arXiv:2602.02161v1 [cs.LG] 2 Feb 2026

用于时间链接预测反事实验证的因果时间交互图生成

Aniq Ur Rahman 和 Justin P. Coon {ANIQ.RAHMAN, JUSTIN.COON }@ENG.OX.AC.UK
牛津大学

摘要

时间链接预测（TLP）模型通常依据预测准确率进行评估，但这类评估无法判断模型是否捕捉到了支配时间交互的因果机制。本文提出一种反事实验证框架，通过生成具有已知真实因果结构的因果时间交互图（CTIGs），对 TLP 模型进行验证。我们首先为连续时间事件序列引入一种结构方程模型，该模型同时支持兴奋性和抑制性效应，随后将这一机制扩展至时间交互图。为比较因果模型，我们提出一种基于跨模型预测误差的距离度量，并通过实证验证以下假设：在一个因果模型上训练的预测器，在距离足够远的模型上进行评估时，性能会下降。最后，我们在以下情形中开展反事实评估：（i）生成模型之间存在受控因果转移；（ii）将时间戳打乱作为一种具有可测量因果距离的随机扰动。我们的框架为因果感知基准测试奠定了基础。

关键词： 时间交互图、时间链接预测、反事实分析

1. 引言

时间交互图（TIG）用于模拟现实世界中的现象，其中实体之间会进行持续一段时间（也可能是瞬时）的交互。社交网络就是一个典型例子：用户通过发送消息或评论彼此的帖子进行互动，由此形成边。

在图学习中，链接预测 指的是观察静态图的其他部分后，预测图中是否存在某些边的任务。在动态场景中，我们研究时间链接预测 （TLP）：观察截至某个时间戳的 TIG 后，模型会预测未来特定时间戳是否存在边。

尽管近年来 TLP 问题本身逐渐受到关注（Kumar 等人，2019年；Rossi 等人，2020年；Yu 等人，2023年），这些模型是否捕捉到了数据中潜在的因果机制，仍鲜有研究（Rahman 等人，2025a）。近期，Rahman 等人（2025b）在反事实场景下评估了TLP模型，得出结论：如果改变边出现的顺序或出现频率，这些模型便无法区分不同的TIG。然而，现实世界的数据集并不能被确证为具有因果性（Pearl，2009年）；尤其是，过去的时间链接对未来交互产生的因果影响通常是未知的。为了解决这一局限，我们提出一种生成因果时间交互图（CTIG）的方法，借助具有已知因果结构的数据，对 TLP模型进行反事实评估。

尽管关于因果时间交互图（CTIGs）的研究有限，但我们可以从成熟的相邻领域因果事件序列（CES）Kim 等人，2011年；Noorbakhsh 和 Rodriguez，2022；Jalaldoust 等人，2022；C¨ uppers 等人，2024年）以及事件预测领域的相关研究中汲取见解

事件预测（Zhao，2021年）。因此，我们首先开发一个用于生成 CES 的模型，然后自然地将其扩展为生成 CTIG，其中每个事件都对应一个边事件。从因果模型生成 CTIG 后，我们希望量化它与由另一因果模型生成的 CTIG 之间的差异，由此引出以下问题：

> 我们如何衡量两个 *CTIG* 生成因果模型之间的距离？

这个问题也为检验以下假设奠定了基础：

> 在某个因果模型生成的数据上训练的预测模型，在另一个因果模型生成的数据上表现会更差，前提是这两个模型之间的距离足够大。

为检验这一假设，我们从不同的因果模型生成 CTIG，并向一个能够访问真实因果参数的 Oracle 查询以进行预测。随后，我们将这一框架扩展到纳入因果扭曲，并使用一些知名的时间链接预测（TLP）模型，重复Rahman 等人（2025b）的反事实实验。

组织结构 在 §2中，我们介绍一种用于生成因果事件序列的结构方程模型，并讨论由此得到的因果模型的属性。随后，我们在 § 3中将所提出的模型与相关文献进行比较。在 § 4中，我们提出一种距离度量，用于量化两个因果模型之间的差异。接下来，在 § 5中，我们扩展 CES 的生成流程以创建因果时间交互图（ CTIGs ）：首先生成节点特征，并以此为基础定义底层因果图。在 § 6中，我们介绍反事实设置，通过实证研究阐明其动机，并将其应用于时间链接预测（TLP）模型。我们提出一种方法，在受控因果转移和时间戳打乱条件下对 TLP 模型进行基准测试。最后，我们在 § 7中总结全文。由于篇幅限制，预备知识、扩展的相关工作以及关于因果模型理论属性的证明均放在附录中。

符号 我们使用 $\emptyset$ 表示空集，使用 $|\mathcal{A}|$ 表示集合 $\mathcal{A}$的基数，并使用 $|\mathcal{A}^{(y)}|$ 表示元素 y 在 $\mathcal{A}$中的出现次数。我们分别使用 $\mathbb{N}^+$ 和 $\mathbb{R}^+$ 表示正自然数集和正实数集。对于 $n \in \mathbb{N}^+$，我们令 $[n] = \{1, 2, \ldots, n\}$；对于 $a,\ b \in \mathbb{R}$使得 $a < b$，我们令 $[a, b)$ $= \{c \in \mathbb{R} \mid a \leq c < b\}$。我们使用大写字母（例如 U）表示变量，使用小写字母（例如 u）表示值，并使用 $\varnothing$表示空值。指示函数表示为 $\mathbb{I}\{\cdot\} \in \{0, 1\}$ ，其参数为一个逻辑陈述。期望 $\widehat{}$和方差分别表示为 $\mathbb{E}[\cdot]$ 和 $\mathbb{V}[\cdot]$。我们用 $\mathbb{E}[\cdot]$表示样本均值。

2. 因果事件序列

考虑一组随时间连续发生的事件 $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$。序列中的每个数据点都可以表示为一对值，其中一个分量标识事件，另一个分量指定事件发生的时间。例如，给定观测事件序列：$(\boldsymbol{a}, t_1)$，$(\boldsymbol{c}, t_2)$，$(\boldsymbol{a}, t_3)$，$(\boldsymbol{b}, t_4)$，$(\boldsymbol{c}, t_5)$……，$(\boldsymbol{a}, t_m)$，可以提出以下预测问题：(1) 我们能否预测下一次事件何时发生？(2) 如果已知下一次事件的发生时间 t_{m+1}，我们能否预测届时会发生什么事件？(3) 我们能否估计事件$e \in \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ 在指定的未来时间区间内发生的概率？(4) 我们能否预测事件 $e \in \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ 在未来某个时间区间内发生的次数？

这些问题凸显了预测连续时间中发生的事件所面临的根本挑战。现在假设存在一个函数 f，它以事件序列为输入，并针对这些预测问题给出答案。由此自然会产生一个问题：

> 是否有理由相信，过去事件的发生会影响未来将发生哪些事件，以及这些事件将在何时发生？

如果过去事件对未来事件没有影响，那么将其作为 f 的输入便是徒劳的。然而，Gong 等人 (2023年) 指出，当数据按顺序到达时，通常会假设存在时间依赖关系。因此，我们提出了一种因果机制，以在连续时间内生成事件序列，从而使这一假设成立。此外，我们基于（*Gong* 等人，2023年）推导出因果机制的以下设计原则：

（*D1*）因果关系需要时间传播，即假设事件 $\boldsymbol{a}$ 导致了另一个事件 $\boldsymbol{b}$，那么，如果 $\boldsymbol{a}$ 在时间 t 发生，事件 $\boldsymbol{b}$可能会在经过一定延迟后发生。

（*D2*）要推断各种事件之间的因果关系，必须同时考虑它们发生的时间顺序以及彼此之间的延迟。

(*D3*) 因果事件以滚动方式沿着单一连续时间线展开。

提出的模型 考虑一系列随时间反复发生的事件，其中包含 n种不同的事件类型。每种事件类型都关联一个触发事件序列，该序列决定其何时可能被激活。我们将 $U_i(t) \in \{0,1\}$ 定义为一个指示变量，用于表示类型为 i 的事件是否在时间 t被触发。

给定过去发生的事件，类型为 i 的事件可能会也可能不会在时间 t发生，我们将其表示为$X_i(t) \in \{0,1\}$。为捕捉历史影响，我们将 $X_i'(t) \in \{0,1\}$ 定义为时间窗口（$t-\bar{\tau}$，t）内是否存在事件 i。因此，因果模型可以函数式地表示为[1]：

$$x_i(t) = f\Big(\underbrace{\{x_j'(t) : \forall j \in \mathcal{P}_i\}}_{\text{past events}}, \underbrace{u_i(t)}_{\text{trigger}}\Big). \tag{1}$$

