Cost-effectively Identifying Causal Effects When Only Response Variable is Observable

Tian-Zuo Wang Xi-Zhu Wu Sheng-Jun Huang Zhi-Hua Zhou

National Key Lab for Novel Software Technology, Nanjing University, China

{wangtz, wuxz, zhouzh}@lamda.nju.edu.cn huangsj@nuaa.edu.cn

Learning And Mining from DatA

http://www.lamda.nju.edu.cn

Problem setting

- **Goal:**

In this paper, we aim to identify the **causal effects** of each variable X_i (covariates) on the **response variable** Y (target/outcome/reward) in Pearl's causal framework.

- **Basic assumptions:**

Causal sufficiency + **Faithfulness**.

- **Input:**

observational data of full variables (X and Y).

- **By:**

Discovering related **causal relations** by introducing **interventions** (causal discovery with both observational and interventional data) and estimate the causal effects by **back-door (adjustment) criterion**.

- **Main difference between ours and previous methods:**

In real tasks, it is hard to observe full variables under intervention. We consider such a setting, that **only response variable** is observed under intervention.

- **Mission:**

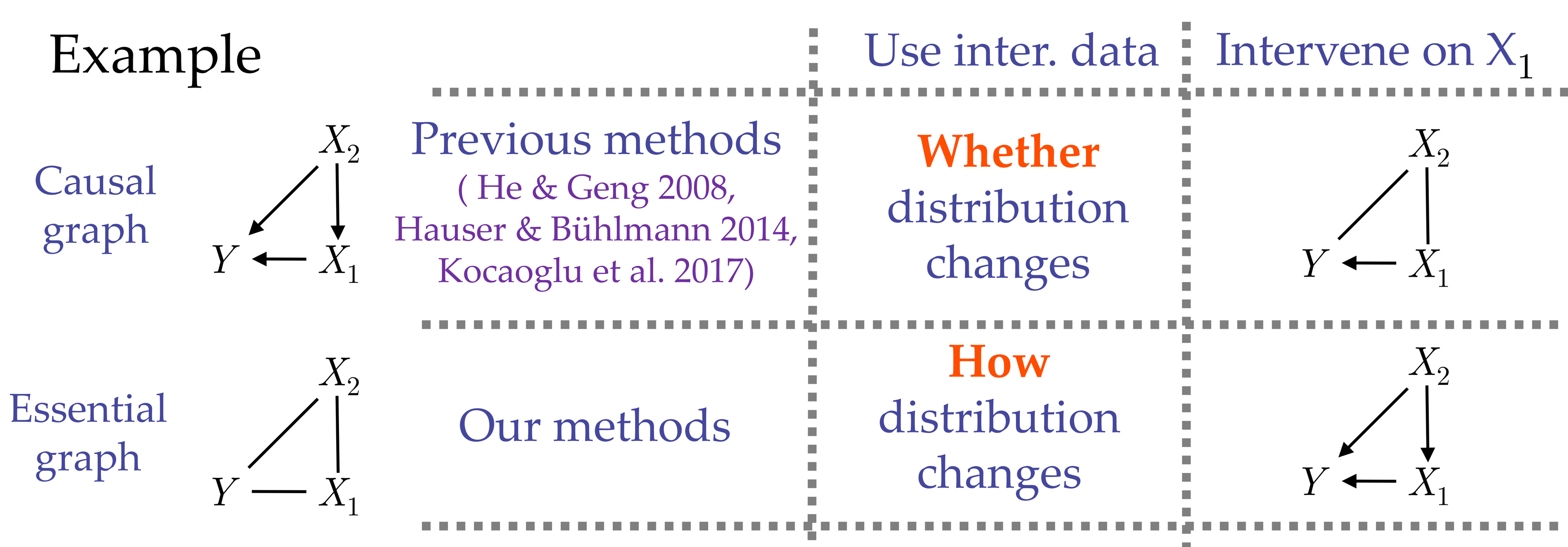
Observational data of full variables

Interventional data of Y

→ Causal effect identification of each X_i on Y

Main innovation

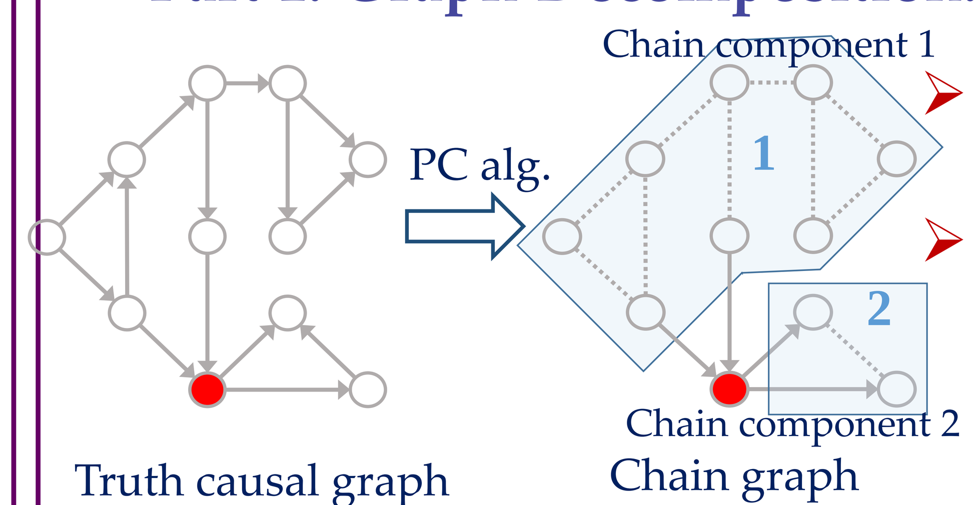
- **How to use the interventional data:**



- **An active intervention strategy to identify causal effects:**

The ACI Algorithm

- **Begin from Essential graph obtained by observational data.**
- **Part 1: Graph Decomposition.**



- Causal discovery in each chain component is independent;
- Ignore the chain component which has no directed path to Y in the chain graph.

