Putting the *Causality* into Continual Causality

Part II: Causal Knowledge Representation, Inference, and Learning

Adèle Helena Ribeiro
adele.ribeiro@uni-marburg.de

AAAI-23 Continual Causality Bridge

February 7th, 2023

Data Science in Biomedicine
Faculty of Mathematics and Computer Science
Philipps-Universität Marburg

HEIDERLAB

Philipps Universität Marburg

Encoding Causal Knowledge

Causal Diagrams

Causal Diagram: Encoder of Structural Knowledge

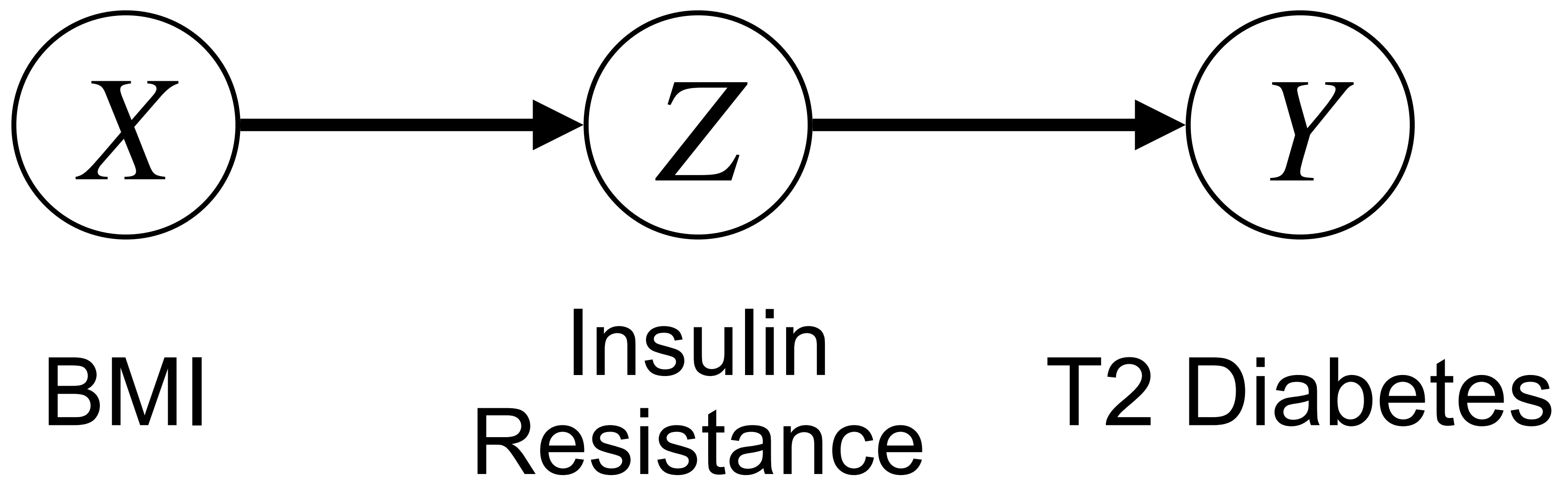
Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_X, U_{XY}) \\ Z \leftarrow f_Z(X, U_Z) \\ Y \leftarrow f_Y(Z, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram

(an Acyclic Directed Mixed Graphs, or ADMG)



An SCM $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in \mathbf{V}$:

- $V_i \rightarrow V_j$, if V_i appears as argument of $f_j \in \mathcal{F}$.

Causal Diagram: Encoder of <u>Structural</u> Knowledge

Philipps Universität Marburg

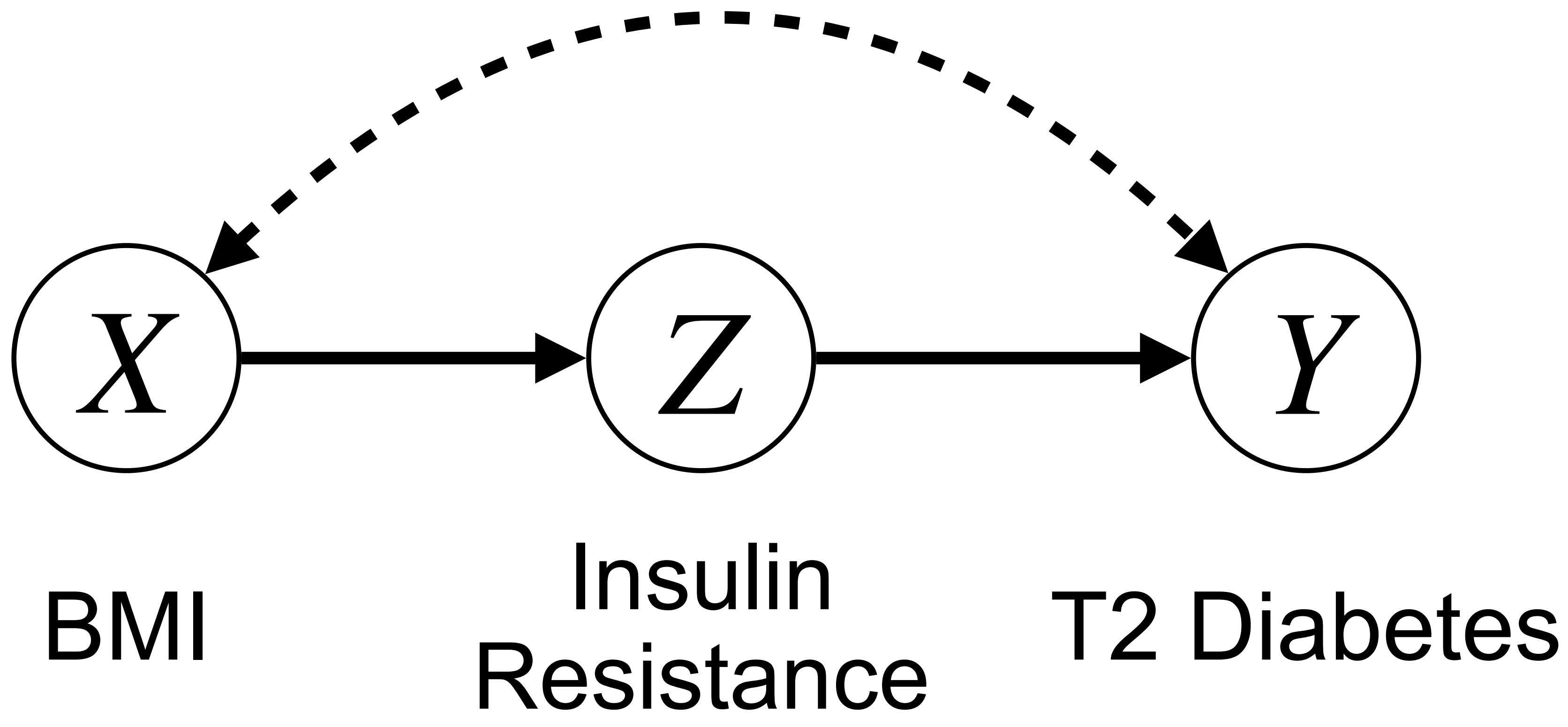
Structural Causal Model (SCM)

$\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$

$$\mathscr{M} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_X, U_Y, U_Z\} \\ \mathscr{F} = \begin{cases} X \leftarrow f_X(U_X, U_{XY}) \\ Z \leftarrow f_Z(X, U_Z) \\ Y \leftarrow f_Y(Z, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram

(an Acyclic Directed Mixed Graphs, or ADMG)



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in \mathbf{V}$:

- $V_i \rightarrow V_j$, if V_i appears as argument of $f_j \in \mathscr{F}$.
- $V_i \dashleftarrow\dashrightarrow V_j$ if the corresponding $U_i, U_j \in \mathbf{U}$ are correlated or f_i , f_j share some argument $U \in \mathbf{U}$.

Causal Diagram: Encoder of Structural Knowledge

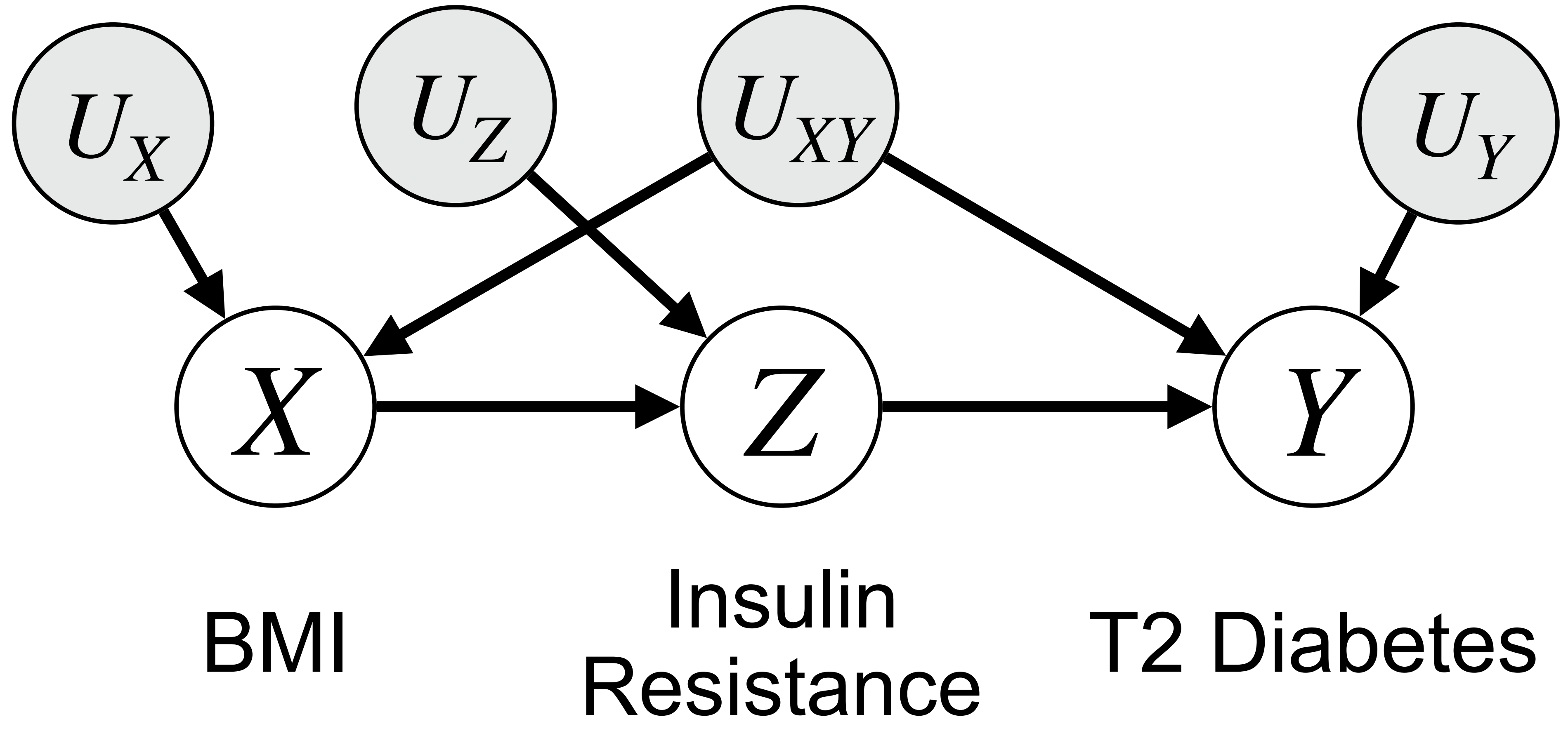
Structural Causal Model (SCM)

$$\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$$

$$\mathscr{M} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_X, U_Y, U_Z\} \\ \mathscr{F} = \begin{cases} X \leftarrow f_X(U_X, U_{XY}) \\ Z \leftarrow f_Z(X, U_Z) \\ Y \leftarrow f_Y(Z, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram

(an Acyclic Directed Mixed Graphs, or ADMG)



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in \mathbf{V}$:

- $V_i \rightarrow V_j$, if V_i appears as argument of $f_j \in \mathscr{F}$.
- $V_i \leftarrow\!\!\!-\!\!\!\rightarrow V_j$ if the corresponding $U_i, U_j \in \mathbf{U}$ are correlated or f_i , f_j share some argument $U \in \mathbf{U}$.

