On the Identifiability of Poisson Branching Structural Causal Model Under Latent Confounding

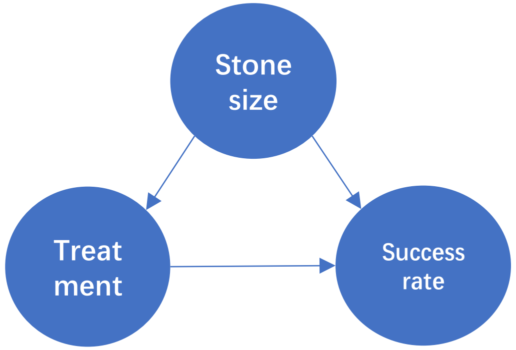
Jie Qiao[1], Zihuai Zeng[1*], Ruichu Cai[1,2], Zhengming Chen[3], Zhifeng Hao[3]*

[1]School of Computer Science, Guangdong University of Technology, China
[2]Peng Cheng Laboratory, Shenzhen 518066, China
[3] College of Science, Shantou University, Shantou 515063, China
[*] Equal contribution

Introduction

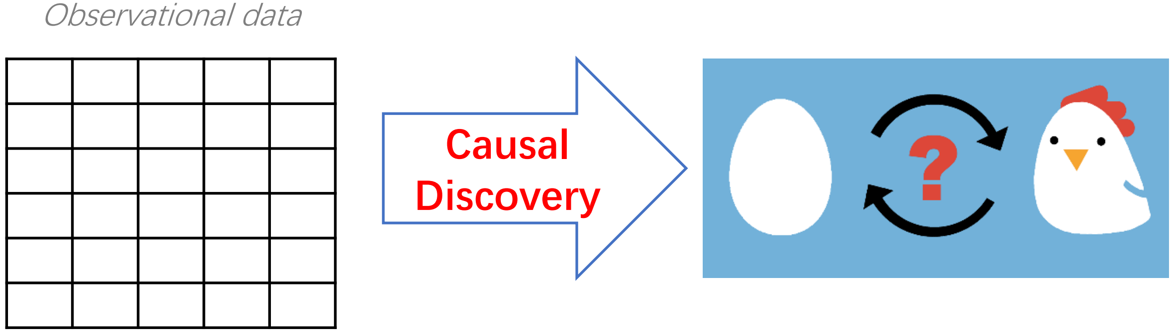
- **Why causal discovery?**
  - Simpson's paradox: Treatment A is better than treatment B in each stone size, however if we do not consider the stone size, Treatment B is better than Treatment A.

Stone size \ Treatment	Treatment A	Treatment B
Small stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**



Causality require us to discover the stone size is the cause of Treatment and success rate.

- **Casual discovery from observational data**
  - Randomized experiments might be unethical or substantial expense.



