

Temporal Inductive Logic Reasoning over Hypergraphs

Yuan Yang¹, Siheng Xiong¹, Ali Payani², James C. Kerce¹ and Faramarz Fekri¹

¹Georgia Institute of Technology

²Cisco

{yyang754@, sxiong45@, clayton.kerce@gtri., faramarz.fekri@ece.}gatech.edu, apayani@cisco.com

Abstract

Inductive logic reasoning is a fundamental task in graph analysis, which aims to generalize patterns from data. This task has been extensively studied for traditional graph representations, such as knowledge graphs (KGs), using techniques like *inductive logic programming* (ILP). Existing ILP methods assume learning from KGs with static facts and binary relations. Beyond KGs, graph structures are widely present in other applications such as procedural instructions, scene graphs, and program executions. While ILP is beneficial for these applications, applying it to those graphs is nontrivial: they are more complex than KGs, which usually involve timestamps and n-ary relations, effectively a type of hypergraph with temporal events. In this work, we propose *temporal inductive logic reasoning* (TILR), an ILP method that reasons on temporal hypergraphs. To enable hypergraph reasoning, we introduce the *multi-start random B-walk*, a novel graph traversal method for hypergraphs. By combining it with a path-consistency algorithm, TILR learns logic rules by generalizing from both temporal and relational data. To address the lack of hypergraph benchmarks, we create and release two temporal hypergraph datasets: YouCook2-HG and nuScenes-HG. Experiments on these benchmarks demonstrate that TILR achieves superior reasoning capability over various strong baselines.

1 Introduction

The task of inductive reasoning concerns generalizing concepts or patterns from data. This task is studied extensively in knowledge graphs (KG) where techniques such as *inductive logic programming* (ILP) are proposed. A typical knowledge graph such as FB15K [Toutanova and Chen, 2015] and WN18 [Bordes *et al.*, 2013] represents commonsense knowledge as a set of nodes and edges, where entities are the nodes and the facts are represented as the edges that connect the entities. For example, `Father(Bob, Amy)` is a fact stating that “Bob is the Father of Amy”; this is represented as an edge of type `Father` connecting the two entities `Bob` and `Amy`. Many ILP techniques are proposed to reason on the

KGs that learn first-order logic rules from the graph. Learning these rules is beneficial as they are interpretable and data efficient [Yang and Song, 2020]. It has applications such as biomedical research, semantic search, data integration and fraud detection.

Apart from KGs, the graph is widely used in other applications to represent structured data, such as temporal events that happened among a set of entities [Boschee *et al.*, 2015; Leetaru and Schrodt, 2013]; the cooking instruction of a video (Figure 1); the abstract syntax tree (AST) of a program; and a scene graph from autonomous driving sensors [Caesar *et al.*, 2020]. Learning explicit rules from these graphs is also beneficial. However, these graphs are more complex than standard graphs, and therefore, existing ILP methods are not readily applicable. Specifically, (1) for temporal data such as video, the facts or *events* are labeled with time intervals indicating their start and end times. While some temporal KGs such as ICEWS [Boschee *et al.*, 2015] also have time labels, events are labeled with only a single time point. This is less expressive than time intervals as it cannot characterize events with durations. For example, “cook the soup while cutting the lettuce”. (2) Many events require n-ary relations, for example “mixing onion, garlic, and oil together”. Such a relation corresponds to an edge that connects to more than two nodes, and a graph with such edges is a *hypergraph*. While it is possible to convert hypergraphs to graphs with various techniques such as clique expansion, the conversions are lossy and can lead to an exponential number of edges.

In this work, we extend traditional ILP methods to hypergraphs whose events are labeled with time intervals. To this end, we first formally define this representation, namely *temporal hypergraph*. Then, we discuss the random walk (RW) algorithm on a hypergraph, as RW is the fundamental mechanism of a family of widely used ILP methods, that is the backward-chaining methods. We revisit the notion of *B-connectivity* on hypergraphs and propose the *multi-start random B-walk* algorithm that explores the hypergraph given a set of starting points. To generalize the temporal relation, we incorporate Allen’s interval algebra [Allen, 1983] and characterize events with the interval operators such as *Before* and *After*, and learn temporal relations by resolving the time constraints with *path consistency algorithm*. Our contributions are as follows:

- We introduce the *temporal hypergraph*, a graph that supports n-ary relations and temporal events with time in-

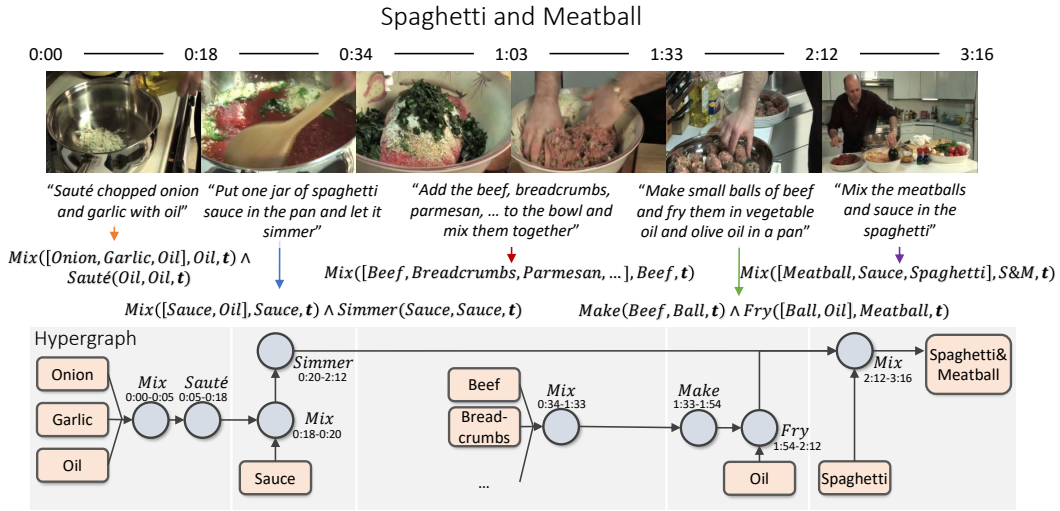


Figure 1: Temporal hypergraph representation of video instructions for making *spaghetti & meatballs*.

tervals. We show that this is a natural and expressive representation for many applications.

