	909	1603	4463	2494	101	2484	57	606
Twitter	459	457	1772	905	32	508	33	88

Table 2. Dataset details for the number of causal and non-causal event pairs, and the number of event, relation, topic, who, when, and where nodes in the heterogeneous hypergraph generated for each dataset.

Experimental setting

We compare the eCHOLGA with the state-of-the-art text-based methods for causal discovery considering BERT, such as NLM-BERT, NLM-BERT++, and CF-Context[2–4] and LLM methods[5]. We compare our method with five large language models using the strategy of Pywhy-LLM library. Different LLMs have been proposed in the last three years, and they have differences that generate advantages and disadvantages for each model[39]. We use the 5-fold cross-validation for our experiments. We use four folds of the causal class to train, the remaining fold to test, and one fold of the non-causal class to test. We use the f_1-macro to compare all models. We use BERT embeddings as representations of event-type nodes and enriched nodes (*topic*, *who*, *when*, and *where*). Furthermore, we use the average of the two event embeddings as the initial representation of the relation node type (node-edges). Finally, we execute the experimental evaluation on a machine with an Ubuntu 24 computer with an i9-14900KF CPU, RTX A5000 (24 GB RAM), and 128 GB RAM. In this computer setting, the LLM gemma2 with 27 billion parameters extracts the components to enrich one event in 2.5 seconds.

We explore in our methodology four families of LLMs open-source: the LLM from meta (LLaMa)[40], from Microsoft (Phi)[41], from Google (Gemma)[42], and from Alibaba Cloud (Qwen)[43]. Each LLM has a number of parameters: Llama 3 (8 and 70 billion of parameters), Phi 3 (14 billion of parameters), Gemma 2 (27 billion of parameters), and Qwen 2 (7 billion of parameters). We compare eCHOLGA with one-class methods since we have an initial representation for each sentence (BERT embedding) and can generate an initial representation for causal relationships (average of causal and effect sentences). In this sense, we explore two one-class methods: One-Class Support Vector Machines (OCSVM)[44] and Isolation Forest (IsoForest)[45]. Finally, we compare eCHOLGA with the eCOLGAT[12]. We use the following parameters for the methods:

- **LLMs**: Temperature = 0 and max_sequence_length = 10240;
- **NLM-BERT, NLM-BERT++**: thresholds = $\{0, 0.3, 0.6, 0.9\}$;
- **CF-Context**: epochs = $\{300\}$, patience = $\{3\}$, walkSize = $\{1\}$, samples = $\{100\}$, $\rho = 0.025$, and joint = $\{0\}$;
- **OCSVM**: kernel = {rbf, poly, sigmoid, linear}, $\nu = \{0.05 * b\}, b \in [1..19]$, and $\gamma = \{$ scale, auto $\}$;
- **IsoForest**: n° of estimators = $\{1, 2, 5, 10, 50, 100, 200, 500\}$, maximum samples and features = $\{0.1 * b\}, b \in [1..10]$;
- **eCOLGAT**: radius = $\{0.35, 0.45, 0.5\}$, epochs = $\{700, 1000, 1500\}$, heads= $\{1, 2, 3\}$, lr = $\{0.001, 0.0001, 0.0005\}$;
- **eCHOLGA**: radius = $\{0.3, 0.4, 0.5\}$, epochs = $\{6000, 3000\}$, and learning rate = $\{0.008\}$. enriched relation edges = {Qwen2.5:14b, Phi4:14b}. α, β, and δ have initial and final values and are increased (α, β) or decreased (δ) considering 100 epochs (During our initial empirical tests, we explored different initial, final, and incremental values for α, β, and δ, in which the best values on at least one dataset were selected to compose the hyperparameter set). final_$\alpha = \{0.8, 0.7, 0.6, 0.5\}$, final_$\beta = \{0, 0.15, 0.2, 0.3\}$, final_$\delta = \{0.5, 0.1, 0.05\}$, initial_$\alpha = \{0\}$, initial_$\beta = \{0\}$, initial_$\delta = \{1\}$, iteration_$\alpha = \{0.02, 0.04, 0.05\}$, iteration_$\beta = \{0.01\}$, iteration_$\delta = \{0.02, 0.04, 0.05\}$.