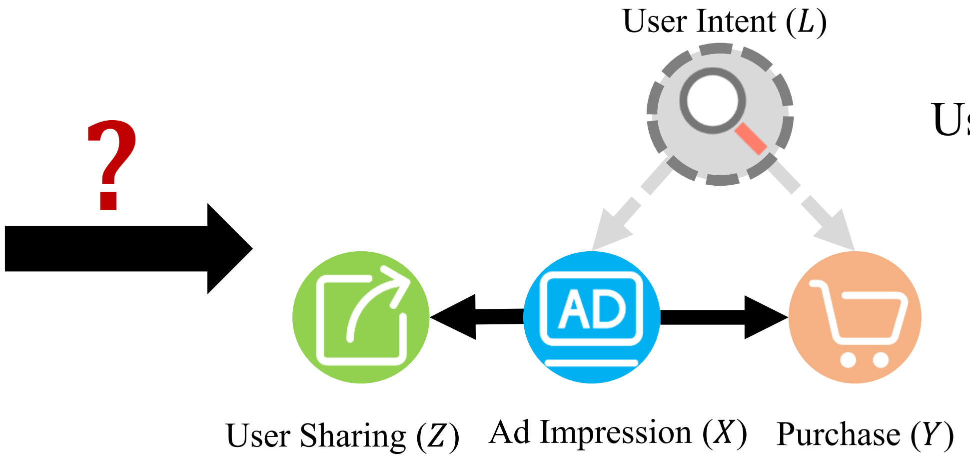
Introduction

Count Data

Observational count data

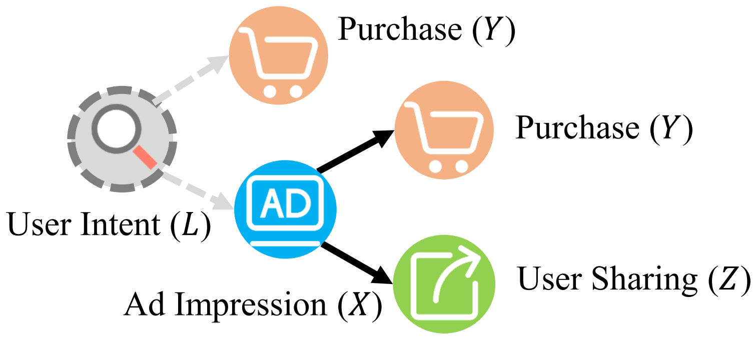
	X	Y	Z
1	2	2	3
2	1	1	2
…	…	…	…
n	1	3	2

Retargeting within digital advertising system

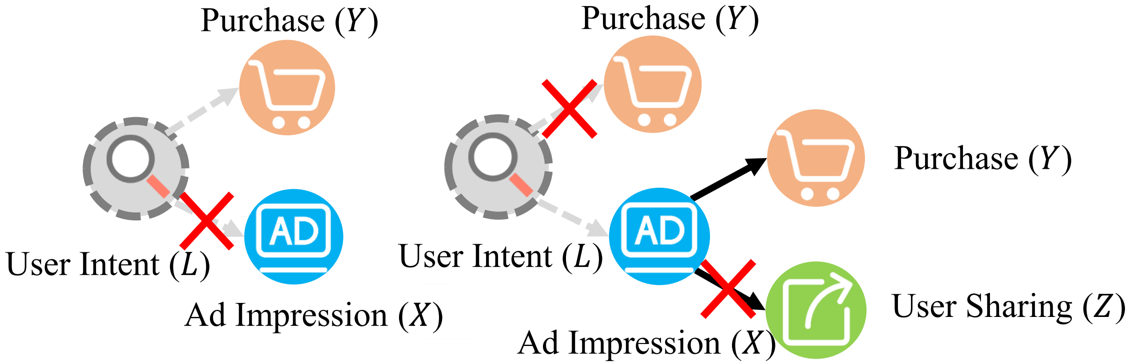


User Intent: searching counts

Branching structure



Each event contributes independently to the occurrence of its child event.

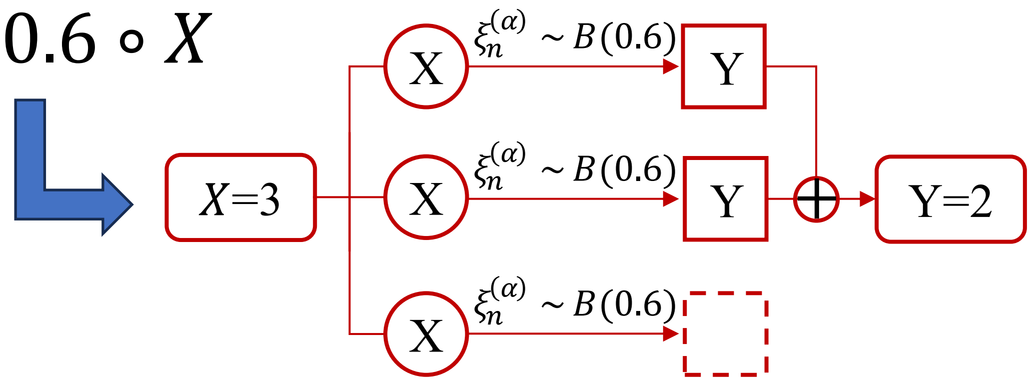


A Purchase event may occur directly from the User Intent, or indirectly through an Ad Impression triggered by User Intent

Introduction

Thinning Operator "∘"

$$Y = 0.6 \circ X$$



$B(0.6)$: Bernoulli distribution with parameter 0.6

$$P(n) = \begin{cases} 0.6, & n = 1 \\ 0.4, & n = 0 \end{cases}$$

A toy example for the thinning operator, each X has a 60% chance of triggering an occurrence of an event Y.

Poisson Branching Structural Causal Model(PB-SCM)

Definition For each observable random variable $X_i \in \mathbf{X}$, let $\epsilon_i \sim Pois(\mu_i)$ be the noise component of X_i, then X_i is generated by:

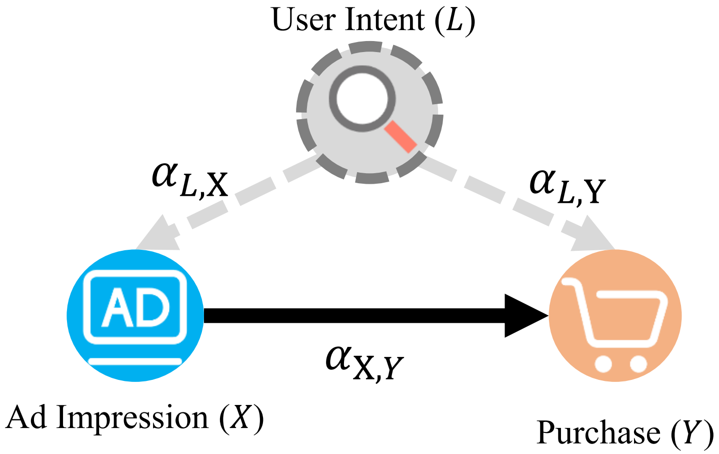
$$X_i = \sum_{k \in Pa(i)} \alpha_{k,i} \circ X_k + \epsilon_i$$

