Single World Intervention Graphs (SWIGs):
Unifying the Counterfactual and Graphical Approaches to Causality

Thomas Richardson
Department of Statistics
University of Washington

24 April 2013

Joint work with James Robins (Harvard School of Public Health)
Outline

- A new unification of graphs and counterfactuals via node-splitting
- Factorization and Modularity
- Contrast with Twin Network approach
- Some Examples and Extensions
The potential outcomes framework: philosophy

Hume (1748) *An Enquiry Concerning Human Understanding*:

*We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, ...*
Hume (1748) *An Enquiry Concerning Human Understanding*:

*We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, ...*

*...where, if the first object had not been the second never had existed.*
Potential outcomes with binary treatment

For binary treatment $X$ and response $Y$, we define two potential outcome variables:

- $Y(x = 0)$: the value of $Y$ that \textit{would} be observed for a given unit \textit{if} assigned $X = 0$;
Potential outcomes with binary treatment

For binary treatment $X$ and response $Y$, we define two potential outcome variables:

- $Y(x = 0)$: the value of $Y$ that \textit{would} be observed for a given unit \textit{if} assigned $X = 0$;
- $Y(x = 1)$: the value of $Y$ that \textit{would} be observed for a given unit \textit{if} assigned $X = 1$;
Potential outcomes with binary treatment

For binary treatment $X$ and response $Y$, we define two potential outcome variables:

- $Y(x = 0)$: the value of $Y$ that would be observed for a given unit \textit{if} assigned $X = 0$;
- $Y(x = 1)$: the value of $Y$ that \textit{would} be observed for a given unit \textit{if} assigned $X = 1$;

\textit{Will also write these as $Y(x_0)$ and $Y(x_1)$.}
## Potential Outcomes

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential Outcomes</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y(x = 0)$</td>
<td>$Y(x = 1)$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
## Assignment to Treatments

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential Outcomes</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y(x = 0)$</td>
<td>$Y(x = 1)$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
# Observed Outcomes from Potential Outcomes

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential Outcomes</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y(x = 0)$</td>
<td>$Y(x = 1)$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
## Potential Outcomes and Missing Data

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential Outcomes</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y(x = 0)$</td>
<td>$Y(x = 1)$</td>
</tr>
<tr>
<td>1</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>
Graphical Approach to Causality (I)

Graph intended to represent direct causal relations.

- Graph with nodes $X$ and $Y$ labeled "No Confounding".
- Graph with nodes $X$, $Y$, and an unobserved variable $H$ labeled "Confounding".

Convention that confounding variables (e.g. $H$) are always included on the graph.

Approach originates in the path diagrams introduced by Sewall Wright in the 1920s.

If $X \rightarrow Y$ then $X$ is said to be a parent of $Y$; $Y$ is a child of $X$. 

- $X$ → $Y$
Graphical Approach to Causality (I)

Graph intended to represent direct causal relations.

- Convention that confounding variables (e.g. \( H \)) are always included on the graph.
Graphical Approach to Causality (I)

- Graph intended to represent direct causal relations.
- Convention that confounding variables (e.g. $H$) are always included on the graph.
- Approach originates in the path diagrams introduced by Sewall Wright in the 1920s.
Graphical Approach to Causality (I)

Graph intended to represent direct causal relations.

Convention that confounding variables (e.g. $H$) are always included on the graph.

Approach originates in the path diagrams introduced by Sewall Wright in the 1920s.

If $X \rightarrow Y$ then $X$ is said to be a parent of $Y$; $Y$ is child of $X$. 
Edges are directed, but are they causal?

No Confounding

\[ P(X, Y) = P(X)P(Y | X) \]

No Confounding

\[ P(X, Y) = P(Y)P(X | Y) \]
Edges are directed, but are they causal?