对于每个类型为 i的事件，我们定义如下结构方程模型 (SEM)[2]：

$$\begin{aligned} x_i(t) = f(\mathcal{P}_i(t), u_i(t)) &= u_i(t) \cdot \hat{f}(\mathcal{P}_i(t)) \\ &= u_i(t) \cdot \mathbb{I}\left\{\sum_{j \in \mathcal{P}_i(t)} \Theta_{i,j} \geq 0\right\} = u_i(t) \cdot \mathbb{I}\left\{\sum_{j \in \mathcal{P}_i} \Theta_{i,j}\, x_j'(t) \geq 0\right\}, \end{aligned} \tag{2}$$

其中，$\mathcal{P}_i(t) \subseteq \mathcal{P}_i \subseteq [n]$ 表示因果图中 $X_i(t)$的父节点，[3]，不包括 $U_i(t)$；$\Theta_{i,\,j} \in [-1,1]$ 是一个参数，我们称之为事件间因果影响。如果$\Theta_{i,\,j} > 0$，则称类型为 j 的事件会兴奋事件 i；如果$\Theta_{i,\,j} < 0$，则称事件 j 会抑制事件 i。Kim 等人（2011年）在对神经元尖峰活动建模时研究了一个事件对另一个事件的兴奋性和抑制性影响。为避免混淆，我们将集合 $\mathcal{P}_i$ 称为结构父节点

1. 尽管变量以大写字母书写，其取值以小写字母书写。2. 与方程 1不同，将输入写作 $\mathcal{P}_i(t)$ 会隐含地提供所有 $j \in \mathcal{P}_i(t)$对应的指示变量 $x_j'(t)$。3. 在因果图中，从变量 A 指向变量 B 的有向边表示 A 是 B的原因。

集合，其定义为 i： $\mathcal{P}_i = \{j : \Theta_{i,j} \neq 0, \forall j \in [n]\}$；而 $\mathcal{P}_i(t) \subseteq \mathcal{P}_i$ 表示时刻 t的 i 的活跃父节点，其定义为 $\mathcal{P}_i(t) = \{j : x'_j(t) = 1, \forall j \in \mathcal{P}_i\}$。稍微滥用一下符号，当我们说 j 是 i的父节点时，我们指的是 $X'_j(t)$ 是 $X_i(t)$的父节点。因此，即使 $i = j$，因果图仍然是无环的，因为 $X'_i(t)$和 $X_i(t)$表示不同的节点。

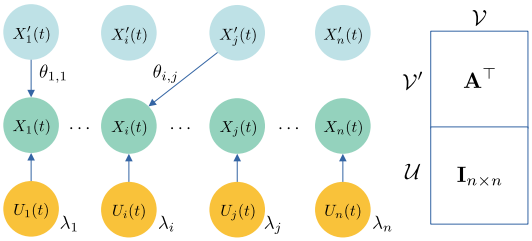


图 1. 在此图中，我们展示了一个样本因果图 $\mathcal{G}$，其中包含三类节点：事件节点 X_i、过去事件指示变量X'_j和触发节点 U_k。从X'_j（t）到 X_i（t）的因果箭头标注了因果影响参数$\Theta_{i,\ j}$。对于触发节点，我们标出了其生成率 λ_k，该生成率遵循齐次泊松过程。

为了指定因果模型，我们描述因果图、触发事件序列和因果影响参数的生成方式：

• 对于因果图，我们关注从 $\mathcal{V}' = \{X'_i(t) : i \in [n]\}$ 到$\mathcal{V} = \{X_i(t) : i \in [n]\}$的二部图，其邻接矩阵$\mathbf{A}$ 从参数化随机图模型中采样得到。[4]

• 对于每个类型为 i的事件，其触发事件序列 Φ_i 通过采样一个齐次泊松点过程（Haenggi, 2012，§ 2.4.1）生成，其强度为 λ_i 该强度可以从另一个支持集为 $\mathbb{R}^+$的分布中随机采样。

• 最后，事件间因果影响参数 $\{\Theta_{i,j} : \mathbf{A}_{i,\ j} = 1,\ \forall i,\ j \in [n]\}$ 从支持集为 $[-1, 1]$ 的分布中采样，并过滤掉 $\{0\}$。

生成因果事件序列的过程详见算法 1。

算法1因果事件序列生成

输入 因果模型（因果图和参数）、总持续时间 T**输出** 事件序列 $\mathcal{S}$

1：将事件序列 $\mathcal{S}$ 初始化为空 2：为每种事件类型生成不超过 T 的触发事件序列 3：收集并排序所有触发事件的时间，形成时间线 $\mathcal{T}$ 4：对于 t 中的每个时间 $\mathcal{T}$，执行 5：确定在时间 t 触发了哪个事件，即 i 6：对于导致 i的每个事件 j，执行 7：检查事件 j 是否在持续时间为 τ的过去时间窗口内发生 8：结束循环 9：使用方程2中的结构方程模型确定事件 i 是否在时间 t发生 10：如果事件发生，则 11：将事件及其时间戳（i，t）添加到序列 $\mathcal{S}$中 12：结束条件判断 13：结束循环

4. 可使用标准随机图模型，例如 Erd˝os–R´enyi（ER）模型（Erdos, 1961）、随机块模型（SBM）（Abbe, 2018）或 Barab´asi–Albert（BA）模型（Barab´asi, 2014，第 5 节）。

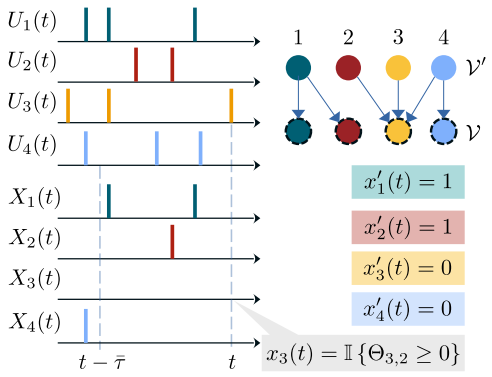


图2. 此图以直观方式展示了算法1在包含4个事件节点的因果图上的运行过程。在时间 t，$u_3(t)$ =1，因此我们计算 $x_3(t) = f(\mathcal{P}_3(t)\ u_3(t))$，并发现其父节点中只有 $x_2'(t) = 1$（$\mathcal{P}_3(t)$）。这使我们能够确定 $x_3(t) = \mathbb{I}\{\Theta_{3,2} \geq 0\}$。为了确定$x_j'(t)$的值，我们只考虑时间窗口 $[t-\bar{\tau}, t)$。

所提出因果模型的性质现在，我们总结所提出因果模型的若干关键理论性质，详细内容可见附录C。

- 因果模型（$\mathscr{G}$，$\Theta_{\mathscr{G}}$）是
  - 对于所有 $\mathbf{A} \in \{0,1\}^{n\times n}$，均为半马尔可夫（命题6）。
  - 对于所有 $\mathbf{A} \in \{0,1\}^{n\times n}$，均非马尔可夫（命题7）。
  - 当 $\mathbf{A} = \mathbf{I}$时为马尔可夫（推论8）。
- $X_i(t)$ 相对于 $X_j'(t)$ 单调，当且仅当$\Theta_{i,\ j} \in [0,1]$（命题9）。
- $X_j'(t)$ 相对于 $X_i(t)$ $\forall i,\ j \in [n]$（命题11）。
- 如果存在某个 $j \in \mathcal{P}_i$，$X_j'(t)$ 相对于 $X_i(t)$是外生的，且 $X_i(t)$ 相对于$X_j'(t)$是单调的，那么必要性与充分性概率(PNS) 是可识别的（定理12）。

3. 相关工作

Noorbakhsh 和 Rodriguez (2022) 提出了一种对点过程中的事件进行有选择地移除的方法，以概率规则为依据对事件序列建模。然而，他们的方法并未明确界定不同事件类型之间的因果关系。作者定义了一种多维 Hawkes 过程（MHP）（Hawkes，1971）。

Jalaldoust 等人 (2022) 提出了一种因果发现算法，用于学习 MHP 中隐含的格兰杰因果网络。为进行评估，作者还通过对 MHP 采样生成一个事件序列，其中 $\lambda_i(t)$ 表示第 i^{th} 维事件的强度函数。

$$\lambda_i(t) = \boldsymbol{\mu}_i + \sum_{j\in\mathcal{P}_i(t)} \int_{-\infty}^{t} \phi_{ij}(t-\tau)\, dU_j(\tau). \tag{3}$$

这里，U_j 表示第 j^{th} 维的计数过程，$U_j(t)$对应于时间 t之前在第 j^{th} 维中发生的事件数量。此外，$\boldsymbol{\mu}_i$ 是底层的泊松强度，ϕ_{ij} 是将第 i维与第 j维因果关联起来的核函数，即$\mathcal{P}_i(t) = \{j : \phi_{ij} \neq 0\}$。

C¨uppers 等人（2024年）提出了一种因果模型，用于生成事件序列，其中，原因的单个事件触发效应的事件，并具有动态延迟。因果图 $\mathscr{G}$ 由一组源节点组成 $\mathcal{U}$和效应节点 $\mathcal{V}$，其中 $\mathcal{U}\cap\mathcal{V} = \emptyset$。该模型假设，类型为 $i \in \mathcal{U}$（原因）的事件在时间 t 会导致类型为 $j \in \mathcal{V}$（效应）的事件在某个