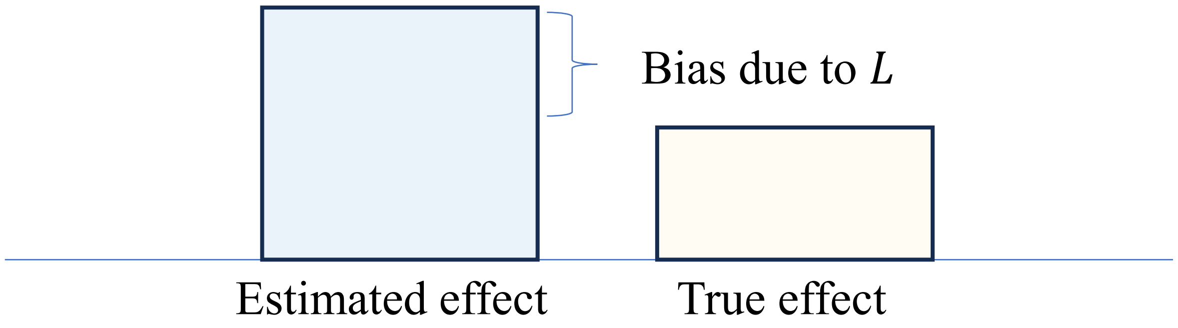
where $\alpha_{j,i} \in (0,1)$ is the causal coefficient from X_j to X_i.

Challenge

The latent variable would lead to :

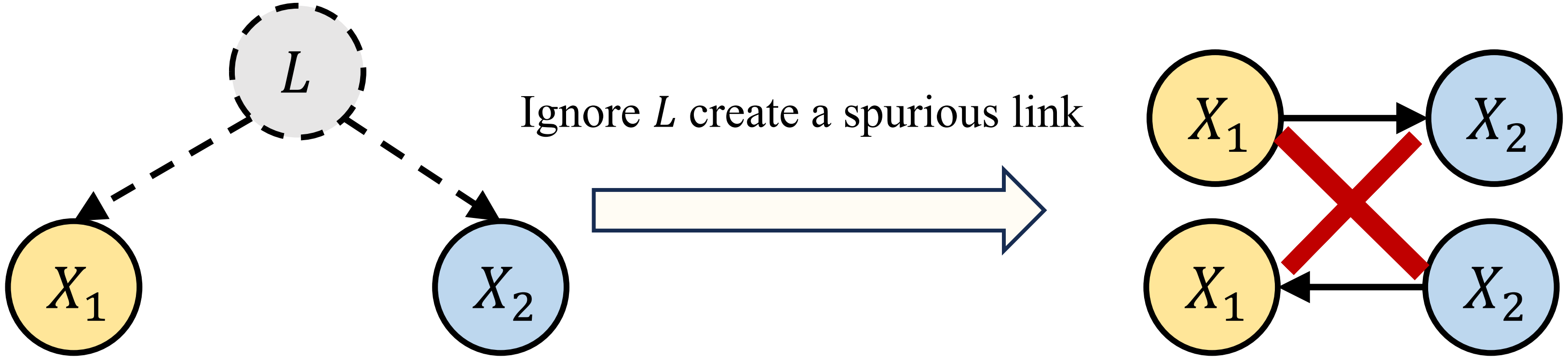
1. Biased effect estimates.





$$Cov(X,Y) = \alpha_{X,Y}Var(X) + (\alpha_{L,X}^2\alpha_{X,Y} + \alpha_{L,X}\alpha_{L,Y})Var(L)$$

2. Spurious causal conclusions



Model

Latent Confounding Poisson Branching Structural Causal Model(LC-PB-SCM)

Definition For each observable random variable $X_i \in \mathbf{X}$, let $\epsilon_i \sim Pois(\mu_i)$ be the noise component of X_i, then X_i is generated by:

$$X_i = \sum_{j \in Pa_{\mathbf{O}}(i)} \alpha_{j,i} \circ X_j + \sum_{k \in Pa_{\mathbf{L}}(i)} \alpha_{k,i} \circ X_k + \epsilon_i$$

where $\alpha_{j,i} \in (0,1)$ is the causal coefficient from X_j to X_i.

Parental Acyclic Directed Mixed Graph

Definition A Parental Acyclic Directed Mixed Graph is an acyclic directed mixed graph defined **over the observed variables**, where the edge set contains directed edge($\rightarrow$) and bi-directed edge ($\leftrightarrow$), the graph satisfiesthe following properties:

Parental Relations: A directed edge $i \rightarrow j$ exists if and only if i is a direct parent of j.

Latent Confounding: A bi-directed edge $i \leftrightarrow j$ exists if and only if i and j share a latent parent.

Acyclicity: The graph contains no directed cycles.