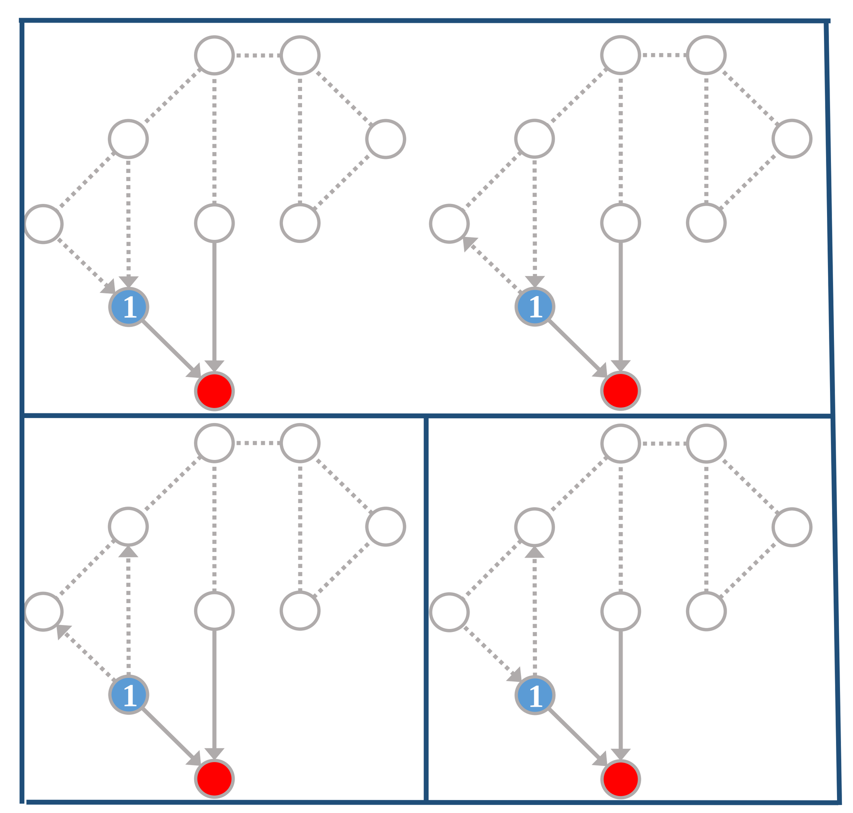
- **Part 2: Structure Inference.**

Back-door criterion

Causal graph ⇄ Causal effect

Out method

Idea: use back-door criterion **reversely**



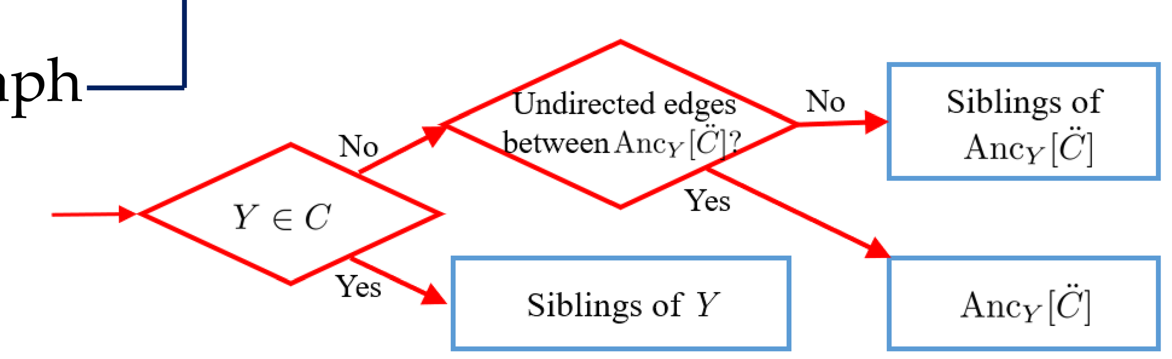
1. **Orient** undirect edges of the intervened variable ;
2. Find **Minimal parental back-door admissible set** for each oriented graph and classify them by the set;
3. **Estimate the causal effects on Y in each class and compare them to interventional data of Y.**

- **Part 3: Intervention Variable Selection.**
  - **Goal:**
    - At least one ancestor edges can be identified;
    - Discover more undirected edges.
  - **Step 1:**