- We propose TILR, an ILP method that learns to generalize on both temporal and higher-order relational data in hypergraphs. This is realized with a novel random B-walk algorithm and a time-constraint propagation algorithm.
- We release two novel temporal hypergraph datasets **YouCook2-HG** and **nuScenes-HG** here, which is created from YouCook2 cooking recipe dataset and nuScenes autonomous driving dataset [Caesar *et al.*, 2020]. We show that TILR outperforms existing embedding-based and ILP baselines significantly on these two benchmarks.

To the best of our knowledge, **YouCook2-HG** and **nuScenes-HG** are the first benchmarks dedicated to temporal hypergraphs, and TILR is the first framework that addresses the ILP problem on hypergraphs with time interval labels.

2 Related Work

ILP Methods. Many ILP methods are proposed for inductive logic reasoning on graphs. These methods are categorized into two types. Forward-chaining methods [Galárraga *et al.*, 2015; Evans and Grefenstette, 2018; Payani and Fekri, 2019] learn by searching in the rule space. It supports inferring difficult tasks but suffers from exponential complexity and does not scale to large graphs. On the other hand, backward-chaining methods [Campero *et al.*, 2018; Yang and Song, 2020; Yang *et al.*, 2017] learn rules by searching in the graph space, while the rule is usually limited to a chain-like path in the graph, this leads to better scalability. For temporal hypergraphs, none of the existing ILP methods are readily applicable. In this work, we propose the ILP method for temporal hypergraphs under the backward-chaining paradigm.

Reasoning on Temporal Knowledge Graphs. Temporal reasoning has been studied extensively on the temporal knowledge graphs (KG)s [Nguyen *et al.*, 2018; Ronca *et al.*, 2018;

Trivedi *et al.*, 2017; Pareja *et al.*, 2020]. Different from the temporal hypergraphs proposed in this work, these temporal KGs typically consist of only unary and binary relations with a single time point tag. Recent methods such as TLogic [Liu *et al.*, 2022] were proposed for solving ILP on these graphs, but *time point algebra* is limited as it cannot represent temporal relations of events with duration. A more expressive representation is proposed in *time interval algebra* with Allen’s interval algebra [Allen, 1983] being the representative schema. TILP [Xiong *et al.*, 2022] adopted such an algebra and improved ILP methods for ordinary graphs. In this work, we show how to incorporate this algebra for learning temporal relations on hypergraphs.

3 Temporal Hypergraphs

An ordinary graph consists of a set of *triples* of the form $P(x_h, x_t)$, where P is the predicate and x_h and x_t are the *head* and *tail* entities. However, many real-world data cannot be fully captured by this formalism. We use Figure 1 as a running example to show this. Consider an instruction video that teaches how to make Spaghetti&Meatball using ingredients such as Onion, Oil, and Spaghetti, which are processed to make the dish in a step-by-step manner. Such information naturally implies a graphical structure, but two aspects cannot be captured by an ordinary graph

- 1) **Higher-order relations.** Actions such as *MixInto* involve multiple entities, which requires n-ary relations to represent, which calls for a *hypergraph* representation where the edges are now *hyperedges*. Note that, however, many techniques convert hypergraphs into ordinary graphs [Carletti *et al.*, 2020], but it is proven that such conversion will lose the higher-order information [Chitra and Raphael, 2019; Hayashi *et al.*, 2020]. Therefore, we focus on developing native algorithms for hypergraphs.
- 2) **Time intervals.** To properly represent events such as “Saute the ingredients from 0:00 to 0:18”, a time interval $\langle t_s, t_e \rangle$ marking the start and the end time is needed.

Temporal Hypergraph. To address the above challenges, we propose the *temporal hypergraph*. Formally, a temporal hypergraph consists of the following components: *Entity*: $x \in \mathcal{X}$ is the unique entity in the temporal data. *Predicate*: $P \in \mathcal{P}$ is the temporal predicate. *Event*: an event is defined as $e := P(x_h, x_t, t)$, where x_h, x_t are the *head* entities and *tail* entities¹, and $t := \langle t_s, t_e \rangle$ consists of the start and end time of an event. Note that an event e is a natural extension to the triple fact in an ordinary graph: a triple $P(x_h, x_t)$ represents an edge with label P that starts from a single entity x_h to another entity x_t ; similarly, $e := P(x_h, x_t, t)$ represents a *hyperedge* with label P that starts from a set of entities x_h and ends at a set of entities x_t during a period of time $t := \langle t_s, t_e \rangle$. For example, `MixInto([Ball, Sauce, Spaghetti], S&M, (2:12,3:16))` in Figure 1 represents an event of “mixing three ingredients, i.e., Meatball, Sauce and Spaghetti, into a dish of Spaghetti&Meatball from 2:12 to 3:16”.

Definition 3.1 (Temporal Hypergraph). *A temporal hypergraph \mathcal{G} is defined as $\mathcal{G} := \langle \mathcal{X}, \mathcal{P}, \mathcal{E} \rangle$, where \mathcal{X} and \mathcal{P} are the space of entities and space of predicates respectively, and $\mathcal{E} = \{e_1, \dots, e_n\}$ is a collection of events, where each event represents a hyperedge $e := P(x_h, x_t, t)$ with timestamps $t := \langle t_s, t_e \rangle$.*

A temporal hypergraph can represent many temporal data that are usually beyond the capacity of ordinary graphs, for example, representing the instructions/procedure to repair/build something, recipes in the video data shown in Figure 1, or detecting the behavior patterns in autonomous driving logs.

4 Temporal Inductive Logic Reasoning

Temporal Inductive Logic Reasoning (TILR) concerns solving the ILP problem on the temporal hypergraph, which involves learning first-order logic (FOL) rules that generalize over the patterns in the graph.