Causal Diagram: Encoder of Structural Knowledge

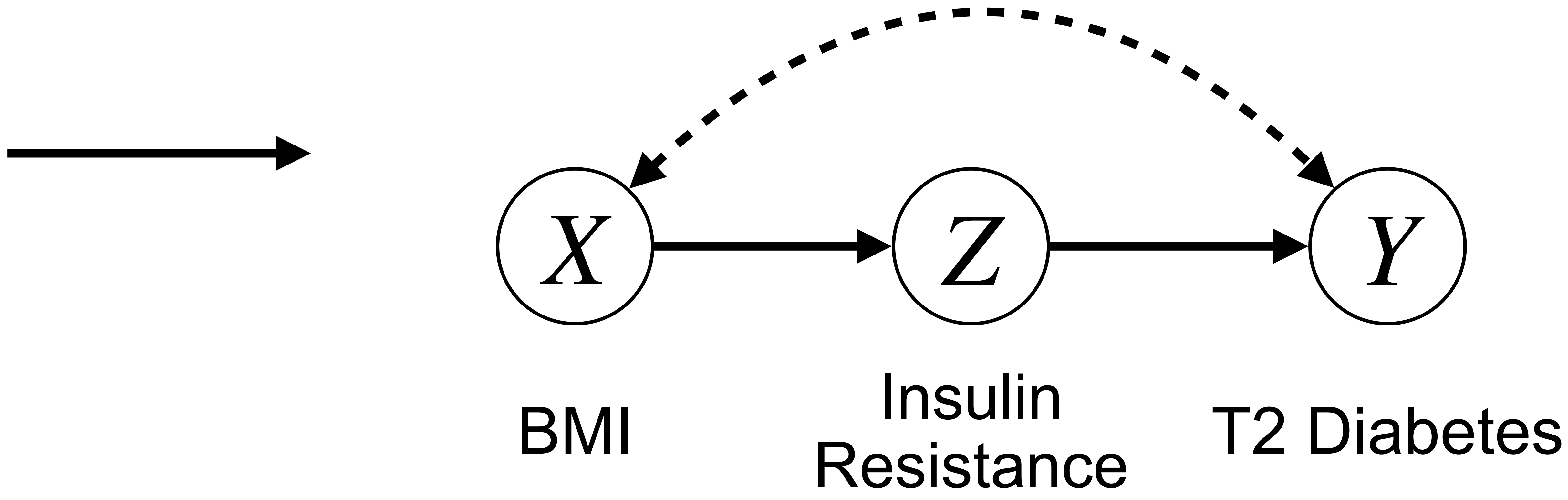
Philipps Universität Marburg

Structural Causal Model (SCM)

$$\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$$

$$\mathscr{M} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_X, U_Y, U_Z\} \\ \mathscr{F} = \begin{cases} X \leftarrow f_X(U_X, U_{XY}) \\ Z \leftarrow f_Z(X, U_Z) \\ Y \leftarrow f_Y(Z, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram
(an Acyclic Directed Mixed Graphs, or ADMG)



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in \mathbf{V}$:

- $V_i \rightarrow V_j$, if V_i appears as argument of $f_j \in \mathscr{F}$.
- $V_i \dashleftarrow\dashrightarrow V_j$ if the corresponding $U_i, U_j \in \mathbf{U}$ are correlated or f_i , f_j share some argument $U \in \mathbf{U}$.

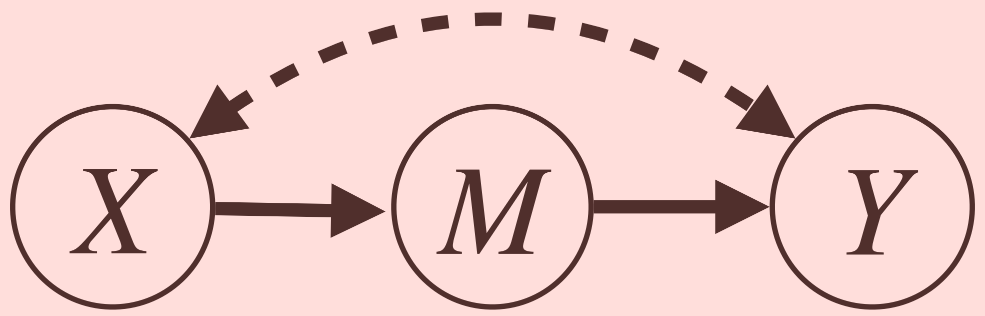
Classical Causal Effect Identification

Structural knowledge available

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, m, y)$ **Observational Distribution**

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_{m} P(m \mid x) \sum_{x'} P(y \mid m, x') P(x')$$

Interventional Distribution ← Available Distributions

👍

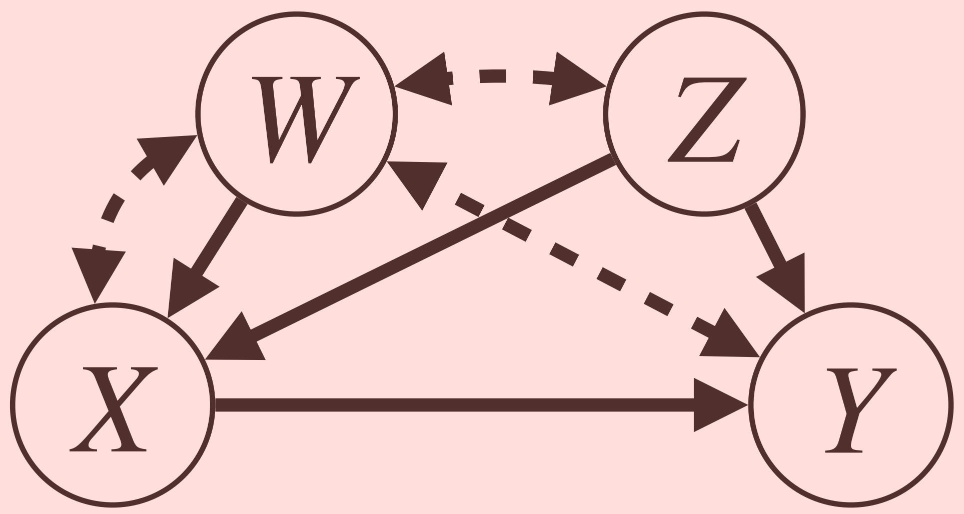
- Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pp. 567–573, Menlo Park, CA, 2002. AAAI Press/MIT Press.

Classical Causal Effect Identification

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, y, z, w)$

Observational Distribution

Inference Engine

Solution

yes / no

Interventional Distribution

Available Distributions

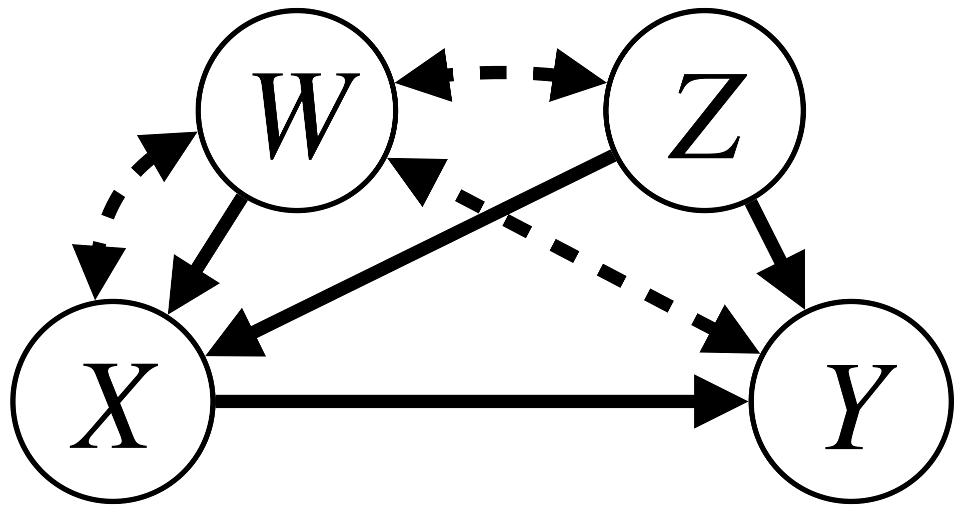
- Tian, J. and Pearl, J. (**2002**) A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI), pp. 567–573, Menlo Park, CA. AAAI Press/MIT Press.

General Causal Effect Identification

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, y, z, w)$ **Observational Distribution**

$P(x, y, z \mid do(w))$ **Experimental Distribution**

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_z P(y \mid x, z, do(w))\big(\alpha_1 P(z) + \alpha_2 P(z \mid do(w))\big)$$

Interventional Distribution ← Available Distributions

👍

- Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. [Link]

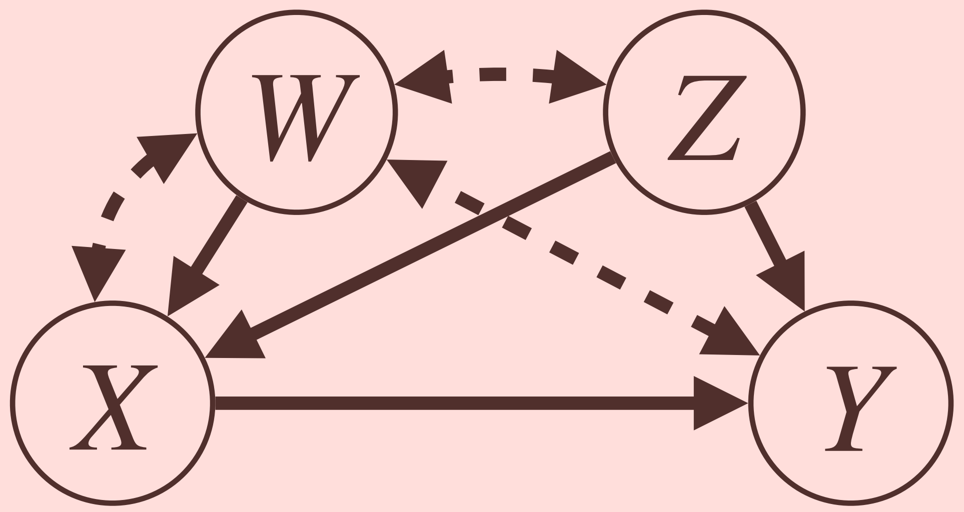
Can we relax the causal assumptions?

What if the required structural knowledge is not available?

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, y, z \mid do(w))$

$P(x, y, z, w)$

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_z P(y \mid x, z, do(w))\left(\alpha_1 P(z) + \alpha_2 P(z \mid do(w))\right)$$

Interventional Distribution ← Available Distributions

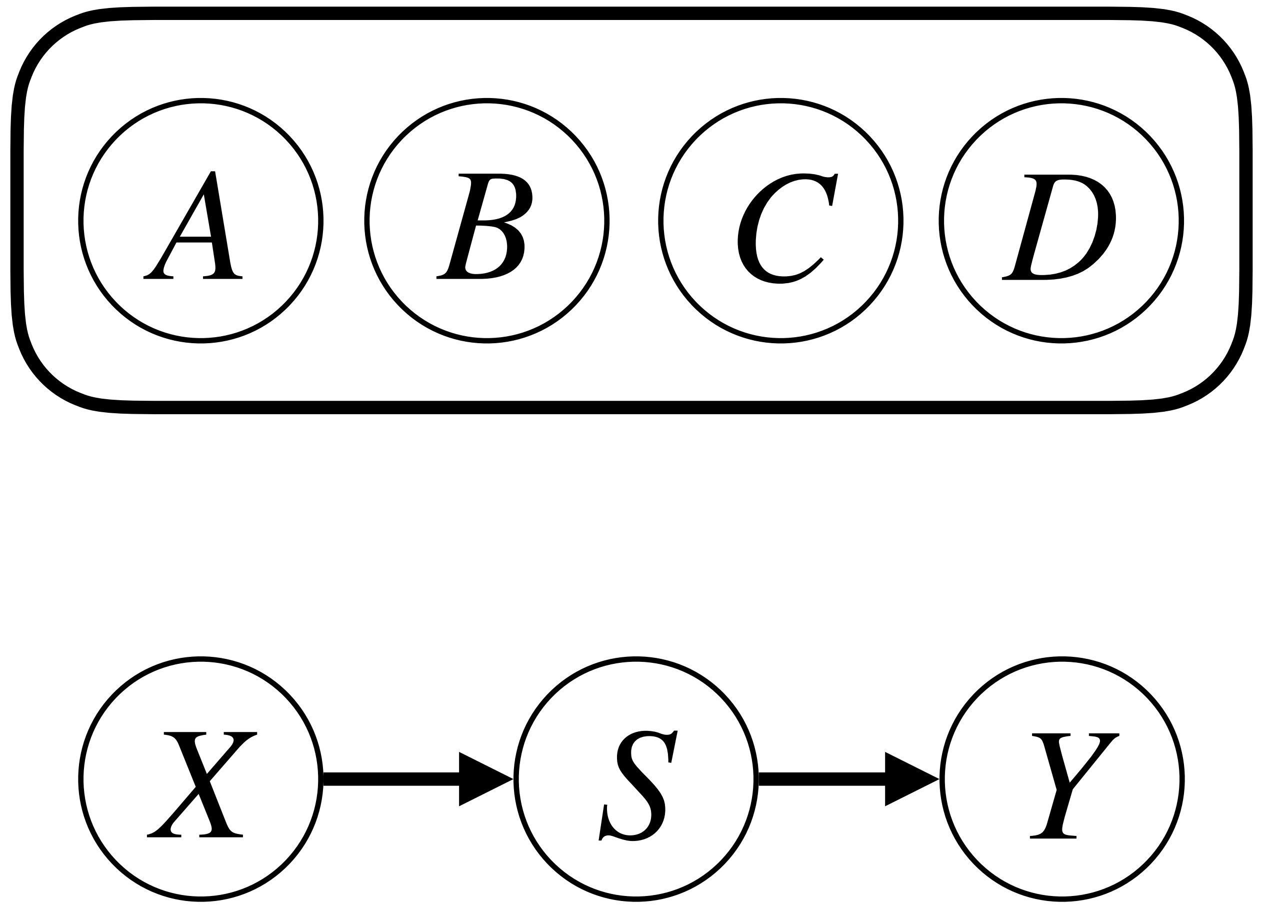
👍

Encoding Causal Knowledge in Partially Understood Domains

Cluster Causal Diagrams

Partially Understood Systems

(A) Age
(B) Blood pressure
(C) Comorbidities
(D) Medication history
(X) Lisinopril
(S) Sleep Quality
(Y) Stroke