Main finding

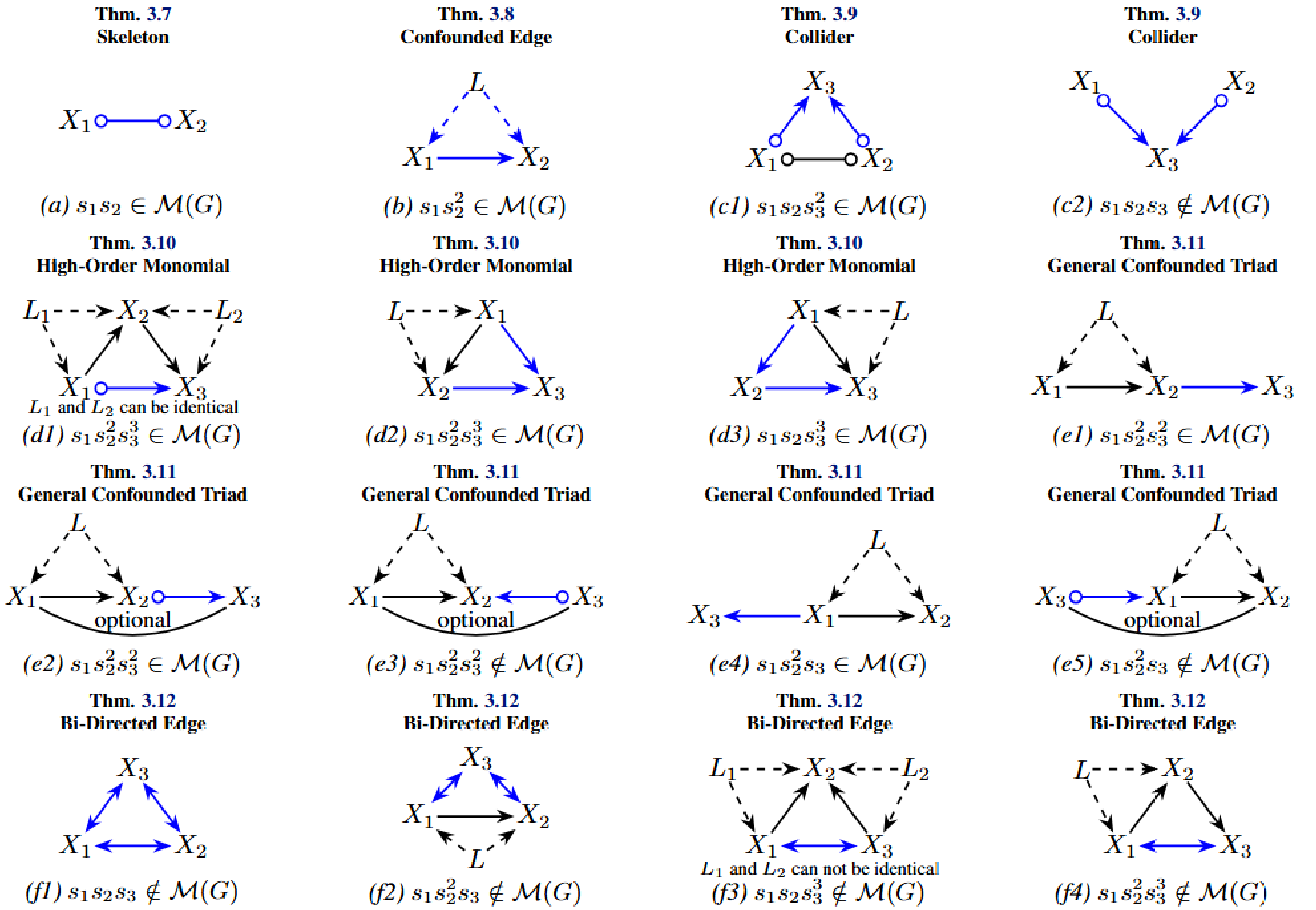


Figure 3. Identifiability of target causal relations (blue edges) in LC-PB-SCM. Monomial conditions are specified in the subfigure titles.

We found 16 identifiable structures based on the analysis on purely observational distribution and show this is complete results under the local 3 variables structure.

Probability Generating Function(PGF)

Definition. Given discrete random vector $\mathbf{X} = [X_1, \ldots, X_d]^T$ taking values in the non-negative integers $\mathbb{Z}^{\geq 0}$. Let $\mathbf{s} = (s_1, \ldots, s_d)$ be a vector of auxiliary indeterminates. The joint probability generating function of $\mathbf{X}$ is defined as:

$$G_{\mathbf{X}}(\mathbf{s}) = \mathbb{E}\left[\prod_{i=1}^{d} s_i^{x_i}\right] = \sum_{x_1,\ldots,x_d} P(x_1, \ldots, x_d) \prod_{i=1}^{d} s_i^{x_i}$$

In PGF, thinning is manifested as variable substitution.

Property F.2 (Binomial Thinning on PGF). Let $Y = \rho \circ X$, where $\circ$ denotes the binomial thinning operator and parameter $\rho \in (0, 1)$. If $G_X(s)$ is the PGF of X, then the PGF of Y is given by $G_Y(s) = G_X(\rho s + 1 - \rho)$. More generally, if a variable X_i contributes to X_j via thinning ρ, this corresponds to a variable substitution in the PGF domain:

$$s_i \mapsto s_i(\rho s_j + (1 - \rho)).$$

Probability Generating Function(PGF)

