On Statistical and Causal Models Associated with Acyclic Directed Mixed Graphs

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

January 14, 2025 @ Online Causal Inference Seminar

- Working paper available at arXiv:2501.03048.
- Clarifies and extends (hopefully) the paper by Thomas S. Richardson et al. (2023). "Nested Markov Properties for Acyclic Directed Mixed Graphs". In: *The Annals of Statistics* 51.1, pp. 334–361.

Acyclic directed mixed graphs (ADMGs)

- ADMGs have directed edges ($\longrightarrow$), bidirected edges ($\longleftrightarrow$), and no directed cycles.
- First used by Sewall Wright a century ago in genetics. Stayed popular in economics (e.g. instrumental variable methods) and social science (e.g. LISREL).

$$Z \longrightarrow X \overset{\longleftrightarrow}{\longrightarrow} Y$$

ADMGs play a critical role in modern causal inference, but a fundamental question is unclear:

What is “’the” ADMG model (statistical or causal)?

What is “the” ADMG model?

This is a tricky question about when we think a mathematical definition is “good”.

Two general arguments

Equivalence When many definitions motivated by apparently different considerations are equivalent to each other, they may describe a natural mathematical concept.

- Examples: $\mathbb{N}$, M-matrices, Hammersley-Clifford (factorization $\Leftrightarrow$ Markov).

Completion When there is a natural definition for a smaller class of objects, we may try to find a “completion” of that definition to a larger class of objects.

- Examples: $\mathbb{R}$ (via Cauchy sequences or Dedekind cuts), Lebesgue measure.

Outline of this talk

1. A survey of different interpretations of ADMGs and their relations.
    - A negative answer solely using the **Equivalence argument**.
2. Completeness of graphical statistical models wrt latent variable explanations.
    - A positive answer using the **Equivalence** and **Completion** arguments.
3. Causal ADMG model and the nested Markov property.
4. Discussion: DAG (model) is a special ADMG (model).

Notation

Probability and statistics

- P (a probability distribution), $\mathbb{P}$ (a collection of P, aka a statistical model).
- $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_d$: a finite-dimensional product measure space.
- $\mathbb{P}(\mathbb{V})$: all probability distributions on $\mathbb{V}$ (with a density function).

Graphs

- $\mathbb{G}_{\mathrm{A}}^*(V)$: all ADMGs with vertex set $V = \{V_1, \ldots, V_d\}$ (acyclic = no directed cycles).
- $\mathbb{G}_{\mathrm{B}}^*(V)$: the subclass of all bidirected graphs.
- $\mathbb{G}_{\mathrm{DA}}^*(V)$: the subclass of all DAGs.

Walks

- $\rightsquigarrow$ means a walk (sequence of connected edges) with no colliders (like $\leftharpoonup\!\!\!\longrightarrow V_j \longleftarrow\!\!\!\rightharpoondown$).
- Half arrowhead means unrestricted status: $\leftharpoonup\!\!\!\longrightarrow \;=\; \longrightarrow$ or $\longleftrightarrow$.
- **not** $J \leftharpoonup\!\!\!\rightsquigarrow * \leftsquigarrow\!\!\!\rightharpoondown K \mid L$ means J and K are m-separated by L ($*$ means ≥ 0 colliders).

Marginalization

Consider $J = V_{\mathcal{J}} \subseteq V$.

- Product spaces: $\text{margin}_J(\mathbb{V}) = \mathbb{V}_{\mathcal{J}} = \prod_{j \in \mathcal{J}} \mathbb{V}_j$.
- Probability distributions: $\text{margin}_J(\mathrm{P})$ returns the marginal distribution of J under P.
- Graphs: $\text{margin}_J : \mathbb{G}^*_{\mathrm{A}}(V) \to \mathbb{G}^*_{\mathrm{A}}(J), \mathrm{G} \mapsto \mathrm{G}'$, where

$$V_j \longrightarrow V_k \textbf{ in } \mathrm{G}' \iff P[V_j \rightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}] \neq \emptyset,$$
$$V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G}' \iff P[V_j \leftrightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}] \neq \emptyset,$$

  where P means the set of corresponding paths.