First-Order Logic. A FOL rule consists of (i) a set of predicates defined in \mathcal{P} , (ii) a set of logical variables such as X and Y , and (iii) logical operators $\{\wedge, \vee, \neg\}$. For example, a FOL rule $R : \text{GrandFatherOf}(X, X') \leftarrow \text{FatherOf}(X, Y) \wedge \text{MotherOf}(Y, X')$ has predicates `GrandFatherOf`, `MotherOf` and `FatherOf`. Terms such as `FatherOf(X, Y)` are referred to as **atoms**, which correspond to the predicates that apply to the logical variables. Each atom can be seen as a Boolean function. For example, for the binary relation `FatherOf`, the atom is a mapping $\mathcal{X} \times \mathcal{X} \mapsto \{0, 1\}$. This function can be evaluated by **instantiating** the logical variables such as X into the object in \mathcal{X} . For example, let $\mathcal{X} = \{\text{Amy}, \text{Bob}\}$, we can evaluate `FatherOf(Bob/X, Amy/Y)` by instantiating X and Y into Bob and Amy respectively. This yields `True` if “Bob is the father of Amy”. The outputs of atoms are combined using logical operations $\{\wedge, \vee, \neg\}$ and the *imply* operation $a \leftarrow b$ is equivalent to $a \vee \neg b$. Thus, when all variables are instantiated, the rule will produce an output as the specified combinations of those from the atoms.

¹Note that in some work such as [Gallo *et al.*, 1993], *head* and *tail* are used in a reversed manner

Inductive Logic Programming (ILP). Given a set of positive and negative queries \mathcal{Q}_+ and \mathcal{Q}_- , the ILP problem concerns learning a set of logic rules that predict (or entail) positive ones and do not predict the negative ones. For example, summarizing a recipe for making *Spaghetti & Meatball* from a set of videos (Figure 1) is an ILP task. Let $\mathcal{G}_1, \dots, \mathcal{G}_n$ be the temporal graphs of the videos, ILP learns a logic rule that predicts the positive video labels $\mathcal{Q}_+ = \{S\&M(\mathcal{G}_1), \dots, S\&M(\mathcal{G}_n)\}$ and does not predict the negative video labels $\mathcal{Q}_- = \{S\&M(\mathcal{G}_i) | i \neq 1, \dots, n\}$. Similarly, ILP can be used to answer event queries, where we learn rules to predict positive events $\mathcal{Q}_+ = \{e_1, \dots, e_n\}$, and not the negative events, which is essentially the link predicate task performed on the temporal hypergraph.

We solve the ILP problem via the backward-chaining approach: given a query q , one seeks to answer the query by finding a relation path in the graph that entails the query [Lao *et al.*, 2011; Yang *et al.*, 2017; Yang and Song, 2020]. In an ordinary graph with a triple query $q = P(x^0, x^n)$, this relation path from x^0 to x^n is represented as a chain-like rule

$$P(X^0, X^n) \leftarrow P^1(X^0, X^1) \wedge \dots \wedge P^n(X^{n-1}, X^n), \quad (1)$$

where X^0, X^n are variables to be instantiated into the query entities x^0 and x^n , and X^1, \dots, X^{n-1} are variables to be instantiated to the entities that exist along the relation path $x^0 \xrightarrow{P^1} x^1 \xrightarrow{P^2} \dots \xrightarrow{P^n} x^n$.

Temporal Inductive Logic Reasoning. We extend this chain-like rule family for temporal hypergraphs to generalize over both the higher-order relational data and the temporal data. Formally, we learn logic rules of the following form:

$$P(\mathbf{X}^0, \mathbf{X}^n, \mathbf{T}) \leftarrow P^1(\mathbf{X}^0, \mathbf{X}^1, \mathbf{T}^1) \psi^1 P^2(\mathbf{X}^1, \mathbf{X}^2, \mathbf{T}^2) \psi^2 \dots \psi^{n-1} P^n(\mathbf{X}^{n-1}, \mathbf{X}^n, \mathbf{T}^n). \quad (2)$$

Similar to Eq.(1), $\mathbf{X}^0, \mathbf{X}^n$ are variables of the head and tail entity sets, and $\mathbf{X}^1, \dots, \mathbf{X}^{n-1}$ are variables for those entities along the relation path which can now be in higher-order. On the other hand, we introduce the temporal operator $\psi \in \{\text{BEFORE}, \text{EQUAL}, \text{MEETS}, \dots\}$ from Allen’s interval algebra [Allen, 1983] (details in §4.2), which is a widely accepted formalism for characterizing the temporal relations between time intervals. With this operator, a logic rule can now generalize over temporal data: similar to logical variables, $\mathbf{T}, \mathbf{T}^1, \dots, \mathbf{T}^n$ are the variables for the timestamps and can be instantiated into values t, t^1, \dots, t^{n-1} , and their temporal relations can be computed by the operators which yield `True` or `False` in a similar way as logical operators.

Definition 4.1 (Temporal Inductive Logic Reasoning). *Given a temporal hypergraph \mathcal{G} , and a set of positive and negative queries \mathcal{Q}_+ and \mathcal{Q}_- , find a model $f(q; \mathcal{G})$ that maps the query q into a logic rule R such that*

$$R(q) = \mathbb{1}[q \in \mathcal{Q}_+], \text{ for } q \in \mathcal{Q},$$

where $\mathcal{Q} = \mathcal{Q}_+ \cup \mathcal{Q}_-$ and $\mathbb{1}[\cdot]$ is an indicator function.

For an event query $q = e := P(x^0, x^n, t)$, $R(x^0, x^n, t)$ denotes a logic rule with head predicate P that entails e is positive given the entities and the time interval. Similarly, for

a graph label query $q = P(\mathcal{G})$, $R(\mathcal{G})$ denotes a logic rule with head predicate P that entails if \mathcal{G} belongs to class P given the entire graph. However, learning the rule family Eq.(2) on the temporal hypergraph is nontrivial and one needs to address two challenges: **(C1)**. How to traverse and learn higher-order relations on hypergraph via random walk? **(C2)** How to generalize over the temporal data while walking on the graph? We address these challenges in the following section.

4.1 Random Walk on Temporal Hypergraphs

In an ordinary graph, chain-like rules Eq.(1) are typically learned via personalized graph random walk [Lao *et al.*, 2011; Lao and Cohen, 2010]. Given a start node x_h , a single run of the random walk involves repeating the following two steps: (i) sample an out-edge of the current node $e \sim \text{OutEdges}(x_h)$, uniformly at random; (ii) move to the tail node x_t and update $x_h \leftarrow x_t$. Given a maximum length n , this yields a sample relation path, and by running random walks multiple times, one collects a set of relation paths R_1, \dots, R_n , which can then be used in various differentiable models to learn the desired chain-like rules (details in §4.3).