Results and discussion

Table 3 presents the results of our study, showing the average f_1 macro across five folds. We evaluate three baseline methods (NLM-BERT, NLM-BERT++, and CF-Context) based on their previous results on the Risk and Twitter datasets. We show results from eight models alongside our proposed method, eCHOLGA, across all four datasets. The highest-performing models are highlighted in bold, while the second-best results are underlined. eCHOLGA achieves state-of-the-art performance in three datasets, while CF-Context outperforms other methods in the Twitter dataset. Gemma2 obtains the second-best results in the Headlines and FinCausal datasets, whereas eCOLGAT and NLM-BERT++ achieve second-best performance in the Risk and Twitter datasets, respectively. Notably, two-step methods (BERT + OCL algorithm) yield the lowest overall results.

Method	Risk	Twitter	Headlines	FinCausal
NLM-BERT	0.669	–	–	–
NLM-BERT++	0.667	<u>0.668</u>	–	–
CF-Context	0.624	**0.673**	–	–
BERT + IsoForest	0.324	0.535	0.488	0.502
BERT + OCSVM	0.672	0.497	0.617	0.544
LLaMa 3 (8b)	0.727	0.570	0.610	0.761
Phi 3 (14b)	0.659	0.509	0.626	0.695
Qwen 2 (7b)	0.691	0.472	0.654	0.719
Gemma 2 (27b)	0.720	0.630	<u>0.734</u>	<u>0.822</u>
LLaMa 3 (70b)	0.726	0.623	0.728	0.767
eCOLGAT	<u>0.771</u>	0.563	0.661	0.512
eCHOLGA	**0.776**	0.564	**0.806**	**0.824**

Table 3. f_1-macro for each method in the four datasets. The best results are in bold, and the second best are underlined. We report the results of NLM-BERT, NLM-BERT++, and CF-Context from the original papers. We update the results from LLMs in the Risk dataset due to the prompt update.

In the Risk dataset, eCOLGAT achieves a competitive second-best result, nearly matching eCHOLGA. The third-best model, LLama 3 (8B), lags by approximately 5%, underscoring the robustness of both one-class and graph-based methods for this dataset. A similar pattern emerges in the FinCausal dataset, with Gemma2 securing the second-best performance. However, LLama 3 (70B) follows closely, performing only 6% worse than eCHOLGA, indicating that this LLM is particularly effective in this dataset. For the Headlines and Twitter datasets, the performance gap between eCHOLGA and the best-performing models is more pronounced. In the Headlines dataset, eCHOLGA outperforms Gemma2 by 7%, demonstrating its effectiveness in this context. However, in the Twitter dataset, a BERT-based method outperforms eCHOLGA by 10%, highlighting that even non-graph-based models can excel in causal discovery. This also reinforces the importance of using BERT embeddings as the initial representation of nodes in graph-based methods like eCHOLGA. Additionally, the informal and concise nature of Twitter texts presents unique challenges for causal discovery, likely contributing to the observed performance differences.

In the Risk dataset, transforming the hypergraph into a heterogeneous hypergraph did not significantly impact performance, as eCOLGAT and eCHOLGA achieved similar results. This can be attributed to how non-causal instances were generated by randomly pairing two sentences, resulting in a connected hypergraph. Nonetheless, even in this already connected scenario, our proposed enrichment strategy demonstrated its potential to enhance performance. This effect was particularly evident in the FinCausal dataset, which also contains synthetically generated non-causal pairs, where eCHOLGA significantly outperformed eCOLGAT. However, the opposite is also true: in a disconnected hypergraph scenario, enrichment does not always guarantee performance gains. This was observed in the Twitter and Headlines datasets, where non-causal instances naturally occur. While enrichment led to substantial improvements in the Headlines dataset, its impact on the Twitter dataset was less pronounced.