时间 $t' \geq t$ 以概率 $\alpha_{i,\ j}$发生。此外，假设类型为 $i \in \mathcal{U}$ 的事件由强度为 λ_i 的齐次泊松过程采样，即 $\mathcal{S}^i = \Phi(\lambda_i)$。类型为 $j \in \mathcal{V}$的一些事件由噪声导致，噪声由另一个强度为 λ_j 的泊松过程描述，即 $\mathcal{W}_j = \Phi(\lambda_j)$。

其中的事件 $\mathcal{S}^i$ 会导致其中出现延迟事件 $\mathcal{S}^{i,\ j}$ 其构造[5] 如下[6]:

$$\mathcal{S}^{i,j} = \left\{ v \cdot (t+d) + (1-v) \cdot \xi : t \in \mathcal{S}^i, d \sim \text{Exp}(\Theta_{i,j}), v \sim \text{Bern}(\alpha_{i,j}) \right\} \setminus \{\xi\}. \tag{4}$$

最后，类型为 $j \in \mathcal{V}$ 的所有事件被定义为集合 $\mathcal{S}^j$ （见（C¨uppers 等人，2024年，公式 3））：

$$\mathcal{S}^j = \left(\bigcup_{i \in \mathcal{P}_j} \mathcal{S}^{i,j} \right) \cup \mathcal{W}_j, \tag{5}$$

其中， $\mathcal{P}_j$ 中的不同事件独立地导致类型为 $j \in \mathcal{V}$的事件，即结构方程通过简单的“或”运算将父变量分组。大多数以往研究假设，原因只会增加效应发生的可能性。然而，现实世界的系统通常包含抑制效应，即一个事件可能会降低另一个事件发生的可能性。例如，Kim 等人（2011年）研究了一个事件可能激发或抑制另一个事件的情况。我们的模型同时纳入了正向和负向影响，使其能够更灵活地适用于各种应用场景。

表1 概述了现有文献中的主要局限，并说明我们提出的因果模型如何解决这些问题。由于篇幅限制，我们已将其推迟至附录B。

4. 比较因果模型

讨论了如何生成因果模型及其事件序列后，我们现在来量化两个因果模型之间的距离。具体而言，我们评估一个模型在由另一模型生成的 CES 上的表现，反之亦然。最后，我们报告两个模型在多个 CES 实现上的平均误差。

我们用 $\mathfrak{C}$ 表示因果模型，其参数由三元组 $(\Lambda,\ \Theta,\ \bar{\tau})$ 表示。给定时间范围 $T \in \mathbb{R}^+$，该模型会生成一个随机触发事件序列 $\Phi \in \mathbb{S}$ 和一个截至时间 T 的事件序列 $\mathcal{S} \in \mathbb{S}$，记作 $(\Phi,\ \mathcal{S}) \sim \mathfrak{C}(T)$，其中 $\mathcal{S} \subseteq \Phi$.

为便于比较不同模型，我们引入一致的记法，以区分它们的参数和生成的序列。考虑两个因果模型 $\mathfrak{C}_A$ 和 $\mathfrak{C}_B$，其中 $(\Phi^A,\ \mathcal{S}^A) \sim \mathfrak{C}_A(T)$，且 $(\Phi^B,\ \mathcal{S}^B) \sim \mathfrak{C}_B(T)$。接着，我们定义函数 $\tilde{f}_B(i,\ t,\ \mathcal{S}^A,\ \Phi^A)$，该函数使用模型 $\mathfrak{C}_B$ 对事件序列 $\mathcal{S}^A$ 进行预测，定义如下：[7]

$$\tilde{f}_B(i, t, \mathcal{S}^A, \Phi^A) = u_i^A(t) \cdot \mathbb{I}\left\{ \sum_{j \in \mathcal{P}_i^B} \Theta_{i,j}^B \, x_j'^A(t) \geq 0 \right\}. \tag{6}$$

我们定义如下距离，以量化 $\mathfrak{C}_B$ 在 $\mathcal{S}^A$ 上的预测误差：

$$d_B(\mathcal{S}^A, \Phi^A) = \frac{1}{n} \sum_{i \in [n]} \mathop{\mathbb{E}}_{t \in \Phi_i^A} \left[\left| \tilde{f}_B(i, t, \mathcal{S}^A, \Phi^A) - \tilde{f}_A(i, t, \mathcal{S}^A, \Phi^A) \right| \right]. \tag{7}$$

5. 我们已将（C¨uppers等人，2024年，式2）及其描述重新表述为方程4。6. 出于实际考虑，ξ 是一个非常大的有限值，可充当 ∞的代理指标。7. 上标 A 在 $u_i^A(t)$ 与 $x_j'^A(t)$ 中表明这些量分别由 Φ_i^A和 $\mathcal{S}^A$导出。

To m为使因果模型 $\mathfrak{C}_A$ 和 $\mathfrak{C}_B$ 之间的距离度量具有对称性，我们定义

$$d_{A,B}(\mathcal{S}^A, \Phi^A, \mathcal{S}^B, \Phi^B) = \sqrt{d_B(\mathcal{S}^A, \Phi^A) \cdot d_A(\mathcal{S}^B, \Phi^B)}. \tag{8}$$

为使距离度量不受采样序列影响，我们对多个样本取均值，将最终距离度量定义为

$$\bar{d}_{A,B} = \mathop{\mathbb{E}}_{\substack{(\Phi^A,\mathcal{S}^A)\sim\mathfrak{C}_A(T) \\ (\Phi^B,\mathcal{S}^B)\sim\mathfrak{C}_B(T)}} \Big[d_{A,B}(\mathcal{S}^A, \Phi^A, \mathcal{S}^B, \Phi^B) \Big]. \tag{9}$$

¯为通过实证评估所提出距离度量的统计稳定性$d_{A,\ B}$，我们考察其方差随有效样本量增加而衰减的情况。图3 提供了实证证据，表明随着观测到的数据增多，该估计量会变得更加稳定。

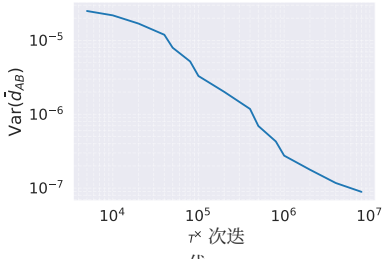


图 3. 两个独立因果模型之间估计距离的经验方差，随有效样本量 T ×次迭代的变化情况。

5. 因果时间交互图

在本节中，我们将事件视为节点之间的交互，这些交互共同构成一个时间交互图。因此，我们不再有 n 个不同的事件，而是拥有一个包含 n 个节点的图，从而总共产生 $E = \binom{n}{2}$ 条可能的边。

我们首先生成节点特征向量$\mathbf{x}_a \in \mathbb{R}^r,\ \forall a \in [n]$，在实践中，它们通过对支撑集为 $\mathbb{R}^r$的分布进行随机采样获得，例如多元正态分布。然后，我们通过边特征向量$\mathbf{z}_{(a,b)} \in \mathbb{R}^r$编码两个节点 a 与 b 之间的关系，该向量由其各自的节点特征$\mathbf{x}_a$ 和$\mathbf{x}_b$ 经由一个对称[8] 函数$\mathbf{z}_{(a,b)} = g(\mathbf{x}_a,\ \mathbf{x}_b)$计算得到，并确保$\mathbf{z}^{(a,b)} = \mathbf{z}^{(b,a)}$。作为一种简单的选择，我们定义

$$g(\mathbf{x}_a, \mathbf{x}_b) = \mathbf{x}_a \odot \mathbf{x}_b. \tag{10}$$

在定义了边特征之后，我们现在将边之间的因果影响建模为这些特征的函数。由于边事件之间的因果影响本质上具有方向性（例如，一次交互可以影响另一次交互，而无需互惠），我们要求这个由 h 表示的函数是非对称的。设两条边的特征分别为$\mathbf{z}$和$\mathbf{z}'$，则

$$h(\mathbf{z}, \mathbf{z}') = \sin\left(\nu_0 \tanh\left(\mathbf{z}^\top \mathbf{B}\, \mathbf{z}'\right)\right), \qquad \nu_0 \in \mathbb{R}, \quad \mathbf{B} \neq \mathbf{B}^\top. \tag{11}$$

正弦函数将输出限制在 $[-1, 1]$，而 ν_0 则增加了可调缩放，使 h更加灵活。为确保 h的不对称性，我们将 $\mathbf{B}$ 构造为斜对称矩阵，如下所示：

$$\hat{\mathbf{B}} = \left[\hat{\mathbf{b}}_i\right]_{i=1}^r,\ \hat{\mathbf{b}}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{r\times r}); \quad \mathbf{B} = \hat{\mathbf{B}} - \hat{\mathbf{B}}^\top. \tag{12}$$