Definition. Given discrete random vector $\mathbf{X} = [X_1, \ldots, X_d]^T$ taking values in the non-negative integers $\mathbb{Z}^{\geq 0}$. Let $\mathbf{s} = (s_1, \ldots, s_d)$ be a vector of auxiliary indeterminates. The joint probability generating function of $\mathbf{X}$ is defined as:

$$G_{\mathbf{X}}(\mathbf{s}) = \mathbb{E}\left[\prod_{i=1}^{d} s_i^{x_i}\right] = \sum_{x_1,\ldots,x_d} P(x_1, \ldots, x_d) \prod_{i=1}^{d} s_i^{x_i}$$

The joint PGF of LC-PB-SCM

Theorem A.2 (Closed-Form Solution for PGF of LC-PB-SCM). *Consider the LC-PB-SCM as defined. For any node $i \in \mathbf{V}$, define an argument function $\phi_i(\mathbf{s})$ recursively as:*

$$\phi_i(\mathbf{s}) = s_i \prod_{k \in Ch(i)} \left(\alpha_{i,k}\phi_k(\mathbf{s}) + 1 - \alpha_{i,k}\right),$$

variable substitution by thinning operator

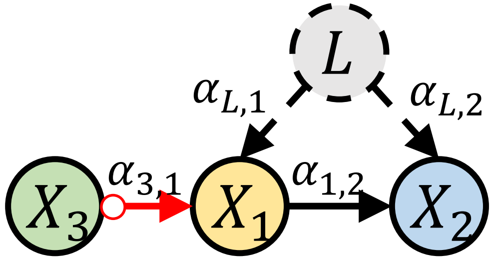
where $\phi_i(\mathbf{s}) = s_i$, if $Ch(i) = \varnothing$, and $\mathbf{s}$ is the vector of PGF arguments. The joint PGF of all variables $\mathbf{X}$ is given by:

$$G_{\mathbf{X}}(\mathbf{s}) = \exp\left(\sum_{i \in \mathbf{V}} \mu_i(\phi_i(\mathbf{s}) - 1)\right).$$

Poisson noise in PGF form

Trie representation

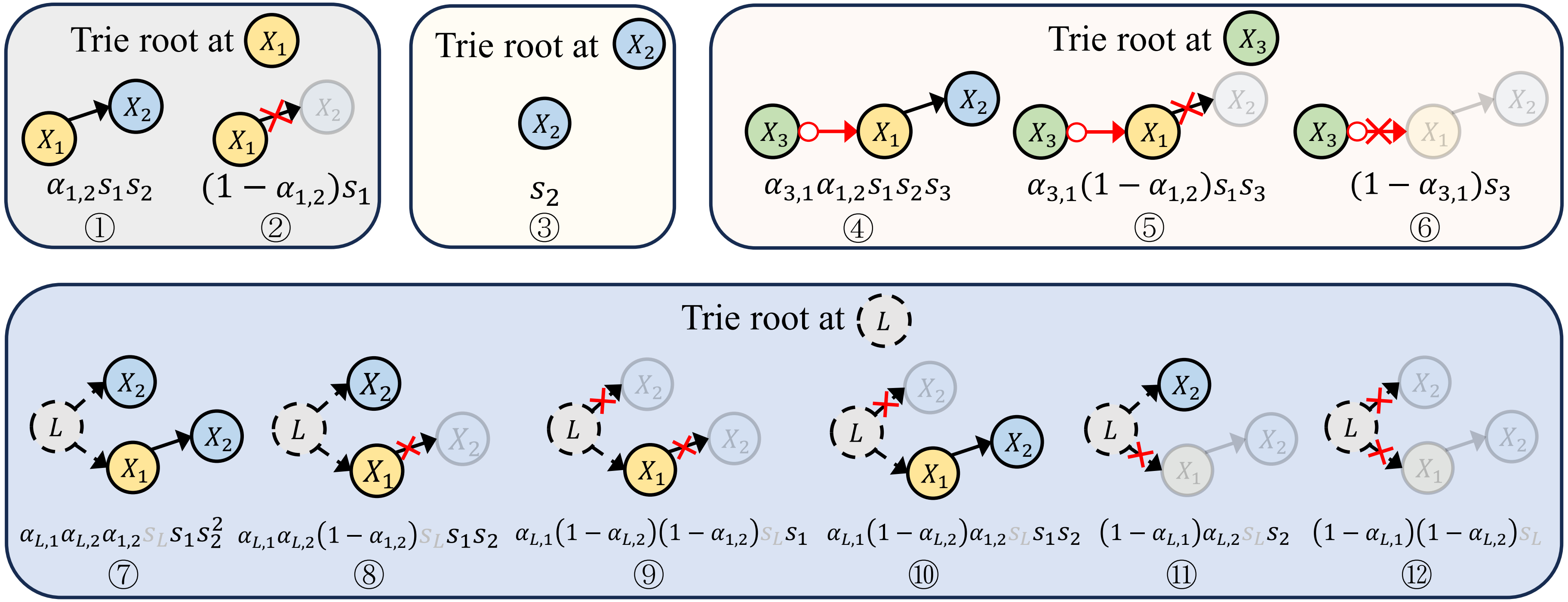
Causal Graph $\mathcal{G}$



Log Probability Generating Function of $\mathcal{G}$

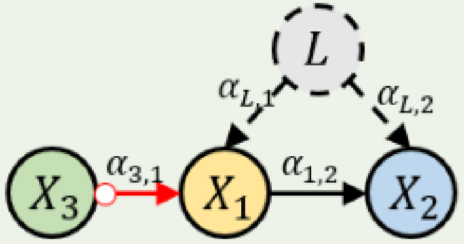
$$\log G_{X_1,X_2,X_3}(s_1, s_2, s_3) = \mu_1(① + ② - 1) + \mu_2(③ - 1) + \mu_3(④ + ⑤ + ⑥ - 1) + \mu_L(⑦ + ⑧ + ⑨ + ⑩ + ⑪ + ⑫ - 1)$$