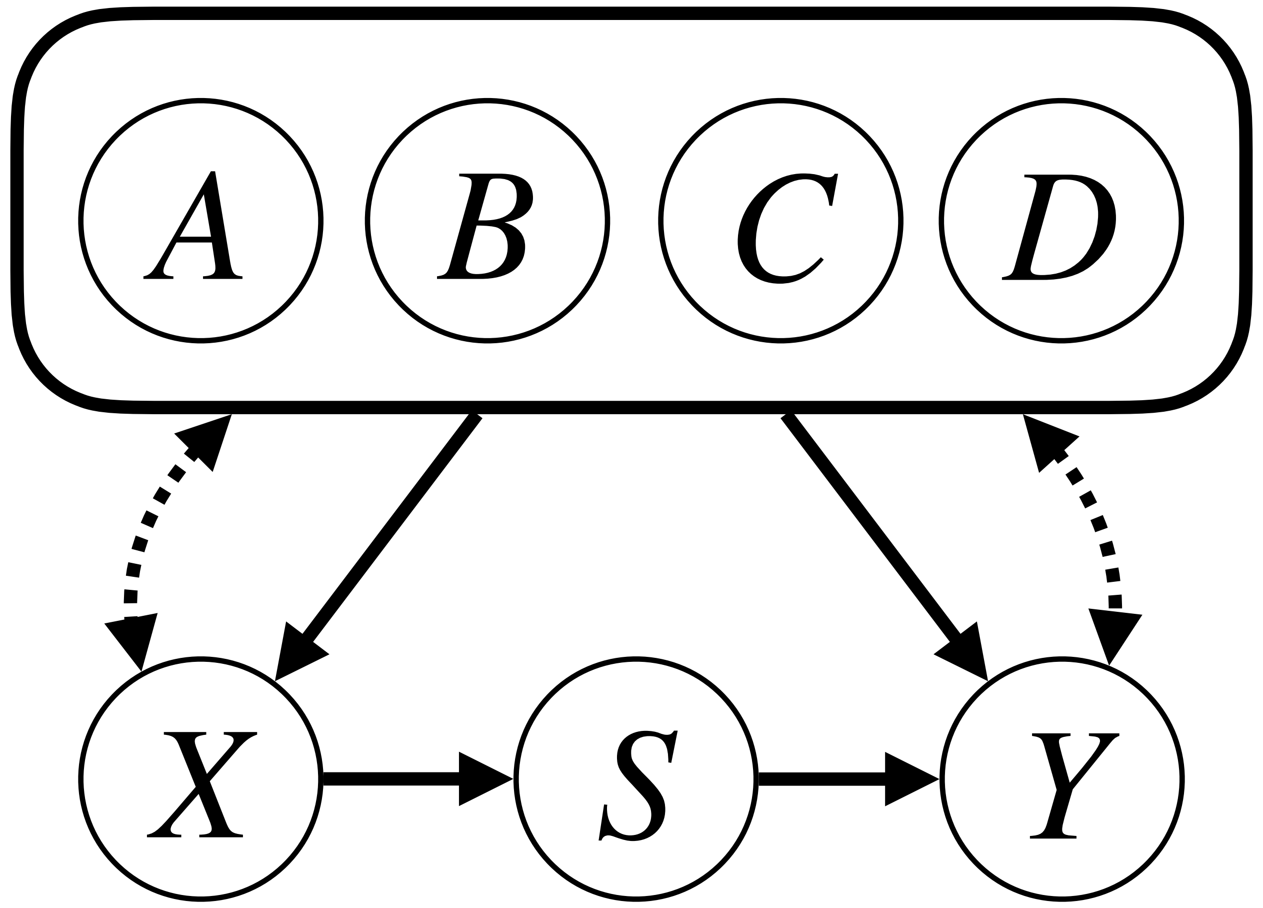


A causal diagram cannot be specified given the existing knowledge!

How can we identify $P(y \mid do(x))$ in this case?

Cluster Causal Diagrams (C-DAGs)

(A) Age
(B) Blood pressure
(C) Comorbidities
(D) Medication history
(X) Lisinopril
(S) Sleep Quality
(Y) Stroke



$\{\{X\}, \{S\}, \{Y\}, \{A, B, C, D\}\}$

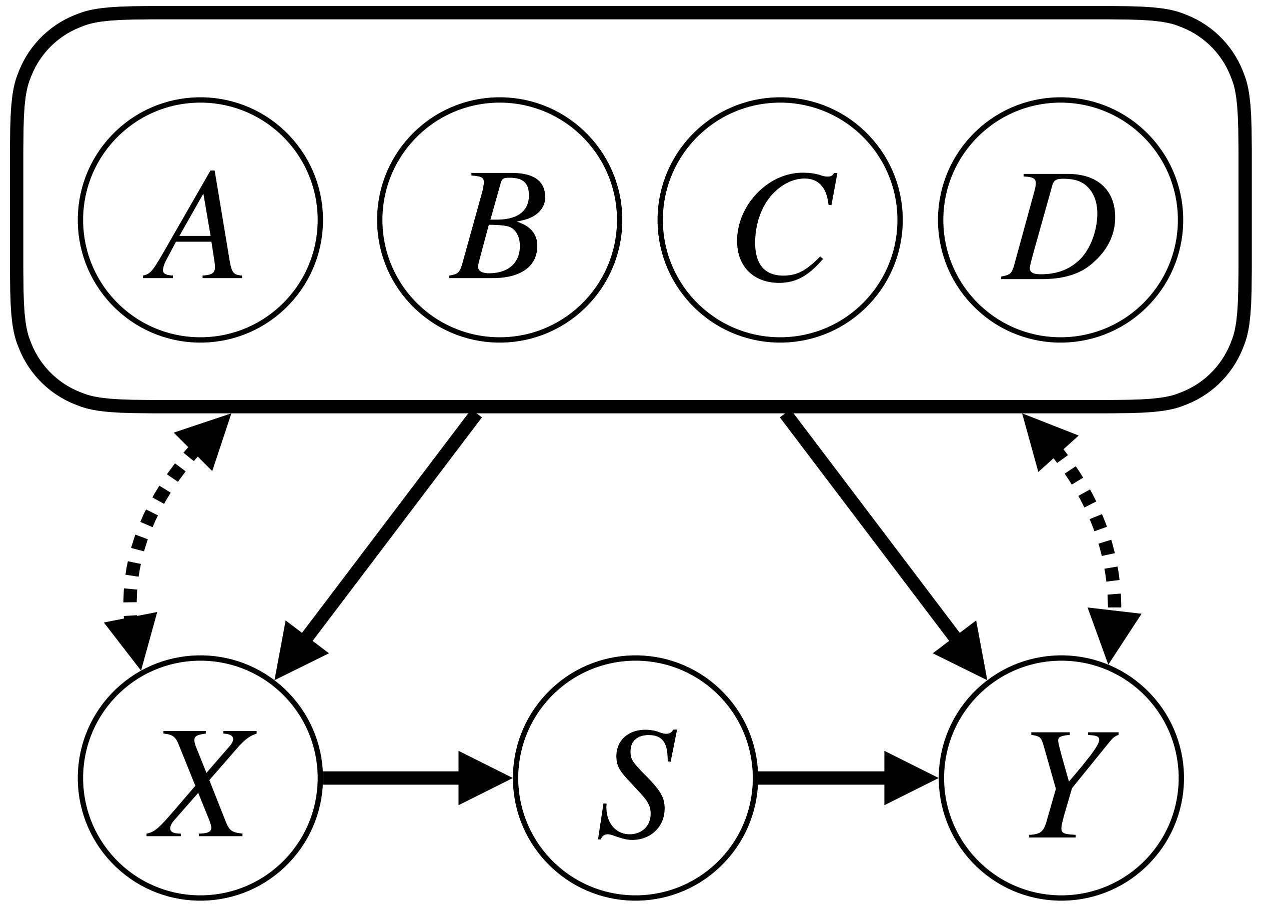
A *cluster causal diagram* $G_{\mathbf{C}}$ over a given partition $\mathbf{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_k\}$ of $\mathbf{V}$ is compatible with a causal diagram G over $\mathbf{V}$ if **for every $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$**:

- $\mathbf{C}_i \rightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \rightarrow V_j$
- $\mathbf{C}_i \dashleftarrow\dashrightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \dashleftarrow\dashrightarrow V_j$

and $G_{\mathbf{C}}$ contains no cycles.

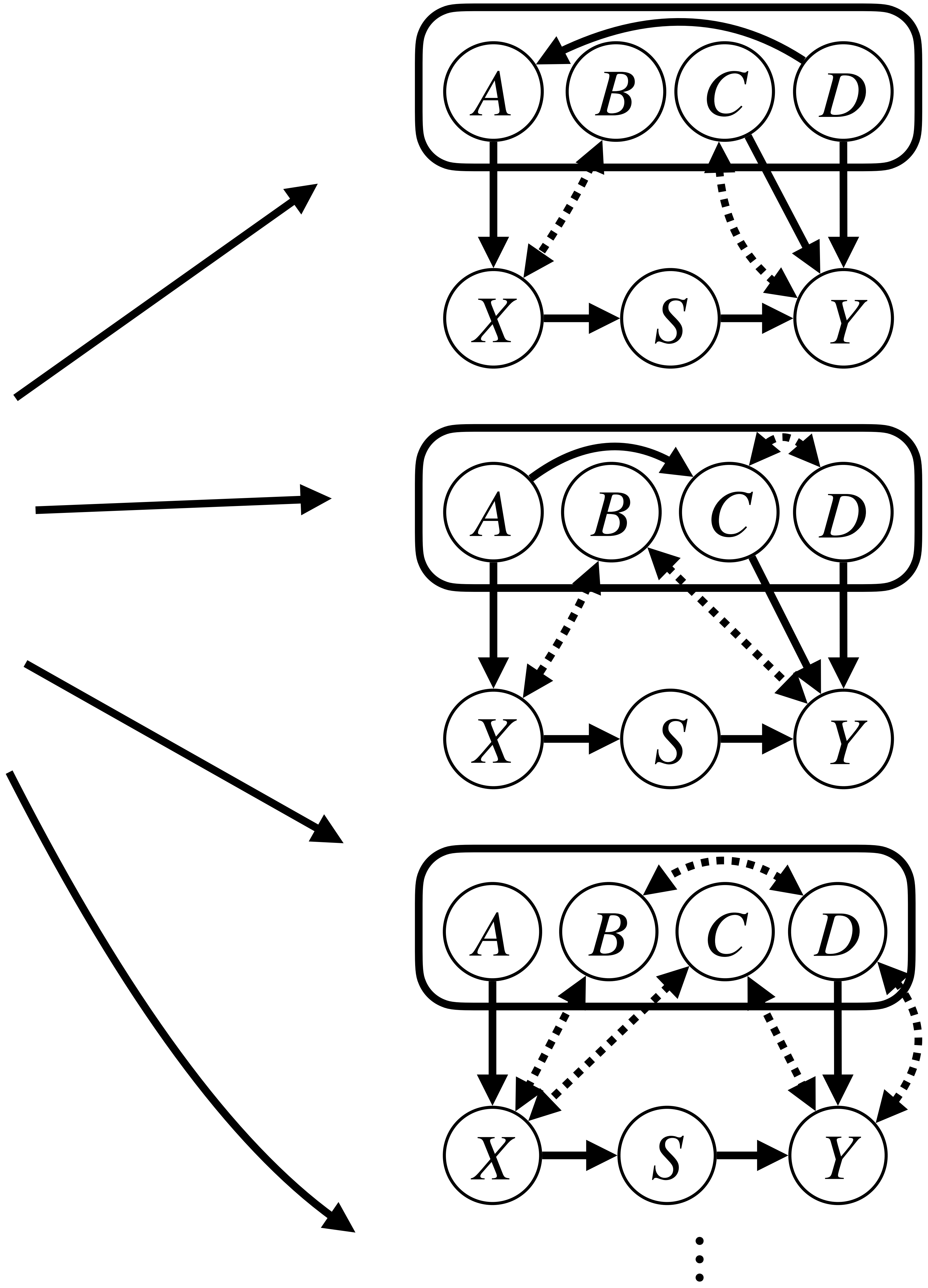
Partially Understood Systems

Many causal diagrams are compatible with the current knowledge!



