

<div>DEFINITION</div> <div> <p><i>Fully Connected Graph</i> $G = (V, E)$</p> </div>	<div>All pairs of nodes are adjacent.</div>
<div>DEFINITION</div> <div> <p><i>Markovian Parents</i></p> </div>	<div> <p>$X_{pa(j)}$ is said to be Markovian parents of X_j if it is a minimal subset of $\{X_1, ..., X_{j-1}\}$ such that $f(x_j \mid x_1, ..., x_{j-1}) = f(x_j \mid x_{pa(j)})$. Then</p> $f(x_1, ..., x_p) = \prod_{j=1}^p f(x_j \mid x_{pa(j)})$ </div>
<div>DEFINITION</div> <div> <p><i>DAG Model or Bayesian Network</i></p> </div>	<div>Combination (G, P), where G is a DAG and P a distribution that factorizes according to G.</div>
<div>DEFINITION</div> <div> <p><i>d-separation</i></p> </div>	<div> <p>A path between i to j is blocked by a set \mathbf{S} not containing either (i, j) if at least one of the following holds:</p> <ul style="list-style-type: none"> There is a non-collider on the path that is in \mathbf{S} There is a collider on the path such that neither this collider nor any descendants are in \mathbf{S} if all paths between $i \in \mathbf{A}$ and $j \in \mathbf{B}$ are blocked by \mathbf{S}, then \mathbf{A} and \mathbf{B} are d-separated. </div>

<div>DEFINITION</div> <div><i>Global Markov Property</i></div>	<div>A dist. P with density p satisfies the global Markov property wrt G if:</div> <div>\mathbf{A} and \mathbf{B} are d-separated by \mathbf{S} in $G \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \mid X_{\mathbf{S}}$ in P</div>
--	---

<div>DEFINITION</div> <div><i>Faithfulness</i></div>	<div>Given G a distr. P on X_V is faithful wrt G if for all pairwise disjoint subsets \mathbf{A}, \mathbf{B} and \mathbf{S} of \mathbf{V}:</div> <div>$X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \mid X_{\mathbf{S}}$ in $P \implies \mathbf{A}$ and \mathbf{B} are d-separated by \mathbf{S} in G</div>
--	---

<div>DEFINITION</div> <div><i>Causal DAG Model or Causal BN</i></div>	<div>(G, P) if for any $\mathbf{W} \subset \mathbf{V}$:</div> <div>$p(x_V \mid do(x_w = x'_w)) = \prod_{i \in V \setminus W} p(x_i \mid x_{pa(i)}) 1\{w_W = x'_W\}$</div> <div>which is called g-formula, manipulation formula or truncated factorization formula.</div>
---	---

<div>DEFINITION</div> <div><i>Confounding</i></div>	<div>$(i, k) \in \mathbf{V}, i \neq k$ and there is a directed path from i to k. Then the causal effect is confounded if:</div> <div>$p(x_k \mid x_i) \neq p(x_k \mid do(x_i))$</div>
---	---

<p>DEFINITION</p> <p><i>Structural Equation Models</i></p>	$X_i \leftarrow h_i(X_{pa(i)}, \epsilon_i), \quad i \in \mathbf{V}$ <p>where the error vector is jointly independent.</p>
--	---

<p>DEFINITION</p> <p><i>Valid Adjustment Sets</i></p>	<p>Sets $\mathbf{Z} \subset \mathbf{V}$ that satisfy:</p> $p(x_k \mid do(x_i)) = \int_{x_z} p(x_k \mid x_i, x_Z)p(x_Z)dx_Z = \mathbb{E}_{x_Z} (p(x_k \mid x_i, x_Z))$ <p>for all $p(\cdot)$ such that (G, P) is a causal Bayesian network.</p>
---	---

<p>DEFINITION</p> <p><i>No Direct Causes</i></p>	<p>$pa(i) = \emptyset$:</p> $p(x_{V \setminus \{i\}} \mid do(x'_i)) = p(x_{V \setminus \{i\}} \mid x'_i)$ <p>meaning the do-operator produces the same results as conditioning. To get the marginal, we integrate out along all variables $k \neq i$ to get</p> $p(x_k \mid do(x'_i)) = p(x_i \mid x'_i)$
--	---