Monomials and related Tries in $\mathcal{G}$

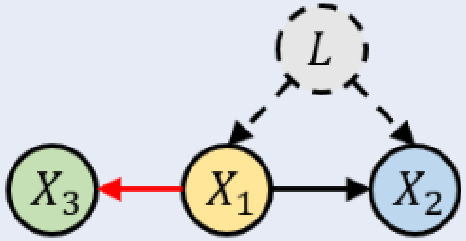


Identifiability

Causal Structure $\mathcal{G}_1$

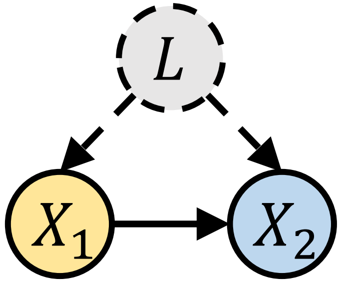


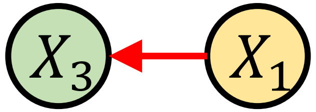
Causal Structure $\mathcal{G}_2$

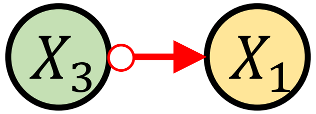


$\mathcal{M}(\mathcal{G}_1) = \{s_1, s_2, s_3, s_1s_2, s_1s_3, s_1s_2s_3, s_1s_2^2\} \neq \mathcal{M}(\mathcal{G}_2) = \{s_1, s_2, s_3, s_1s_2, s_1s_3, s_1s_2s_3, s_1s_2^2\} \cup \{s_1s_2^2s_3\}$

Not all the monomial are required, only some specific monomials need to be identify

With structure 

if $s_1s_2^2s_3 \in \mathcal{M}(G)$, we can identify 

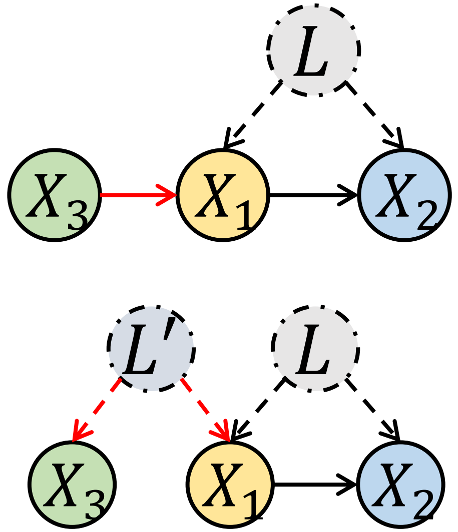
if $s_1s_2^2s_3 \notin \mathcal{M}(G)$, we can identify 

Identifiability

Monomial equivalence

Definition 3.6 (Monomial Equivalence)**.** Let $\mathcal{M}(\mathcal{G})$ be the set of monomials appearing in the log-probability generating function (log-PGF) associated with an LC-PB-SCM over a DAG $\mathcal{G}$ such that $\mathcal{M}(\mathcal{G}) = \{m(\mathcal{T}') \mid \mathcal{T}' \in \mathbb{S}(\mathcal{T}_{\mathcal{G}}(v_k)), k \in \mathbf{V}\}$ where $m(\mathcal{T}')$ denotes the corresponding monomial as define at theorem 3.5. Two causal graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be monomial equivalent if and only if they have the identical set of monomials, i.e., $\mathcal{M}(\mathcal{G}_1) = \mathcal{M}(\mathcal{G}_2)$.

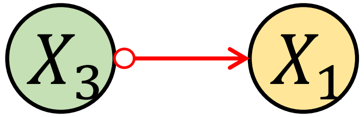
Indistinguishable equivalences



Identical monomial set

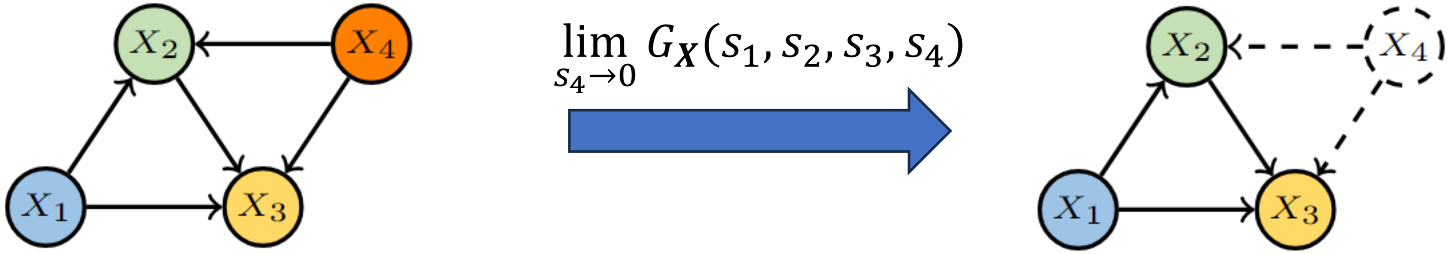
$s_1, s_2, s_3, s_1 s_2, s_1 s_3,$
$s_1 s_2 s_3, s_1 s_2^2$

