Confounder Selection via (Iterative) Graph Expansion

Qingyuan Zhao

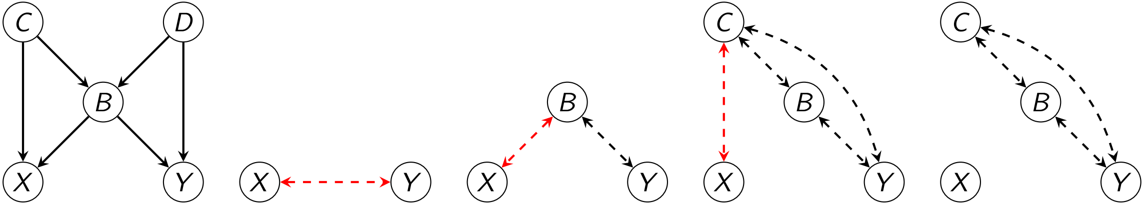
Statistical Laboratory, University of Cambridge

September 17, 2023 @ Pacific Causal Inference Conference

Based on:

1. Confounder selection: objectives and approaches. arXiv:2208.13871 (with F Richard Guo and Anton Rask Lundborg)
2. Confounder selection via iterative graph expansion. arXiv:2309.06053 (with F Richard Guo)

How it works in a nutshell: Illustration with “butterfly bias”



Outline

Outline

Confounder selection

- Arguably the most important task in observational studies.
- Many criteria and methods, often loosely stated, sometimes ill-advised.

Example

Austin (2011) claimed that there are four choices:

1. all measured baseline covariates;
2. all baseline covariates that are associated with treatment assignment;
3. all covariates that affect the outcome (i.e., the potential confounders),
4. all covariates that affect both treatment assignment and the outcome (i.e., the true confounders).

Citing simulation studies, he concluded that "there were merits to including **only the potential confounders** or **the true confounders** in the propensity score model."

Two common heuristics

The conjunction heuristic (a.k.a. the common cause principle)

Contrlling for all covariates "related" to both the treatment and the outcome.

- Very common in practice (Glymour, Weuve, and Chen 2008) and methodological development (Koch et al. 2020; Shortreed and Ertefaie 2017).
- Well known that this may select too few.

The pre-treatment heuristic

Controlling for all covariates that precede the treatment temporally.

- Defended in Rubin (2009): "I cannot think of a credible real-life situation where I would intentionally allow substantially different observed distributions of a true covariate in the treatment and control groups."
- Counter-examples from graphical models: e.g. M-bias (Greenland, Pearl, and Robins 1999).

Graphical approaches

Theorem (Back-door criterion (Pearl 1993, 2009))

Given a treatment X and an outcome Y, a set of covariates S controls for confounding if

1. *S contains no descendant of X (in the notation below, $X \not\rightsquigarrow S$);*
2. *S blocks all back-door paths from X to Y (in the notation below, $X \not\leftrightsquigarrow\leftarrow Y \mid S$).*

- Limitation: requires "full" structural knowledge.

Graphical approaches

Theorem (Back-door criterion (Pearl 1993, 2009))

Given a treatment X and an outcome Y, a set of covariates S controls for confounding if

1. *S contains no descendant of X (in the notation below, $X \not\rightsquigarrow S$);*
2. *S blocks all back-door paths from X to Y (in the notation below, $X \not\leftrightsquigarrow\!\!\leftrightsquigarrow Y \mid S$).*

- Limitation: requires "full" structural knowledge.

Theorem (Disjunctive criterion (VanderWeele and Shpitser 2011))

Suppose the causal graph is faithful. If at least one subset of S controls for confounding, then $S \cap [\mathrm{An}(X) \cup \mathrm{An}(Y)]$ controls for confounding.

- Limitation: verifying the assumption can be as difficult as the task of confounder selection itself.

Definition of confounding

What does confounding mean?

- It is easy to define "no confounding", but surprisingly difficult to define "confounding".
- The dictionary definition of confound is "to confuse and very much surprise someone, so that they are unable to explain or deal with a situation".
- In the causal literature, confounding usually means spurious association due to common causes.

Definition of confounding

What does confounding mean?

- It is easy to define "no confounding", but surprisingly difficult to define "confounding".
- The dictionary definition of confound is "to confuse and very much surprise someone, so that they are unable to explain or deal with a situation".
- In the causal literature, confounding usually means spurious association due to common causes.