Ancestral subsets

- $J \subseteq V$ is *ancestral* in G if it contains all its ancestors:

$$\{V_k \in V : V_k \rightsquigarrow J \textbf{ in } \mathrm{G}\} \subseteq J.$$

- If J is ancestral, then $\text{margin}_J(\mathrm{G}) = \mathrm{G}_J$ is the subgraph of G restricted to J.

Outline

Ovewview of statistical models associated with ADMG

1. **Global Markov (GM)**.
2. **Unconditional Markov (UM)**.
3. Ordered local Markov (LM): see the paper.
4. **Nested Markov (NM)**.
5. Augmentation (A) criterion (generalizes moralization): see the paper.
6. **Pairwise expansion (PE), clique expansion (CE), noise expansion (NE)**.
7. **Nonparametric equations (E)**.
8. Factorization (F)/exogenous factorization (EF): applies to DAGs/unconfounded ADMGs.

Global Markov (GM) and unconditional Markov (UM)

For $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$, define

$$\mathbb{P}_{\mathsf{GM}}(\mathsf{G}, \mathbb{V})$$
$$=\{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \textbf{not } J \rightsquigarrow * \leftsquigarrow K \mid L \textbf{ in } \mathsf{G} \implies J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P} \text{ for all disjoint } J, K, L \subset V\}.$$

- Every **m-separation** in G implies a **conditional independence** in P.

$$\mathbb{P}_{\mathsf{UM}}(\mathsf{G}, \mathbb{V}) = \{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \textbf{not } J \leftrightsquigarrow K \textbf{ in } \mathsf{G} \implies J \perp\!\!\!\perp K \textbf{ under } \mathsf{P} \text{ for all disjoint } J, K \subset V\}.$$

- Every **unconditional m-separation** in G implies a **marginal independence** in P.

Pairwise (PE), clique (CE), and noise (NE) expansions

For $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$,

- $\text{expand}_{\mathsf{P}}(\mathsf{G})$ replaces a bidirected edge $V_j \longleftrightarrow V_k$ with $V_j \longleftarrow E_{jk} \longrightarrow V_k$.
- $\text{expand}_{\mathsf{C}}(\mathsf{G})$ replaces a bidirected clique $C \subseteq V$ (meaning $V_j \longleftrightarrow V_k$ for all $V_j, V_k \in C$) with $E_C \longrightarrow V_j$, $V_j \in C$.
- $\text{expand}_{\mathsf{N}}(\mathsf{G})$ replaces a bidirected edge $V_j \longleftrightarrow V_k$ with $V_j \longleftarrow E_j \longleftrightarrow E_k \longrightarrow V_k$.

The corresponding statistical models are defined as

$$\mathbb{P}_{\mathsf{PE}}(\mathsf{G}, \mathbb{V}) = \text{margin}_V \left(\mathbb{P}_{\mathsf{GM}}\big(\text{expand}_{\mathsf{P}}(\mathsf{G}), \mathbb{V} \times [0,1]^{|\mathcal{B}|} \big) \right),$$

$$\mathbb{P}_{\mathsf{CE}}(\mathsf{G}, \mathbb{V}) = \text{margin}_V \left(\mathbb{P}_{\mathsf{GM}}\big(\text{expand}_{\mathsf{C}}(\mathsf{G}), \mathbb{V} \times [0,1]^{|\mathcal{C}(\mathsf{G})|} \big) \right),$$

$$\mathbb{P}_{\mathsf{NE}}(\mathsf{G}, \mathbb{V}) = \text{margin}_V \left(\mathbb{P}_{\mathsf{GM}}\big(\text{expand}_{\mathsf{N}}(\mathsf{G}), \mathbb{V} \times [0,1]^{|V|} \big) \right).$$

Nonparametric equations (E)