B-Graph and B-Connectivity. For temporal hypergraphs, however, while there exist many random walk techniques for hypergraphs [Chitra and Raphael, 2019; Chan *et al.*, 2018; Li *et al.*, 2020], none of the existing work is designed for ILP and concerns an important property, i.e., the B-connectivity. To see this, consider an event with `Mix` relation in the recipe example in Figure 1, i.e., `Mix([Onion, Garlic, Oil], Oil, t)`. This is a hyperedge with 3 head nodes and 1 tail node. A fundamental issue arises in developing traversal algorithms for exploring the graph through this edge, that is, “*can we reach the tail node if we haven’t visited all the head nodes?*”. Intuitively, the answer is no, as we consider the actual transformational nature of the event: “*we Mix the ingredients only when we have collected all three of them*”.

Specifically, let `Mix`(x_h, x_t, t) be a `Mix` event with x_h denotes the ingredients and x_t (since $\dim(x_t) = 1$) denotes the mixed output. Suppose at the i th step of a random walk, \mathbf{x}^i is the set of nodes that we have visited. Then, there exists a natural constraint that “ x_t is reachable via `Mix` if and only if $x_h \subseteq \mathbf{x}^i$ ”, that is, the tail nodes cannot be reached until all the head nodes are reached. This property is referred to as *B-connectivity* [Gallo *et al.*, 1993]. This problem can be further formulated with the notion of *B-graph* which is a specific subset of directed hypergraphs. Formally, let $P(\mathbf{x}_h, \mathbf{x}_t, \mathbf{t})$ be a hyperedge of predicate P . A hyperedge is a *B-edge* if $\dim(\mathbf{x}_h) \geq 1$ and $\dim(\mathbf{x}_t) = 1$, meaning P is a many-to-one relation. Nodes are *B-connected*, if there exists a path consisting of B-edges that connects them, and the path is referred to as a *B-path*. A hypergraph with B-edges is a *B-graph*. Figure 2 shows an example hypergraph with 3 paths, where R_1 and R_3 are valid B-paths and R_2 is the invalid one since it does not visit x_3 before x_6 .

Multi-Start Random B-walk. We propose a novel random walk method on the B-graph, i.e., the multi-start random B-walk (MRBW) to address challenge **(C1)**. The MRBW has two unique properties compared to the traditional random walk: 1) it maintains a set of *active nodes* \mathbf{X} , which keeps track of

Algorithm 1: Multi-start Random B-walk

Input: Graph \mathcal{G} , start nodes $\mathbf{X}^0 = \{x_1, x_2, \dots\}$
Init: Path $R = []$, active nodes $\mathbf{X} = \{\}$
 Set all start nodes as active $\mathbf{X} \leftarrow \mathbf{X}^0$
while $|\mathbf{X}| > 0$ **do**
 Get all viable edges $\mathbf{E} \leftarrow B\text{-Connected}(\mathbf{X}; \mathcal{G})$
 Sample $e := P(\mathbf{x}_h, \mathbf{x}_t, \mathbf{t}) \sim \text{Uniform}(\mathbf{E})$
 Set x_t as active $\mathbf{X} \leftarrow \mathbf{X} \cup \{x_t\}$
 for $x_h \in \mathbf{x}_h$ **do**
 if $\text{AllOutEdgesVisited}(x_h)$ **then**
 Set x_h inactive $\mathbf{X} \leftarrow \mathbf{X} / \{x_h\}$
 end
 end
 Add the edge to path $R \leftarrow R + [e]$
end
 Return R

the nodes that have been visited but still have unvisited out-edges; 2) for every hop, it samples an edge uniformly from all edges that are B-connected to \mathbf{X} , and then updates the active nodes accordingly. Algorithm 1 shows how to sample a B-path via MRBW. The function $B\text{-Connected}(\mathbf{X}; \mathcal{G})$ collects all nodes that are B-connected to the active nodes \mathbf{X} and the corresponding B-edges \mathbf{E} . Then, it randomly chooses an edge e and hops to its tail node x_t . After each hop, the function $\text{AllOutEdgesVisited}(x_h)$ checks if all out-edges of each $x_h \in \mathbf{x}_h$ have been visited and will remove it from the active nodes if true.

Specifically, to sample logic rules Eq.(2), we perform MRBW over B-paths with predicates sequence $P^1 \dots P^n$. One important feature to characterize the likelihood of such a path is the distribution of random walk given the specified predicates [Lao *et al.*, 2011]. Formally, let \mathbf{x}^{i-1} be the nodes reached at $i - 1$ th step, and P^i be the predicate by which transition is performed at i th step. If it is an ordinary graph, then $\mathbf{x}^{i-1} = x^{i-1}$ and we have path probability to \mathbf{x}^i as

$$p(\mathbf{x}^i | \mathbf{x}^{i-1}; P^i) = \frac{N_{P^i}(\mathbf{x}^{i-1}, \mathbf{x}^i)}{|N_{P^i}(\mathbf{x}^{i-1}, \cdot)|} \cdot p(\mathbf{x}^{i-1} | \mathbf{x}^{i-2}; P^{i-1}),$$

where N_{P^i} is an indicator function that checks if \mathbf{x}^i and \mathbf{x}^{i-1} are connected by edge of type P^i and \cdot means arbitrary nodes. For a hypergraph where $\dim(\mathbf{x}^{i-1}), \dim(\mathbf{x}^i) \geq 1$, we extend the above, such that for $\forall \mathbf{x}^i \in \mathbf{x}^i$ we have

$$p_{i-1}^{RW}(\mathbf{x}^i) = \sum_{\mathbf{x}_h \in \mathbf{x}^{i-1}} \frac{N_{P^i}(\mathbf{x}_h, \mathbf{x}^i)}{|N_{P^i}(\mathbf{x}_h, \cdot)|} \prod_{x_h \in \mathbf{x}_h} p_{i-2}^{RW}(x_h), \quad (3)$$

where $p_{i-1}^{RW}(\cdot)$ denotes $p(\cdot | \mathbf{x}^{i-1}, P^i)$. Eq.(3) specifies that for every head nodes $\mathbf{x}_h \in \mathbf{x}^{i-1}$ that are connected to \mathbf{x}^i , the probability of them to land on \mathbf{x}^i is the product of $p_{i-2}^{RW}(x_h)$ for every individual node, and $p_{i-1}^{RW}(\mathbf{x}^i)$ is the sum of probabilities of all such head nodes divided by their out degrees. Finally, for path distribution over \mathbf{x}_i we have

$$p_{i-1}^{RW}(\mathbf{x}^i) = \prod_{\mathbf{x}^i \in \mathbf{x}^i} p_{i-1}^{RW}(\mathbf{x}^i). \quad (4)$$