Our strongest competitor was LLMs. Although LLMs are considered unsupervised methods, their immense computational power stems from having been trained on trillions of texts, potentially including causal and non-causal event pairs. This makes outperforming them a challenging task. Nevertheless, eCHOLGA outperformed LLMs in two datasets (Headlines and Risk) and achieved competitive, near-equivalent performance in the other two (FinCausal and Twitter). Notably, only in one of the evaluated scenarios LLMs outperform eCHOLGA, highlighting that despite their robustness, they also face challenges in event causal discovery. Another key observation is that increasing the number of parameters in LLMs does not lead to better results. Moreover, models with more parameters require significantly more computational resources, which is a disadvantage.

eCOLGAT served as the basis for eCHOLGA. In the Headlines and FinCausal datasets, eCHOLGA outperformed eCOLGAT by 25–30% in macro f_1 score. This demonstrates the impact of hypergraph enrichment and the new triple heterogeneous loss, two key improvements in eCHOLGA over eCOLGAT, that led to better performance. In contrast, the results of eCOLGAT and eCHOLGA were similar for the Risk and Twitter datasets. While enrichment did not significantly improve performance in these cases, it offers additional advantages beyond accuracy. A richer hypergraph structure enables better explainability techniques to interpret causality. Furthermore, with more diverse information embedded in the graph, new composite loss functions can be explored to enhance learning, an opportunity that is more limited in homogeneous hypergraphs.

The use of LLMs for event enrichment has shown consistently positive results, both in our experiments and in related literature[46,47]. Studies have demonstrated that LLMs are particularly effective in extracting 5W1H components, Who, What, When, Where, Why, and How, from unstructured text, which are crucial elements for representing events in structured forms[46,47]. While such models can occasionally produce incorrect inferences, their overall accuracy and ability to capture semantically meaningful relationships generally outweigh these errors. Nevertheless, before applying the enrichment strategy, it is advisable to conduct preliminary evaluations to assess the LLM's reliability for the specific domain or dataset. In our case, this approach proved beneficial because, despite potential inaccuracies, the LLM-driven enrichment led to clear performance improvements, confirming that the advantages of using LLMs for enrichment surpass their limitations.

Figure 4 illustrates the interpretability results for the Headlines, FinCausal, and Risk datasets. These figures present the learned representations generated by eCHOLGA, showcasing its interpretability capabilities. Specifically, we visualize the representations of all nodes and node-edges in our heterogeneous hypergraph, highlighting the learning process. Since our learned representations are three-dimensional, it is possible to create a real-time video of the learning process without additional post-processing. Each figure contains eight plots. In the top four plots, each color represents a different node type. In the bottom four, yellow points indicate non-causal relation nodes, while purple points represent causal relation nodes. It is important to note that the color schemes in the top and bottom plots do not correspond to the same node types. We selected four key stages of the learning process to illustrate the evolution of the embeddings: (1) the initial stage (0–1% of training), (2) an early intermediate stage (25% of training), (3) a late intermediate stage (75%), and (4) the final stage (model convergence).

The interpretability of eCHOLGA allows us to observe the learning dynamics, as demonstrated in Fig. 4. Initially, eCHOLGA prioritizes node type prediction, as the impact factor of the cross-entropy loss ($\mathscr{L}_{ce}$) is set higher at the beginning. As training progresses, the one-class loss ($\mathscr{L}_{ocl}$) gains more influence, causing relation nodes, particularly causal relations, to move toward the sphere. Meanwhile, the reconstruction loss ($\mathscr{L}_{gae}$) also becomes more significant, leading to a more structured node distribution. In the final training stages, all three losses work together: causal relations are clustered near the sphere (driven by $\mathscr{L}_{ocl}$), non-causal relations remain outside the sphere (regulated by $\mathscr{L}_{gae}$), and node types remain well-separated, further reinforcing the one-class loss ($\mathscr{L}_{ce}$).