No Confounding

\( P(X, Y) = P(X)P(Y \mid X) \)

No Confounding

\( P(X, Y) = P(Y)P(X \mid Y) \)

Neither factorization places any restriction on \( P(X, Y) \).
Linking the two approaches

\[ X \perp Y(x_0) \quad \& \quad X \perp Y(x_1) \]

- Elephant in the room:

Unobserved

\[ X \not\perp Y(x_0) \quad \& \quad X \not\perp Y(x_1) \]
Linking the two approaches

X ⊥ ⊥ Y(x₀) & X ⊥ ⊥ Y(x₁)

X ⊥ Y(x₀) & X ⊥ Y(x₁)

Elephant in the room:
*The variables Y(x₀) and Y(x₁) do not appear on these graphs!!*
Node splitting: Setting $X$ to 0

\[ P(X = \tilde{x}, \ Y = \tilde{y}) \ = \ P(X = \tilde{x}) \ P(Y = \tilde{y} \mid X = \tilde{x}) \]

Can now ‘read’ the independence: $X \perp \perp Y(x = 0)$. 

Node splitting: Setting $X$ to 0

\[
P(X = \tilde{x}, Y = \tilde{y}) = P(X = \tilde{x})P(Y = \tilde{y} \mid X = \tilde{x})
\]

Can now ‘read’ the independence: $X \perp \perp Y(x = 0)$.
Also associate a new factorization:

\[
P(X = \tilde{x}, Y(x = 0) = \tilde{y}) = P(X = \tilde{x})P(Y(x = 0) = \tilde{y})
\]
Node splitting: Setting $X$ to 0

\[ P(X = \tilde{x}, Y = \tilde{y}) = P(X = \tilde{x})P(Y = \tilde{y} \mid X = \tilde{x}) \]

\[ \xrightarrow[\Rightarrow]{P(\ Y \mid X = 0)} \]

Can now ‘read’ the independence: $X \perp \perp Y(x = 0)$. Also associate a new factorization:

\[ P(\tilde{x}, Y(x = 0) = \tilde{y}) = P(X = \tilde{x})P(Y(x = 0) = \tilde{y}) \]

where:

\[ P(\tilde{y}) = P(Y = \tilde{y} \mid X = 0). \]

This last equation links a term in the original factorization to the new factorization. We term this the ‘modularity assumption’.
Node splitting: Setting $X$ to 1

\[ P(X = \tilde{x}, Y = \tilde{y}) = P(X = \tilde{x})P(Y = \tilde{y} | X = \tilde{x}) \]

Can now ‘read’ the independence: \( X \perp \!\!\! \perp Y(\chi = 1). \)
Node splitting: Setting $X$ to 1

\[ P(X = \tilde{x}, Y = \tilde{y}) = P(X = \tilde{x})P(Y = \tilde{y} \mid X = \tilde{x}) \]

Can now ‘read’ the independence: $X \perp \perp Y(x = 1)$.
Also associate a new factorization:

\[ P(X = \tilde{x}, Y(x = 1) = \tilde{y}) = P(X = \tilde{x})P(Y(x = 1) = \tilde{y}) \]
Node splitting: Setting $X$ to 1

\[
P(X = \tilde{x}, Y = \tilde{y}) = P(X = \tilde{x})P(Y = \tilde{y} | X = \tilde{x})
\]

Can now ‘read’ the independence: $X \perp \perp Y(x=1)$.
Also associate a new factorization:

\[
P (X = \tilde{x}, Y(x=1) = \tilde{y}) = P(X = \tilde{x})P (Y(x=1) = \tilde{y})
\]

where:

\[
P (Y(x=1) = y) = P(Y = y | X = 1).
\]
Crucial point: $Y(x=0)$ and $Y(x=1)$ are never on the same graph. Although we have:

$$X \perp \!\!\!\!\!\!\perp Y(x=0) \quad \text{and} \quad X \perp \!\!\!\!\!\!\perp Y(x=1)$$

we do not have

$$X \not\perp \!\!\!\!\!\!\perp Y(x=0), Y(x=1)$$
Crucial point: \( Y(x=0) \) and \( Y(x=1) \) are never on the same graph. Although we have:

\[
X \perp \perp Y(x=0) \quad \text{and} \quad X \perp \perp Y(x=1)
\]

we do not have

\[
X \perp Y(x=0), Y(x=1)
\]