For $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$, $\mathbb{P}_{\mathsf{E}}(\mathsf{G}, \mathbb{V})$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ such that the following has P-probability 1:

$$V_j = f_j(V_{\mathrm{pa}_{\mathsf{G}}(j)}, E_j),\ j \in [d],^{1} \tag{1}$$

where

- $f_j : \mathbb{V}_{\mathrm{pa}_{\mathsf{G}}(j)} \times [0,1] \to \mathbb{V}_j$, $j \in [d]$;
- $(E_1, \ldots, E_d) \in [0,1]^d$ has a distribution that is UM wrt the bidirected component of G:

$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathsf{G} \Longrightarrow E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}} \textbf{ under } \mathsf{Q},\ \text{for all disjoint } \mathcal{J}, \mathcal{K} \subset [d]. \tag{2}$$

Remarks

- (1) only uses $\longrightarrow$ and (2) only uses $\longleftrightarrow$ in G.
- Closely related to Pearl's **semi-Markovian (causal) model** that does not write down (2).

---

[1] Parent set $\mathrm{pa}_{\mathsf{G}}(j) = \{k \in [d] : V_k \longrightarrow V_j \textbf{ in } \mathsf{G}\}$.

Nested Markov (NM)

The nested Markov property means **the fixed probability distribution is global Markov wrt the fixed graph along all fixable sequences** (Richardson et al. 2023).

- This is closely related to (nonparametric) causal identification.
- **Fixability** of a vertex $V_j \in V$ and the **fixing** operator fix_{V_j} will be defined later.

Relations between ADMG models: the **Equivalence** argument **fails**.

Theorem 1.1 (General ADMGs)

For $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$, we have ($\Rightarrow$ means $\subseteq$ and $\Leftrightarrow$ means $=$ for corresponding statistical models)

$$\begin{array}{ccccc} \text{Pairwise expansion (PE)} & & & & \\ \Downarrow & & & & \\ \text{Clique expansion (CE)} & & & & \\ \Downarrow & & & & \\ \text{Noise expansion (NE)} & \Longleftrightarrow & \text{Nonparametric equation (E)} & & \\ \Downarrow & & & & \\ \text{Nested Markov (NM)} & & & & \\ \Downarrow & & & & \\ \text{Ordered local Markov (LM)} & \Longleftrightarrow & \text{Global Markov (GM)} & \Longleftrightarrow & \text{Augmentation (A)} \\ \Downarrow & & & & \\ \text{Unconditional Markov (UM)} & & & & \end{array}$$

- Most of these are trivial or known. The most nontrivial is NE $\Rightarrow$ NM (end of talk).
- Top half are generative and bottom half are constraint-based.

Equivalence succeeds for simpler subclasses

Theorem 1.2 (DAGs)

For G $\in \mathbb{G}^*_{\mathsf{DA}}(V)$, we have

$$\begin{array}{l} \text{PE} \Leftrightarrow \text{CE} \Leftrightarrow \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{Factorization (F)} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\ \Downarrow \\ \text{UM} \end{array}$$

Theorem 1.3 (Bidirected graphs)

For G $\in BDG(V)$, we have

$$\begin{array}{l} \text{PE} \\ \Downarrow \\ \text{CE} \\ \Downarrow \\ \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \Leftrightarrow \text{UM} \end{array}$$

Outline

A definition of completeness

- An "interpretation" of a ADMG is a collection $\mathbb{P}(\mathrm{G})$ of probability distributions.
- For $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ and $V' \subseteq V$, denote $\mathrm{expand}_{V'}(\mathrm{G}) = \{\mathrm{G}' \in \mathbb{G}_{\mathrm{A}}^*(V') : \mathrm{margin}_V(\mathrm{G}') = \mathrm{G}\}$.
- For each vertex set V, let $\mathbb{G}_0(V)$ be a subclass of ADMGs.