We propose an ablation study to assess the impact of the different components of our composed loss function and the strategy used to enrich the edges between relation nodes, as discussed in Section "eCHOLGA: edge classification through heterogeneous one-cLAss graph autoencoder". We evaluate four configurations: (i) using

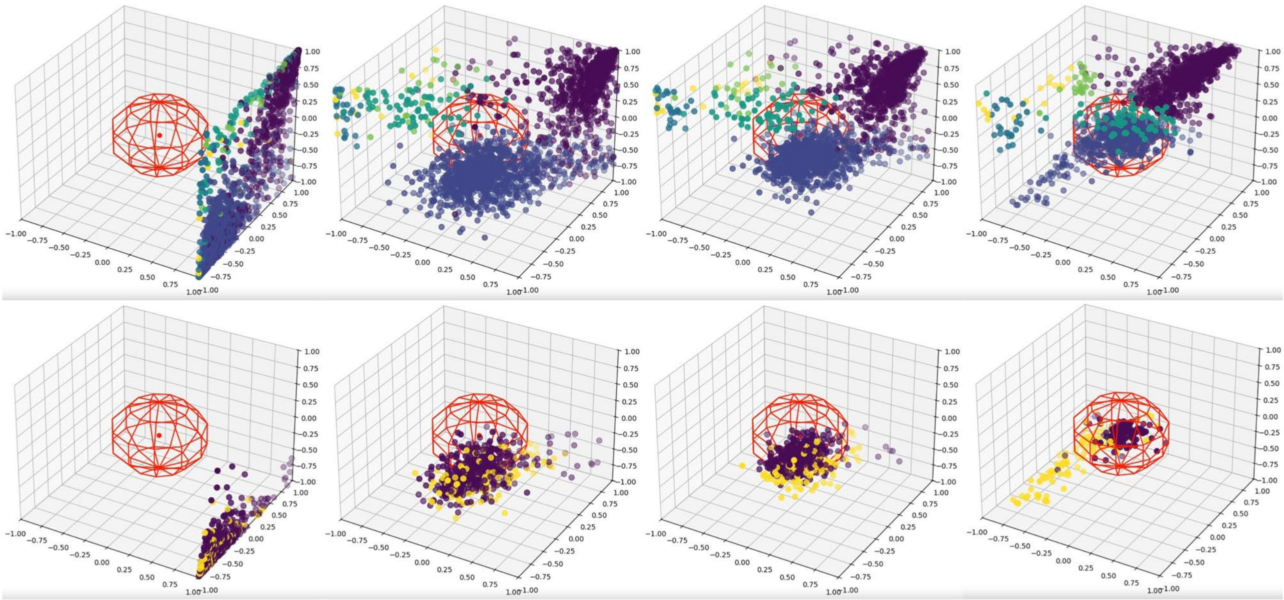

Fig. 4. Interpretability analysis of the three-dimensional representations of eCHOLGA in the FinCausal dataset. In the top four plots, each color indicates one node type, and in the bottom four plots, the colors indicate the causal (purple) and the non-causal (yellow) classes.

Scenarios	Risk	Twitter	Headlines	FinCausal
$\mathscr{L}_{ocl}$	0.324	0.334	0.264	0.312
$\mathscr{L}_{ocl} + \mathscr{L}_{rec}$	0.657	0.496	0.264	0.394
$\mathscr{L}_{ocl} + \mathscr{L}_{rec} + \mathscr{L}_{ce}$	0.619	0.488	0.572	0.415
$\mathscr{L}_{ocl} + \mathscr{L}_{rec} + \mathscr{L}_{ce}$ + enriched edges	0.776	0.564	0.806	0.824

Table 4. Ablation study for eCHOLGA considering its loss components and the enriched node-edges relations.

only the OCL function ($\mathscr{L}_{ocl}$), similar to the method of Wang et al.[33], (ii) using a combination of one-class and reconstruction losses ($\mathscr{L}_{ocl} + \mathscr{L}_{rec}$), similar to the method of Gôlo et al.[13], (iii) using all three components without node-edge relations strategy ($\mathscr{L}_{ocl} + \mathscr{L}_{rec} + \mathscr{L}_{ce}$), our proposal without enriched edges between node-edges through LLMs, and (iv) using the full setup of eCHOLGA (results from Table 3), which includes all loss components along with the strategy for adding edges between relation-type nodes. Table 4 summarizes the ablation results. The best values are shown in bold.