8. 我们假设交互是无向的；若交互是有向的，则 $g(\cdot,\ \cdot)$ 必须是不对称的。

边 j 对 i 的因果影响可直接定义为 $\Theta_{i,\,j} = h(\mathbf{z}_i,\ \mathbf{z}j)$。然而，从设计角度来看，这种表述存在局限。为了进一步推广该模型，我们对 h 的输出进行阈值化，以控制生成的因果图的稀疏性 。此外，我们允许将某些边指定为非因果，这意味着它们既不会对其他边产生因果影响，也不会受到其他边的影响，因此在数据中表现为伪噪声。

我们首先对影响函数 h 的输出进行阈值化，以控制稀疏性。具体而言，对于阈值 $\nu_1 \in (0,\ 1)$，我们定义函数 $\psi(x) = x \cdot \mathbb{I}\{|x| \geq \nu_1\}$.，随后˜定义因果影响矩阵 Θ（在引入非因果边之前），其元素为

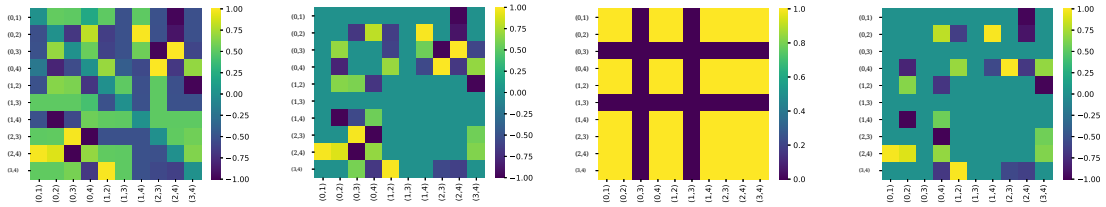
$$\tilde{\Theta}_{i,j} = \psi\left(h(\mathbf{z}_i, \mathbf{z}_j)\right). \tag{13}$$

在现实场景中，一些边对应伪事件或无关事件，与其他边不存在因果关系。为对此进行建模，我们从 $[E]$ 中均匀随机抽取 l 条不同的边，并将这些 $l < E$ 边指定为非因果边。为此，我们构造一个对称二元掩码 $\mathbf{M}_l \in \{0,1\}^{E\times E}$，并将对应的 l 行和列中的元素设为 0。随后，因果影响矩阵 Θ可表示为:

$$\Theta = \tilde{\Theta} \odot \mathbf{M}_l. \tag{14}$$

最后，可通过$\mathbf{A}_{i,\ j} = \mathbb{I}\{\Theta_{i,j} \neq 0\}$.得到因果图$\mathbf{A}$

在图 4中，我们展示了 CTIG 模型在一个包含 5 个节点的图上实例化时，因果影响矩阵 Θ 的逐步构建过程。



(一) $\mathbf{H}$ $\nu_0 = 100$ (b) $\tilde{\Theta}$, $\nu_1 = 0.55$ (c) $\mathbf{M}_l, l = 2$ (d) Θ

Fi图 4：为包含 $n = 5$ 和 $r = 5$的图构造 CTIG 因果参数。

6. 反事实分析

动机 沿用 § 4中的记号，我们将 真实 因果模型记为 $\mathfrak{C}_0 = (\Lambda_0, \Theta_0)$，将估计 模型记为 $\mathfrak{C}_\star = (\Lambda_\star, \Theta_\star)$，将 失真 因果模型记为 $\mathfrak{C}_\dagger = (\Lambda_\dagger, \Theta_\dagger)$。。

估计模型的距离函数为 $d_\star$： $\mathbb{S} \times \mathbb{S} \to [0, 1]$，它将因果模型 $\mathfrak{C}$ 生成的事件序列 $\mathcal{S} \in \mathbb{S}$ 及其触发序列 $\Phi \in \mathbb{S}$ 映射为一个标量性能误差。因此，估计模型在真实因果模型生成的序列上的性能为 $1 - d_\star(\mathcal{S}^0, \Phi^0)$，而其在失真模型生成的序列上的性能为 $1 - d_\star(\mathcal{S}^\dagger, \Phi^\dagger)$。

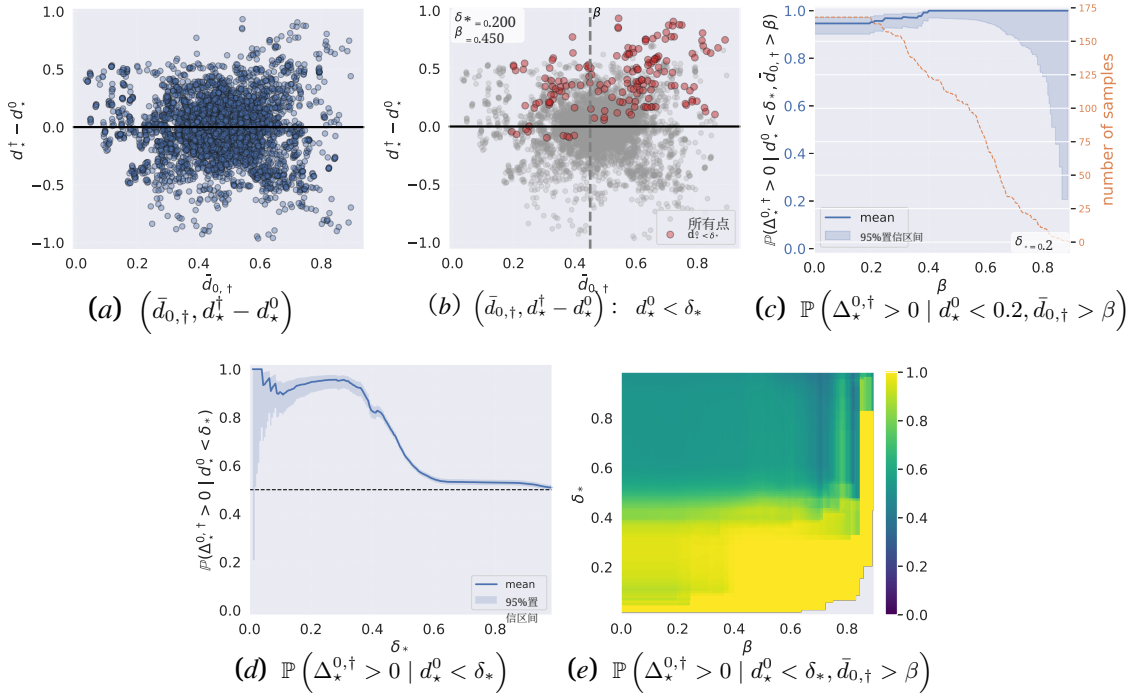
比较估计模型在失真序列与真实序列上的性能，我们将差距记为 $\Delta_\star(\mathcal{S}^0, \Phi^0, \mathcal{S}^\dagger, \Phi^\dagger)$，其定义如下

$$\underbrace{\Delta_\star(\mathcal{S}^0, \Phi^0, \mathcal{S}^\dagger, \Phi^\dagger)}_{\Delta_\star^{0,\dagger}} = \left(1 - d_\star(\mathcal{S}^0, \Phi^0)\right) - \left(1 - d_\star(\mathcal{S}^\dagger, \Phi^\dagger)\right) = \underbrace{d_\star(\mathcal{S}^\dagger, \Phi^\dagger)}_{d_\star^\dagger} - \underbrace{d_\star(\mathcal{S}^0, \Phi^0)}_{d_\star^0}.$$

接下来，我们定义映射 $\Delta_\star(\mathcal{S}^0, \Phi^0, \mathcal{S}^\dagger, \Phi^\dagger) \mapsto \bar{d}_{0,\dagger}$，适用于实现 $(\Phi^0, \mathcal{S}^0) \sim \mathfrak{C}_0(T)$ 和 $(\Phi^\dagger, \mathcal{S}^\dagger) \sim \mathfrak{C}_\dagger(T)$，其中 $\bar{d}_{0,\dagger} \in [0,1]$ 表示真实因果模型与失真因果模型之间的距离（见公式 9）。该映射可用于分析性能差距与因果模型 $\mathfrak{C}_0$ 和 $\mathfrak{C}_\dagger$之间距离的对应关系。

假设 6.1 对于任何满足 $d^0_\star \in [0, \delta_*)$ 的因果模型估计 $\mathfrak{C}_\star$，都存在 $\beta \in (0,1)$，使得对于所有具有 $\bar{d}_{0,\dagger} > \beta$ 的失真模型 $\mathfrak{C}_\dagger$，性能差距满足 $\Delta^{0}_{\star,\dagger} > 0$.