<p>DEFINITION</p> <p><i>Reweighting</i></p>	$p(x_{V \setminus \{i\}} \mid do(x'_i)) = \frac{p(x_V)}{p(x_i \mid x_{pa(i)})} x'_i = p(X_{V \setminus \{i, pa(i)\}} \mid x_i, x_{pa(i)})p(x_{pa(i)})$ <p>While $\forall k \notin \{i, pa(i)\}$, integrating out all other variables other than X_i and X_k yields $p(x_k \mid do(x_i)) = \int_{x_{pa(i)}} p(x_k \mid x_i, x_{pa(i)})p(x_{pa(i)})dx_{pa(i)}$ which is known as adjusting for $X_{pa(i)}$</p>
---	--

<p>DEFINITION</p> <p><i>Backdoor Criterion</i></p>	<p>Let $G = (V, E)$ be a DAG and $i, k \in \mathbf{V}, i \neq k$. A set $\mathbf{Z} \subset \mathbf{V}$ (w.o i and k) satisfies the bc wrt (i, k) in G if:</p> <ul style="list-style-type: none"> $\mathbf{Z} \cap desc(i) = \emptyset$ \mathbf{Z} blocks all backdoor paths from i to k in G, i.e all paths between i and k that start with arrow into i ($i \leftarrow \dots k$). the criterion is sufficient for adjustment
<p>DEFINITION</p> <p><i>Adjustment Criterion</i></p>	<p>Let $G = (V, E)$ be a DAG and $i, k \in \mathbf{V}, i \neq k$. A set $\mathbf{Z} \in \mathbf{V}$ (w.o i and k) satisfies the ac wrt (i, k) in G if:</p> <ul style="list-style-type: none"> \mathbf{Z} does not contain any descendants of nodes $r \neq i$ on a directed path from i to k in G \mathbf{Z} blocks all paths between i and k in G that are not directed from i to k ($i \leftarrow \dots k$). the criterion is necessary and sufficient for adjustment for total causal effects (\iff) via adjustment. It is only sufficient for the identification of total causal effects (some effects are identified by other means, e.g frontdoor criterion)
<p>DEFINITION</p> <p><i>Adjustment Sets</i></p>	<ul style="list-style-type: none"> $pa(i)$ is a valid adjustment set Any set \mathbf{Z} satisfying the backdoor criterion relative to (i, k) in G is a valid adjustment set A set \mathbf{Z} is a valid adjustment set if and only if it satisfies the adjustment criterion relative to (i, k) in G
<p>DEFINITION</p> <p><i>Asymptotically Optimal Valid Adjustment Set</i></p>	<p>let $k \in desc(i)$</p> <ul style="list-style-type: none"> causal nodes: $cn(i, k)$: nodes $r \neq i$ on a directed path from i to k in G forbidden nodes: $forb(i, k)$: descendants of causal nodes and node i $\mathbf{O}(i, k) = pa(cn(i, k)) \setminus forb(i, k)$ such that we have $a.var(\hat{\tau}_{ik}^O) \leq a.var(\hat{\tau}_{ik}^Z)$

<p>DEFINITION</p> <p><i>Direct Causal Effect</i></p>	<p>Do not want the total effect, but only the direct causal effect. Controlled direct effect in a linear SEM:</p> $CDE = \mathbb{E}(X_k \mid do(x_i), do(pa(k) \setminus \{i\})) - \mathbb{E}(X_k \mid do(x_i+1), do(pa(k) \setminus \{i\}))$ <p>We can choose a level for $pa(i) \setminus \{i\}$, e.g report a CDE $\forall l \in \mathbb{L}$</p>
--	---

<p>DEFINITION</p> <p><i>Frontdoor Criterion</i></p>	<p>A confounder is unobservable which we cannot use the the criteria to choose a valid adjustment criterion. A set $\boldsymbol{M} \subset \boldsymbol{V}$ (w.o (i, k)) satisfies the fc wrt (i, k) in G if:</p> <ul style="list-style-type: none"> \boldsymbol{M} blocks all directed paths from i to k in G There are no unblocked backdoor paths from i to \boldsymbol{M} in G i blocks all backdoor paths from \boldsymbol{M} to k in G $p(x_k \mid do(x'_i)) = \int_{x_M} p(x_M \mid x'_i) \int_{x_i} p(x_k \mid x_i, x_M) p(x_i) dx_i dx_M$ <p>where the first integral is the effect of the path $X \rightarrow M \rightarrow Y$ and the second one is the backdoor path that is blocked by X and thus the backdoor adjustment formula.</p>
---	--