Select a set as the graph

  - **Step 2:**

Select the variable with the maximum sibling in the set



Theoretical analysis

- **Identifiability**

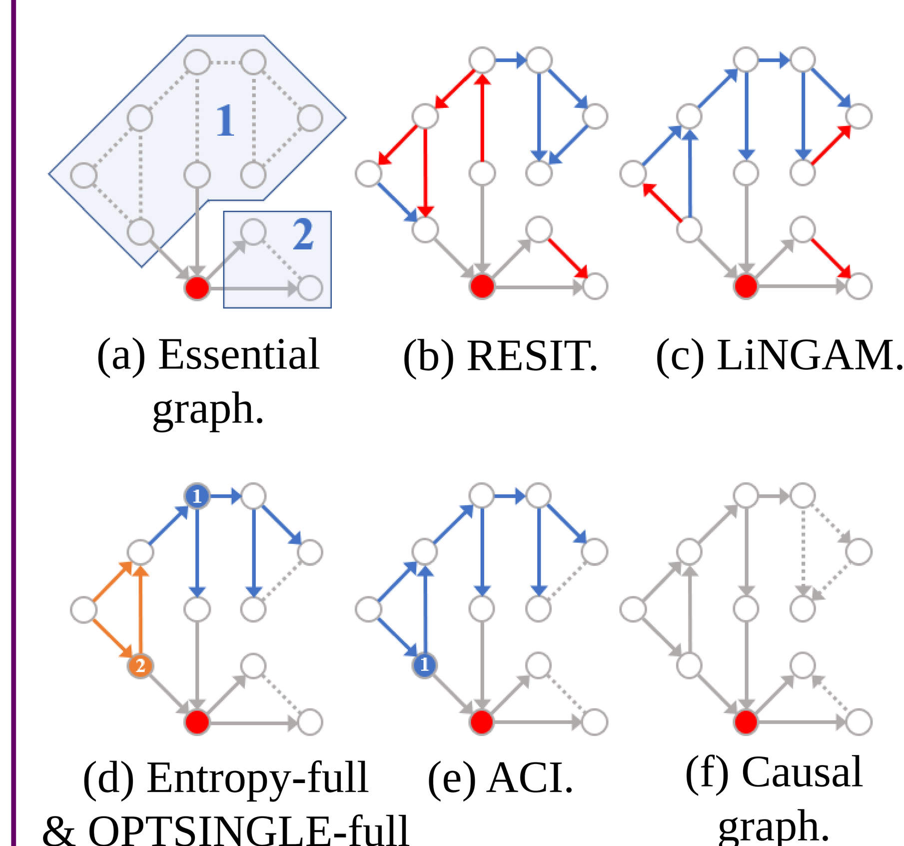
Under the **interventional-faithfulness** assumption, the **ancestor causal structure** is identifiable, which leads to the identifiability of the causal effect of each variable on Y.

- **Intervention times analysis**

	Ratio of Inter. times	Ours	Eberhardt (2007)
Complete causal graph	Y is at the last position	2/3	2/3
	Y is in random position	5/6	1/3

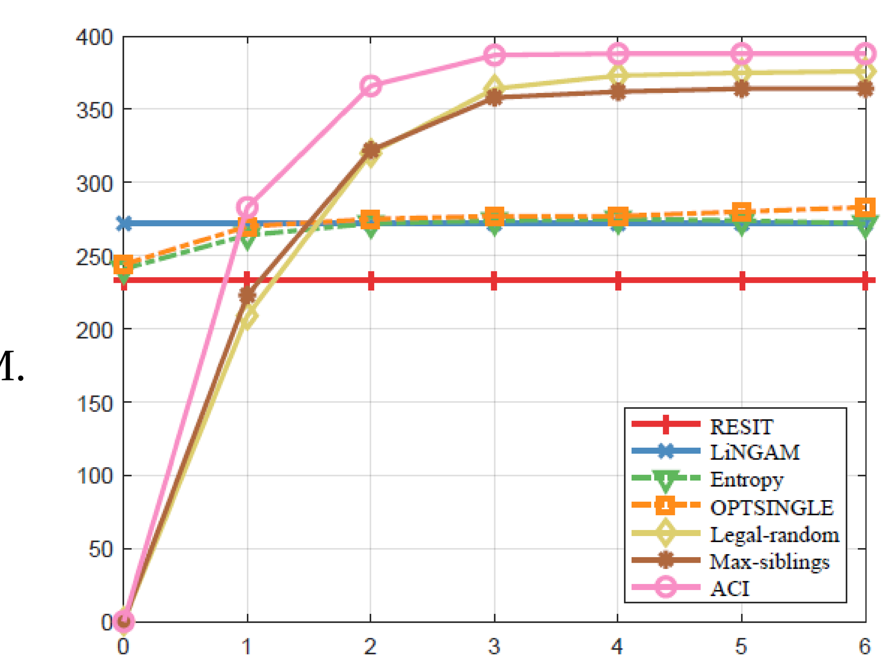
Experiments

Experimental process:



(a) Essential graph. (b) RESIT. (c) LiNGAM. (d) Entropy-full & OPTSINGLE-full (e) ACI. (f) Causal graph.

Simulation



The **number** of newly identified ancestor edges as the **intervention times** grow

Contact:

Paper

Personal web

Email: wangtz@lamda.nju.edu.cn, wangtz1994@gmail.com