换言之，如果失真模型与真实因果模型 $\mathfrak{C}_\dagger$ 之间的距离 $\mathfrak{C}_0$ 足够大，并且 $\mathfrak{C}_\star$ 估计模型足够准确，即其在真实序列上的误差满足 $0 \le d^0_\star < \delta_*$，那么与失真序列相比，估计模型在真实序列上的表现更好。图5给出了这一讨论的实证可视化。



(a) $\left(\bar{d}_{0,\dagger}, d^\dagger_\star - d^0_\star\right)$　(b) $\left(\bar{d}_{0,\dagger}, d^\dagger_\star - d^0_\star\right)$: $d^0_\star < \delta_\star$　(c) $\mathbb{P}\left(\Delta^{0,\dagger}_\star > 0 \mid d^0_\star < 0.2, \bar{d}_{0,\dagger} > \beta\right)$

(d) $\mathbb{P}\left(\Delta^{0,\dagger}_\star > 0 \mid d^0_\star < \delta_*\right)$　(e) $\mathbb{P}\left(\Delta^{0,\dagger}_\star > 0 \mid d^0_\star < \delta_*, \bar{d}_{0,\dagger} > \beta\right)$

图5：因果扭曲下性能差距的实证评估。各面板展示了性能差距 $\Delta^{0}_{\star,\dagger}$及其相关概率如何随因果距离 $\bar{d}_{0,\dagger}$和准确度阈值 δ_*变化，从而为假设 6.1提供支持。 $n = 7$、$T = 10^3$。

在图 5 中，我们通过实证考察性能差距 $\Delta^{0}_{\star,\dagger} = d^\dagger_\star - d^0_\star$与真实因果模型和失真因果模型之间的距离 $d_{0,\dagger}$之间的关系。图 5(a) 展示了性能差距与 $\bar{d}_{0,\dagger}$之间的散点图。在图 5(b) 中，我们将关注范围限定在估计模型足够准确的情形，即 $d^0_\star < \delta_*$ 满足 $\delta_* = 0.2$，并指出，对于满足 $d_{0,\dagger} > 0.45$ 的扭曲，性能差距严格为正。基于这一观察，图 5(c)遍历条件 $d_{0,\dagger} > \beta$ 中的 β 取值，并表明，性能差距为正的概率会随着 β 增大。在图 5(d) 中，我们改为改变 δ_*，并观察到，随着准确度约束放宽，概率 $\mathbb{P}(\Delta^{0}_{\star,\dagger} > 0 \mid d^0_\star < \delta_*)$ 从接近 1 降至 0.5。

最后，图 5(*e*) 展示了关于 $(\beta,\ \delta_*)$的热图，说明对于较大的 δ_*，仅增加 β 不足以保证性能差距始终为正。

应用 假设6.1 指出，如果估计量 $\mathfrak{C}_\star$ 足够准确，即 $d^0_\star \le \delta_*$，¯则存在 $\beta \in (0,\ 1)$，使得对于每一个失真因果模型 $\mathfrak{C}_\dagger$，当$d_{0,\ \dagger} > \beta$时，相应的性能差距以概率1满足 $\Delta^{0,\ \star\ \dagger} > 0$，该概率是在联合抽样（$\mathcal{S}^0,\ \Phi^0$）$\sim \mathfrak{C}_0\,(T)$ 与（$\mathcal{S}^\dagger,\ \Phi^\dagger$）$\sim \mathfrak{C}_\dagger\,(T)$ 下取得的。形式上，这可以写为

$$\mathbb{P}\left(\Delta^{0,\dagger}_\star > 0 \mid d^0_\star < \delta_*, \bar{d}_{0,\dagger} > \beta\right) = 1. \tag{15}$$

¯在许多实际场景中，计算$d_{0,\ \dagger}$ 可能不可行。例如，人们可能只能观察到事件序列，而生成它们的因果模型的参数仍然未知。此外，真实因果模型 $\mathfrak{C}_0$ 可能不可用，或者因果失真序列$\mathcal{S}^\dagger$ 可能由替代机制产生，例如对真实序列 $\mathcal{S}^0$ 通过外部程序加以破坏，从而使对 $d^0_\dagger$以及随之而来的$\bar{d}_{0,\ \dagger}$的直接计算变得不可能。因此，式15 中的陈述在实践中无法直接验证。

然而，我们仍然可以将 $d^0_\star$ 定义为估计模型在真实因果模型生成的序列¯上的预测误差。在缺少$d_{0,\ \dagger}$的情况下，因此我们使用概率

$$p^{0,\dagger}_\star(\delta_*) \triangleq \mathbb{P}\left(\Delta^{0,\dagger}_\star > 0 \mid d^0_\star < \delta_*\right)$$

作为量化 $\mathcal{S}^\dagger$ 中扭曲程度的代理指标 。此外，对于固定的 δ_*，该量能够比较不同的预测器：在数据相同的情况下，取得更高 $p^{0,\ \star\ \dagger}(\delta_*)$ 值的预测器被认为表现更好。

负采样 在实践中，触发事件通常是不可观测的[9]；因此，我们通过负采样构造负事件集合。[10] 然后我们定义

$$\Phi \ = \ \mathcal{S} \ \cup \ \{(j,t) : (i,t) \in \mathcal{S},\ j \in [E] \setminus \{i\}\}\,. \tag{16}$$

请注意，$\Phi \setminus \mathcal{S}$ 中的事件并不保证对于所考虑的因果模型而言是不可行的。[11] 因此，不应将 Φ 解读为真实触发事件的集合。在传导式设置下，负样本仅从观测到的边集合中均匀抽取，因此$j \sim \mathrm{Unif}\,(\{k : (k,\ t) \in \mathcal{S}\} \setminus \{i\})$ 。

实验 A 我们考虑两个因果事件序列：$\mathcal{S}^0$ 由因果模型 $\mathfrak{C}_0(T)$生成，以及 $\mathcal{S}^\dagger$ 由因果模型 $\mathfrak{C}_\dagger(T)$生成。我们将[12] 序列 $\mathcal{S}^0$ 按时间顺序划分为训练集和测试集，即 $\mathcal{S}^{\mathrm{train}} = \mathcal{S}^0(0,\ \tau_*)$ 和 $\mathcal{S}^{\mathrm{test}} = \mathcal{S}^0(\tau_*,\ T)$，其中 $\tau_* \in (0,\ T)$。随后，我们构建反事实测试集 $\bar{\mathcal{S}}^{\mathrm{test}} = \mathcal{S}^\dagger(\tau_*,\ T)$。在传导式设置下，分别为训练和测试生成负样本。

令 Y表示预测模型的性能指标。令U表示训练条件，即用于训练的数据集，并令 X 表示测试数据集。我们采用反事实式

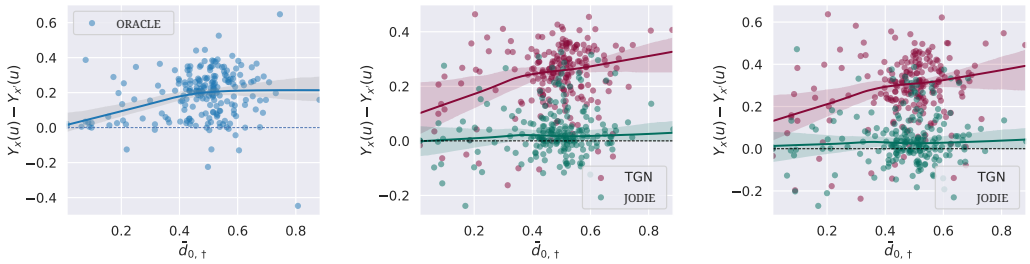
9. TLP 天然适用于正例-未标记学习（Mansouri 和 Ben-David，2025）设置。10. 对于每个 $(i,t) \in \mathcal{S}$，使用 $j \in [E] \setminus \{i\}$ 生成一个负事件 (j,t)。11. 如果事件 (i,t) 对因果模型而言是可行的，则称其为可行事件，即 $\hat{}\ f(\mathcal{P}_i(t)) = 1$；参见公式 2。12. 我们将 $\mathcal{S}$ 在时间区间 $[a,b)$ 上的限制记为 $\mathcal{S}(a,b) = \{(i,t) \in \mathcal{S} :\ t \in [a,b)\}$.