<p>DEFINITION</p> <p><i>Instrumental Variables</i></p>	<p>Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{I}$ observed while \boldsymbol{U} is unobserved and we are interested in $\boldsymbol{X} \rightarrow \boldsymbol{Y}$.</p> <ul style="list-style-type: none"> $\boldsymbol{I} \not\perp\!\!\!\perp \boldsymbol{X} \mid \boldsymbol{S}$ Every path from \boldsymbol{I} to \boldsymbol{Y} that is not blocked by \boldsymbol{S} has an arrow pointing into X: $\boldsymbol{I} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{S}, do(\boldsymbol{X})$ which implies $\boldsymbol{I} \perp\!\!\!\perp \boldsymbol{U} \mid \boldsymbol{S}$ and hence I affects Y only through X once we control for \boldsymbol{S}.
--	--

<p>DEFINITION</p> <p><i>Transportability</i></p>	<p>Transport formula: Condition: z-specific effects are invariant across populations, e.g</p> $\mathbb{P}^{LA}(y \mid do(t), z) = \mathbb{P}^{NYC}(y \mid do(t), z) = \mathbb{P}(y \mid do(t), z)$ $\mathbb{P}^*(y \mid do(t)) = \sum_{z \in Z} \mathbb{P}(y \mid do(t), z) \mathbb{P}^*(z)$
--	--

DEFINITION

Potential Outcomes

- $Y_i(x = 0)$ or $(Y_i(x_0))$ is the value of Y that would be observed for a given unit i if assigned $X = 0$ and vice versa $Y_i(x_1)$
- ACE for binary setting: $ACE = \mathbb{E}(Y(x_1) - Y(x_0)) = \mathbb{P}(helped) - \mathbb{P}(hurt)$

DEFINITION

Inverse Probability Weighting

$$\hat{\mathbb{E}}(Y \mid do(X = 1)) = n^{-1} \sum_{i \in I} Y_i 1\{X_i = 1\} w_i$$

where $w_i = \frac{1}{\pi_i} = \frac{1}{\hat{P}(X=1|Z_i)}$ or more generally

$$\hat{\mathbb{E}}(l(x_V) \mid do(X_k \leftarrow \tilde{N}_k)) = n^{-1} \sum_{i \in I} l(x_V^i) w_i$$

$$\text{where } w_i = \frac{\tilde{p}(x_k^i | x_{pa(k)}^i)}{p(x_k^i | x_{pa(k)}^i)}$$

DEFINITION

Summary Markov Properties

A DAG $G = (V, E)$ and a distr. P on X_V are said to satisfy:

- global Markov property wrt G if for all pairwise disjoint subsets \mathbf{A}, \mathbf{B} and \mathbf{S} of \mathbf{V} :

$$\mathbf{A} \text{ and } \mathbf{B} \text{ are d-separated by } \mathbf{S} \text{ in } G \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \mid X_{\mathbf{S}} \text{ in } P$$

- local Markov property wrt G if for all $j \in \mathbf{V}$:

$$X_j \perp\!\!\!\perp X_{\text{non-desc}(j) \setminus pa(j)} \mid X_{pa(j)}$$

- the Markov factorization property wrt G if:

$$p(x_V) = \prod_{j \in \mathbf{V}} p(x_j \mid x_{pa(j)})$$

- If P has a density wrt to a product measure, then all three Markov properties are equivalent (P is Markov wrt G) and equivalently G is an independence (I-map) map of P

DEFINITION

Minimal I-map

Every distribution is Markov wrt a full DAG or a full DAG is an I-map of any distribution. A DAG $G = (V, E)$ is a minimal I-map of P if:

- G is an I-map of P and
- $G' = (V, E)$ with $' \subset E$ is not an I-map of P
- P is then said to satisfy causal minimality wrt G