Identify



Identifiability

The PGF has a local property, allowing for the efficient identification of **local structures.**



By setting $\lim_{s_4 \to 0} G_{\boldsymbol{X}}(s_1, s_2, s_3, s_4)$, we retain only the monomials involving the target variables

Focus on local structures, No need for high-order information

Practical algorithm

Algorithm 1: Causal Discovery for LC-PB-SCM

Input: Count Data $\mathcal{D}$
Output: PPADMG $\mathcal{P}$
Initialize $\mathcal{P}$ with vertices $\mathbf{V}$.

① `// Phase 1: Skeleton Identification`
foreach *pair* (X_i, X_j) **do**
  if *Test*$\left(\frac{\partial^2 \log \hat{G}_{\mathbf{X}}^{\{i,j\}}(\mathbf{s})}{\partial s_i \partial s_j} \neq 0\right)$ **then**
    Add edge $X_i – X_j$

② `// Phase 2: Latent Confounded Edge`
foreach *adjacent pairs in* $\mathcal{P}$ **do**
  if *Test*$\left(\frac{\partial^3 \log \hat{G}_{\mathbf{X}}^{\{i,j\}}(\mathbf{s})}{\partial s_i \partial s_j^2} \neq 0\right)$ **then**
    Orient $X_i \to X_j$ together with
    $X_i \leftarrow\!\!-\!- L -\!-\!\!\rightarrow X_j$

③ `// Phase 3: Triples Orientation`
foreach *connected triples in* $\mathcal{P}$ **do**
  Orient edges based on Theorems 3.9 to 3.12;
return $\mathcal{P}$