记法，写作 $Y_{X=x}(U=u)=y$，其含义为：在条件 u下训练，并在数据集 x 上评估的模型，其性能为 y。为简洁起见，也写作 $Y_x(u)=y$。

为在我们的设定中实例化这一记法，我们固定训练条件 $u \equiv \mathcal{S}^{\text{train}}$，并考虑两种替代测试条件，即 $x \equiv \mathcal{S}^{\text{test}}$ 和 $x' \equiv \bar{\mathcal{S}}^{\text{test}}$。在固定 u 的情况下，来自另一因果模型的数据所产生的效应由性能差异 $Y_x(u) - Y_{x'}(u)$表示。我们尤其关注其符号，并通过 $\mathbb{P}(Y_x(u) - Y_{x'}(u) > 0)$量化，该指标与 $\mathbb{P}\left(\Delta_\star^{0,\dagger} > 0\right)$一致。

图 6 展示了经验关系：替换 $\mathcal{S}^{\text{test}}$为 $\bar{\mathcal{S}}^{\text{test}}$ 所引起的因果偏移，与由此产生的性能差异 $Y_x(u) - Y_{x'}(u)$之间的关系；结果汇总了多个独立生成的模型对和序列实现。[13] 在图 6(*a*)中，我们报告了 `Oracle` 在反事实评估下取得的性能差距。尽管总体趋势随着因果距离增大而愈发趋于正向，但仍存在少量离群点。正如前文所述，这些偏差是意料之中的，因为负采样并不能保证所有采样得到的负事件都不可行。

其余面板评估了两个TLP模型[14], `TGN` (Rossi 等人，2020年)和`JODIE`(Kumar 等人，2019年)。图6(*b*) 报告了使用平均精度（AP）的结果，而图6(*c*) 报告了使用ROC曲线下面积（AUC）的结果。`JODIE` 的性能在因果距离不断增加的过程中基本保持平坦且接近于零，表明其对因果失真的敏感性有限。相比之下，`TGN` 呈现出明显的正向趋势，性能差距随着因果距离的增大而增加，表明其在原始测试集上的表现始终优于其失真对应版本。这一行为支持如下结论：`TGN`在我们的反事实测试中具有因果敏感性，而`JODIE`则没有。



(*a*) 准确率　(*b*) 平均精度　(*c*) AUC

图6：TLP模型的反事实评估。每个面板展示了性能差距的散点图，并叠加了一条LOWESS平滑曲线（比例0.95），以突出整体趋势。

为完整起见，我们还报告了各模型在原始测试集上的性能 $Y_x(u)\mathcal{S}^{\text{test}}$。Oracle 的平均准确率为 0.693（95%置信区间 $[0.678, 0.708]$），而 TGN 和 JODIE 的平均精度分别为 0.763（95%置信区间 $[0.751, 0.775]$）和 0.563（95%置信区间$[0.551, 0.575]$）。

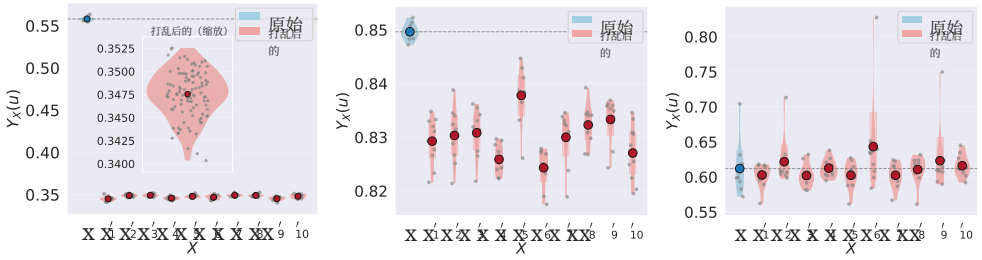
实验B 我们沿用前一个实验的设置。不过，我们不再根据另一种因果模型构造反事实测试集 $\bar{\mathcal{S}}^{\text{test}}$，而是直接通过随机打乱原始测试序列 $\mathcal{S}^{\text{test}}$ 中的事件时间戳来对其进行扰动，从而破坏时间结构，同时

13. 每个点对应 $(\mathcal{S}^0, \mathcal{S}^\dagger)$ 的一次实现和一个训练好的预测器；横轴表示因果距离 $\bar{d}_{0,\dagger}$，纵轴表示性能差距 $Y_x(u) - Y_{x'}(u)$。14. 将基准测试扩展到更多 TLP模型留待未来工作。

保留已观测交互的集合。随后，我们重复实验 A 中相同的评估流程，在 $\mathcal{S}^{\text{train}}$ 上训练模型，并在 $\mathcal{S}^{\text{test}}$的原始版本和打乱后版本上进行评估。Rahman 等人 (2025b)此前也曾考虑过基于时间戳打乱的类似反事实设置，但未能使用底层因果模型。相比之下，我们的设置可以根据已知的因果模型生成测试序列，并提供一种有理论依据的距离度量，用于量化打乱后的序列偏离原始序列的程度。

由于打乱是一种随机过程，我们生成了测试数据的多个打乱后实现，其中x'_i 表示第 i^{th} 个实现。原始序列与打乱后序列之间的因果距离（由因果模型 $\mathfrak{C}_0$测得）均值为 0.4322，95% 置信区间为 $[0.4302,\ 0.4342]$。这使我们能够将打乱视为一种量化的因果扭曲，而非临时采用的压力测试。

在图 7中，我们报告了所达到的性能指标的小提琴图 $Y_X(u)$，其中横轴对应测试数据集 X。原始测试数据上获得的指标值 x 与打乱后的测试集上获得的指标值 $x'_1, \ldots, x'_{10}$之间存在明显差异，这表明时间扭曲会导致性能持续下降。对于 $Y_x(u)$ 和 $Y_{x'_i}(u)$，`Oracle` 和 `TGN` 模型呈现出明显差异，表明它们对时间扭曲敏感，而 `JODIE` 则没有出现这种差异。这些结果再次支持以下结论：`TGN` 通过了反事实测试，而 `JODIE` 则没有。



(*a*) 准确率，Oracle　(*b*) 平均精度，TGN　(*c*) 平均精度，JODIE

图 7：时间戳打乱时模型性能的小提琴图。

7. 结论

在本研究中，我们提出了一种生成因果事件序列的算法，以及一种用于衡量生成这些序列的参数模型之间距离的指标。随后，我们通过提出一种构造方法，将这一机制扩展到时间交互图：节点特征会在边事件上诱导出因果图，从而生成真实结构已知的因果时间交互图（CTIGs）。借助这一设置，我们以两种互补的方式开展反事实评估：一是测试模型在由不同因果模型生成的 CTIG 上的表现；二是通过打乱时间戳引入随机时间扭曲，其严重程度可用所提出的距离加以量化。在这两种设置中，我们都观察到因果扭曲会导致预测性能下降；尤其是，TGN 对诱导的扭曲表现出明显的敏感性，而 JODIE 的表现则基本保持平稳。我们的评估框架提供了一个受控试验平台，用于开发和评估能够利用底层因果结构、而非表面相关性的 TLP模型。此外，该方法还可扩展至其他关注模型在因果变化下性能的预测领域。

参考文献

Emmanuel Abbe。社区检测与随机块模型：近期进展。机器学习研究期刊，18(177):1–86，2018。

Albert-L´aszl´o Barab´asi。《网络科学》一书。网络科学，625，2014。Joscha C¨uppers、Sascha Xu、Ahmed Musa 和 Jilles Vreeken。通过局部因果归因从事件序列中进行因果发现。载于第三十八届神经信息处理系统年会，2024年。Paul Erdos。随机图的演化。国际统计学会通报，38：343–347，1961。

Tianwei Gong、Tobias Gerstenberg、Ralf Mayrhofer 和 Neil R. Bramley。连续时间中的主动因果结构学习。认知心理学，140：101542，2023年2月。ISSN 0010-0285。doi：10.1016/j.cogpsych.2022.101542。

Martin Haenggi。《无线网络随机几何》。剑桥大学出版社，2012。

Alan G Hawkes。若干自激和相互激发点过程的谱。《*Biometrika*》，58(1):83–90，1971。

Amirkasra Jalaldoust、Kateˇrina Hlav´aˇckov´a-Schindler 和 Claudia Plant。通过最小描述长度对霍克斯过程进行因果发现。载于*AAAI*人工智能会议论文集，第36卷，第6978–6987页，2022。

Sanggyun Kim、David Putrino、Soumya Ghosh 和 Emery N Brown。用于群体神经元尖峰活动点过程模型的格兰杰因果关系度量。*PLoS*计算生物学，7(3)：e1001110，2011年。Srijan Kumar、Xikun Zhang 和 Jure Leskovec。预测时序交互网络中的动态嵌入轨迹。载于 *ACM SIGKDD*知识发现与数据挖掘国际会议第*25*届会议论文集，第1269–1278页，2019年。Farnam Mansouri 和 Shai Ben-David。从正例与未标记样本中学习：有限样本规模界。*arXiv*预印本 *arXiv:2507.07354*，2025年。Kimia Noorbakhsh 和 Manuel Rodriguez。反事实时间点过程。神经信息处理系统进展，35：24810–24823，2022。Judea Pearl。因果性。剑桥大学出版社，2009年。Aniq Ur Rahman、Ahmed A Elhag 和 Justin P Coon。时间图学习入门。*ACM*计算调查，58(5)：1–28，2025a。Aniq Ur Rahman、Alexander Modell 和 Justin Coon。通过反事实分析重新思考时间链接预测的评估。载于《我简直不敢相信它没有更好：应用深度学习的挑战》，2025b。网址：https://openreview.net/forum?id=TKydQh6koc。