Through the ablation study, we observe that using only a one-class loss in a heterogeneous graph is not sufficient to predict causality with high performance. When the reconstruction loss is added, performance improves, but it still falls short of the combination of all three losses. Using the three losses together improved results in 50% of the datasets, and in the remaining 50% where performance did not improve, it also did not degrade significantly. This suggests that the third loss generally contributes positively to learning in heterogeneous graphs. Finally, the best results were achieved by implementing the triplet loss and the enrichment of the heterogeneous hypergraph by adding edges between relation-type nodes (original graph edges transformed into nodes). Therefore, we identify the step of constructing edges between relation-type nodes as one of the key stages of eCHOLGA, highlighting the importance of these connections for learning. This finding also indicates that the GNN's aggregation steps benefited from these enriched edges, which, when added, improved the GNN's performance.

Conclusion and future work

In this study, we introduce eCHOLGA, a novel end-to-end and interpretable method for causal discovery between event pairs. eCHOLGA leverages graph neural networks, heterogeneous hypergraphs, and one-class learning to enhance causal inference. It employs a hypergraph structure to improve edge representation learning via graph neural networks. To further enrich and optimize learning, we propose transforming the hypergraph into a heterogeneous hypergraph, mitigating issues that arise when hypergraphs are disconnected. Our approach introduces a new heterogeneous triple-loss function, incorporating: (i) a sphere loss function to enforce one-class learning, (ii) a reconstruction loss function, and (iii) a heterogeneous loss function to predict node types. The sphere loss function enables eCHOLGA to operate as a one-class learning method, confining the model's focus to causal relations without requiring exhaustive non-causal relation coverage. Additionally, eCHOLGA utilizes three-dimensional representations, providing interpretability in the causal discovery process for event pairs.

Our results demonstrate that eCHOLGA outperforms eleven competing methods, including LLM-based, BERT-based, and one-class-based approaches, in three out of four datasets. To illustrate the interpretability of our model, we visualize the three-dimensional learned embeddings, revealing how eCHOLGA integrates multiple loss functions. These plots show how eCHOLGA separates node types while encapsulating causal instances within the sphere, where node-type and reconstruction losses act as constraints to the one-class loss. This visualization enables a deeper understanding of eCHOLGA's learning process. Selecting an appropriate causal discovery method depends on the user's needs. If a user prioritizes ease of use and does not require model training or parameter tuning, LLMs are the most convenient choice, as they perform well without additional training. However, if the goal is to understand and refine results while maintaining interpretability, eCHOLGA is the superior choice. With minimal training and labeling of only causal event relations, eCHOLGA delivers competitive or even superior results to LLMs while offering interpretability.

eCHOLGA also has some limitations. The method relies on LLMs to establish connections between relation nodes, meaning that poor LLM performance can negatively impact eCHOLGA's results, as observed in the Twitter dataset. Thus, the approach as a whole depends on the quality of the chosen LLM and its performance. To address this, future work will explore alternative strategies for connecting similar relation nodes, reducing dependence on LLMs. Potential solutions include integrating LLMs with additional methods through consensus or ensemble strategies. Additionally, we plan to investigate alternative heterogeneous GNN layers to further enhance eCHOLGA's performance. We highlight that it is pertinent to investigate in future work which combinations of hyperparameters of different LLMs generate better heterogeneous hypergraphs with better cost-benefit, considering different methods to detect causality. Finally, Future work could explore domain adaptation techniques for causal graphs, enabling transfer of causal knowledge across domains or datasets[48].

Data availability

Source codes and datasets: https://github.com/GoloMarcos/eCHOLGA. There are no Competing Interests.