Can be seen as an *equivalence class* of causal diagrams, where any relationships are allowed among the variables within each cluster.

Can we infer causal effects without deciding on any one particular causal diagram?

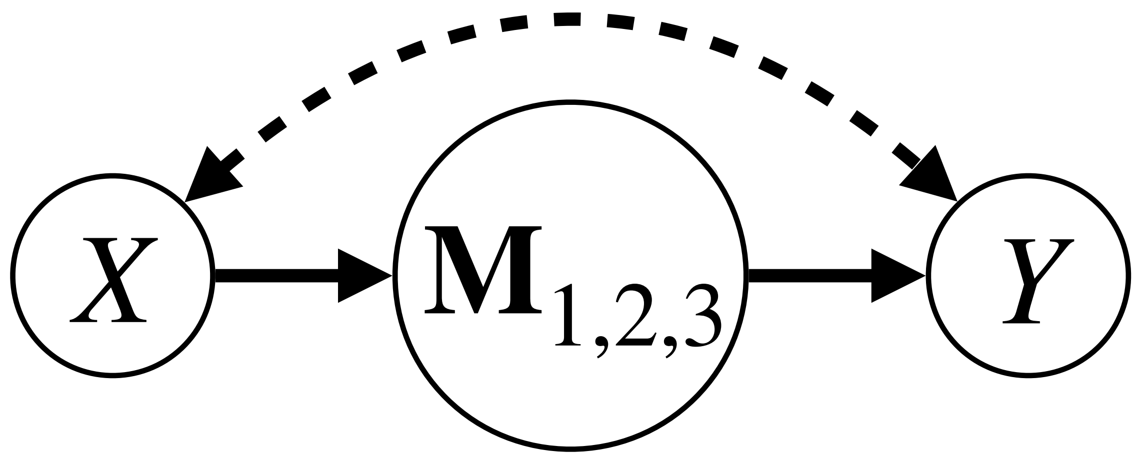


Identification of Causal Effects from C-DAGs

1 **Query**

$P(y \mid do(x))$

2 **C-DAG**



3 **Data**

$P(x, m_1, m_2, m_3, y)$

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_{\mathbf{m}_{123}} P(\mathbf{m}_{123} \mid x) \sum_{x'} P(y \mid \mathbf{m}_{123}, x') P(x')$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

Anand, T. V.*, **Ribeiro, A. H.***, Tian, J. , Bareinboim, E. (2022). Causal Effect Identification in Cluster DAGs. *Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)*

Causal Effect Identification

Graphical Criteria, Do-Calculus, and ID-Algorithm

Identification via Backdoor Adjustment

Let $\mathbf{X}$ be a set of treatment variables and $\mathbf{Y}$ a set of outcome variables in the causal graph G.

If there exists a set $\mathbf{Z}$ such that:

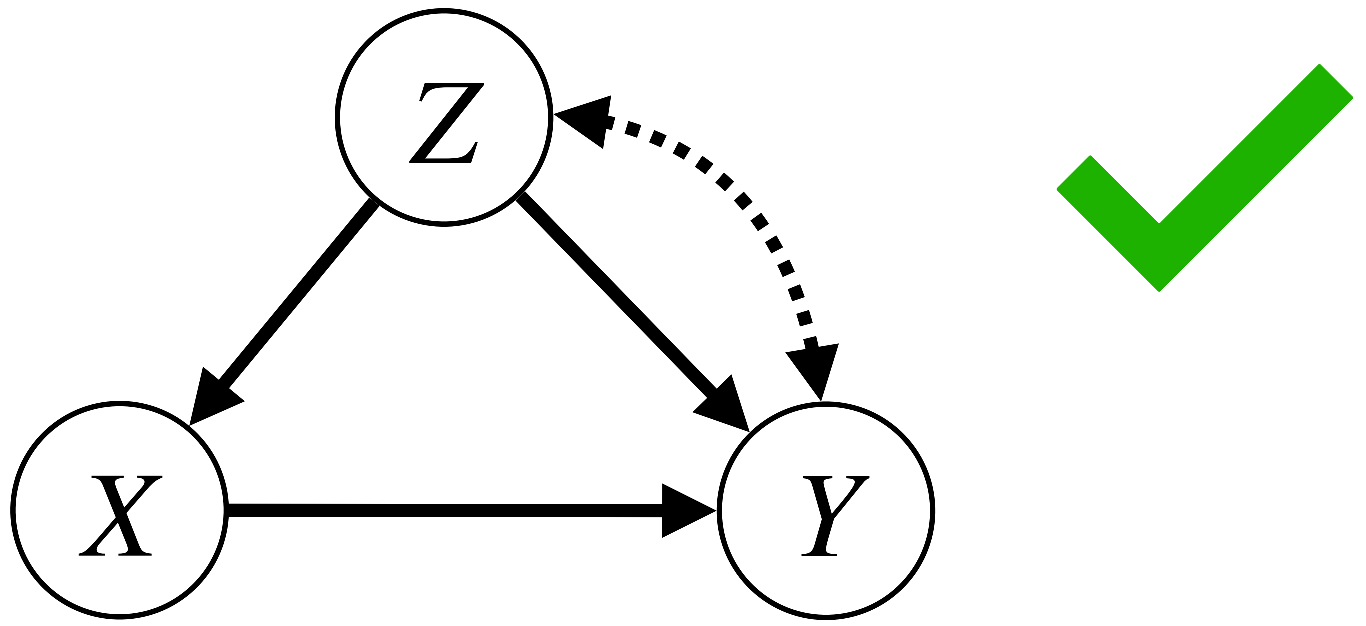
1. for every $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$, $\mathbf{Z}$ *blocks* every path between X and Y that has an arrow into X, and

2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ (all variables in $\mathbf{Z}$ are pre-treatment)

Then, $\mathbf{Z}$ satisfies the ***backdoor criterion*** and, then the effect of $\mathbf{X}$ on $\mathbf{Y}$ is given by:

$$P(\mathbf{y} \,|\, do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \,|\, \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

$$\mathbf{X} = \{X\}$$
$$\mathbf{Y} = \{Y\}$$
$$\mathbf{Z} = \{Z\}$$



Identification via Backdoor Adjustment

Let $\mathbf{X}$ be a set of treatment variables and $\mathbf{Y}$ a set of outcome variables in the causal graph G.

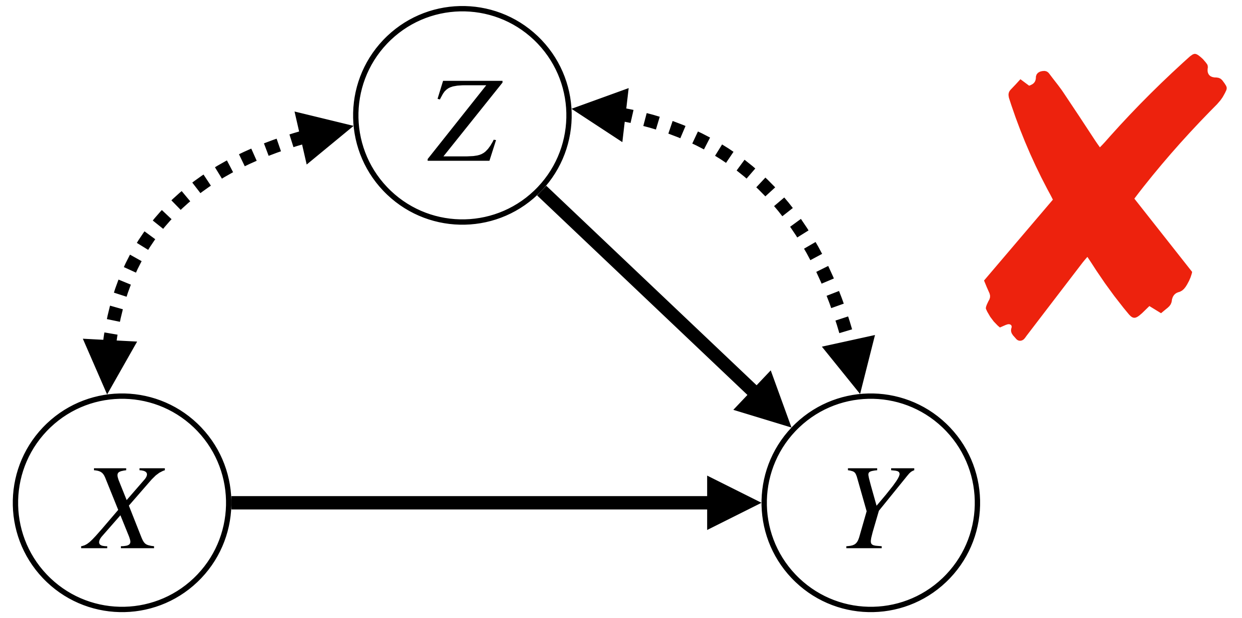
If there exists a set $\mathbf{Z}$ such that:

1. for every $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$, $\mathbf{Z}$ blocks every path between X and Y that has an arrow into X, and
2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ (all variables in $\mathbf{Z}$ are pre-treatment)

Then, $\mathbf{Z}$ satisfies the ***backdoor criterion*** and, then the effect of $\mathbf{X}$ on $\mathbf{Y}$ is given by:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

$$\mathbf{X} = \{X\}$$
$$\mathbf{Y} = \{Y\}$$
$$\mathbf{Z} = \{Z\}$$



Identification via Front-Door Adjustment

Let $\mathbf{X}$ be a set of treatment variables and $\mathbf{Y}$ a set of outcome variables in the causal graph G.

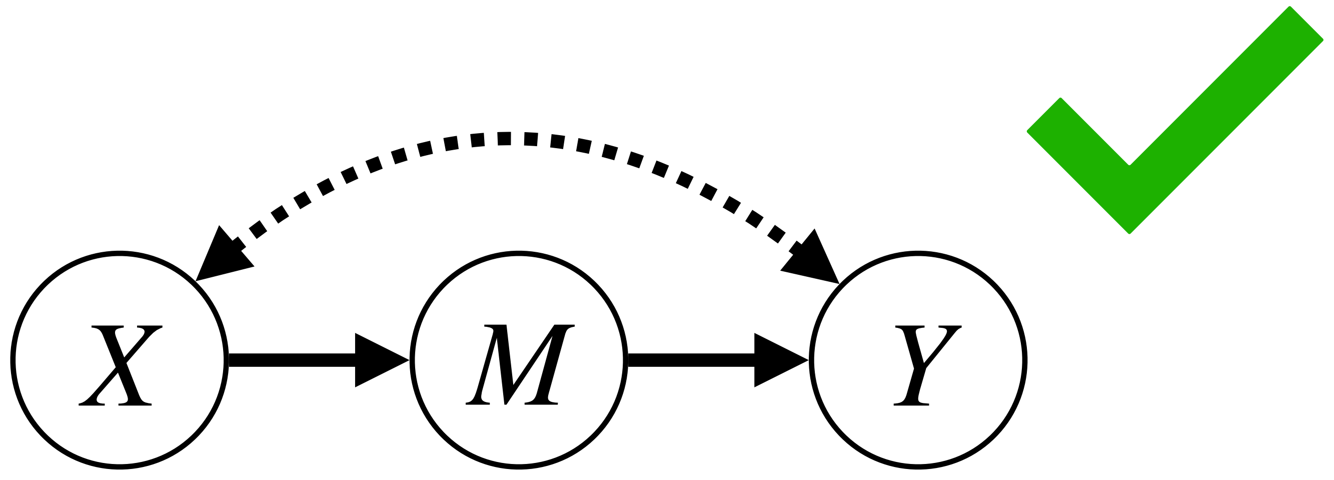
If there exists a set $\mathbf{M}$ such that:

1. $\mathbf{M}$ intercepts all directed paths from any vertex $X \in \mathbf{X}$ to any vertex $Y \in \mathbf{Y}$;
2. There is no unblocked back-door path from any vertex $X \in \mathbf{X}$ to vertex $M \in \mathbf{M}$; and
3. All back-door paths from any vertex $M \in \mathbf{M}$ to any vertex $Y \in \mathbf{Y}$ are blocked by $\mathbf{X}$.

Then, $\mathbf{M}$ satisfies the ***front-door criterion*** and, then the effect of $\mathbf{X}$ on $\mathbf{Y}$ is given by:

$$P(\mathbf{y} \,|\, do(\mathbf{x})) = \sum_{\mathbf{m}} P(\mathbf{m} \,|\, \mathbf{x}) \sum_{\mathbf{x}'} P(\mathbf{y} \,|\, \mathbf{m}, \mathbf{x}') P(\mathbf{x}')$$

$\mathbf{X} = \{X\}$
$\mathbf{Y} = \{Y\}$
$\mathbf{M} = \{M\}$



Do-Calculus (a.k.a. Causal Calculus)

Pearl, 1995

Graphical conditions implying invariances between observational ($\mathscr{L}_1$) and interventional ($\mathscr{L}_2$) distributions

Theorem: Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ be any disjoint subjects of variables.