Emanuele Rossi、Ben Chamberlain、Fabrizio Frasca、Davide Eynard、Federico Monti 和 Michael Bronstein。用于动态图深度学习的时序图网络。*arXiv*预印本 *arXiv:2006.10637*，2020年。

Le Yu、Leilei Sun、Bowen Du 和 Weifeng Lv。迈向更优的动态图学习：新架构与统一库。神经信息处理系统进展，36：67686– 67700，2023年。

Liang Zhao。大数据时代的事件预测：系统性综述。*ACM*计算调查 *(CSUR)*，54(5)：1–37，2021年。

附录 A. 预备知识

定义 1（点过程，（**Haenggi，2012，§ 2.2**）点过程是由可数个点组成的随机集合，这些点位于某个测度空间中，通常是欧几里得空间 $\mathbb{R}^d$。相应的σ-代数由博雷尔集 $\mathcal{B}^d$构成，测度为勒贝格测度。

简单来说，点过程是一个可数随机集合 $\Phi = \{x_1, x_2, \cdots\} \subseteq \mathbb{R}^d$，其元素为随机变量 $x_i \in \mathbb{R}^d$。对于集合 $B \subseteq \mathbb{R}^d$，$\Phi \cap B$ 的基数记为$N(B) \in \mathbb{N}$，而 $N(\cdot)$称为计数测度。

定义 2（一维泊松点过程，（**Haenggi，2012，§ 2.4.1**）参数为 $\lambda \in \mathbb{R}^+$ 的一维泊松点过程（*PPP*）是定义在 $\mathbb{R}$ 上的点过程，满足

1. 对于每个有界区间 $[a, b)$，区间 $N([a, b)]) \in \mathbb{N}$ 中的点数服从均值为 $\lambda(b-a)$的泊松分布，即 $P(N([a, b)]) = k) = e^{-\lambda(b-a)} {}_{k}(\lambda(b-a))k \quad !$，并且 ————

2. 若 $[a_1, b_1)$ $[a_2, b_2)$ $\cdots$、$[a_m, b_m)$ 是互不相交的有界区间，则这些区间中的点数 $N([a_1, b_1))$ $\cdots$、$N([a_m, b_m))$ 是相互独立的随机变量。

我们将从参数为 λ 的泊松点过程中抽取的样本记为 $\Phi(\lambda) \subset \mathbb{R}$，且（$\Phi(\lambda)$ <）满足全序性质，即 $\Phi(\lambda) = \{t_1, t_2, \cdots, t_m\}$，使得对于某个 $m \in \mathbb{N}^+$.，$t_i < t_{i+1}$，$\forall i \in [m]$

定义 *3*（因果模型，（*Pearl，2009*年，*§ 2.2*，*§ 7.1*））设 $\mathcal{U} = \{U_1, \cdots, U_n\}$ 是一组由模型外部因素决定的背景变量，$\mathcal{V} = \{X_1, \cdots, X_n\}$ 是一组由模型中的变量决定的内生变量，即 $\mathcal{U} \cup \mathcal{V}$ 中的变量。因果模型是一个二元组 $\mathfrak{M} = (\mathcal{G}, \Theta_{\mathcal{G}})$，由一个与 $\mathcal{G}$ 兼容的因果图[15] $\mathcal{G}$ 和一组参数 $\Theta_{\mathcal{G}}$[16] 组成。参数 $\Theta_{\mathcal{G}}$ 为每个变量 $X_i \in \mathcal{V}$ 指定一个函数 $x_i = f_i(\mathcal{P}_i(t)\ u_i)$，并为每个 u_i 指定一个概率测度 $\mathbb{P}(u_i)$；其中 $\mathcal{P}_i$ 是 X_i 的父节点，且每个$U_i \in \mathcal{U}$ 都根据 $\mathbb{P}(u_i)$ 随机分布。

15. 因果图是一种有向图，其中从节点 X 到 Y的边表示 X 是 Y的原因。16. $\Theta_{\mathcal{G}}$ 中的元素指定了函数 f_i：$U_i \cup \mathcal{P}_i \mapsto X_i$，$\forall i \in [n]$。

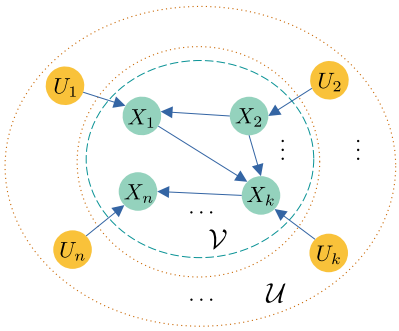


图8：因果模型。

定义 4（单调性，（**Pearl，2009年，§ 9.2**））变量 Y 在因果模型中相对于变量 X 被称为单调，当且仅当

$$\mathbb{I}\{Y = 0 \mid do(X = 1)\} \cdot \mathbb{I}\{Y = 1 \mid do(X = 0)\} = 0.$$

定义 5（外生性，（**Pearl，2009年，§ 7.4**））在因果模型（$\mathscr{G}$，$\Theta_{\mathscr{G}}$）中，变量 X 相对于 Y 是外生的，当且仅当 X 和 Y 在 $\mathscr{G}$ 中没有共同祖先。

附录 B. 相关工作

下表重点列出了现有文献中的主要局限，并说明我们提出的因果模型如何解决这些问题。

表1：文献中的局限性及其通过我们提出的因果模型得到的解决方案。

	局限性	解决方案
Noorbakhsh 和 Rodriguez (2022):		
(*L1*)	一个变量的稀疏化过程是独- 立于其他变量。	我们将接受概率定义为一个 其他变量的函数，通过结- 结构方程。
C¨uppers 等人(2024年):		
(*L2*)	类型为 i 、发生于时间 t 的事件，最多只能 导致某个时间发生的一个类型为 j 的事件 $t' > t$。	事件的影响会持续有限的 在此持续时间内，它可以导致一个事件 i 多次。
(*L3*)	一个事件是单一原因的结果，而 多个原因共同作用的更现实情形 以产生某种效应的情况尚未得到探索。	在结构方程模型方程2 中我们可以看到，多个—— 多个事件共同导致一个事件。
C¨uppers 等人 (2024年); Noorbakhsh 和 Rodriguez(2022); Jalaldoust 等人 (2022):		
(*L4*)	因果关系严格是 正向的，即，原因只能导致 效应的发生。	在这项工作中，我们同时研究正向和 负向因果关系，即，兴奋性和 抑制性。

附录 C. 因果模型的性质

命题 6 对于所有 $\mathbf{A} \in \{0,1\}^{n\times n}$，因果模型$(\mathscr{G}, \Theta_{\mathscr{G}})$ 都是半马尔可夫的。

证明 因果图 $\mathscr{G} \subset \{\mathcal{U}\times\mathcal{V}\}\cup\{\mathcal{V}'\times\mathcal{V}\}$ 由两个单向二分子图组成，一个从 $\mathcal{U}$ 指向 $\mathcal{V}$，另一个从 $\mathcal{V}'$ 指向 $\mathcal{V}$。从任意节点出发，都无法到达 $\mathcal{U}$ 或 $\mathcal{V}'$中的节点。此外，从 $\mathcal{V}$ 中的节点出发，无法返回该节点，因为 $\mathcal{V}$中的节点没有指向外部的边。因此，不可能在$\mathscr{G}$中形成有向环，由此可得 $\mathscr{G}$是无环的，因此因果模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 是半马尔可夫的（Pearl，2009年，§ 3.2）。 ■

命题 7 因果模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 对于所有$\mathbf{A} \in \{0,1\}^{n\times n}$都不是马尔可夫的。

证明 在所提出的模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 中，集合 $\mathcal{U}\cup\mathcal{V}'$ 表示背景变量，而 $\mathcal{V}$则构成内生变量。模型要满足马尔可夫性，必须是半马尔可夫的，此外，背景变量还必须相互独立 (Pearl，2009年，§ 3.2)。因此，半马尔可夫模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 在以下条件下也满足马尔可夫性

(C1) $U_i(t) \perp\!\!\!\perp U_j(t)\ \forall i,\ j \in [n],$
$i \neq j$, (C2) $X_i'(t) \perp\!\!\!\perp U_j(t)\ \forall i,$
$j \in [n]$, (C3) $X_i'(t) \perp\!\!\!\perp X_j'(t)\ \forall i,$
$j \in [n],\ \ i \neq j$。

验证 *(C1)*：我们将 $U_i(t) \perp\!\!\!\perp U_j(t)$ 表示为

$$\mathbb{I}\{t \in \Phi(\lambda_i)\} \perp\!\!\!\perp \mathbb{I}\{t \in \Phi(\lambda_j)\},$$

这等价于验证 $\Phi(\lambda_i) \perp\!\!\!\perp \Phi(\lambda_j)$ 是否成立。由于这两个泊松点过程相互独立，(C1) 成立为真。