Had we tried to construct a single graph containing both \( Y(x=0) \) and \( Y(x=1) \) this would have been impossible. (Why?)
Crucial point: $Y(x=0)$ and $Y(x=1)$ are never on the same graph. Although we have:

$$X \perp \perp Y(x=0) \text{ and } X \perp \perp Y(x=1)$$

we do not have

$$X \perp Y(x=0), Y(x=1)$$

Had we tried to construct a single graph containing both $Y(x=0)$ and $Y(x=1)$ this would have been impossible. (Why?)

$\Rightarrow$ **Single-World Intervention Graphs** (SWIGs).
Representing both graphs via a ‘template’

Represent both graphs via a template.

Formally this is a ‘graph valued function’:

- Takes as input a specific value $x^*$
Representing both graphs via a ‘template’

Represent both graphs via a template.

Formally this is a ‘graph valued function’:
- Takes as input a specific value $x^*$
- Returns as output a SWIG $g(x^*)$. 
Representing both graphs via a ‘template’

Represent both graphs via a template.

Formally this is a ‘graph valued function’:
- Takes as input a specific value $x^*$
- Returns as output a SWIG $g(x^*)$.

Each instantiation of the template is a SWIG $g(x^*)$ that represents a different margin: $P(X, Y(x^*))$ with red nodes $x^*$ becoming constants.
Q: *How could we identify whether someone would choose to take treatment, i.e. have $X = 1$, and at the same time find out what happens to such a person if they don’t take treatment $Y(x = 0)$?*
Q: How could we identify whether someone would choose to take treatment, i.e. have $X = 1$, and at the same time find out what happens to such a person if they don’t take treatment $Y(x = 0)$?

A: Consider an experiment in which, whenever a patient is observed to swallow the drug have $X = 1$, we instantly intervene by administering a safe ‘emetic’ that causes the pill to be regurgitated before any drug can enter the bloodstream.
Q: How could we identify whether someone would choose to take treatment, i.e. have $X = 1$, and at the same time find out what happens to such a person if they don’t take treatment $Y(x = 0)$?

A: Consider an experiment in which, whenever a patient is observed to swallow the drug have $X = 1$, we instantly intervene by administering a safe ‘emetic’ that causes the pill to be regurgitated before any drug can enter the bloodstream. Since we assume the emetic has no side effects, the patient’s recorded outcome is then $Y(x = 0)$. 
Query: does this causal graph imply?

\[ Y(x_0, x_1) \perp X_1(x_0) \mid Z(x_0), X_0, \]
Query does this graph imply:

\[ Y(x_0, x_1) \perp \!
\!
\perp X_1(x_0) \mid Z(x_0), X_0 \]

Answer: \[ Y(x_0, x_1) \perp \!
\!
\perp X_1(x_0) \mid Z(x_0), X_0 \]. More on this shortly...
Query does this graph imply:

\[ Y(x_0, x_1) \perp \!\!\!\!\!\!\perp X_1(x_0) \mid Z(x_0), X_0 \]

Answer: Yes – applying d-separation to the SWIG on the right we see that there is no d-connecting path from \( Y(x_0, x_1) \) given \( Z(x_0) \).
Simple solution

Query does this graph imply:

\[ Y(x_0, x_1) \perp \!\!\!\!\!\!\perp X_1(x_0) \mid Z(x_0), X_0 \]

Answer: Yes – applying d-separation to the SWIG on the right we see that there is no d-connecting path from \( Y(x_0, x_1) \) given \( Z(x_0) \).