Rule 1 (Insertion/Deletion of Observations):

$$P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}}}$$

Rule 2 (Actions/Observations Exchange):

$$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$$

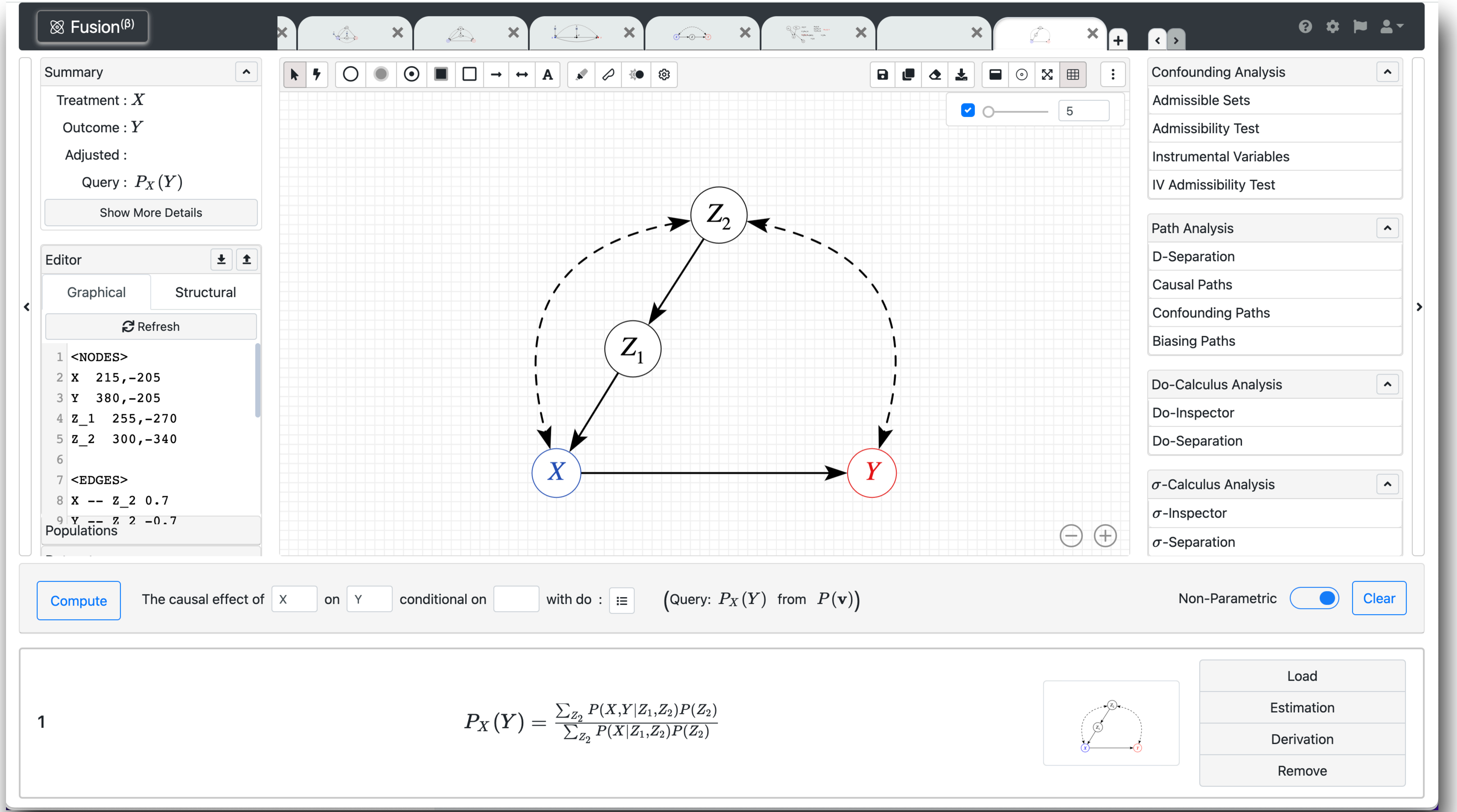
Rule 3 (Insertion/Deletion of Actions):

$$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}, \overline{\mathbf{Z}(\mathbf{W})}}}$$

$G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$: graph G after removing the incoming arrows into $\mathbf{X}$ and the outgoing arrows from $\mathbf{Z}$;

$\mathbf{Z}(\mathbf{W})$: set of $\mathbf{Z}$-nodes that are not ancestors of any $\mathbf{W}$-node in $G_{\overline{\mathbf{X}}}$.

http://causalfusion.net

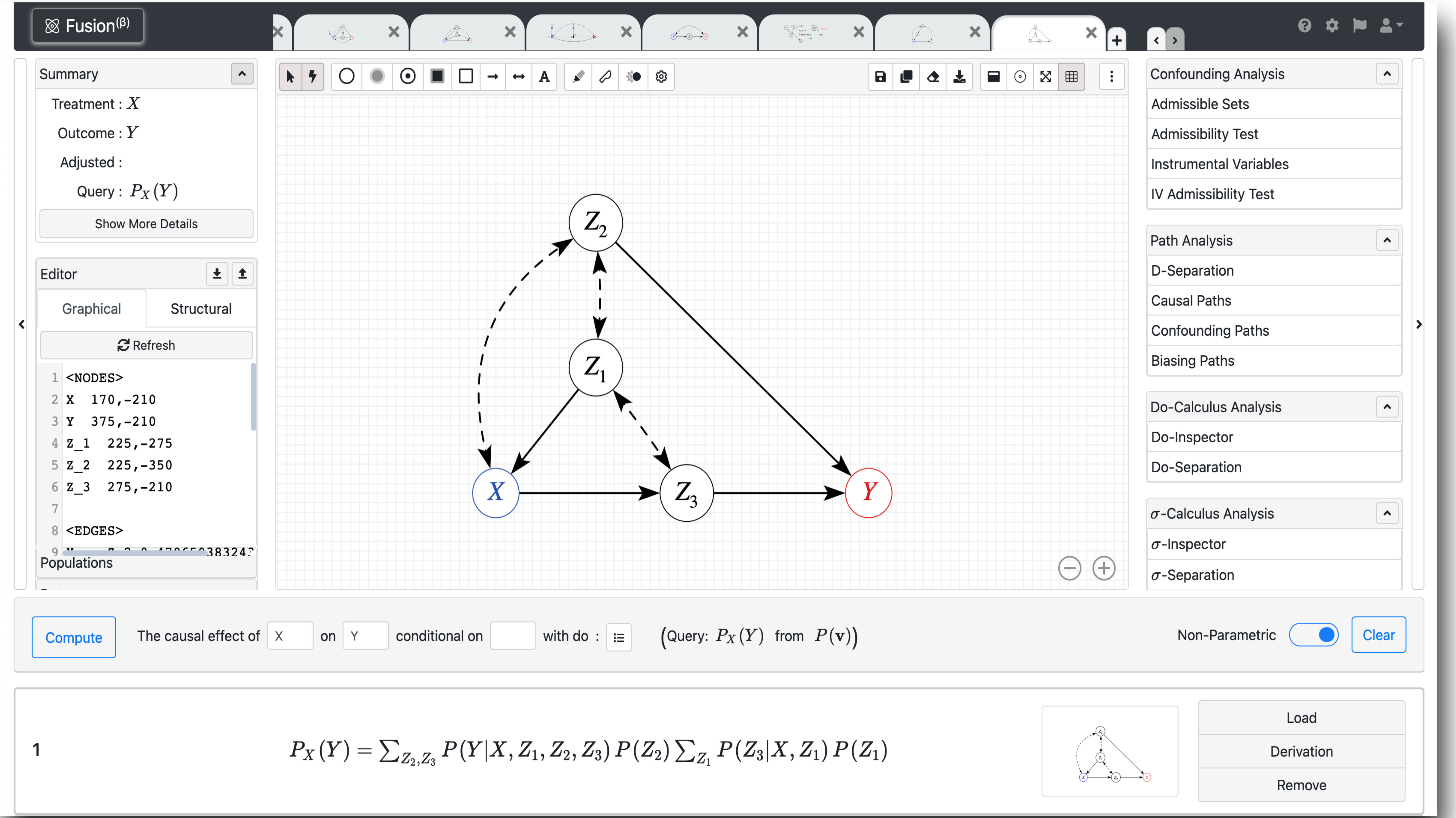


Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. Link

http://causalfusion.net

Philipps Universität Marburg

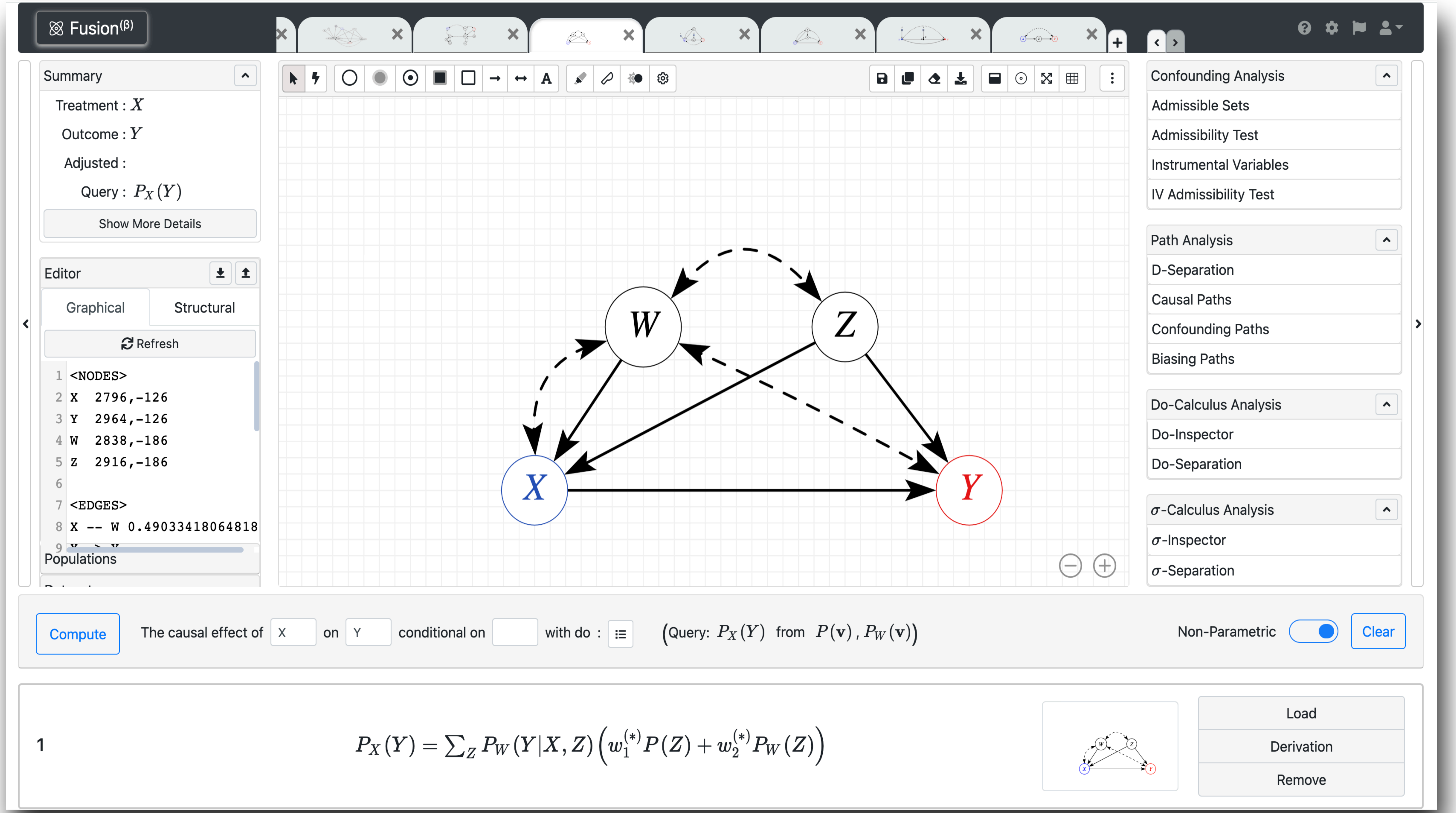
HEIDERLAB



Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. Link

http://causalfusion.net

Philipps Universität Marburg



Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. Link

Purely Data-Driven Causal Inference

Causal Discovery and Identification under Markov Equivalence

Learning Causal Structures from Data

What if no knowledge is available for constructing a C-DAG?

Can we learn a causal diagram $\mathscr{G}$ from observational data?