<div>DEFINITION</div> <div>Perfect Map</div>	<p>If a distr. P is Markov and faithful wrt DAG G, then G is said to be a perfect map of P.</p>
<div>DEFINITION</div> <div>Markov Equivalence</div>	<p>Two DAGs G_1 and G_2 are Markov equivalent if they describe the same set of d-separation relationships, i.e for all pairwise disjoint subsets \mathbf{A}, \mathbf{B} and \mathbf{S} of \mathbf{V}, we have:</p> <p>\mathbf{A} and \mathbf{B} are d-separated by \mathbf{S} in $G_1 \iff \mathbf{A}$ and \mathbf{B} are d-separated by \mathbf{S} in G_2</p> <p>A perfect map is unique up to Markov equivalence.</p>
<div>DEFINITION</div> <div>CPDAG</div>	<ul style="list-style-type: none"> All DAGs in a Markov equivalence class have the same skeleton and the same v-structures CPDAG <ul style="list-style-type: none"> $i \rightarrow j$ iff $i \rightarrow j$ in all DAGs from the class $i - j$ iff there is a DAG in the Markov equivalence class with $i \rightarrow j$ and one with $j \rightarrow i$
<div>DEFINITION</div> <div>Adjacency</div>	<p>Let $G = (V, E)$ be a DAG. Let $i, j \in V$ such that $i \neq j$. Then:</p> <ul style="list-style-type: none"> If $i \in adj(j)$ in G, they cannot be d-separated by any subset of the remaining variables If $i \notin adj(j)$ in G, they are d-separated by $\mathbf{S} \in \{pa(i), pa(j)\}$ (or) $i \in adj(j) \iff$ they cannot be d-separated by any subset of the remaining nodes If they can be d-separated by any subset of the remaining nodes, they can be d-separated by $pa(i)$ or $pa(j)$

DEFINITION

SGS Algorithm

Constraint based and assumes Markov and faithfulness to estimate CPDAGs.

1. determine the skeleton
2. no edge $i - j \iff i \perp j \mid \mathbf{S} \subset x_{V \setminus \{i,j\}} \iff X_i \perp\!\!\!\perp X_j \mid X_{\mathbf{S}}$ with $\mathbf{S} \subset x_{V \setminus \{i,j\}}$

Starts with the complete graph and for all pairs $i \neq j$ assesses conditional independence of X_i and X_j given $X_{\mathbf{S}}$ for **all** subsets \mathbf{S} of the remaining nodes and removes an edge if cond. independence is found. Starts with 1 and sets of $|Z_1| = 1$ and gradually increases setsize and tests all edges using the corresponding set as covariates.

DEFINITION

PC Algorithm

Constraint based and assumes Markov and faithfulness to estimate CPDAGs.

1. determine the skeleton
2. no edge between i and $j \iff i$ and j are d-separated by $pa(i, G)$ or $pa(j, G) \iff i$ and j are d-separated by a subset \mathbf{S}' of $adj(i, G)$ or of $adj(j, G) \iff X_i \perp\!\!\!\perp X_j \mid X_{\mathbf{S}'}$ for a subset \mathbf{S}' of $adj(i, G)$ or of $adj(j, G)$

Starts with the complete graph and for $k = 0, 1, \dots, p - 2$ considers all pairs of adjacent vertices (i, j) and removes edge if X_i and X_j are conditionally independent given some subset of size k of $adj(i, G)$ or of $adj(j, G)$. Skeleton phase means $|\mathbf{Z}| = 0$

DEFINITION

Properties of the BIC

- $S(G, D) = 2(D \mid \hat{\theta}, G) - \log n \times (\#parameters)$
- Decomposability: $S(G, D) = \sum_{j \in J} s(X_j, X_{pa(j)})$
- local consistency: S increases iff adding an edge that eliminates an independence constraint that does not hold in P and vice versa

DEFINITION

Greedy Equivalence Search

Optimizes BIC, searches over MEC and assumes Markov and faithfulness.

Define $\epsilon^+(\epsilon)$ as: an equivalence class ϵ' is in $\epsilon^+(\epsilon) \iff$ there is some DAG $G \in \epsilon$ to which we can add a single edge that results in a DAG $G' \in \epsilon'$

1. write down $\epsilon(G)$
2. add the different possible directed edges to each member $G \in \epsilon(G)$
3. construct the CPDAG for the possible DAGs and add the remaining possible DAGs from that CPDAG

\implies GES starts with empty CPDAG, adds edges until a local max is achieved and then removes until a local maximum is reached. Consistent procedure.