- VanderWeele and Shpitser (2013) reviewed several notions of confounding variable ("confounder") and showed that none of them satisfies:
  1. Controlling for all confounders suffices to control for confounding;
  2. Each confounder in some contexts helps eliminate or reduce confounding bias.
- They provided a new definition: "a pre-exposure covariate C for which there exists a set of other covariates X such that effect of the exposure on the outcome is unconfounded conditional on (X, C) but such that for no proper subset of (X, C) is the effect of the exposure on the outcome unconfounded given the subset."
- But this is difficult (or almost impossible?) to use.

Outline

Basic setting

- An **acyclic directed mixed graph** (ADMG) $\mathcal{G}$ describes the causal structure of all relevant variables V.
- An ADMG is a graph with directed ($\rightarrow$) and bidirected edges ($\leftrightarrow$) that has no directed cycle.
  - Directed edges means (direct) causal influences;
  - Bidirected edges means existence of latent common causes (a.k.a. endogeneity).
- $\mathcal{G}$ is not known, but we can make repeated queries about certain structures in $\mathcal{G}$. (Imagine a meeting between a statistician and a subject expert.)

Latent projection and m-separation

- Notation: $\rightsquigarrow$ means a path/walk with no collider (i.e., no part like $\to C \leftarrow$).
- ADMG is desirable for causal modelling because it is closed under **latent projection** ($\mathrm{U} = \mathrm{V} \setminus \tilde{\mathrm{V}}$):

$$A \left\{ \begin{matrix} \to \\ \leftarrow \\ \leftrightarrow \end{matrix} \right\} B\ [\mathcal{G}(\tilde{\mathrm{V}})] \quad \Longleftrightarrow \quad A \left\{ \begin{matrix} \overset{\text{via } \mathrm{U}}{\rightsquigarrow} \\ \overset{\text{via } \mathrm{U}}{\leftsquigarrow} \\ \overset{\text{via } \mathrm{U}}{\leftrightsquigarrow} \end{matrix} \right\} B\ [\mathcal{G}].$$

- Also useful is **m-separation** (Richardson 2003), an extension of Pearl's d-separation:

$$A \leftrightsquigarrow \not\leftrightsquigarrow B \mid C\ [\mathcal{G}] \Rightarrow A \perp\!\!\!\perp B \mid C,$$

  where a path from A to B is m-connected given C if every non-collider on the path is not in C, and every collider on the path is in C or has a descendant in C.
- It is well-known that **m-separation is preserved by latent projection**: for disjoint $A, B, C \subset \tilde{\mathrm{V}}$,

$$A \leftrightsquigarrow \not\leftrightsquigarrow B \mid C\ [\mathcal{G}(\tilde{\mathrm{V}})] \quad \Longleftrightarrow \quad A \leftrightsquigarrow \not\leftrightsquigarrow B \mid C\ [\mathcal{G}].$$

Sufficient adjustment sets

- **Directed/causal path:** $A \rightsquigarrow B$.
- **Confounding arc**: $A \leftrightsquigarrow B$, either looks like $A \leftsquigarrow U_1 \rightsquigarrow B$ or $A \leftsquigarrow U_1 \leftrightarrow U_2 \rightsquigarrow B$.
- **Confounding path**: $A \leftrightsquigarrow * \leftrightsquigarrow B$.

Definition (Adjustment set)

- A set $C \subseteq \mathrm{V} \setminus \{A, B\}$ is an **adjustment set** for $A, B \in \mathrm{V}$ if $A \not\rightsquigarrow C$ and $B \not\rightsquigarrow C$.
- An adjustment set C for A and B is **sufficient** if $A \leftrightsquigarrow \not* \leftrightsquigarrow B \mid C$.
- Moreover, C is **minimal sufficient** if none of its proper subsets is still sufficient.

Lemma (Symmetric back-door criterion)

- *If S is a sufficient adjustment set for (X, Y), then it satisfies the back-door criterion.*
- *Moreover, if there exists a directed path from X to Y and S satisfies the back-door criterion, then S is a sufficient adjustment set.*

Primary adjustment sets

- So confounder selection is essentially about blocking confounding paths.
- This is complicated because

$$A \leftrightsquigarrow \not\ast \leftrightsquigarrow B \mid C \not\Rightarrow A \leftrightsquigarrow \not\ast \leftrightsquigarrow B \mid \tilde{C} \text{ for } C \subset \tilde{C}.$$

- But observe that

$$A \not\leftrightsquigarrow B \mid C \Rightarrow A \not\leftrightsquigarrow B \mid \tilde{C} \text{ for } C \subset \tilde{C}.$$

- This motivates us to **block one confounding arc at a time**.