More on this shortly...
Single World Intervention Template Construction (1)

Given a graph $G$, a subset of vertices $A = \{A_1, \ldots, A_k\}$ to be intervened on, we form $G(a)$ in two steps:

(1) **(Node splitting):** For every $A \in A$ split the node into a *random* node $\lambda$ and a *fixed* node $a$:

![Diagram of node splitting](image)

*Splitting:* Schematic Illustrating the Splitting of Node $A$

- The random half inherits all edges directed into $A$ in $G$;
- The fixed half inherits all edges directed out of $A$ in $G$. 
(2) Relabel descendants of fixed nodes:
A Single World Intervention Graph (SWIG) $G(a^*)$ is obtained from the Template $G(a)$ by simply substituting specific values $a^*$ for the variables $a$ in $G(a)$.

For example, we replace $G(x)$ with $G(x=0)$.

- Although the random variables in a SWIG take different values for different units, fixed variables take the same value.
- Changing the value of a fixed variable corresponds to constructing a new graph and considering a different population, e.g. $P(X, Y(x=0))$ vs. $P(X, Y(x=1))$.
- It is only instantiated graph $G(x^*)$ that represents $P(\forall(\tilde{x})$, not the template $G(x)$. 


Inferential Problem (II)

Pearl (2009), Ex. 11.3.3, claims the causal DAG above does not imply:

\[ Y(x_0, x_1) \perp\!
\perp X_1 | Z(x_0), X_0 = x_0. \]  

The SWIG shows that (1) does hold; Pearl is incorrect. Specifically, we see from the SWIG:

\[ Y(x_0, x_1) \perp\!
\perp X_1 | Z(x_0), X_0 = x_0. \]  

This last condition is then equivalent to (1) via consistency.
Pearl (2009), Ex. 11.3.3, claims the causal DAG above does not imply:

\[ Y(x_0, x_1) \perp \perp X_1 \mid Z, X_0 = x_0. \]  

(1)
Pearl (2009), Ex. 11.3.3, claims the causal DAG above does not imply:

\[ Y(x_0, x_1) \perp \perp X_1 \mid Z, X_0 = x_0. \]  \hspace{1cm} (1)

The SWIG shows that (1) does hold; Pearl is incorrect.
Pearl (2009), Ex. 11.3.3, claims the causal DAG above does not imply:

\[ Y(x_0, x_1) \perp \perp X_1 \mid Z, X_0 = x_0. \]  

(1)

The SWIG shows that (1) does hold; Pearl is incorrect. Specifically, we see from the SWIG:

\[ Y(x_0, x_1) \perp \perp X_1(x_0) \mid Z(x_0), X_0, \]  

(2)

\[ \Rightarrow \quad Y(x_0, x_1) \perp \perp X_1(x_0) \mid Z(x_0), X_0 = x_0. \]  

(3)

This last condition is then equivalent to (1) via consistency. (Pearl infers a claim of Robins is false since if true then (1) would hold.)
Pearl (2009), Ex. 11.3.3, claims the causal DAG above does not imply:

\[ Y(x_0, x_1) \perp\!\!\!\!\!\!\perp X_1 \mid Z, X_0 = x_0. \]  

(1)

The SWIG shows that (1) does hold; Pearl is incorrect. Specifically, we see from the SWIG:

\[ Y(x_0, x_1) \perp\!\!\!\!\!\!\perp X_1(x_0) \mid Z(x_0), X_0, \]  

(2)

\[ \Rightarrow \quad Y(x_0, x_1) \perp\!\!\!\!\!\!\perp X_1(x_0) \mid Z(x_0), X_0 = x_0. \]  

(3)

This last condition is then equivalent to (1) via consistency. (Pearl infers a claim of Robins is false since if true then (1) would hold).
Twin network approach for the same problem

\[
\begin{align*}
X_0 & \xrightarrow{H} X_1 & \xrightarrow{Z} Y \\
X_0 & \xrightarrow{H} Z & \xrightarrow{Y} U \\
X_1 & \xrightarrow{Z} \xrightarrow{U_Y} Y \\
X_0 & \xrightarrow{H} X_1 & \xrightarrow{Z} \xrightarrow{U_Z} H(x_0, x_1) \\
& \xrightarrow{U_H} x_0 & \xrightarrow{H(x_0, x_1)} Z(x_0, x_1) & \xrightarrow{Y(x_0, x_1)} Y(x_0, x_1)
\end{align*}
\]

The twin network fails to reveal that \( Y(x_0, x_1) \perp \perp X_1 \mid Z, X_0 = x_0 \).

