LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY

Sofia Triantafillou and Ioannis Tsamardinos

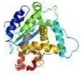
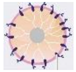
Computer Science Department, University of Crete

June 2016

OUTLINE

1. Integrative causal discovery
   i. Motivation.
   ii. Causal models.
   iii. m-separation.
   iv. Reverse engineering causal models (single data set).
   v. Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.
   vi. Modeling interventions/selection.
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

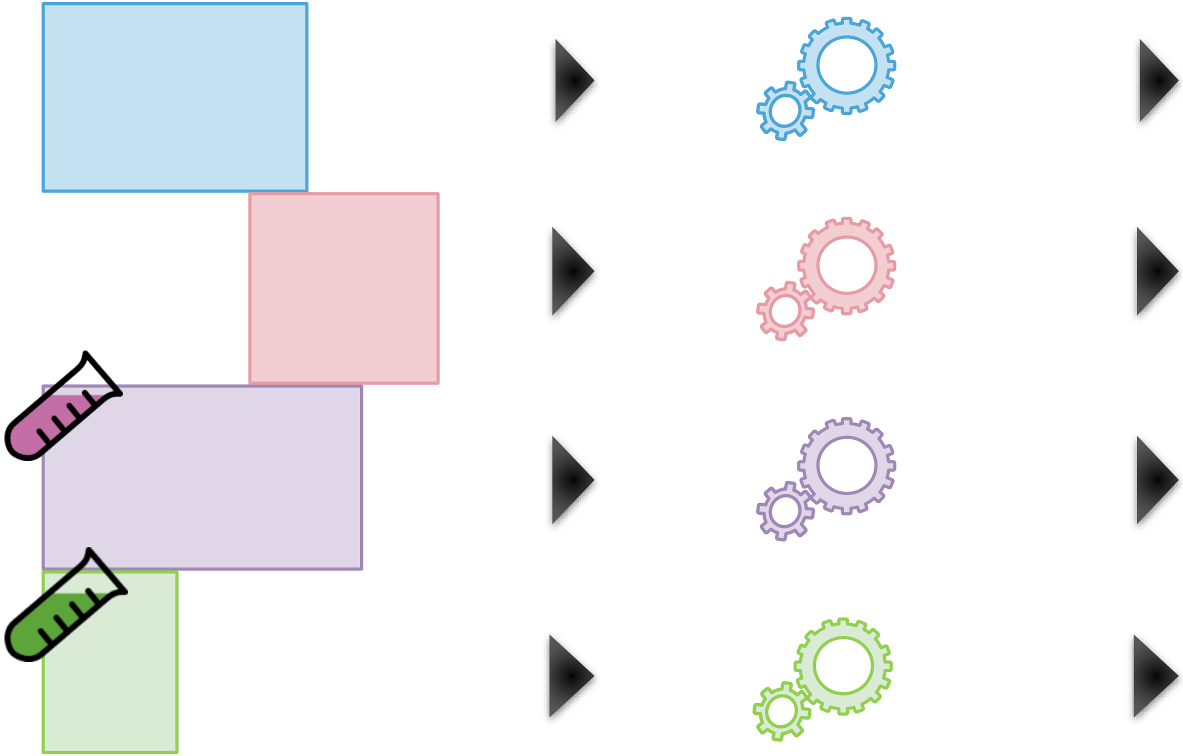
HETEROGENEOUS DATA SETS MEASURING THE SAME SYSTEM UNDER STUDY

Variables / Study	Thrombosis 	Contraceptives 	Protein C 	Breast Cancer 	Protein Y 	Protein Z 
1 observational data	Yes	No	10.5	Yes	-	-
	No	Yes	5.3	No	-	-
					-	-
	No	Yes	0.01	No	-	-
2 observational data	-	-	-	Yes	0.03	9.3
	-	-	-			
	-	-	-	No	3.4	22.2
3 experimental data	No	No	0 (Control)	No	3.4	-
	Yes	No	0 (Control)	Yes	2.2	-
					-	-
	Yes	Yes	5.0 (Treat.)	Yes	7.1	-
	No	Yes	5.0 (Treat.)	No	8.9	-
4 experimental data	No	No (Ctrl)	-	-	-	-
	No	No (Ctrl)	-	-	-	-
			-	-	-	-
	Yes	Yes(Treat)	-	-	-	-