Definition (Primary adjustment set)

We say an adjustment set $C \subseteq \mathrm{V}$ for $A, B \in \mathrm{V}$ is **primary** given another adjustment set $S \subseteq \mathrm{V}$ if $A \not\leftrightsquigarrow B \mid S \cup C$ holds; further, we say C is **minimal primary**, if none of its proper subsets is primary.

Lemma

If C is a minimal primary adjustment set for A and B given S, then $C \cap S = \emptyset$ and $C \subseteq \mathrm{an}(A) \cup \mathrm{an}(B)$.

Confounding paths under latent projection

Theorem (Latent projection preserves some refined m-connections)

Let $\mathcal{G}$ be an ADMG on vertex set V. For any distinct $A, B \in \mathrm{V}$, $C \subseteq V \setminus \{A, B\}$, and vertex set $\tilde{V}$ such that $V \supseteq \tilde{V} \supseteq \{A, B\} \cup C$, we have

$$A \left\{ \begin{array}{c} \rightsquigarrow \\ \leftrightsquigarrow \\ {\color{red}\leftrightsquigarrow * \leftrightsquigarrow} \end{array} \right\} B \mid C \; [\mathcal{G}] \quad \Longleftrightarrow \quad A \left\{ \begin{array}{c} \rightsquigarrow \\ \leftrightsquigarrow \\ {\color{red}\leftrightsquigarrow * \leftrightsquigarrow} \end{array} \right\} B \mid C \; [\mathcal{G}(\tilde{V})].$$

Confounding paths under latent projection

Theorem (Latent projection preserves some refined m-connections)

Let $\mathcal{G}$ be an ADMG on vertex set V. For any distinct $A, B \in \mathrm{V}$, $C \subseteq V \setminus \{A, B\}$, and vertex set $\tilde{V}$ such that $V \supseteq \tilde{V} \supseteq \{A, B\} \cup C$, we have

$$A \left\{ \begin{matrix} \rightsquigarrow \\ \leftrightsquigarrow \\ \color{red}{\leftrightsquigarrow * \leftrightsquigarrow} \end{matrix} \right\} B \mid C \; [\mathcal{G}] \quad \iff \quad A \left\{ \begin{matrix} \rightsquigarrow \\ \leftrightsquigarrow \\ \color{red}{\leftrightsquigarrow * \leftrightsquigarrow} \end{matrix} \right\} B \mid C \; [\mathcal{G}(\tilde{V})].$$

Corollary (**District criterion**)

$$A \left\{ \begin{matrix} \rightsquigarrow \\ \leftrightsquigarrow \\ \color{red}{\leftrightsquigarrow * \leftrightsquigarrow} \\ \rightsquigarrow * \leftsquigarrow \end{matrix} \right\} B \mid C \; [\mathcal{G}] \quad \iff \quad A \left\{ \begin{matrix} \rightarrow \\ \leftrightarrow \\ \color{red}{\leftrightarrow * \leftrightarrow} \\ \longrightarrow * \longleftarrow \end{matrix} \right\} B \; [\mathcal{G}(\{A, B\} \cup C)].$$

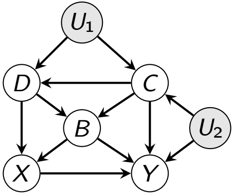
Compared to the (symmetric) backdoor criterion, this is more useful for **verification**.

Iterative graph expansion (recursive version)

procedure ConfounderSelect(X, Y)
    $\mathcal{R} = \emptyset$
    procedure GraphExpand(S, $\mathcal{B}_y$, $\mathcal{B}_n$)
        $\bar{S} = S \cup X \cup Y$
        if $X \leftrightarrow * \leftrightarrow Y$ by edges in $\mathcal{B}_y$ **then**
            return
        else if $X \leftrightarrow \not{*} \leftrightarrow Y$ by edges in $(\bar{S} \times \bar{S}) \setminus \mathcal{B}_n$ **then**
            $\mathcal{R} = \mathcal{R} \cup \{S\}$
            return
        end if
        $\pi = (A \leftrightarrow B) = \text{SelectEdge}(X, Y, S, \mathcal{B}_y, \mathcal{B}_n)$
        for C in FindPrimary(π, $S \setminus \{A, B\}$) **do**
            GraphExpand($S \cup C$, $\mathcal{B}_y$, $\mathcal{B}_n \cup \{\pi\}$)
        end for
        GraphExpand(S, $\mathcal{B}_y \cup \{\pi\}$, $\mathcal{B}_n$)
    end procedure
    GraphExpand($\emptyset$, $\emptyset$, $\emptyset$)
    return $\mathcal{R}$
end procedure

It is more flexible (and efficient) to implement this using a priority queue. This is illustrated below.