In non-parametric settings, we can't learn the true causal diagram,
but we can learn a graphical representation of all *compatible* causal diagrams,
called **Markov equivalence class!**

Markov Equivalence Class

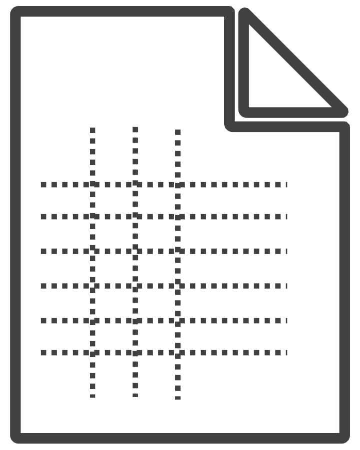
$$\mathcal{M}_1 = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_x, U_Y\} \\ \mathcal{F} = \begin{cases} f_X(U_X) \\ f_Y(X, U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$\vdots$

$$\mathcal{M}_{N-1} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_x, U_Y, U_{X,Y}\} \\ \mathcal{F} = \begin{cases} f_X(Y, U_X, U_{X,Y}) \\ f_Y(U_Y, U_{X,Y}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$$\mathcal{M}_N = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_x, U_Y\} \\ \mathcal{F} = \begin{cases} f_X(U_X) \\ f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

→ Data



→ Conditional (in)dependencies

$P(\mathbf{v})$

$X \not\perp\!\!\!\perp Y$

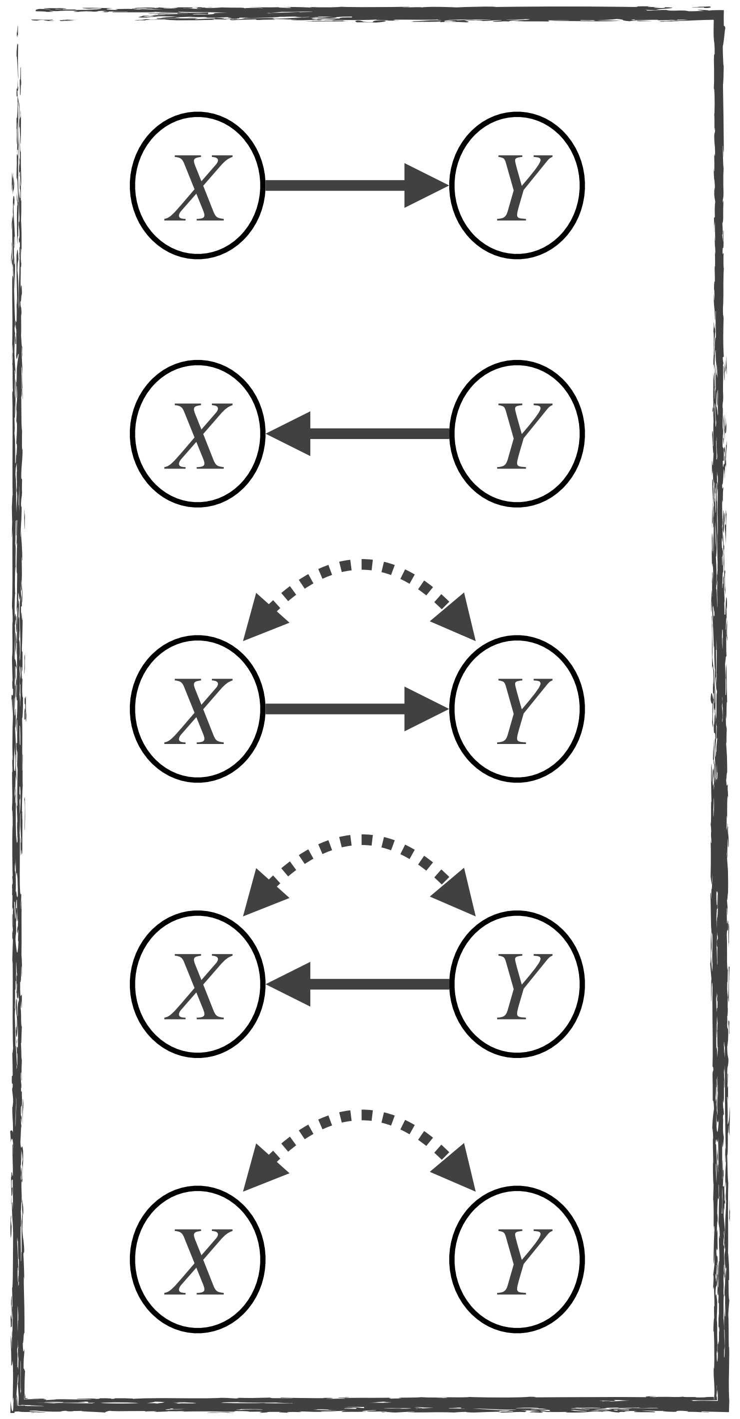
→

$$P(x, y) = \sum_{u_x, u_y} P(x \mid y)P(y)P(u_x, u_y)$$

$$P(x, y) = \sum_{u_x, u_y} P(y \mid x)P(x)P(u_x, u_y)$$

Markov Equivalence Class

(class of models implying the same set of conditional independencies)



Correlation does not imply causation!

Constraint-Based Causal Discovery

Goal: Learn a graphical representation of the Markov Equivalence Class from observational data.

Assumptions: the observed distribution is the marginal of a distribution P that satisfies the following conditions for the true causal diagram G (an **ADMG**):

1) **I-Map / Semi-Markov Condition:** for any disjoint subsets $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z})_G \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z})_P.$$

G is an ***I-Map of*** P

P is ***semi-Markov relative*** to G.

2) **Faithfulness Condition:** for any disjoint subsets $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z})_P \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z})_G.$$

P is ***faithful*** to G

Note: Estimation of the marginal distribution from limited data requires and **additional assumption**:

3) An adequate conditional independence test is available.

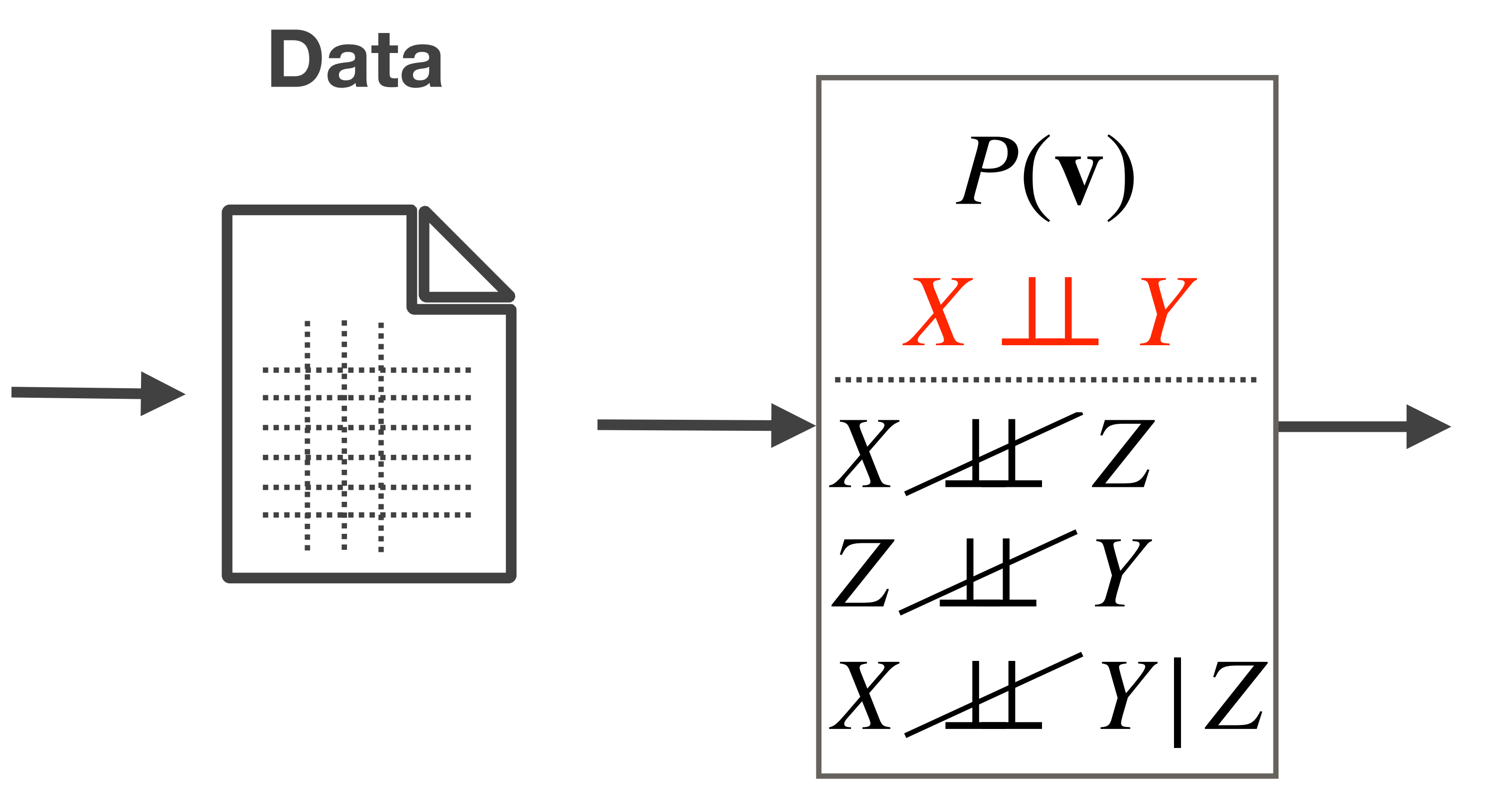
Learning Structural Invariances

$$\mathcal{M}_1 = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_x, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_X) \\ Z \leftarrow f_Z(X, Y, U_Z) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

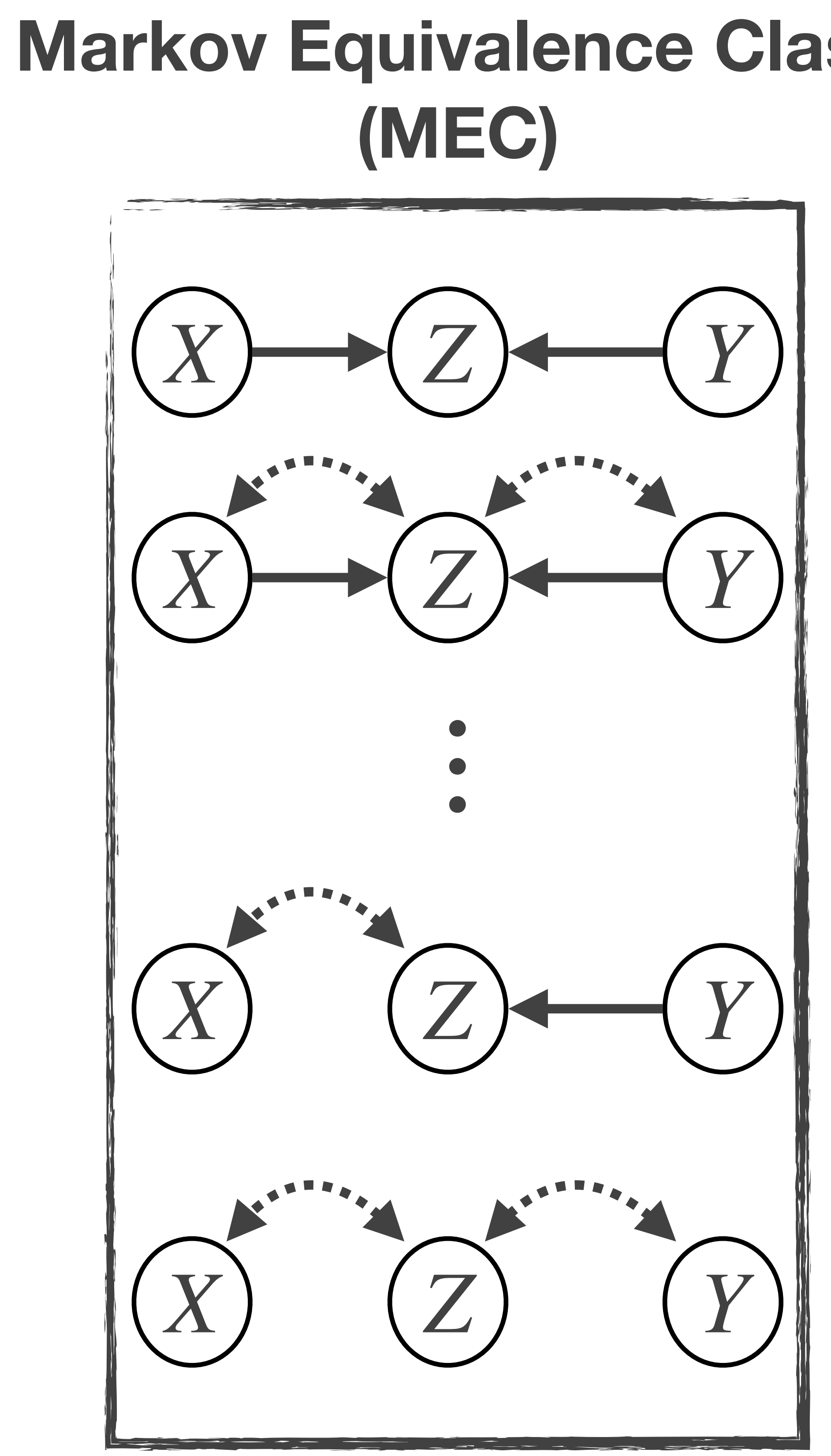
$$\vdots$$

$$\mathcal{M}_{N-1} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(Y, U_{XZ}, U_Z) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$$\mathcal{M}_N = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(U_{XZ}, U_{YZ}, U_Z) \\ Y \leftarrow f_Y(U_{YZ}, U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



$$P(x, y, z) = P(z \mid x, y)P(x \mid y)P(y) = P(z \mid x, y)P(x)P(y)$$



Learning Structural Invariances

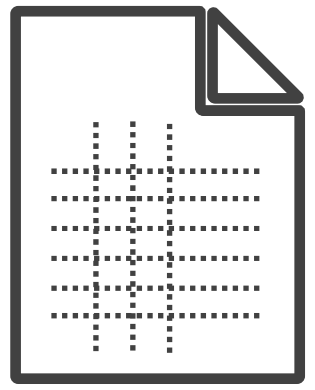
$$\mathcal{M}_1 = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_x, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_X) \\ Z \leftarrow f_Z(X, Y, U_Z) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$\vdots$

$$\mathcal{M}_{N-1} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(Y, U_{XZ}, U_Z) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$$\mathcal{M}_N = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(U_{XZ}, U_{YZ}, U_Z) \\ Y \leftarrow f_Y(U_{YZ}, U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Data