Received: 30 May 2025; Accepted: 12 November 2025
Published online: 21 November 2025

References

1. Kayesh, H., Islam, M. S. & Wang, J. Event causality detection in tweets by context word extension and neural networks. In *Int. Conf. on parallel and distributed computing, applications and technologies* 352–357 (IEEE, 2019).
2. Hassanzadeh, O. et al. Answering binary causal questions through large-scale text mining an evaluation using cause-effect pairs from human experts. *IJCAI*, 5003–5009 (2019).
3. Kayesh, H. et al. Answering binary causal questions: A transfer learning based approach. In *2020 International Joint Conference on Neural Networks (IJCNN)* 1–9 (IEEE, 2020).
4. Kayesh, H., Islam, M. S. & Wang, J. Answering binary causal questions using role-oriented concept embedding. *IEEE Trans. Artif. Intell.* **4**, 1426–1436 (2023).
5. Liu, X. *et al.* Large language models and causal inference in collaboration: A comprehensive survey. arXiv:2403.09606 (2024).
6. Zanga, A., Ozkirimli, E. & Stella, F. A survey on causal discovery: Theory and practice. *Int. J. Approximate Reason.* **151**, 101–129 (2022).
7. Jo, J. et al. Edge representation learning with hypergraphs. In *NeurIPS* 7534–7546 (NeurIPS foundation, 2021).
8. Sasaki, H., Fujii, M., Sakaji, H. & Masuyama, S. Enhancing risk analysis with gnn: Edge classification in risk causality from securities reports. *Int. J. Inf. Manag. Data Insights* **4**, 100217 (2024).
9. Bing, R. et al. Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications. *Artif. Intell. Rev.* **56**, 8003–8042 (2023).
10. Seliya, N., Abdollah Zadeh, A. & Khoshgoftaar, T. M. A. literature review on one-class classification and its potential applications in big data. *J. Big Data* **8**, 1–31 (2021).
11. Emmert-Streib, F. & Dehmer, M. Taxonomy of machine learning paradigms: A data-centric perspective. *Wiley Interdis. Rev.: Data Min. Knowl. Disc.* **12**, e1470 (2022).
12. Gôlo, M. P. S. & Marcacini, R. M. One-class learning for text causal discovery through hypergraph neural networks. *II Workshop on Causal Discovery (CaDis)* , 1–12 (2024).
13. Gôlo, M. P. S., de Medeiros Junior, J. G. B., Silva, D. F. & Marcacini, R. M. One-class graph autoencoder: A new end-to-end, low-dimensional, and interpretable approach for node classification. *Inf. Sci.* **708**, 122060 (2025).
14. Zhang, F., Fan, H., Wang, R., Li, Z. & Liang, T. Deep dual support vector data description for anomaly detection on attributed networks. *Int. J. Intell. Syst.* **37**, 1509–1528 (2022).
15. Wan, G., Wu, Y., Hu, M., Chu, Z. & Li, S. Bridging causal discovery and large language models: A comprehensive survey of integrative approaches and future directions. arXiv:2402.11068 (2024).
16. Minghim, R., Milos, E., Provia, K. *et al.* Ggnn@ causal news corpus 2022: Gated graph neural networks for causal event classification from social-political news articles. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, 85–90 (2022).
17. Sakaji, H. & Izumi, K. Financial causality extraction based on universal dependencies and clue expressions. *N. Gener. Comput.* **41**, 839–857 (2023).
18. Ju, W. et al. A comprehensive survey on deep graph representation learning. *Neural Netw.* **173**, 106207 (2024).
19. Zhao, Y. et al. Dynamic hypergraph structure learning for traffic flow forecasting. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)* 2303–2316 (IEEE, 2023).
20. Ju, W. et al. Hypergraph-enhanced dual semi-supervised graph classification. In *International Conference on Machine Learning* 22594–22604 (PMLR, 2024).
21. Yi, S. et al. Hypergraph consistency learning with relational distillation. *IEEE Trans. Multimed.* **27**, 7028–7039 (2025).
22. Gao, Y., Feng, Y., Ji, S. & Ji, R. Hgnn+: General hypergraph neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 3181–3199 (2022).
23. Wu, H. et al. Collaborative contrastive learning for hypergraph node classification. *Pattern Recogn.* **146**, 109995 (2024).
24. Wu, H., Yan, Y. & Ng, M.K.-P. Hypergraph collaborative network on vertices and hyperedges. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 3245–3258 (2022).