An example

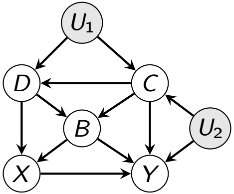


Step 0

$$\mathcal{Q} = \left[\boxed{X \dashleftarrow\dashrightarrow Y} \right], \quad \mathcal{R} = \{\}.$$

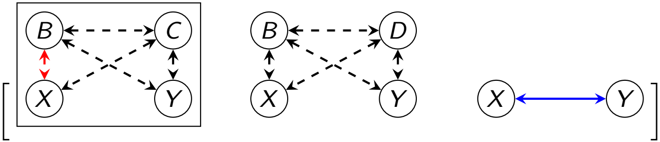
$$\text{FindPrimary}(X, Y) = \{\{B, C\}, \{B, D\}\}$$

An example



Step 1

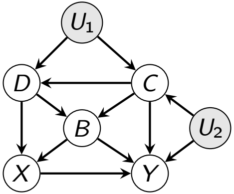
$\mathcal{Q} =$



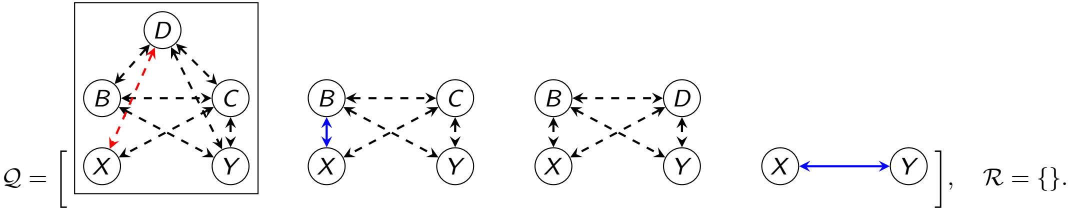
$\mathcal{R} = \{\}.$

$$\text{FindPrimary}(B, X) = \{\{D\}\}$$

An example

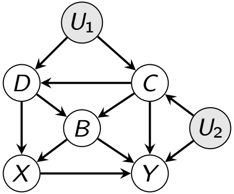


Step 2



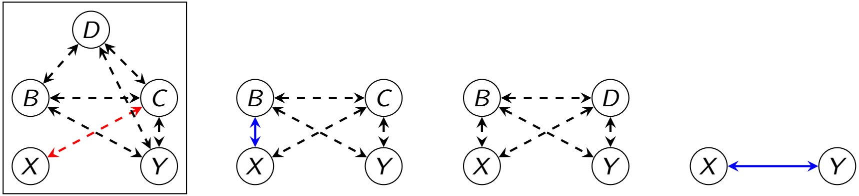
$$\text{FindPrimary}(D, X) = \{\emptyset\}$$

An example



Step 3

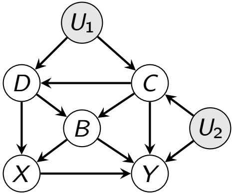
$\mathcal{Q} = \Big[$



$\Big], \quad \mathcal{R} = \{\}.$

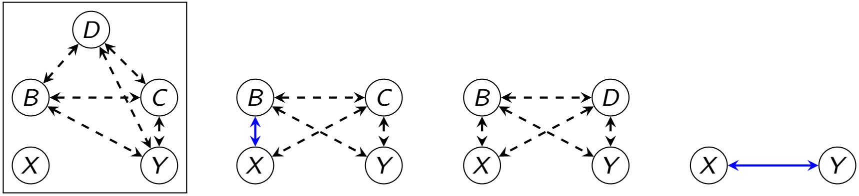
$\text{FindPrimary}(X, C) = \{\emptyset\}$