With Eq.(4), we can now compute the likelihood of MRBW over a specified path $P^1 \dots P^n$ for any logic rules of the

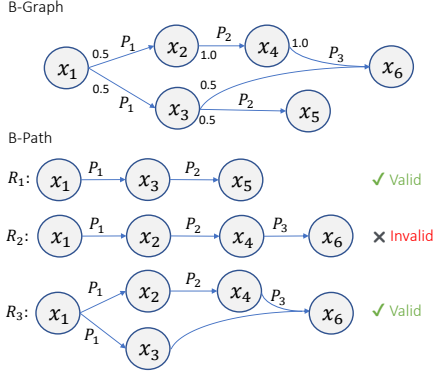


Figure 2: Example B-graph and B-paths.

form Eq.(2). In §4.3, we build the differentiable model that utilizes this as the feature.

4.2 Temporal Relation Generalization

Here, we address challenge (C2) and incorporate Allen’s Interval Algebra [Allen, 1983] for generalizing over temporal data. Allen’s Interval Algebra consists of 6 asymmetric interval relations {BEFORE, DURING, MEETS, OVERLAPS, STARTS, FINISHES}, the corresponding reverse relations {AFTER, CONTAINS, MET-BY, OVERLAPPED-BY, STARTED-BY, FINISHED-BY}, and one symmetric relation EQUAL. Similar to the logical operators, each interval relation is a unique mapping of timestamps to the Boolean value True or False. For example, for two intervals t, t' , BEFORE checks if t ends earlier than where t' starts, e.g., $\langle 0 : 00, 0 : 03 \rangle$ BEFORE $\langle 0 : 04, 0 : 10 \rangle = \text{True}$. All 13 relations are self-explanatory and the complete list of the operations is provided in [Allen, 1983].

TILR incorporates interval algebra into the logic rule. Recall the temporal rule family Eq.(2), instead of treating temporal relations as a separate operation from conjunction \wedge . we consider them as composite conjunction of the following form

$$P^i(\mathbf{X}^{i-1}, \mathbf{X}^i, \mathbf{T}^i) \psi^i P^{i+1}(\mathbf{X}^i, \mathbf{X}^{i+1}, \mathbf{T}^{i+1}) = P^i(\mathbf{X}^{i-1}, \mathbf{X}^i) \wedge P^{i+1}(\mathbf{X}^i, \mathbf{X}^{i+1}) \wedge \psi^i(\mathbf{T}^i, \mathbf{T}^{i+1}),$$

where $\psi^i \in \{\text{BEFORE, EQUAL, MEETS, ...}\}$. To generalize the temporal relations from the data, one keeps track of the set of applicable temporal relations for each pair of events in the rule. Whenever a positive query is matched by the rule, one updates the temporal relations to satisfy the time constraints posed in the corresponding subgraph of the query. We incorporate and modify the path-consistency (PC) algorithm [Allen, 1983] for MRBW. Algorithm 2 shows the process of resolving the temporal relations for a given path R . For every pair of events (e_1, e_2) in the path, one first obtains the pairwise temporal relation via function $TempRel(e_1, e_2)$ which returns the satisfied relation. Then, it checks the path consistency between the three edges of the path. This is done with $PC3(e_1, e_2, e_3)$ which is the PC3 algorithm.

Algorithm 2: Path-consistency for temporal relation generalization.

Input: Path $R = [e^1, \dots, e^T]$
Init: Temporal relation table $B[\cdot, \cdot] = \emptyset$
for $(e_1, e_2) \in R \times R$ **do**
 Update the temporal relation
 $B[e_1, e_2] \leftarrow TempRel(e_1, e_2)$
 for $e_3 \in R / \{e_1, e_2\}$ **do**
 Resolve path consistency
 $B[e_1, e_3], B[e_2, e_3] \leftarrow PC3(e_1, e_2, e_3)$
end
end
 Return B

4.3 Differentiable TILR for Hypergraph Reasoning

Through MRBW and the PC algorithm, one can reliably traverse the temporal hypergraph and sample reasoning path, enabling us to develop models for solving the ILP task. Here, we propose TILR $_{\theta}$, a differentiable ILP model that reasons on the temporal hypergraph.

Let R be a temporal relation B-path, $x^0 \xrightarrow{P^1} \dots \xrightarrow{P^n} x^n$ that traverses from x^0 to x^n via relation $P^1 \dots P^n$ and under the temporal constraints $\psi^1, \dots, \psi^{n-1}$. We seek to model R with a distribution p_R as the path and temporal constrained random walk. Let $\mathcal{T}^i = \{t | P^i(x_h, x_t, t) \in \mathcal{G}\}$ be the space of time intervals in graph \mathcal{G} with respect to predicate P^i at i -th step. Given two such spaces $\mathcal{T}^{i-1}, \mathcal{T}^i$, the space of time interval pairs that satisfy temporal constraint ψ^{i-1} is computed as $\psi^{i-1}(\mathcal{T}^{i-1}, \mathcal{T}^i) = \{ \langle t^{i-1}, t^i \rangle | \psi^{i-1}(t^{i-1}, t^i) = \text{True}; t^{i-1} \in \mathcal{T}^{i-1}, t^i \in \mathcal{T}^i \}$, which represents the set of interval pairs where ψ^{i-1} yields True. Finally, the probability of traversing from step $i-1$ to step i via under ψ^{i-1} can be computed as