Definition

A collection of models $\mathbb{P}(\mathrm{G})$ for different $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ is **complete** (wrt $\mathbb{G}_0$) if

$$\mathbb{P}(\mathrm{G}) = \bigcup_{V' \supset V} \bigcup_{\mathrm{G}'} \mathrm{margin}_V(\mathbb{P}(\mathrm{G}')),$$

where the second union is over $\mathrm{G}' \in \mathrm{expand}_{V'}(\mathrm{G}) \cap \mathbb{G}_0(V')$.

- Roughly speaking, an ADMG means a **unspecified expansion of itself in the $\mathbb{G}_0$ subclass** (if the model is complete).

Unconfounded ADMGs and completeness

- We say an ADMG is **unconfounded** ($\mathsf{G} \in \mathbb{G}^*_{\mathsf{UA}}(V)$) if

$$V_j \longleftrightarrow V_k \textbf{ in } \mathsf{G} \Longrightarrow V_l \not\longrightarrow V_j, \quad \text{for all distinct } V_j, V_k, V_l \in V.$$

- Simple semantics: **exogenous** variables linked by $\longleftrightarrow$ and **endogenous** variables by $\longrightarrow$.

Theorem 1.4 (Unconfounded ADMGs generalize DAGs and bidirected graphs)

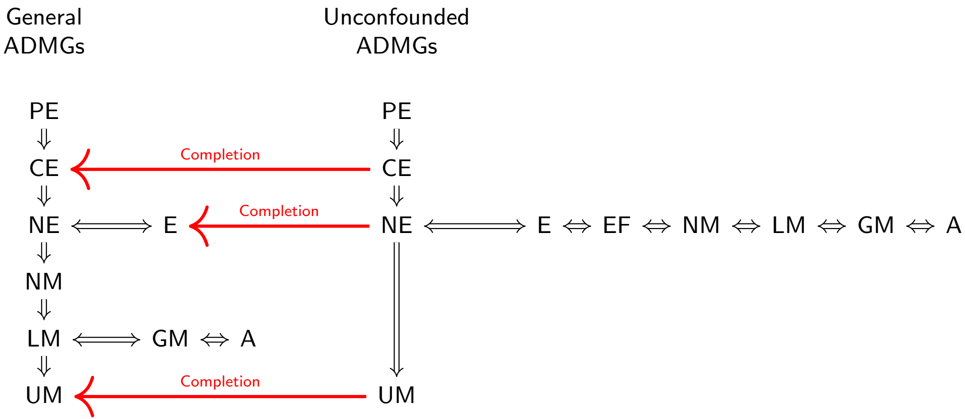
For $\mathsf{G} \in \mathbb{G}^*_{\mathsf{UA}}(V)$, we have

$$\begin{array}{l} \text{PE} \\ \Downarrow \\ \text{CE} \\ \Downarrow \\ \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{Exogenous Factorization (EF)} \Leftrightarrow \text{NM} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\ \Downarrow \\ \text{UM} \end{array}$$

Theorem 2

When $\mathbb{G}_0(V) = \mathbb{G}^*_{\mathsf{UA}}(V)$ for all V, only the CE, NE/E, and UM models are complete.

A visualization of Equivalence + Completion



- PE and CE are “intrinsically directed” and UM is “intrinsically bidirected”.
- **NE/E seems “just right”** if $\longleftrightarrow$ is on an “equal footing” with $\longrightarrow$.

Outline

Causal Markov model

- If E/NE is the "right" statistical model, **what is the "right" causal model?**
- A causal model means a collection of distributions on the **potential outcome schedule**:

$$V(\cdot) = (V_j(v_{\mathcal{I}}) : j \in [d], \mathcal{I} \subseteq [d], v_{\mathcal{I}} \in \mathbb{V}_{\mathcal{I}}).$$

Definition

We say a distribution P of $V(\cdot)$ is **causal Markov** wrt $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ (write $\mathrm{P} \in \mathbb{CP}(\mathrm{G}, \mathbb{V})$) if

1. The potential outcomes are consistent:

$$V_j(v_{\mathcal{I}}) = V_j(v_{\mathrm{pa}(j) \cap \mathcal{I}}, V_{\mathrm{pa}(j) \setminus \mathcal{I}}(v_{\mathcal{I}})), \text{ for all } j \in [d], \mathcal{I} \subseteq [d], v \in \mathbb{V}. \quad (3)$$