An example



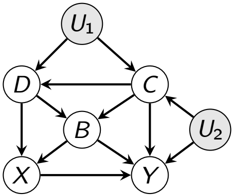
Step 4

$\mathcal{Q} = \Bigg[$

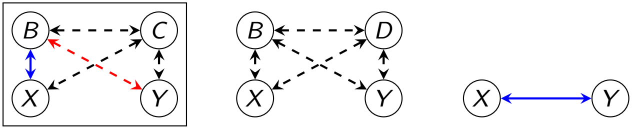


$\Bigg], \quad \mathcal{R} = \{\{B, C, D\}\}.$

An example



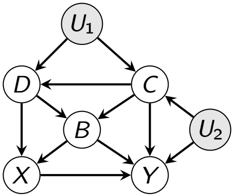
Step 5

$\mathcal{Q} = \Big[$ 

$\Big], \quad \mathcal{R} = \{\{B, C, D\}\}.$

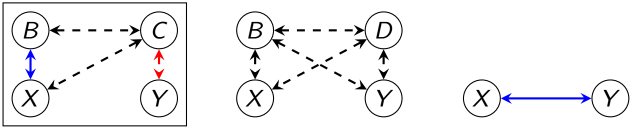
$$\text{FindPrimary}(B, Y) = \{\emptyset\}$$

An example



Step 6

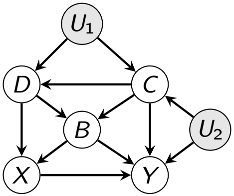
$\mathcal{Q} = \Big[$



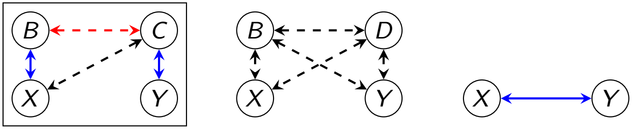
$\Big], \quad \mathcal{R} = \{\{B, C, D\}\}.$

$\text{FindPrimary}(C, Y) = \emptyset$

An example



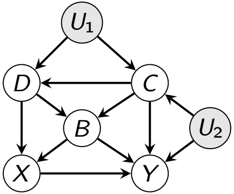
Step 7

$\mathcal{Q} = \Big[$ 

$\Big]$, $\quad \mathcal{R} = \{\{B, C, D\}\}$.

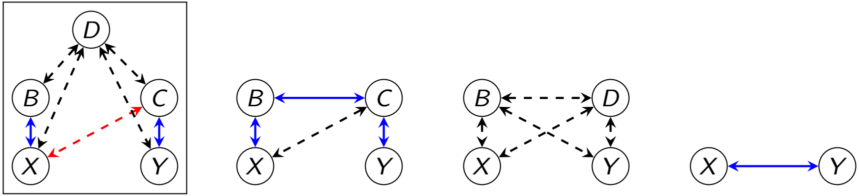
$$\text{FindPrimary}(B, C) = \{\{D\}\}$$

An example



Step 8

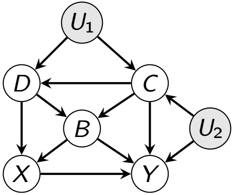
$$\mathcal{Q} = \Bigg[$$



$$\Bigg], \quad \mathcal{R} = \{\{B, C, D\}\}.$$

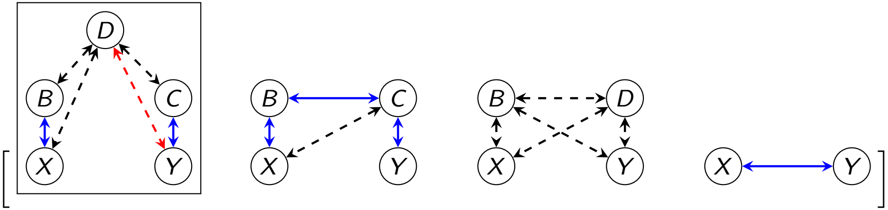
$$\text{FindPrimary}(X, C) = \{\emptyset\}$$

An example



Step 9

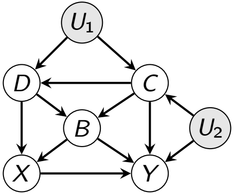
$\mathcal{Q} =$



$\mathcal{R} = \{\{B, C, D\}\}.$

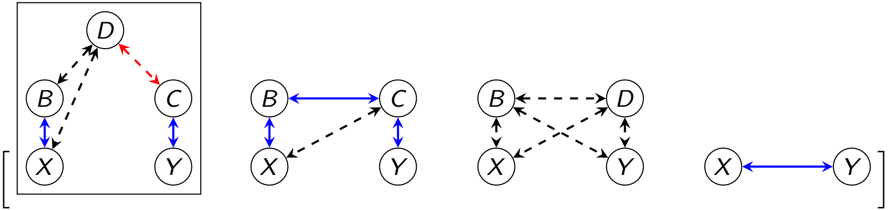
$$\text{FindPrimary}(D, Y) = \{\emptyset\}$$