$$p(P^i | P^{i-1}) = |\psi^{i-1}(\mathcal{T}^{i-1}, \mathcal{T}^i)| / |\psi^{i-1}(\mathcal{T}^{i-1}, \cdot)|, \quad (5)$$

where \cdot indicates arbitrary spaces in \mathcal{G} . We can now define distribution p_R over R with Eq.(4) and Eq.(5) in a recursive fashion:

$$p_R(x^i | x^0) = p_{i-1}^{RW}(x^i) \cdot p(P^i | P^{i-1}). \quad (6)$$

Recall Definition 4.1, to answer a query with respect to x^0 to x^n , we sample a set of paths $R_1 \dots R_L$ and learn a parameterized model

$$f_{\theta}(q; \mathcal{G}) = \sum_{j=1}^L \theta_j^p \sum_{i=1}^n \theta_i^s \cdot p_{R_j}(x^i | x^0), \quad (7)$$

where θ^p are the weights for choosing the appropriate paths and θ^s are the weights for combining the path features over $p_{R_j}(x^i | x^0)$ the n steps respectively. To train such a model, we collect queries $\mathcal{D} = \{q_i\}$ from the training split of the hypergraph \mathcal{G} by sampling the events and corresponding paths. We optimize the model by minimizing a cross-entropy loss

$$\mathcal{L}(\theta; \mathcal{G}) = \sum_{q \in \mathcal{D}} y_i \log p(q | \mathcal{G}) + (1 - y_i) \log(1 - p(q | \mathcal{G})),$$

where $y_i \in 0, 1$ indicates if the query is positive, and $p(q | \mathcal{G}) = 1 / (1 + e^{-f_{\theta}(q; \mathcal{G})})$ is the conditional probability of q computed from Eq.(7) with sigmoid function.

5 Experiment

We conduct experiments to answer the following questions: **(Q1)** can temporal hypergraphs preserve more information over ordinary graphs such that they lead to better performance in real-world tasks? **(Q2)** Can TILR generalize better and interpretable logic rules from hypergraphs than traditional ILP methods? **(Q3)** How do MRBW, PC algorithm, and parametrized model contribute to the performance of TILR?

To do so, we create two temporal hypergraph datasets **YouCook2-HG** and **nuScenes-HG** and evaluate TILR on the reasoning tasks of recipe summarization and driver behavior explanation.

5.1 Recipe Summarization

YouCook2 cooking recipe dataset [Zhou *et al.*, 2018] consists of instructional videos of 89 cooking recipes such as *spaghetti & meatballs*. Each recipe has 22 videos and each video is annotated by a sequence of natural language sentences that describe the procedure steps, as shown in Figure 1. The dataset serves as a challenging benchmark for evaluating the generalizability of both temporal and relational data: instructions of the same class can have varied procedures and ingredients. For example, one can make *BLT sandwich* by first putting the lettuce on the bread and then the ham, or in the reverse order. This requires the ILP method to learn generalized temporal relations for the temporal events.

Predicate Type	# Predicates	# Facts		Examples
		Per Graph	Total	
Relation	Unary	208	58	Cut, fry
	Binary	51	29	Put
	N-ary	1	69	MixInto
Class	1136	257	45881	Oil, Onion, Noodle

Table 1: Stats. of the hypergraph benchmark **YouCook2-HG**.

YouCook2-HG Construction. We construct the temporal hypergraph from the instruction sentences. For each clip, we run NLTK tools and extract the nouns and verbs from the sentences. We collect verbs as temporal relations. After pre-processing and grouping synonyms, there are 208 unary relations, 51 binary relations, and 1 n-ary relation. There are many possible n-ary relations in the raw data that describe the same action, such as “put together” and “stir together”. Given the low frequency of each data, we consider merging them into a single n-ary relation, that is `MixInto`. Statistics are shown in Table 1.

Clique-Expansion Graphs. To evaluate **(Q1)** and **(Q2)**, an ordinary graph version of the **YouCook2-HG** needs to be created such that we can evaluate TILR on both the hypergraphs and the ordinary graphs with traditional baselines. To do so, we use the clique expansion algorithm to convert the hypergraph to an ordinary graph: for every hyperedge $e := P(x_h, x_t, \langle t_s, t_e \rangle) \in \mathcal{E}$, it creates edges for each pair of head and tail nodes $\mathcal{E}_{\text{ord}} = \{P(x_h, x_t, t_s) | x_h \in \mathbf{x}_h, x_t \in \mathbf{x}_t\}$. Note that, for time intervals, we drop the end time t_e and use t_s as the time point label. This way, the clique-expansion graphs share the same structure as those traditional temporal graphs [Boschae *et al.*, 2015].

Task. The goal of recipe summarization is to generalize a structured procedure for each distinct recipe from the hypergraph data as that in Figure 1. We formalize this problem as a *graph classification* task. Formally, Let $P(\mathcal{G})$ be the label of a hypergraph \mathcal{G} , where P denotes its recipe type label (e.g. *BLT sandwich*). Let P^* be the target type for which we want to summarize, then the positive set of P^* is the set of all graphs with the same label, and the negative set is the rest of the graphs with different labels

$$\mathcal{Q}_+ = \{P(\mathcal{G}) | P = P^*\}, \quad \mathcal{Q}_- = \{P(\mathcal{G}) | P \neq P^*\}. \quad (8)$$

Recall the objective in Definition 4.1, we learn rules for recipe type P^* that can classify graphs in \mathcal{Q}_+ as positive and those in \mathcal{Q}_- as negative. Furthermore, we also constrain the temporal rules to cover the full timespan of the graphs, ensuring rules fully represent the underlying recipe.

Methods and Baselines. We evaluate three modes of TILR with increasing complexity, so that it serves as an ablation study on the proposed modules: 1) TILR: this mode performs vanilla MRBW without the path-consistency (PC) algorithm; it proposes the best rule for each recipe by counting the occurrences and picking the most frequent one. 2) TILR-PC: this mode is the same as MRBW but with the PC algorithm implemented. 3) TILR $_{\theta}$ -PC: this mode performs MRBW and PC algorithm during search and learns rule via the parametric model introduced in §4.3.