This 'extra' independence holds in spite of d-connection because (by consistency) when \( X_0 = x_0 \), then \( Z = Z(x_0) = Z(x_0, x_1) \).

Note that \( Y(x_0, x_1) \not\perp \perp X_1 \mid Z, X_0 \neq x_0 \).

Shpitser & Pearl (2007) introduce a pre-processing step to address this.
Twin network approach for the same problem

The twin network fails to reveal that \( Y(x_0, x_1) \perp X_1 \mid Z, X_0 = x_0 \).
Twin network approach for the same problem

The twin network fails to reveal that $Y(x_0, x_1) \perp \perp X_1 \mid Z, X_0 = x_0$. This ‘extra’ independence holds in spite of d-connection because (by consistency) when $X_0 = x_0$, then $Z = Z(x_0) = Z(x_0, x_1)$. Note that $Y(x_0, x_1) \not\perp \not\perp X_1 \mid Z, X_0 \neq x_0$. 

Shpitser & Pearl (2007) introduce a pre-processing step to address this.
Twin network approach for the same problem

The twin network fails to reveal that $Y(x_0, x_1) \perp \perp X_1 | Z, X_0 = x_0$. This ‘extra’ independence holds in spite of d-connection because (by consistency) when $X_0 = x_0$, then $Z = Z(x_0) = Z(x_0, x_1)$.

Note that $Y(x_0, x_1) \not\perp \not\perp X_1 | Z, X_0 \neq x_0$.

Shpitser & Pearl (2007) introduce a pre-processing step to address this.
Mediation graph (I)

Intervention on Z alone.

\[
\begin{align*}
\text{factorization:} & \quad P(Z, X(\tilde{z}), Y(\tilde{z})) = P(Z)P(X(\tilde{z}))P(Y(\tilde{z}) \mid X(\tilde{z})) \\
\text{modularity:} & \quad \begin{align*}
P(X(\tilde{z})=x) &= P(X=x \mid Z=\tilde{z}), \\
P(Y(\tilde{z})=y \mid X(\tilde{z})=x) &= P(Y=y \mid X=x, Z=\tilde{z}).
\end{align*}
\end{align*}
\]

d-separation gives:

\[Z \perp \perp X(\tilde{z}), Y(\tilde{z}).\]
Mediation graph (II)

Intervention on Z and X:

\[ \begin{array}{c}
Z \quad X \quad Y \\
\end{array} \quad \Rightarrow \quad \begin{array}{c}
Z \quad \tilde{z} \quad X(\tilde{z}) \quad \tilde{x} \\
\end{array} \quad \begin{array}{c}
Y(\tilde{x}, \tilde{z}) \\
\end{array} \]

factorization:

\[ P(Z, X(\tilde{z}), Y(\tilde{x}, \tilde{z})) = P(Z)P(X(\tilde{z}))P(Y(\tilde{x}, \tilde{z})) \]

modularity:

\[ P(X(\tilde{z})=\chi) = P(X=\chi \mid Z=\tilde{z}), \]
\[ P(Y(\tilde{x}, \tilde{z})=y) = P(Y=y \mid X=\tilde{x}, Z=\tilde{z}). \]

d-separation gives:

\[ Z \perp \perp X(\tilde{z}) \perp \perp Y(\tilde{x}, \tilde{z}) \]
No direct effect graph (I)

\[ \begin{array}{c}
Z & \rightarrow & X & \rightarrow & Y \\
\Rightarrow & & Z \vdash & X(\tilde{z}) & \rightarrow & Y(\tilde{z})
\end{array} \]

factorization:

\[ P(Z, X(\tilde{z}), Y(\tilde{z})) = P(Z)P(X(\tilde{z}))P(Y(\tilde{z}) \mid X(\tilde{z})) \]