验证 *(C2)*：我们将 $X_i'(t)$ 和 $U_j(t)$ 表示为泊松点过程的形式：

$$\begin{aligned} X_i'(t) &= \mathbb{I}\left\{\textstyle\sum_{\bar{t}\in(t-\bar{\tau},t)} X_i(\bar{t}) > 0\right\} \\ &= \mathbb{I}\left\{\sum_{\bar{t}\in(t-\bar{\tau},t)} f\left(\bigcup_{j\in\mathcal{P}_i(\bar{t})} \mathbb{I}\{|\Phi(\lambda_j)\cap(\bar{t}-\bar{\tau},\bar{t})|>0\}, U_i(\bar{t})\right) > 0\right\}. \\ U_j(t) &= \mathbb{I}\{t \in \Phi(\lambda_j)\} = \mathbb{I}\{\lim_{\epsilon\downarrow 0} |\Phi(\lambda_j)\cap[t,t+\epsilon)| > 0\}. \end{aligned}$$

$X_i'(t)$ 的值取决于 $\Phi(\lambda_k)\cap(t-2\bar{\tau},\ t)\ \forall k \in \mathcal{P}_i$ 以及 $\Phi(\lambda_i)\cap(t-\bar{\tau},\ t)$。类似地，对于某个 $\epsilon\downarrow 0$，$U_j(t)$ 的值取决于 $\Phi(\lambda_j)\cap[t,t+\epsilon)$。由于 $(t-2\bar{\tau},\ t)$ 与 $\lim_{\epsilon\downarrow 0}[t,t+\epsilon)$ 之间没有重叠，(C2) 成立为真。

验证 *(C3)*：根据上一段，我们可以说 $X_i'(t)$ 依赖于 $\Phi(\lambda_k)\cap(t-2\bar{\tau},t)\ \forall k \in \mathcal{P}_i(t)$ 和 $\Phi(\lambda_i)\cap(t-\bar{\tau},t)$，并且 $X_j'(t)$ 依赖于 $\Phi(\lambda_m)\cap(t-2\bar{\tau},t)\ \forall m \in \mathcal{P}_j(t)$ 和 $\Phi(\lambda_j)\cap(t-\bar{\tau},t)$。

直观而言，这意味着每个 $X_i'(t)$都会受到其父事件的近期活动以及其自身在各自时间窗口内近期出现的影响。要使 $X_i'(t)$和 $X_j'(t)$

相互独立，它们对应的影响事件集合在任何时刻 t都不得重叠。换言之，它们的依赖域必须随时间保持互不相交。

因此，只有当 $X_i'(t) \perp\!\!\!\perp X_j'(t)$ 时，

$$\begin{aligned}&\{\mathcal{P}_i(t) \cup \{i : X_i'(t) = 1\}\} \cap \{\mathcal{P}_j(t) \cup \{j : X_j'(t) = 1\}\} = \emptyset, \quad \forall i \neq j, \forall t \in \mathbb{R}^+ \\ &\implies \{\mathcal{P}_i \cup \{i\}\} \cap \{\mathcal{P}_j \cup \{j\}\} = \emptyset, \quad \forall i \neq j \\ &\implies \mathcal{P}_i \subseteq \{i\}, \forall i \in [n].\end{aligned}$$

该蕴含关系在以下假设下成立：对于某个 t，时变父集 $\mathcal{P}_i(t)$ 和 $\mathcal{P}_j(t)$ 分别与其结构对应项 $\mathcal{P}_i$ 和 $\mathcal{P}_j$ 相同，并且指标 $X_i'(t) = 1$和 $X_j'(t) = 1$ 可以同时满足。该条件实际上要求潜在因果影响因素集合完全互不重叠。然而，这种互斥性限制很强，且仅在 $\mathbf{A} = \mathbf{I}$ 时才可能出现。因此，我们得出结论：(C3) 对 $\mathbf{A}$ 的所有取值都不成立 。

由于 (C3) 并非总是成立，我们得出结论：对于所有 $\mathbf{A}$，因果模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 都不是马尔可夫的。 ■

推论8 当 $\mathbf{A} = \mathbf{I}$ 时，因果模型 $(\mathscr{G}, \Theta_{\mathscr{G}})$ 是马尔可夫的。

命题 9 $X_i\ (t)$ 相对于 $X_j'\ (t)$ 是单调的，当且仅当 $\Theta_{i,\ j} \in [0, 1]$。

证明 根据定义 4，我们可以说 $X_i\ (t)$ 相对于 $X_j'\ (t)$ 是单调的，当且仅当

$$\mathbb{I}\big\{X_i(t) = 0 \mid do\big(X_j'(t) = 1\big)\big\}\,\mathbb{I}\big\{X_i(t) = 1 \mid do\big(X_j'(t) = 0\big)\big\} = 0.$$

上述方程必须对所有子集 $\mathcal{C}_i \subseteq \mathcal{P}_i \setminus \{j\}$ 成立，即

$$\mathbb{I}\left\{\Theta_{i,j} + \sum_{k \in \mathcal{C}_i} \Theta_{i,k} < 0\right\}\mathbb{I}\left\{\sum_{k \in \mathcal{C}_i} \Theta_{i,k} \geq 0\right\} = 0, \quad \forall \mathcal{C}_i \subseteq \mathcal{P}_i \setminus \{j\}. \tag{17}$$

情况 *1*：$\sum_{k \in \mathcal{C}_i} \Theta_{i,\ k} \geq 0$。为确保方程 17 成立，我们要求

$$\Theta_{i,j} + \sum_{k \in \mathcal{C}_i} \Theta_{i,k} \geq 0.$$

对所有这样的 $\mathcal{C}_i$ 取交集，我们得到

$$\begin{aligned}\Theta_{i,j} &\in \bigcap_{\mathcal{C}_i : \sum_{k \in \mathcal{C}_i} \Theta_{i,k} \geq 0} \left[-\sum_{k \in \mathcal{C}_i} \Theta_{i,k},\ 1\right] \\ &= \left[\max\left\{-\sum_{k \in \mathcal{C}_i} \Theta_{i,k} : \sum_{k \in \mathcal{C}_i} \Theta_{i,k} \geq 0,\ \forall \mathcal{C}_i \subseteq \mathcal{P}_i \setminus \{j\}\right\}, 1\right].\end{aligned}$$

对于 $\mathcal{C}_i = \emptyset$，我们有$\sum_{k \in \mathcal{C}_i} \Theta_{i,k} = 0$，从而得到

$$\Theta_{i,j} \in [0, 1].$$

情形 *2*：$\sum_{k\in\mathcal{C}_i}\Theta_{i,k}<0$。在这种情况下，$\mathbb{I}\left\{\sum_{k\in\mathcal{C}_i}\Theta_{i,k}\geq 0\right\}=0$，因此方程 17 对所有 $\Theta_{i,j}$ 自动成立。因此，这些子集除了假设的定义域 $\Theta_{i,j}\in[-1,1]$ 外，不施加任何额外限制。

将两种情况的结果相交，得到 $\Theta_{i,\ j}\in[0,1]$，由此完成证明。 ■

推论10 如果对于某些 $i\in[n]$，$\Theta_{i,\ j}\in[0,1]$，$\forall j\in\mathcal{P}_i$，那么变量 $X_i(t)$ 相对于所有 $X'_j(t)$ $\forall j\in\mathcal{P}_i$ 都是单调的。然而，结构方程可简化为 $x_i(t)=u_i(t)$，且 $X'_j(t)$ 对 $X_i(t)$ 的因果影响消失[17] $\forall j\in\mathcal{P}_i$。

命题 11 $X'_j(t)$ 相对于 $X_i(t)$ $\forall i,j\in[n]$ 是外生的。

证明 首先，对于所有 $\mathbf{A}\in\{0,1\}^{n\times n}$，由边 $\mathcal{V}'\times\mathcal{V}$ 构成的子图是一个单向二部图。其次，$X'_j(t)\in\mathcal{V}'$ 没有祖先节点。因此，$X_i(t)$ 和 $X'_j(t)$ 没有共同祖先 $\forall i,\ j\in[n]$ 由此得证（参见定义 5）。 ■

定理12（可识别性）如果对于某个 $j\in\mathcal{P}_i$ $X'_j(t)$，其相对于 $X_i(t)$ 是外生的，且 $X_i(t)$ 相对于 $X'_j(t)$ 是单调的，则必要性与充分性概率（PNS）是可识别的，并可通过以下方式确定：

$$\mathrm{PNS}_{i,j}=\mathbb{P}\left(X_i=1\mid X'_j=1\right)-\mathbb{P}\left(X_i=1\mid X'_j=0\right).$$

证明 详细证明请参见（Pearl，2009年，§ 9.2.3）。 ■

17. 为排除平凡可行性，我们验证了 $|S|\neq|\Phi|$ 对所有数据集均成立（等式 $|S|=|\Phi|$ 将表明存在退化）；验证脚本将包含在公开代码发布中。