25. Mohammed, S., Getahun, F. & Chbeir, R. Semantic event relationships identification and representation using hypergraph in multimedia digital ecosystem. *J. Intell. Inf. Syst.* **60**, 463–493 (2023).
26. Reimers, N. & Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 1–11 (Association for Computational Linguistics, 2019).
27. Ma, J. *et al.* Learning causal effects on hypergraphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1202–1212 (2022).
28. Jin, M. et al. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *Trans. Pattern Anal. Mach. Intell.* **46**, 10466–10485 (2024).
29. Mattos, J. P. R. & Marcacini, R. M. Semi-supervised graph attention networks for event representation learning. In *2021 IEEE International Conference on Data Mining (ICDM)* 1234–1239 (IEEE, 2021).
30. Hamborg, F., Lachnit, S., Schubotz, M., Hepp, T. & Gipp, B. Giveme5w: main event retrieval from news articles by extraction of the five journalistic w questions. In *International conference on information* 356–366 (Springer, 2018).
31. Grootendorst, M. Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint (2022).
32. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*, 1–14 (Open Review, Toulon, France, 2017).
33. Wang, X. et al. One-class graph neural networks for anomaly detection in attributed networks. *Neural Comput. Appl.* **33**, 12073–12085 (2021).
34. Kipf, T. N. & Welling, M. Variational graph auto-encoders. *stat* **1050**, 21 (2016).
35. Qin, Y., Wang, X., Zhang, Z., Chen, H. & Zhu, W. Multi-task graph neural architecture search with task-aware collaboration and curriculum. In *Thirty-seventh Conference on Neural Information Processing Systems*, 1–13 (2023).
36. Sohrabi, S. et al. Ibm scenario planning advisor: Plan recognition as ai planning in practice. *AI Commun.* **32**, 1–13 (2019).
37. Gusev, I. & Tikhonov, A. Headlinecause: A dataset of news headlines for detecting causalities. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 6153–6161 (2022).
38. Kayesh, H., Islam, M. S. & Wang, J. On event causality detection in tweets. arXiv:1901.03526 (2019).
39. Zhao, W. X. *et al.* A survey of large language models. arXiv (2023).
40. Meta. The llama 3 herd of models (2024). arXiv:2407.21783.
41. Microsoft. Phi-3 technical report: A highly capable language model locally on your phone (2024). arXiv:2404.14219.
42. Gemma. Gemma 2: Improving open language models at a practical size (2024). arXiv:2408.00118.
43. Alibaba. Qwen technical report (2023). arXiv:2309.16609.
44. Alam, S., Sonbhadra, S. K., Agarwal, S. & Nagabhushan, P. One-class support vector classifiers: A survey. *Knowl.-Based Syst.* **196**, 105754 (2020).
45. Liu, F. T., Ting, K. M. & Zhou, Z.-H. Isolation forest. In *int. conf. on data mining* 413–422 (IEEE, 2008).
46. Grande, E. & Begga, A. Syntax savants-ua at iberlef 2024: Leveraging flan-t5-xxl for automatic 5w1h identification in texts. In *Proc. of the Iberian Languages Evaluation Forum (IberLEF 2024)* (CEUR, 2024).
47. Akter, M. & Santu, S. K. K. Fans: A facet-based narrative similarity metric. arXiv:2309.04823 (2023).
48. Luo, J. et al. Gala: Graph diffusion-based alignment with jigsaw for source-free domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**, 9038–9051 (2024).

Acknowledgements
We also acknowledge support from the Google Latin America PhD Fellowship. We would like to thank the Google mentor, Mara Finkelstein, for her valuable feedback and support in reviewing this manuscript.

Author contributions
M. P. S. Gôlo: Writing, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. R. M. Marcacini: Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization.

Funding
This work was partially supported by CAPES (Process No. 88887.671481/2022-00), CNPq (Process No. 316507/2023-7), and the São Paulo Research Foundation (FAPESP), Brazil (Grants 2024/15430-5, 2022/09091-8, 2023/10100-4, 2019/07665-4, and 2013/07375-0).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.P.S.G. or R.M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.