modularity:

\[ P(X(\tilde{z}) = x) = P(X = x \mid Z = \tilde{z}), \]
\[ P(Y(\tilde{z}) = y \mid X(\tilde{z}) = x) = P(Y = y \mid X = x). \]

d-separation gives:

\[ Z \perp \perp X(\tilde{z}), Y(\tilde{z}) \]
No direct effect graph (II)

\[ Z \rightarrow X \rightarrow Y \implies Z \perp \perp X(\tilde{z}) \perp \perp Y(\tilde{x}) \]

factorization:
\[ P(Z, X(\tilde{z}), Y(\tilde{x})) = P(Z)P(X(\tilde{z}))P(Y(\tilde{x})) \]

modularity:
\[ P(X(\tilde{z}) = x) = P(X = x | Z = \tilde{z}), \]
\[ P(Y(\tilde{x}) = y) = P(Y = y | X = \tilde{x}). \]

d-separation gives:
\[ Z \perp \perp X(\tilde{z}) \perp \perp Y(\tilde{x}) \]
Here we can read directly from the template that $X \not\perp\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/X] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/x] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/Y(x)] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/H] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/X] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/x] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/Y(x)] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/H] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/X] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/x] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/Y(x)] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/H] /
\!
\!
\!
\!
\!
\!
\!
\!
\!
\![/X] /
\!
Here we can read directly from the template that

\[ X \perp Y(\tilde{x}) \mid L. \]

It follows that:

\[ P(Y(\tilde{x}) = y) = \sum_l P(\tilde{Y} = y \mid L = l, X = \tilde{x})P(L = l). \quad (4) \]
Contrast with approach via removing edges

This 'explains' why $L$ is sufficient to control confounding under the null (where $X$ has no effect on $Y$) but not under the alternative.
Contrast with approach via removing edges

This ‘explains’ why $L$ is sufficient to control confounding under the null (where $X$ has no effect on $Y$) but not under the alternative.
Adjusting for confounding

\[ X \perp Y(\tilde{x}) | L. \]

Proof of identification:

\[
P[Y(\tilde{x}) = y] = \sum_l P[Y(\tilde{x}) = y | L = l]P(L = l)
\]

\[= \sum_l P[Y(\tilde{x}) = y | L = l, X = \tilde{x}]P(L = l) \text{ indep} \]

\[= \sum_l P[Y = y | L = l, X = \tilde{x}]P(L = l) \text{ modularity} \]
Here we can read directly from the template that

\[ X \perp \! \! \! \perp Y(x) \mid L. \]
Here we can read directly from the template that

$$X \perp \perp Y(x) \mid L.$$
Sequentially randomized experiment (I)

A and C are treatments;

H is unobserved;

B is a time varying confounder;

D is the final response;

Treatment C is assigned randomly conditional on the observed history, A and B;

Want to know $P(D(\tilde{a}, \tilde{c}))$. 
Sequentially randomized experiment (II)

\[ \tilde{a} \xrightarrow[]{} A \xrightarrow[]{} B(\tilde{a}) \xrightarrow[]{} C(\tilde{a}) \xrightarrow[]{} D(\tilde{a}, \tilde{c}) \]

\[ \tilde{c} \]

\[ H \]

\text{d-separation:}

\[ A \perp \perp B(\tilde{a}), C(\tilde{a}), D(\tilde{a}, \tilde{c}) \]
\[ C(\tilde{a}) \perp \perp D(\tilde{a}, \tilde{c}) \mid B(\tilde{a}), A \]

General result of Robins (1986) then implies:

\[ P(D(\tilde{a}, \tilde{c}) = d) = \sum_b P(B = b \mid A = \tilde{a})P(D = d \mid A = \tilde{a}, B = b, C = \tilde{c}). \]
Dynamic regimes

\[ P(Y(g)) \] is identified.
P(\(Y(g)\)) is not identified.
We saw earlier that the causal DAG $X \to Y$ implied:

$$X \perp \perp Y(x_0) \quad \text{and} \quad X \perp \perp Y(x_1)$$

However, *joint* independence relations such as:

$$X \perp Y(x_0), Y(x_1)$$

never follow from our SWIG transformation: There is no way via node-splitting to construct a graph with both $Y(x_0)$, and $Y(x_1)$. This has important consequences for the identification of direct effects.
Mediation graph (II)