An example



Step 10

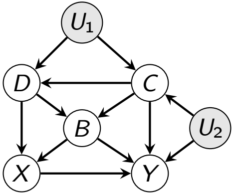
$\mathcal{Q} =$



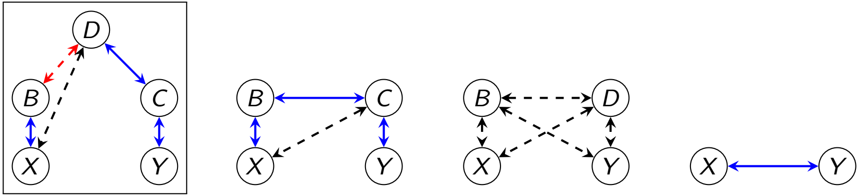
$\mathcal{R} = \{\{B, C, D\}\}.$

$\text{FindPrimary}(D, C) = \emptyset$

An example



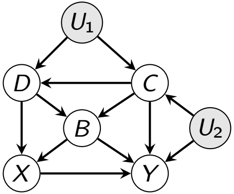
Step 11

$\mathcal{Q} = \Bigg[$ 

$\Bigg]$, $\quad \mathcal{R} = \{\{B, C, D\}\}$.

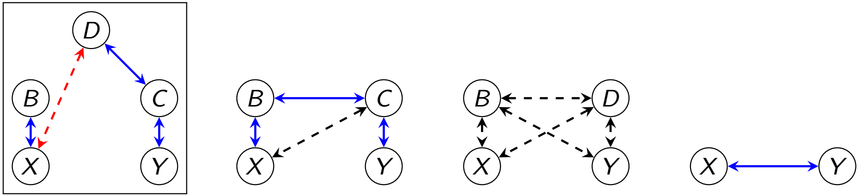
FindPrimary$(B, D) = \{\emptyset\}$

An example



Step 12

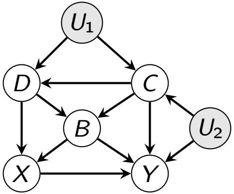
$\mathcal{Q} = \Bigg[$



$\Bigg], \quad \mathcal{R} = \{\{B, C, D\}\}.$

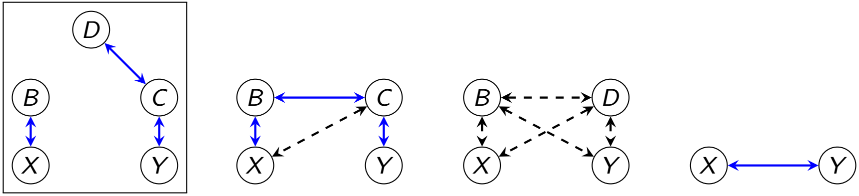
$\text{FindPrimary}(D, X) = \{\emptyset\}$

An example



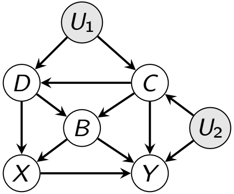
Step 13

$\mathcal{Q} = \Big[$

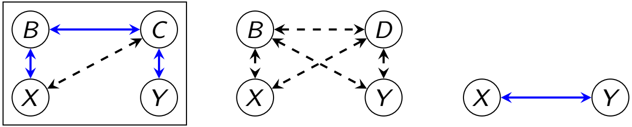


$\Big], \quad \mathcal{R} = \{\{B, C, D\}\}.$

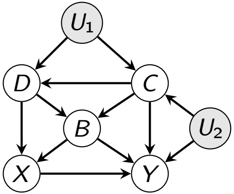
An example



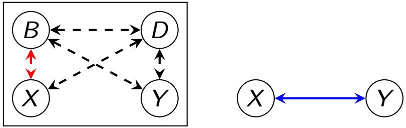
Step 14

$\mathcal{Q} = \Big[$  $\Big]$, $\quad \mathcal{R} = \{\{B, C, D\}\}.$

An example

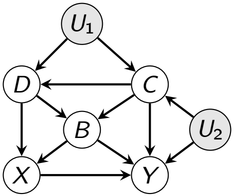


Step 15

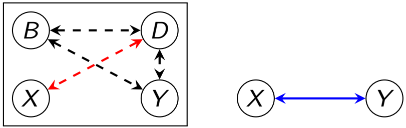
$\mathcal{Q} = \Bigg[$  $\Bigg]$, $\mathcal{R} = \{\{B, C, D\}\}.$