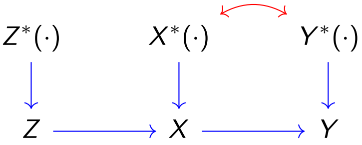
2. The distribution of basic potential outcomes are Markov wrt bidirected part of G:

$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow V_{\mathcal{J}}(v) \perp\!\!\!\perp V_{\mathcal{K}}(v) \textbf{ under } \mathrm{P} \text{ for all } v \in \mathbb{V}. \quad (4)$$

- (3) only uses $\longrightarrow$ (for causality) and (4) only uses $\longleftrightarrow$ (for exogenous correlation).

An illustration

means



- $Z^*(\cdot) = (Z(z, x, y) : z, x, y \in \mathbb{R})$ means the **basic potential outcomes** of Z. Similar for $X^*(\cdot)$ and $Y^*(\cdot)$.
- We use basic p.o. as noise in the E model and interpret equations causally by consistency.
- The noise expansion decouples $\longrightarrow$ (causality) and $\longleftrightarrow$ (exogenous correlation).

Properties of the causal Markov model

Suppose $\mathsf{P} \in \mathbb{CP}(\mathsf{G}, \mathbb{V})$ for some $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$.

Proposition 1 (Extended consistency)

For all disjoint $V_{\mathcal{I}}, V_{\mathcal{I}'} \subset V$, we have

$$\mathsf{P}(V(v_{\mathcal{I}}, v_{\mathcal{I}'}) = V(v_{\mathcal{I}}) \mid V_{\mathcal{I}'}(v_{\mathcal{I}}) = v_{\mathcal{I}'}) = 1.$$

Definition

Let $\mathsf{G}(v_{\mathcal{I}})$ be obtained by removing all edges in $V_{\mathcal{I}} \longrightarrow V$ and relabeling V_j as $V_j(v_{\mathcal{I}})$.

- Basically SWIG with no fixed vertices.

Proposition 2 (Markov property of potential outcomes)

We have $\text{margin}_{V(v_{\mathcal{I}})}(\mathsf{P}) \in \mathbb{P}_{\mathsf{GM}}(\mathsf{G}(v_{\mathcal{I}}), \mathbb{V})$ for all $V_{\mathcal{I}} \subseteq V$ and $v \in \mathbb{V}$.

Nested Markov (NM) property

- V_j is called **fixable** if there exists no V_k such that $V_j \rightsquigarrow V_k$ and $V_j \longleftrightarrow * \longleftrightarrow V_k$.
- NM requires that if V_j is fixable, the next distribution is global Markov wrt $\mathsf{G}_{V_{-j}}$:

$$(\mathrm{fix}_{V_j=v_j}(\mathsf{p}))(v_{-j}) = \frac{\mathsf{p}(v)}{\mathsf{p}(v_j \mid v_{\mathrm{mb}\mathsf{g}_{\mathsf{G}}(j)})},$$

  and **this needs to hold recursively**.[2]
- Importantly, this is a property of **statistical (not causal) models**.

Some remarkable results in Richardson et al. (2023)

- The order of fixing does not matter:

$$\mathrm{fix}_{V_1=v_1} \circ \mathrm{fix}_{V_2=v_2}(\mathsf{p}) = \mathrm{fix}_{V_2=v_2} \circ \mathrm{fix}_{V_1=v_1}(\mathsf{p}) \text{ for } \mathsf{P} \in \mathbb{P}_{\mathrm{NM}}(\mathsf{G}),$$

  as long the sequences (V_1, V_2) and (V_2, V_1) are both fixable.
- CE $\Rightarrow$ NM in general ADMGs. Proof is based on DAG factorization and fairly long.

[2]In personal communications, Thomas Richardson pointed out that the actual nested Markov model makes more assumptions (Verma constraints/no directed effects). See his discussion.