Intervention on $Z$ and $X$:

\[ \begin{align*}
  &Z \xrightarrow{} X \xrightarrow{} Y \\
  \Rightarrow &\quad Z \xrightarrow{} \tilde{Z} \xrightarrow{} X(\tilde{Z}) \xrightarrow{} \tilde{X} \xrightarrow{} Y(\tilde{X}, \tilde{Z})
\end{align*} \]

$d$-separation gives:

\[ Z \perp \perp X(\tilde{Z}) \perp \perp Y(\tilde{X}, \tilde{Z}) \]

Pearl associates additional independence relations with this graph:

\[ \begin{align*}
  Y(z_1, x) &\perp \perp X(z_0), Z \\
  Y(z_0, x) &\perp \perp X(z_1), Z
\end{align*} \]

Extra assumptions imply identification of the PDE

\[ E[Y(z_1, X(z_0)) - Y(z_0, X(z_0))] (!) \]
Contrast with NPSEM with independent errors

Pearl (2000, 2009) approach to unification is based on an non-parametric structural equation model with independent errors. This has the consequence that there are many ‘cross-world’ counterfactual independence assumptions:

<table>
<thead>
<tr>
<th>No. Actual Vars.</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim. P(V)</td>
<td>3</td>
<td>7</td>
<td>15</td>
<td>$2^K - 1$</td>
</tr>
<tr>
<td>No. Counterfactual Vars.</td>
<td>3</td>
<td>7</td>
<td>15</td>
<td>$2^K - 1$</td>
</tr>
<tr>
<td>Dim. Counterfactual Dist.</td>
<td>7</td>
<td>127</td>
<td>32767</td>
<td>$2^{(2^K - 1)} - 1$</td>
</tr>
<tr>
<td>Dim. FFRCISTG</td>
<td>5</td>
<td>113</td>
<td>32697</td>
<td>$(2^{(2^K - 1)} - 1) - \sum_{j=1}^{K-1} (4^j - 2^j)$</td>
</tr>
<tr>
<td>Dim. NPSEM-IE</td>
<td>4</td>
<td>19</td>
<td>274</td>
<td>$\sum_{j=0}^{K-1} (2^{2^j} - 1)$</td>
</tr>
<tr>
<td>No. untestable indep. constrnts in NPSEM-IE</td>
<td>1</td>
<td>94</td>
<td>32423</td>
<td>$O(2^{2^K-2})$</td>
</tr>
</tbody>
</table>

Table: Dimensions of counterfactual models associated with complete graphs with binary variables.
Summary

- A simple approach to unifying graphs and counterfactuals via node-splitting
Summary

- A simple approach to unifying graphs and counterfactuals via node-splitting
- The approach works via linking the factorizations associated with the two graphs
Summary

- A simple approach to unifying graphs and counterfactuals via node-splitting
- The approach works via linking the factorizations associated with the two graphs
- The approach provides a language that allows counterfactual and graphical people to communicate
Summary

- A simple approach to unifying graphs and counterfactuals via node-splitting
- The approach works via linking the factorizations associated with the two graphs
- The approach provides a language that allows counterfactual and graphical people to communicate
- The approach leads to many fewer untestable independence assumptions than in the NPSEM-IE approach of Pearl.
Summary

- A simple approach to unifying graphs and counterfactuals via node-splitting
- The approach works via linking the factorizations associated with the two graphs
- The approach provides a language that allows counterfactual and graphical people to communicate
- The approach leads to many fewer untestable independence assumptions than in the NPSEM-IE approach of Pearl.
- The approach also provides a way to combine information on the absence of individual and population level direct effects.
Thank You!