We compare TILR with baseline methods that reason on the clique-expansion graphs (as TILR is the only method that reasons on hypergraphs): 1) GNN-GCN [Kipf and Welling, 2016] and GNN-Cheb [Defferrard *et al.*, 2016]. GNN methods are evaluated on the clique-expansion graph with the time labels. Specifically, we set GNN methods to the inductive setting, where the entities all share the same one-hot vector. This way GNNs learn to generalize to classify graphs with unseen entities. 2) CTDNE [Nguyen *et al.*, 2018]. A graph embedding method that utilizes the time point labels in the clique-expansion graph. 3) traditional ILP methods: random walk (RW), NeuralLP [Yang *et al.*, 2017], NLIL [Yang and Song, 2020], and Drum [Sadeghian *et al.*, 2019]. We implement the RW method as a simplistic baseline that performs personalized random walks on the clique-expansion graph while ignoring the time labels; it counts the frequency of paths that answer the query. All experiments are done on a PC with i7-8700K and one GTX1080ti. We use the Mean Reciprocal Rank (MRR) and Hits@3 and Hits@10 as the metrics.

Results. The results are shown in Table 2. We find TILR yields the best performance on the hypergraphs, which is 30% higher than the ordinary graph counterparts. This suggests the proposed temporal hypergraph representation can better capture the higher-order information, which addresses question **(Q1)**. On the other hand, we also find TILR outperforms all baseline methods on clique-expansion graphs as well. This suggests that TILR, with the introduction of time intervals, generalizes better than other baselines in the temporal domain, as others can at most utilize time point information. To further inspect the generalizability of TILR, we show example rules in Table 3. This shows that TILR can generalize meaningful and interpretable rules over both logic and temporal domains,

Method	YouCook2-HG						nuScenes-HG					
	Clique-Expansion			Hypergraph			Clique-Expansion			Hypergraph		
	MRR	Hits@3	Hits@10	MRR	Hits@3	Hits@10	MRR	Hits@3	Hits@10	MRR	Hits@3	Hits@10
GNN-GCN	0.32	28.1	36.9	-	-	-	0.40	32.2	50.4	-	-	-
GNN-Cheb	0.37	32.7	43.7	-	-	-	0.41	32.6	55.6	-	-	-
CTDNE	0.15	11.9	16.6	-	-	-	0.08	9.7	11.3	-	-	-
RW	0.19	18.8	20.6	-	-	-	0.15	11.0	17.4	-	-	-
NeuralLP	0.36	35.4	39.5	-	-	-	0.47	39.4	63.1	-	-	-
NLIL	0.38	36.3	6.8	-	-	-	0.45	38.3	58.2	-	-	-
Drum	<u>0.40</u>	<u>38.5</u>	45.2	-	-	-	<u>0.51</u>	<u>45.5</u>	66.8	-	-	-
TILR	0.25	20.4	23.7	0.35	29.4	36.3	0.14	10.1	19.5	0.19	20.8	31.4
TILR-PC	0.30	27.9	31.1	<u>0.60</u>	<u>55.8</u>	<u>59.1</u>	0.18	15.1	22.9	<u>0.22</u>	<u>24.1</u>	<u>35.5</u>
TILR _θ -PC	0.44	40.2	47.3	0.72	76.0	79.4	0.52	47.2	<u>65.2</u>	0.64	66.0	81.1

Table 2: MRR and Hits of TILR and baselines on YouCook2 recipe summarization and nuScenes behavior explanation benchmarks.

Recipe	Example temporal rules
BLT Sandwich	Put(Mayo, Bread1)
	BEFORE MixInto((Bread1, lettuce, tomato), M)
	BEFORE Cover(Bread2, M)
Onion Ring	Remove(Layers, Onion)
	AND Cut(Onion, Onion)
	BEFORE MixInto((Onion, flour), M)
	BEFORE Put(M, Fryer)
Spaghetti & Meatball	MixInto((Garlic, Beef), M1)
	BEFORE Saute(M1, M1)
	AND MixInto((Tomato Paste, cheese), M2)
	BEFORE MixInto((M1, M2, Spaghetti), M3)

 Table 3: Example rules that summarize the recipes. We simplify the notations. The class M denotes the intermediate products.

addressing question (Q2). We investigate (Q3) in §5.2 with both tasks’ results.

5.2 Driving Behavior Explanation

The nuScenes autonomous driving dataset [Caesar *et al.*, 2020] consists of 750 driving scenes in the training set. Each scene is a 20s video clip; each frame in the clip is annotated with 2D bounding boxes and class labels. The dataset also provides lidar and ego information such as absolute position, brake, throttle, and acceleration.

Predicate Type	# Predicates	# Facts		Examples
		Per Graph	Total	
Relation	Unary	13	4631	ego.throttle ego.brake
	Binary	6	1859	in_front behind
	N-ary	1	1692	Between
Class	23	257	45881	vehicle.car human.ped.adult ped.crossing

Table 4: Stats. of the hypergraph benchmark nuScenes-HG.

nuScenes-HG Construction. We construct the hypergraph for each driving scene. We use the 23 object classes as class labels, including `vehicle.car` and `human.pedestrian.adult`; We convert ego information and attributes such as `vehicle.moving` and

`pedestrian.moving` into unary predicates. The original dataset does not provide relations for the objects. Here, we create the relative spatial relations by inferring from the absolute spatial information. This includes `In-front`, `Behind`, `Between` and etc. The statistics are shown in Table 4. Clique-expansion graphs are created for nuScenes-HG the same way as that in **YouCook2-HG**.

Task. We apply TILR to explain the driving behavior with temporal logic rules. Such an explanation is beneficial as it can provide insight into why certain behavior happens, for example, one can explain in what situation will the driver hit the brake; the evidence is likely to be there is a pedestrian crossing the street. Formally, we consider the positive queries as events of `ego.brake` and `ego.throttle` and the rest of the events as the negative queries. Let $e = P(x^0, x^n, t)$ be the event and \mathcal{G} be the temporal graph, we have

$$Q_+ = \{e | P \in \{\text{ego.brake}, \text{ego.throttle}\}, e \in \mathcal{G}\},$$

$$Q_- = \{e | P \notin \{\text{ego.brake}, \text{ego.throttle}\}, e \in \mathcal{G}\}.$$

Results. We show results in Table 2. TILR performs the best on both clique-expansion graphs and hypergraphs, which is similar to that in **YouCook2-HG**, suggesting temporal hypergraph and the proposed TILR are the better representation and ILP method for learning rules from real-world data with temporal and higher-order relations.