NM and causality

Proposition 3 (Causal identification via fixing)

Suppose $\mathsf{P} \in \mathbb{CP}(\mathsf{G}, \mathbb{V})$ for some $\mathsf{G} \in \mathbb{G}^*_{\mathsf{A}}(V)$. Then

$$
\begin{aligned}
& V_j \in V \text{ is fixable in } \mathsf{G} \\
\iff & \textbf{not } V_j \longleftrightarrow * \longleftrightarrow V_{\mathrm{de}_{\mathsf{G}}(j)} \mid V_{\mathrm{nd}_{\mathsf{G}}(j)} \\
\iff & \textbf{not } V_j(v_j) \leftrightsquigarrow * \leftrightsquigarrow V_{\mathrm{de}_{\mathsf{G}}(j)}(v_j) \mid V_{\mathrm{nd}_{\mathsf{G}}(j)}(v_j) \textbf{ in } \mathsf{G}(v_j) \\
\implies & \mathrm{margin}_{V_{-j}(v_j)}(\mathsf{P}) = \mathrm{fix}_{V_j = v_j}(\mathsf{P}_V).
\end{aligned}
$$

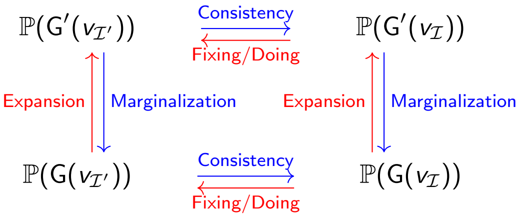
A simple proof of E/NE $\Rightarrow$ NM (Theorem 3 in the paper)

- Consider $\mathsf{P}_V \in \mathbb{P}_{\mathsf{E}}(\mathsf{G}, \mathbb{V})$.
- By interpreting the equations causally, there exists $\mathsf{P} \in \mathbb{CP}(\mathsf{G}, \mathbb{V})$ s.t. $\mathrm{margin}_V(\mathsf{P}) = \mathsf{P}_V$.
- By Propositions 2 and 3, if V_j is fixiable, then $\mathrm{fix}_{V_j = v_j}(\mathsf{P}_V) \in \mathbb{P}_{\mathsf{GM}}(\mathrm{margin}_{V_{-j}(v_j)}(\mathsf{G}(v_j)), \mathbb{V})$
- Notice that $\mathrm{margin}_{V_{-j}(v_j)}(\mathsf{G}(v_j))$ is isomorphic to $\mathsf{G}_{V_{-j}}$, so $\mathrm{fix}_{V_j = v_j}(\mathsf{P}_V) \in \mathbb{P}_{\mathsf{GM}}(\mathsf{G}_{V_{-j}}, \mathbb{V}_{-j})$.
- Now repeatedly apply this argument.

Outline

Understanding ADMG models

- Consider $V_{\mathcal{I}} \subset V_{\mathcal{I}'} \subseteq V \subset V'$.



- When $V_{\mathcal{I}' \setminus \mathcal{I}}$ is fixable, this commutative diagram holds for
  - $\mathbb{P} = \mathbb{P}_{\mathsf{CE}}$ (Richardson et al. 2023, Lemma 43);
  - $\mathbb{P} = \mathbb{P}_{\mathsf{NE}}$ (proved above).

Lots of theory, what's the takeaway?

Use ADMGs, not DAGs

- In theory and practice, ADMG is usually treated as unspecified DAG with latent variables.
- But this is counter-intuitive: DAG is a special ADMG.

Time to treat DAG model as a special ADMG model

ADMG-based causal inference is better because:

1. Philosophically, there are no mysterious latent variables or latent causes.
2. Mathematically, the ADMG-native model $\mathbb{P}_{\mathsf{NE}}$ is preferred by Equivalence + Completion.
3. Practically, ADMGs users are instinctively encouraged to **think about the missing edges** which really drive causal identification.
   - No confounding is about missing $\longleftrightarrow$.
   - IV and proximal inference are mainly about missing $\longrightarrow$.