<p>DEFINITION</p> <p><i>RESIT</i></p>	<p>Special case of RESIT for two variables</p> <ol style="list-style-type: none"> 1. Regress Y on X with some (possibly non-linear) regression technique 2. test if $Y - \hat{f}_Y(X)$ is independent of X 3. Repeat the procedure with exchanging the roles of Y, X 4. If independence is not rejected for one direction but is for the other one, the one that is rejected is assumed to be the causal direction (noise of $Y - \hat{f}_Y(X)$ should be independent of X if $X \in desc(Y)$)
---------------------------------------	--

<p>DEFINITION</p> <p><i>LiNGAM</i></p>	<ul style="list-style-type: none"> • no faithfulness needed • $X := BX + \epsilon$ with $diag(B) = 0$ due to acyclicity • a bijective mapping $\pi : \{1, ..., p\} \rightarrow \{1, ..., p\}$ is called a causal ordering if it satisfies $\pi(i) < \pi(j)$ if $j \in desc(i)$ which can be used to make B strictly lower triangular
--	--

<p>DEFINITION</p> <p><i>LiNGAM Algorithm</i></p>	<p>$X = BX + \epsilon$ vs. $X = AS$ for ICA meaning $A = (I - B)^{-1}$ and $S = \epsilon$</p> <ol style="list-style-type: none"> 1. estimate $W = A^{-1} = (I - B)$ up to permutation, scaling and sign of the columns of the true matrix 2. find unique permutation of rows that yields \tilde{W} w.o any zeros on the diagonal by $\text{argmin} \sum_{i \in I} 1/ \tilde{W}_{ii}$ 3. divide each row of \tilde{W} by its diagonal element to yield \tilde{W}' with only ones on the diagonal 4. Compute $\tilde{B} = I - \tilde{W}'$ 5. find causal order by using the Choleski factorization on to get $\hat{B} = \tilde{P}\tilde{B}\tilde{P}^T$ as close as possible to strictly lower triangular
--	---

<p>DEFINITION</p> <p><i>Causal Invariance</i></p>	<p>Let $X_V^e = (X_1^e, ..., X_d^e) \sim P^e$ for some $e \in \epsilon$. Let $pa(Y) := S^*$ and finding that set in the goal. No faithfulness assumed.</p> <ul style="list-style-type: none"> • invariance: $Y^e \mid X_{S^*}^e = x = Y^f \mid X_{S^*}^f = x \forall e, f \in \epsilon$ • $H_{0,S^*}(\epsilon) : \exists \gamma^* \in \mathbb{R}^p$ with $supp(\gamma^*) = S^*$ that satisfies $\forall e \in \epsilon : X^e$ has an arbitrary distribution and $Y^e = X^e \gamma^* + \epsilon^e, \epsilon_\epsilon^e$ and $\epsilon^e \perp\!\!\!\perp X_{S^*}^e$ • test H_{0,S^*} for all subsets of the predictors to estimate $\hat{S} := \cap_{S: H_{0,S^*} \text{ not rej.}} S$
---	---

DEFINITION

Causal Invariance 2.0

- Theorem: $\mathbb{P}(\hat{S} \subseteq S^*) \geq \mathbb{P}(H_{0,S^*} \text{ not rejected}) \geq 1 - \alpha$
- $H_0 : \forall S \subseteq \{1, \dots, p\} : \text{Fit a linear regression using set } S \text{ of variables and data from all environments. Let } R = Y - \hat{f}(X_S) \text{ and test the hypothesis that the mean and variances of } R \text{ are identical across all } e \in \epsilon. \text{ Let } p = 2 \times \min\{p_1, p_2\}. \text{ If } p < \alpha, \text{ reject } S.$
- let $\epsilon_1 \subset \epsilon_2 : \emptyset \subseteq S(\epsilon_1) \subseteq S(\epsilon_2) \subseteq S^*$

DEFINITION

Intervention Stable

- A set S is intervention stable if it d-separates the response from all interventions, i.e $Y \perp I \mid X_S \forall I$ active in ϵ
- S is called a set of plausible causal predictors under a set of environments $e \in \epsilon$ if for all environments $e, f \in \epsilon$ and all $x : Y^e \mid X_S^e = x = Y^f \mid X_S^f = x$
- the true S^* is a pl. causal but there are also other sets satisfying that property.
- Assumption: S intervention stable $\leftarrow S$ is a set of plausible causal predictors