Conditional (in)dependencies

$P(\mathbf{v})$

$X \perp\!\!\!\perp Y$

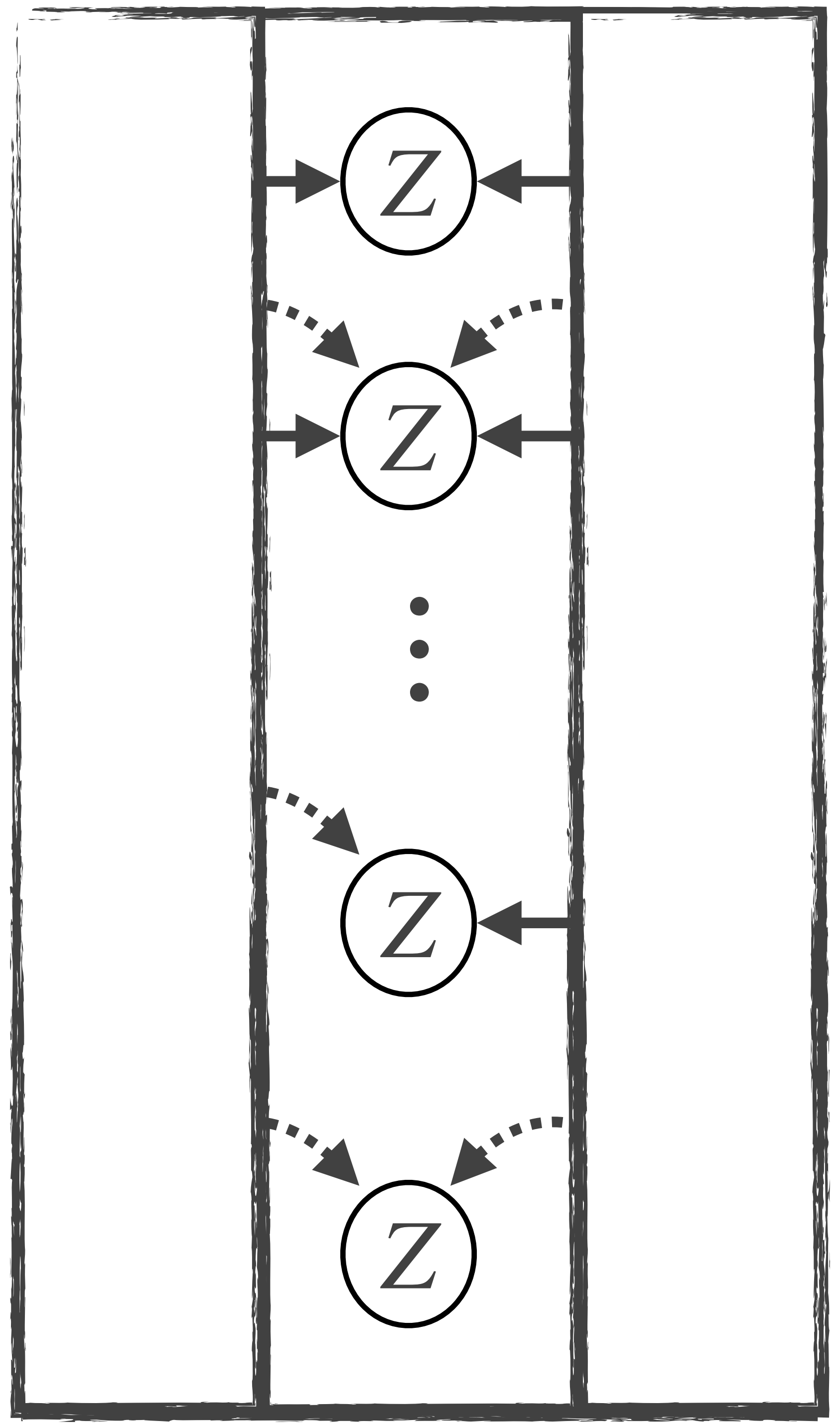
$X \not\perp\!\!\!\perp Z$

$Z \not\perp\!\!\!\perp Y$

$X \not\perp\!\!\!\perp Y \mid Z$

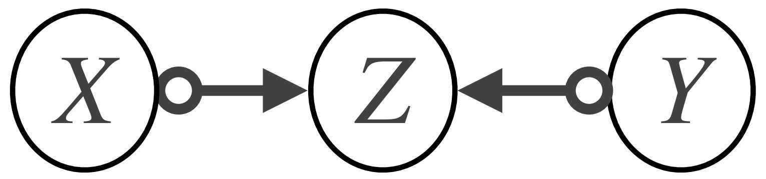
$$P(x, y, z) = P(z \mid x, y) P(x \mid y) P(y)$$
$$= P(z \mid x, y) P(x) P(y)$$

Markov Equivalence Class (MEC)



Causal Discovery

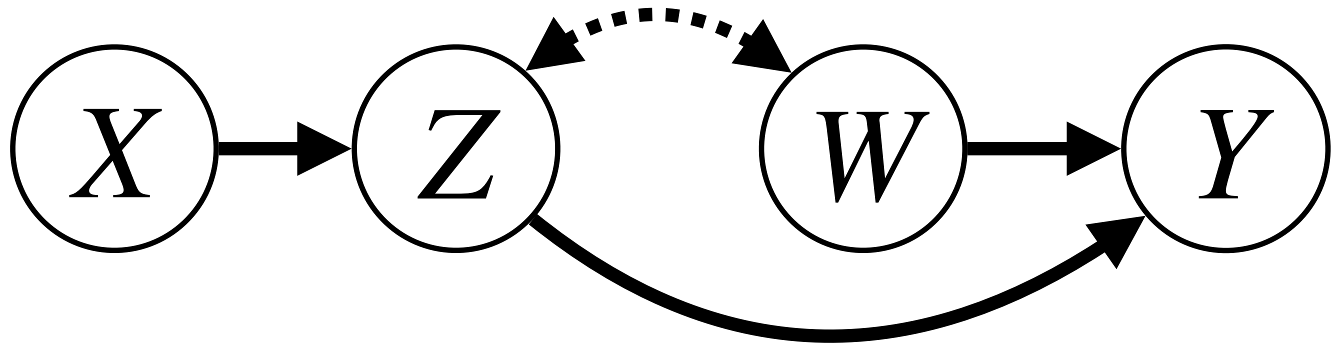
MEC Representation

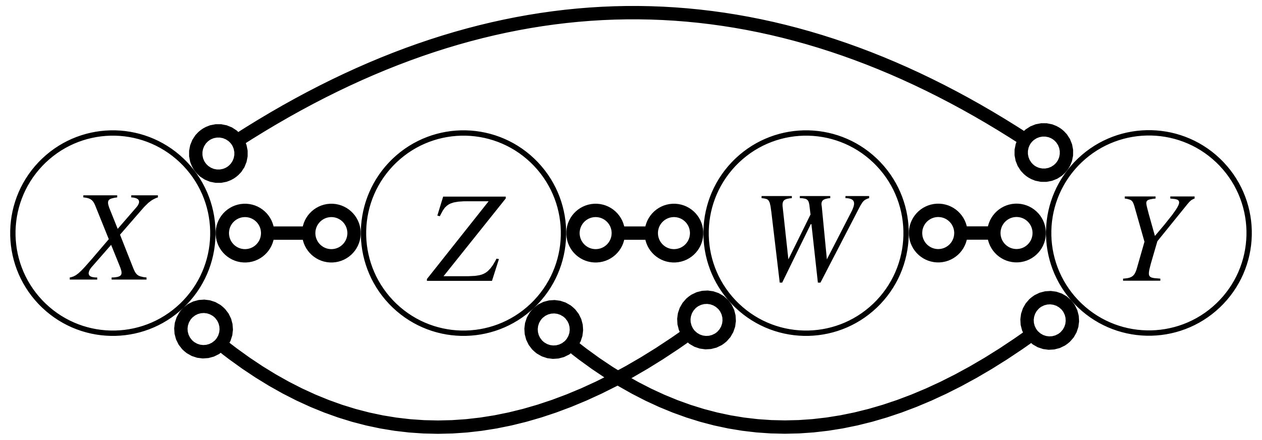


E.g., PC and FCI algorithms

Fast Causal Inference (FCI) Algorithm

True (unknown) causal diagram



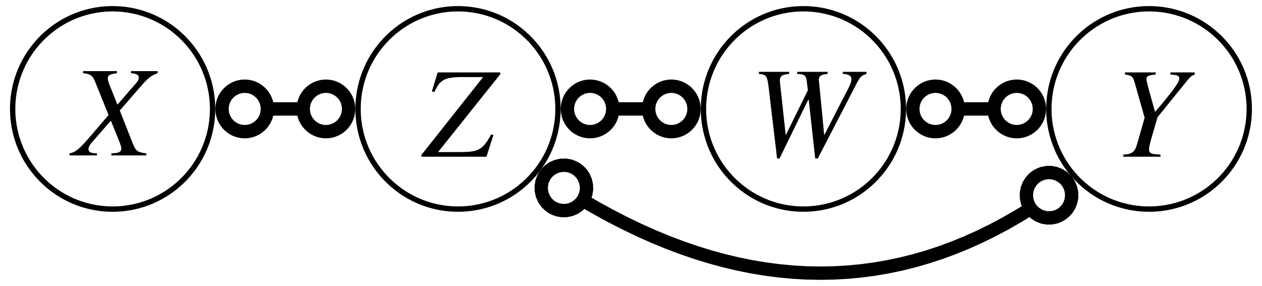


Complete Graph

Conditional Independence Tests

$X \perp\!\!\!\perp W$

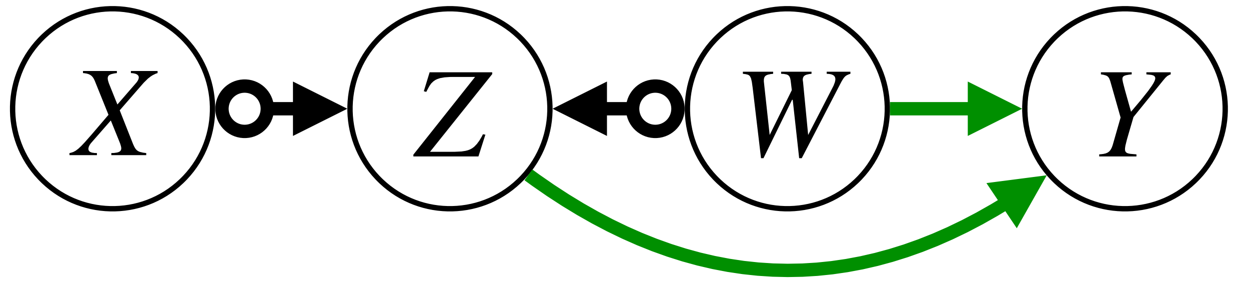
$X \perp\!\!\!\perp Y \mid Z, W$



Skeleton

FCI Rules

$(R1) - (R10)$



Partial Ancestral Graph (PAG)

Arrowhead $\Longrightarrow$ non-ancestrality

Tail $\Longrightarrow$ ancestrally

Circle $\Longrightarrow$ non-invariance

$A \longleftrightarrow B$ — spurious association

$A \longmapsto B$ — selection bias

Z is not an ancestor of X or W.