Ablation Study. For both benchmarks, using PC and parameter model improves the performance, suggesting they are essential components of TILR, resolving question (Q3). Furthermore, we find that TILR_θ-PC performs significantly better than two other modes in this benchmark. This is because nuScenes are more diverse and many situations can lead to the target event, making it difficult to learn rules by counting.

6 Conclusion

We propose TILR, a framework that extends ILP technique to beyond KGs. We introduce the temporal hypergraph and show it is a powerful representation for many applications; we then propose the MRBW algorithm, which, combined with the path-consistency algorithm, serves as the foundation of TILR. In experiments, we create and release two dedicated temporal hypergraph datasets and evaluate TILR with strong baselines, which shows superior reasoning capabilities.

Acknowledgements

This work was supported by a sponsored research award by Cisco Research.

References

- [Allen, 1983] James F Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.
- [Boschee *et al.*, 2015] Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. ICEWS Coded Event Data, 2015.
- [Caesar *et al.*, 2020] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020.
- [Campero *et al.*, 2018] Andres Campero, Aldo Pareja, Tim Klinger, Josh Tenenbaum, and Sebastian Riedel. Logical rule induction and theory learning using neural theorem proving. *arXiv preprint arXiv:1809.02193*, 2018.
- [Carletti *et al.*, 2020] Timoteo Carletti, Federico Battiston, Giulia Cencetti, and Duccio Fanelli. Random walks on hypergraphs. *Physical review E*, 101(2):022308, 2020.
- [Chan *et al.*, 2018] T-H Hubert Chan, Anand Louis, Zhihao Gavin Tang, and Chenzi Zhang. Spectral properties of hypergraph laplacian and approximation algorithms. *Journal of the ACM (JACM)*, 65(3):1–48, 2018.
- [Chitra and Raphael, 2019] Uthsav Chitra and Benjamin Raphael. Random walks on hypergraphs with edge-dependent vertex weights. In *International Conference on Machine Learning*, pages 1172–1181. PMLR, 2019.
- [Defferrard *et al.*, 2016] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.
- [Evans and Grefenstette, 2018] Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018.
- [Galárraga *et al.*, 2015] Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M Suchanek. Fast rule mining in ontological knowledge bases with amie+. *The VLDB Journal—The International Journal on Very Large Data Bases*, 24(6):707–730, 2015.
- [Gallo *et al.*, 1993] Giorgio Gallo, Giustino Longo, Stefano Pallottino, and Sang Nguyen. Directed hypergraphs and applications. *Discrete applied mathematics*, 42(2-3):177–201, 1993.
- [Hayashi *et al.*, 2020] Koby Hayashi, Sinan G Aksoy, Cheong Hee Park, and Haesun Park. Hypergraph random walks, laplacians, and clustering. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 495–504, 2020.
- [Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [Lao and Cohen, 2010] Ni Lao and William W Cohen. Relational retrieval using a combination of path-constrained random walks. *Machine learning*, 81(1):53–67, 2010.
- [Lao *et al.*, 2011] Ni Lao, Tom Mitchell, and William Cohen. Random walk inference and learning in a large scale knowledge base. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 529–539, 2011.
- [Leetaru and Schrodt, 2013] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013.
- [Li *et al.*, 2020] Pan Li, Niao He, and Olgica Milenkovic. Quadratic decomposable submodular function minimization: Theory and practice. *J. Mach. Learn. Res.*, 21:106–1, 2020.
- [Liu *et al.*, 2022] Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. *arXiv preprint arXiv:2112.08025*, 2022.
- [Nguyen *et al.*, 2018] Giang Hoang Nguyen, John Boaz Lee, Ryan A Rossi, Nesreen K Ahmed, Eunye Koh, and Sungchul Kim. Continuous-time dynamic network embeddings. In *Companion Proceedings of the The Web Conference 2018*, pages 969–976, 2018.
- [Pareja *et al.*, 2020] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 5363–5370, 2020.
- [Payani and Fekri, 2019] Ali Payani and Faramarz Fekri. Inductive logic programming via differentiable deep neural logic networks. *arXiv preprint arXiv:1906.03523*, 2019.
- [Ronca *et al.*, 2018] Alessandro Ronca, Mark Kaminski, Bernardo Cuenca Grau, Boris Motik, and Ian Horrocks. Stream reasoning in temporal datalog. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [Sadeghian *et al.*, 2019] Ali Sadeghian, Mohammadreza Armandpour, Patrick Ding, and Daisy Zhe Wang. Drum: End-to-end differentiable rule mining on knowledge graphs. *Advances in Neural Information Processing Systems*, 32, 2019.
- [Toutanova and Chen, 2015] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on*

Continuous Vector Space Models and their Compositionality, pages 57–66, 2015.

- [Trivedi *et al.*, 2017] Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *international conference on machine learning*, pages 3462–3471. PMLR, 2017.
- [Xiong *et al.*, 2022] Siheng Xiong, Yuan Yang, Faramarz Fekri, and James Clayton Kerce. Tilp: Differentiable learning of temporal logical rules on knowledge graphs. In *The Eleventh International Conference on Learning Representations*, 2022.
- [Yang and Song, 2020] Yuan Yang and Le Song. Learn to explain efficiently via neural logic inductive learning. In *International Conference on Learning Representations*, 2020.
- [Yang *et al.*, 2017] Fan Yang, Zhilin Yang, and William W Cohen. Differentiable learning of logical rules for knowledge base reasoning. In *Advances in Neural Information Processing Systems*, pages 2319–2328, 2017.
- [Zhou *et al.*, 2018] Luwei Zhou, Chenliang Xu, and Jason J Corso. Towards automatic learning of procedures from web instructional videos. In *AAAI Conference on Artificial Intelligence*, pages 7590–7598, 2018.