DEFINITION

General Remarks

- Every P factorizes according to a full DAG, i.e $f(x_j)_{j=1}^p = f(x_1)f(x_2 \mid x_1) \dots f(x_p \mid x_{p-1}, \dots, x_1)$ while there are $p!$ possibilities.
- A dist. P with density p satisfies the global Markov property wrt $G \iff p$ factorizes according to $G \implies$ d-separations in G imply conditional independencies in P
- DAGs with n nodes: $G(n) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} G(n-k)$
- Mathematical definition of causal effect: a, b such that $p(y \mid do(X = a)) \neq p(y \mid do(X = b))$

- Formulations of total causal effect
 - there is a total causal effect from X to $Y \iff$
 - $\exists x', \tilde{x}$ such that $p(y \mid do(X = x')) \neq p(y \mid do(X = \tilde{x})) \iff$
 - $\exists x'$ such that $p(y \mid do(X = x')) \neq p(y)$
 - $X \not\perp\!\!\!\perp Y$ in $p(x, y \mid do(X = \tilde{N}_x))$ if $Var(\tilde{N}_x) > 0$ meaning $X \rightarrow Y$
- path method estimates the total **average** causal effect of X on Y
- Identifiability: A model is identifiable when it cannot be changed without there also being *some* change in the distribution of the observable variables (about

true distribution - not finite data). An interventional distribution is **identifiable** if it can be computed from the observational distribution and the graph structure G

- Positivity: $p(x_i, x_Z) > 0 \forall x_i \in X_i, x_Z \in X_Z$. *Everything is at least possible in the observational distribution.*
- Simplification in the Gaussian setting: $\mathbb{E}(X_k \mid do(x_i = x'_i + 1)) - \mathbb{E}(X_k \mid do(x_i = x'_i)) = \gamma$ where γ is the coefficient of a linear regression of X_k on (X_i, X_Z) . This holds for arbitrary error distributions but the above proof is not as easy.
- M -bias: adjusting for pre-treatment variables be-

<p>cause the pre-treatment variable forms the middle of an M-like structure</p> <ul style="list-style-type: none"> • If a valid adjustment set exists, $\mathbf{O}(i, k)$ is one. • Direct causal effect: given by the edge weights in a linear SEM (controlled direct effect and natural direct effect coincide). In a non-linear SEM the level of $pa(k) \setminus \{i\}$ which is e.g the department. • consistency for β^{OLS} for IV: $\hat{\beta}^{OLS}$ is inconsistent unless $\delta = 0$ and/or $\gamma = 0$ with those being the effects $U - X$ and $U - Y$ • $\hat{\beta}^{IV} = \frac{Cov(I, Y)}{Cov(I, X)}$ while I is a weak instrument if $Cov(I, X)$ is small. 	<ul style="list-style-type: none"> • ACE with binary treatment and binary response: $ACE = \mathbb{E}(Y(x_1) - Y(x_0)) = \mathbb{P}(helped) - \mathbb{P}(hurt)$ • smd for matching: $smd = \left \frac{\hat{Z}_T - \hat{Z}_C}{\sqrt{\frac{s_T^2 - s_C^2}{2}}} \right$ while $smd < 0.1$ is adequate, $smd \in [0.1, 0.2]$ is not too alarming and bigger means serious imbalance. • A distribution does not have a unique I-map, nor a unique minimal I-map and different minimal I-maps of a distribution can have different skeletons • Not for every distribution P there exists a DAG G such that G is a perfect map of P • there are no d-separated nodes in a fully connected
--	---

<p>DAG</p> <ul style="list-style-type: none"> • a full DAG is an I-map of any distribution • PC algorithm: if an edge $i - j$ is removed at any point, they are not adjacent, at any point in the algorithm, the current skeleton is a supergraph of the skeleton of G and if an edge $i - j$ is not removed, then i and j are adjacent. • partial correlation: $\rho_{X, Y Z} = \frac{\rho_{X, Y} - \rho_{X, Z}\rho_{Z, Y}}{\sqrt{(1 - \rho_{X, Z}^2)(1 - \rho_{Z, Y}^2)}}$ which is the correlation of residuals after regressing X on Z and Y on Z. We generally cannot use that number to infer something about the true cond. independencies. 	<ul style="list-style-type: none"> • Because of acyclicity, there is always a causal ordering which is not necessarily unique. • S is ICA is identifiable up to permutation, scaling and sign of the columns
---	---

--	--

--	--