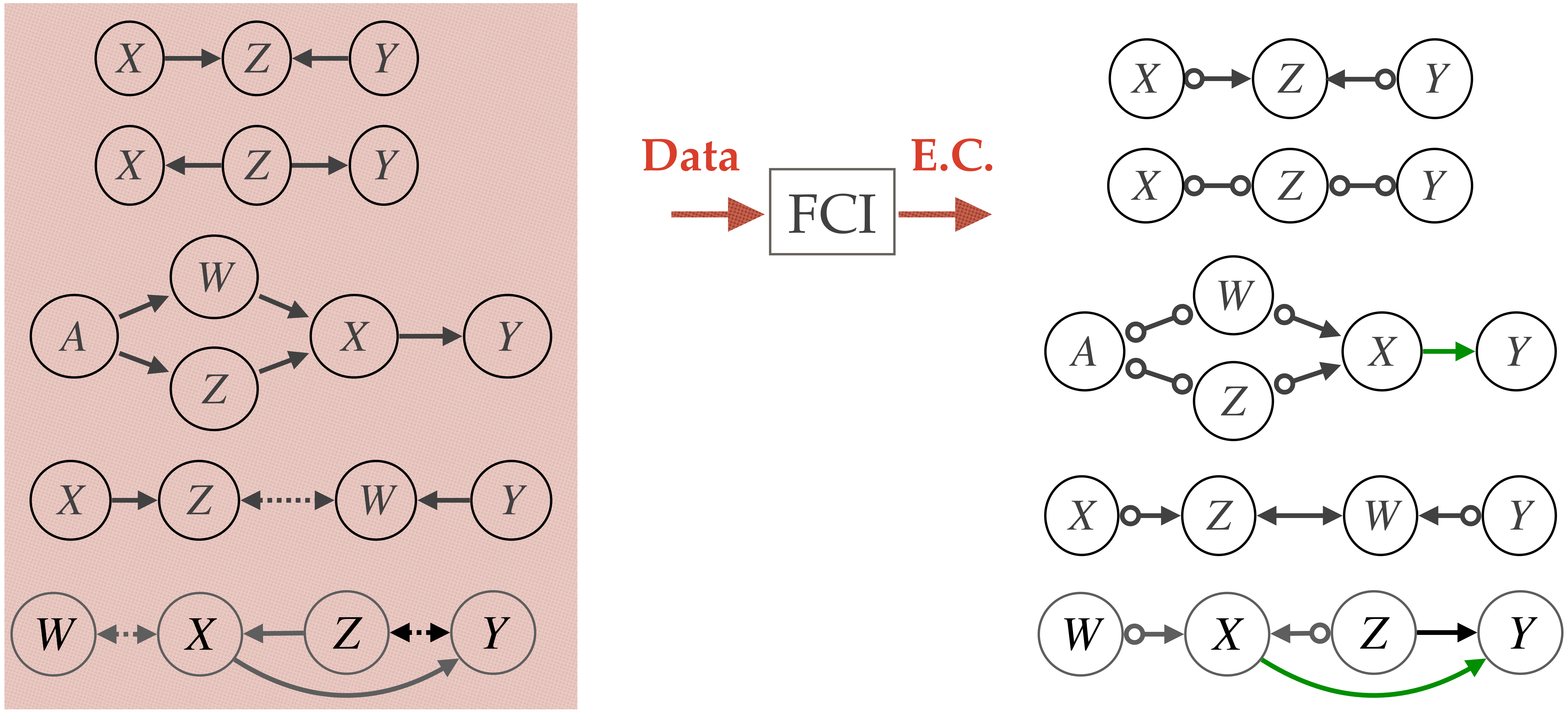
Z and W are ancestors (and definite causes) of Y.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. Link

Causal Structure Learning

Given an **adequate conditional independence test**, structure learning algorithms (e.g. PC/IC, FCI, etc) learn a representation of the Markov equivalence class:

Underlying Causal Diagram | **Partial Ancestral Graph**

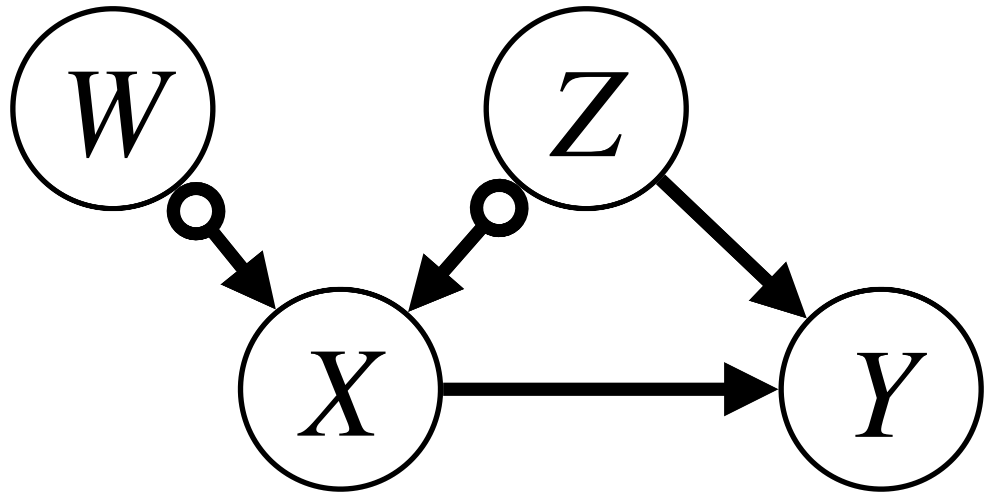


Identification of Causal Effects from PAGs

1 **Query**

$P(y \mid do(x))$

2 **PAG**



3 **Data**

$P(x, y, z, w)$

Inference Engine

Solution

yes / **no**

$$P(y \mid do(x)) = \sum_{z} P\left(y \mid x, z\right) P\left(z\right)$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

Jaber A., **Ribeiro A. H.,** Zhang, J., Bareinboim, E. Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS 2022. (Link)

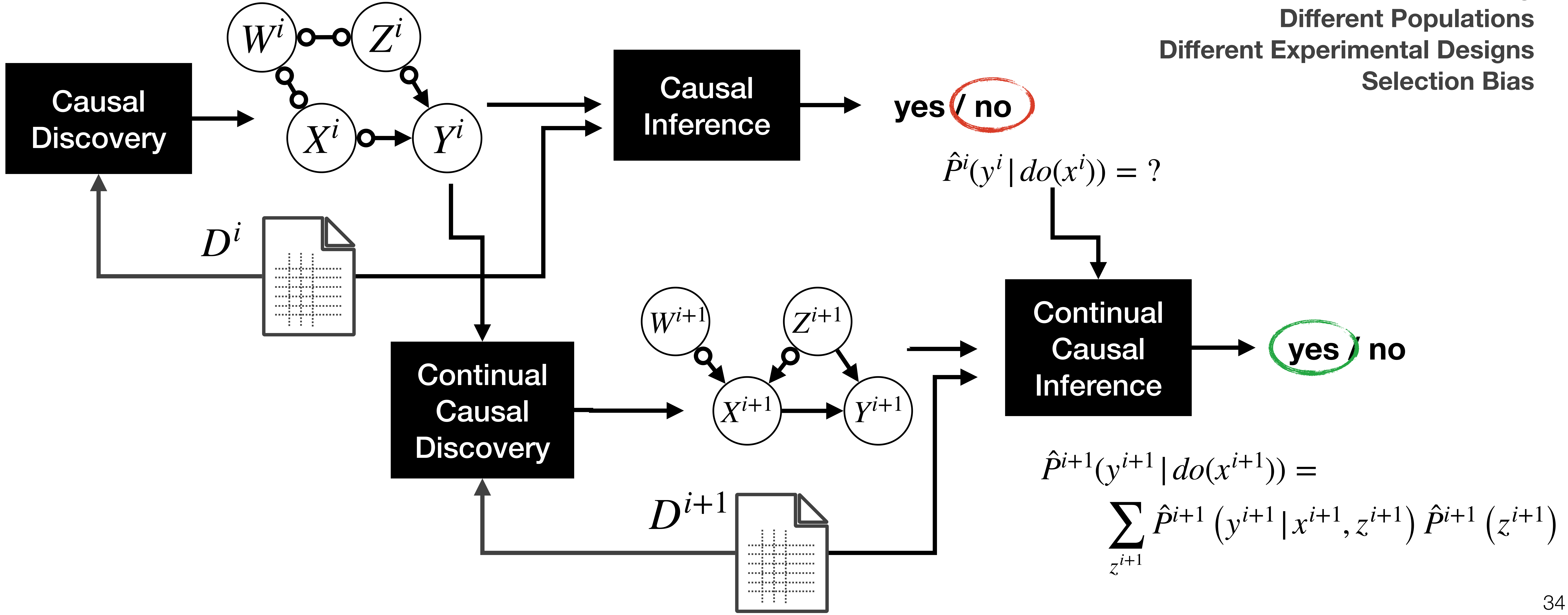
Causal Challenges in Continual Causality

Continual Causal Discovery and Inference

Query: $P(y \mid do(x))$ = **?**

Challenges: Causal Data Fusion / Transfer Learning

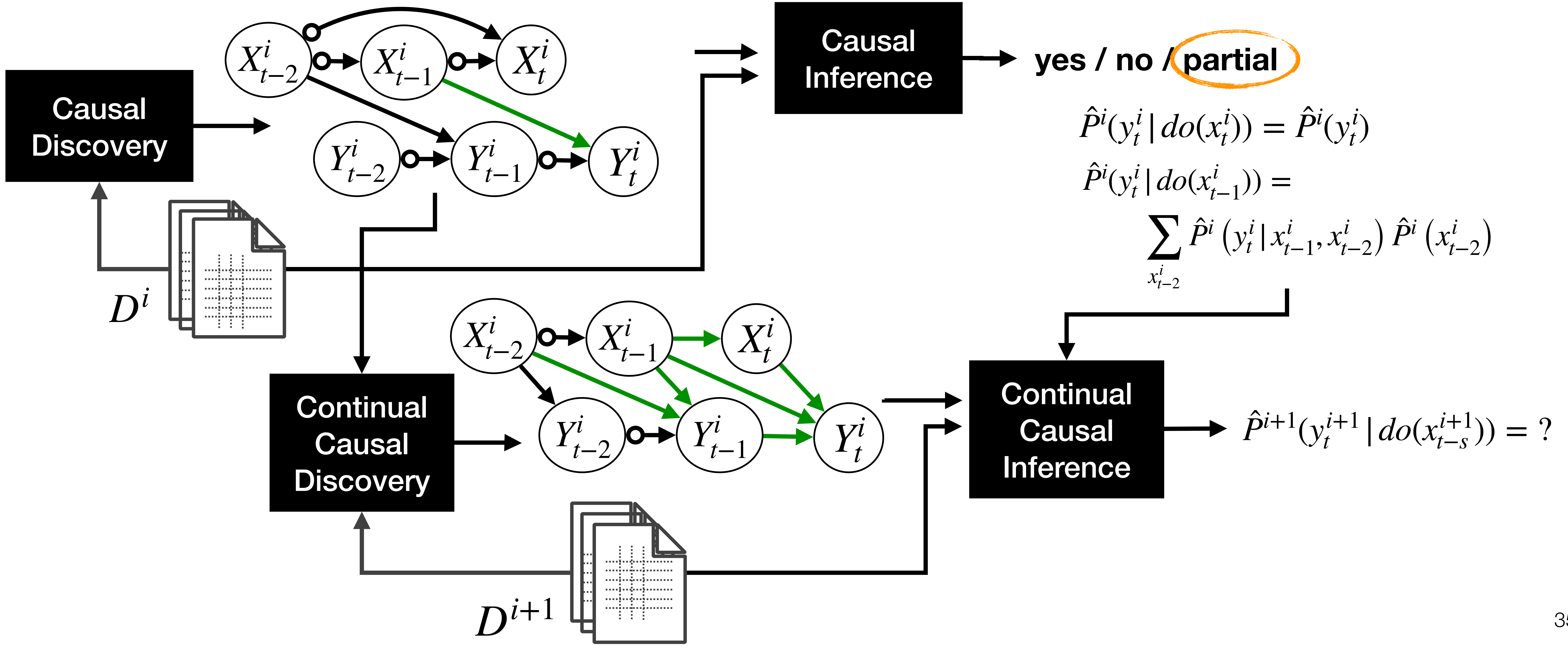
Partial Observability
Different Populations
Different Experimental Designs
Selection Bias



Continual Dynamic Causal Systems

Query: $s \geq 0,\ P(y_t \mid do(x_{t-s}))$ **= ?**

Challenges: Temporal Dependence, Non-Stationarity



Thank You!

adele.ribeiro@uni-marburg.de

Questions?