$\text{FindPrimary}(B, X) = \{\emptyset\}$

An example

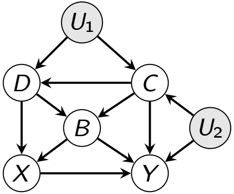


Step 16

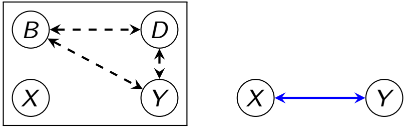
$\mathcal{Q} = \Big[$  $\Big]$, $\quad \mathcal{R} = \{\{B, C, D\}\}.$

$$\text{FindPrimary}(D, X) = \{\emptyset\}$$

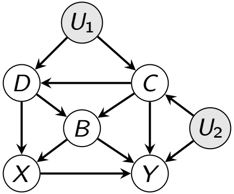
An example



Step 17

$\mathcal{Q} = \Big[$  $\Big]$, $\quad \mathcal{R} = \{\{B, C, D\}, \color{blue}{\{B, D\}}\}.$

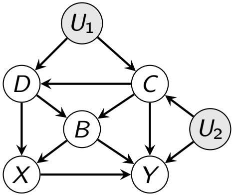
An example



Step 18

$$\mathcal{Q} = \left[\ \boxed{X \longleftrightarrow Y}\ \right], \quad \mathcal{R} = \{\{B, C, D\},\ \{B, D\}\}.$$

An example



Step 19

$$\mathcal{Q} = \Big[\;\Big], \quad \mathcal{R} = \{\{B, C, D\},\ \{B, D\}\}.$$

The algorithm terminates and returns $\{\{B, C, D\},\ \{B, D\}\}$, which includes the only minimal sufficient adjustment set $\{B, D\}$.

Soundness and completeness

Theorem

Consider any two vertices X, Y in an ADMG $\mathcal{G}$.

1. **Soundness (primary $\Rightarrow$ sufficiency)**: *Suppose every $C \in \mathrm{FindPrimary}((A, B); S')$ is a primary adjustment set for A, B given S' in $\mathcal{G}$. Then every element in the output of* $\mathrm{ConfounderSelect}(X, Y)$ *is a sufficient adjustment set for (X, Y).*
2. **Completeness (all minimal primary $\Rightarrow$ all minimal sufficiency)**: *Suppose further that* $\mathrm{FindPrimary}((A, B); S')$ *contains all minimal primary adjustment sets for (A, B) given S' in $\mathcal{G}$. Then the output of* $\mathrm{ConfounderSelect}(X, Y)$ *contains all minimal sufficient adjustment sets for (X, Y).*

Note: Primary adjustment sets can be found by iterating a more basic routine that queries if there exists an "unblocked" common ancestor.

Discussion

Summary

- The basic idea: reduce confounder selection to iteratively apply the following steps
  - Enumrate primary adjustment sets;
  - Expand the graph with each primary adjustment set.
- This provides a systematic way to solicit domain knowledge about endogeneity ($S_1 \leftrightarrow S_2$).
- It does not use causal structure between the adjustment variables ($S_1 \rightarrow S_2$ or $S_1 \leftarrow S_2$).
- Compared to previous approaches, graph expansion does not require full structure knowledge or a given set of observed covariates.

Discussion

Summary

- The basic idea: reduce confounder selection to iteratively apply the following steps
  - Enumrate primary adjustment sets;
  - Expand the graph with each primary adjustment set.
- This provides a systematic way to solicit domain knowledge about endogeneity ($S_1 \leftrightarrow S_2$).
- It does not use causal structure between the adjustment variables ($S_1 \rightarrow S_2$ or $S_1 \leftarrow S_2$).
- Compared to previous approaches, graph expansion does not require full structure knowledge or a given set of observed covariates.

- Rubin (2008): "for objective causal inference, **design trumps analysis**".
- More generally: How can graphs improve the design of observational studies?