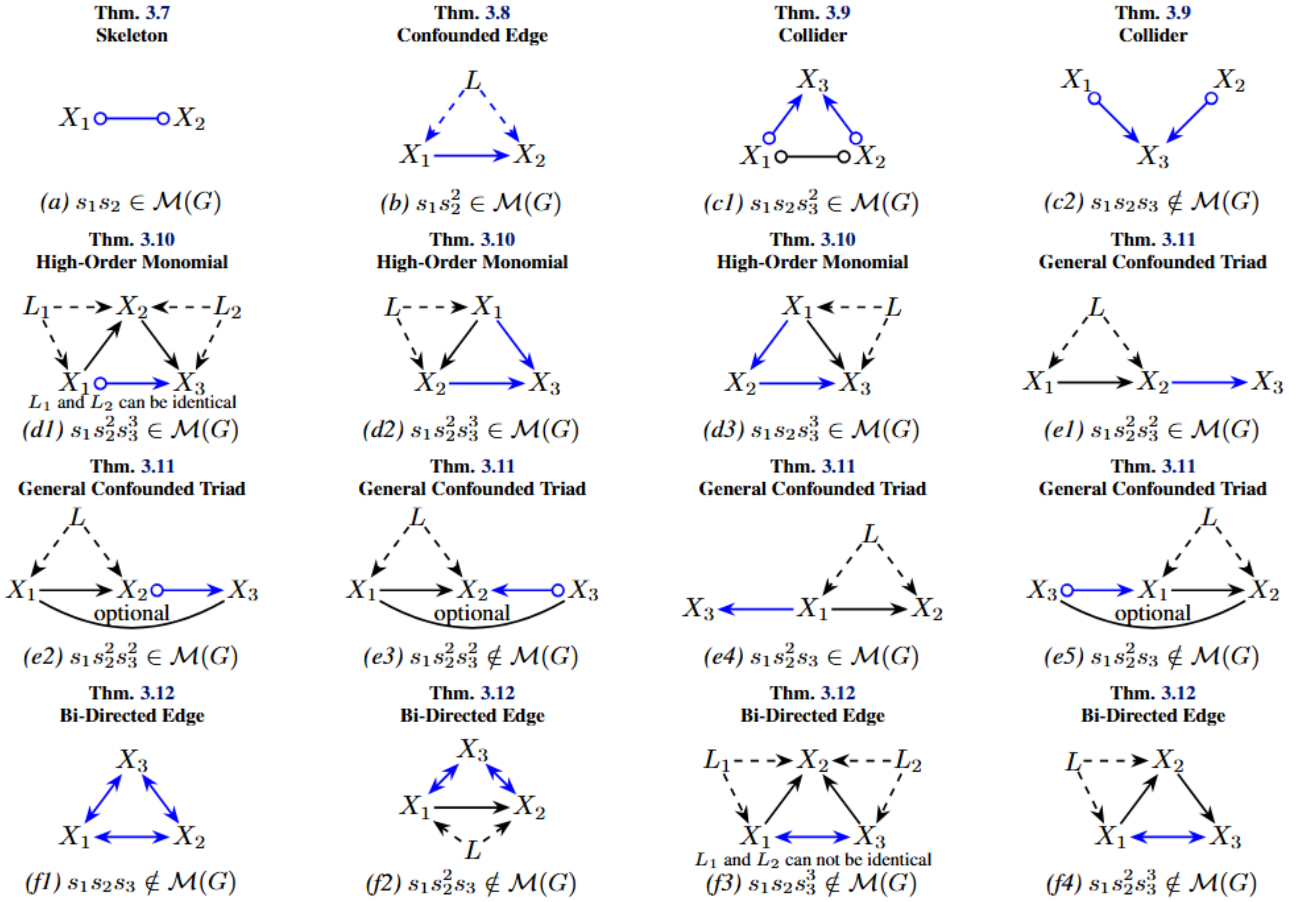


Figure 3. Identifiability of target causal relations (blue edges) in LC-PB-SCM. Monomial conditions are specified in the subfigure titles.

① Skeleton Identification
② Latent Confounded Edge
③ Triples Orientation

Experiment

Random graph experiment

Table 3. Performance comparison of random graphs consisting of 10 variables. Metrics are reported as F1..

Sample Size	LC-PB-SCM	FCI	PBSCM	GES	OCD
5000	**0.60±0.06**	0.58±0.08	0.06±0.09	0.38±0.07	0.17±0.07
10000	**0.68±0.06**	0.59±0.09	0.24±0.10	0.44±0.10	0.20±0.09
30000	**0.73±0.05**	0.61±0.09	0.45±0.08	0.48±0.09	0.22±0.08
50000	**0.75±0.06**	0.63±0.09	0.50±0.06	0.50±0.11	0.35±0.06
100000	**0.77±0.06**	0.64±0.08	0.57±0.04	0.52±0.08	0.28±0.14

Case study

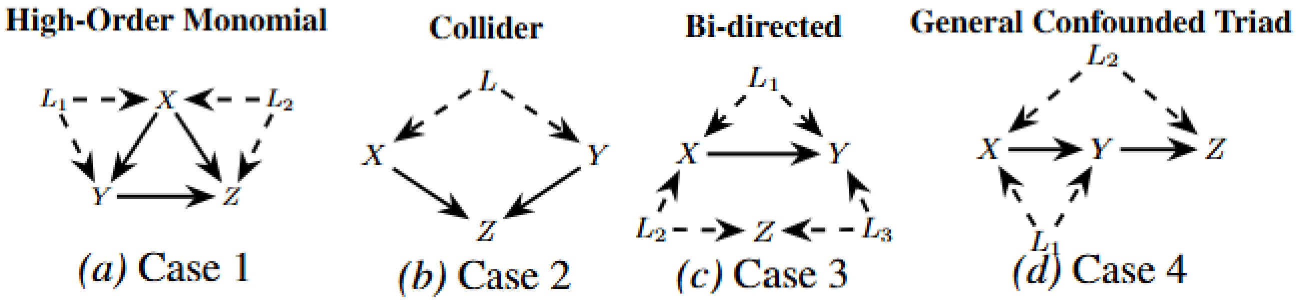


Table 1. Performance comparison measured by F1 scores.

Method	Case 1	Case 2	Case 3	Case 4
Ours	**1.00±0.00**	**0.72±0.19**	**0.84±0.05**	**0.87±0.15**
FCI	0.00±0.00	0.53±0.00	0.00±0.00	0.00±0.00
PB-SCM	0.47±0.00	0.42±0.00	0.47±0.12	0.41±0.07
GES	0.55±0.00	0.33±0.00	0.37±0.00	0.44 ±0.02
OCD	0.46±0.11	0.33±0.03	0.66±0.08	0.58±0.13

Real World Experiment.

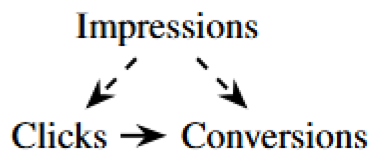
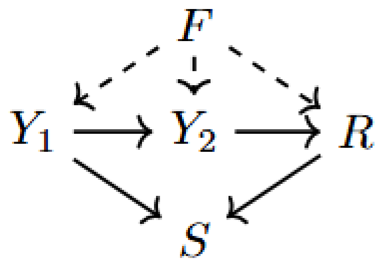


Figure 6. Causal Graph of Marketing Funnel

Our method accurately recovered the causal direction $Clicks \rightarrow Conversions$.



F: Foul, Y1: Yellow card, Y2: Second yellow card, R: Red card, S: Substitution.

we perform structure learning on the skeleton of the true causal graph, and our method successfully recovers the correct edge orientations of $Y_1 \rightarrow Y_2, Y_2 \rightarrow R, Y_1 \rightarrow S, R \rightarrow S$

Thanks!