ISOLATED ANALYSIS

Analyze data

Publish results



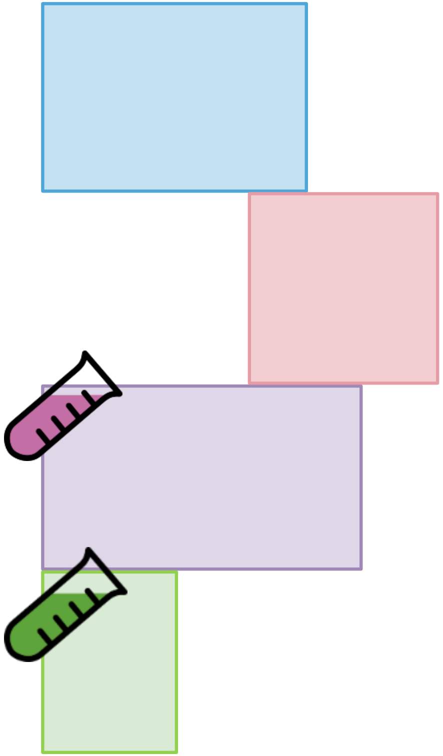
"…The use of contraceptives is correlated with Thrombosis, negatively correlated with Breast Cancer and levels of Protein E …"

"…Protein E is a risk factor for Breast Cancer…"

"…Drugs reducing protein C reduced the probability of Breast Cancer and lowered the levels of Protein E…"

"… In the randomized control trial, women taking contraceptives had 30% more chances of being diagnosed with thrombosis … "

INTEGRATIVE CAUSAL ANALYSIS

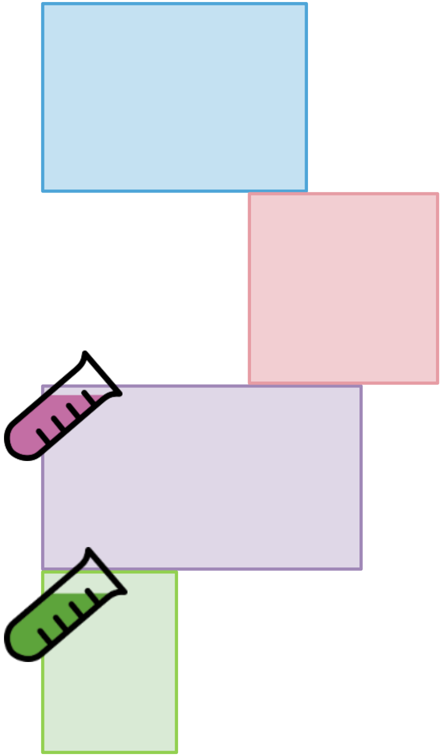


Data can not be pooled together:

Missing variables cannot be treated as missing values.

They come from **different experimental/sampling conditions (different distributions).**

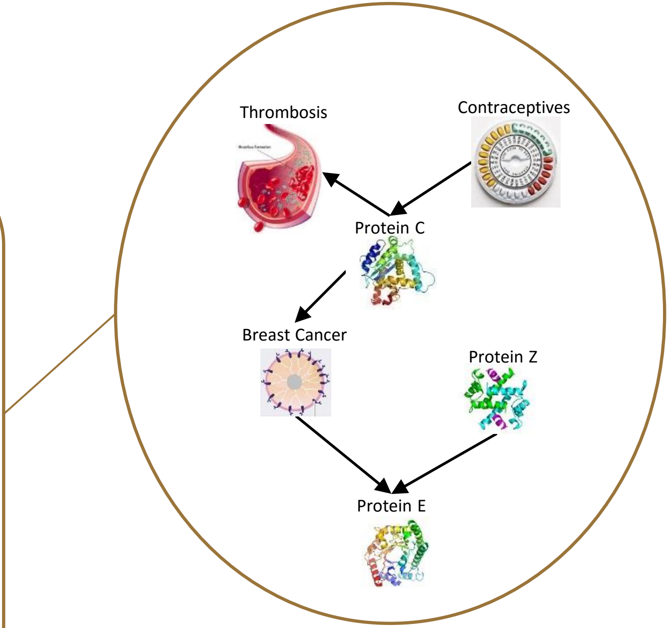
INTEGRATIVE CAUSAL ANALYSIS



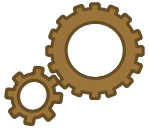
Data can not be pooled together:

Missing variables cannot be treated as missing values.

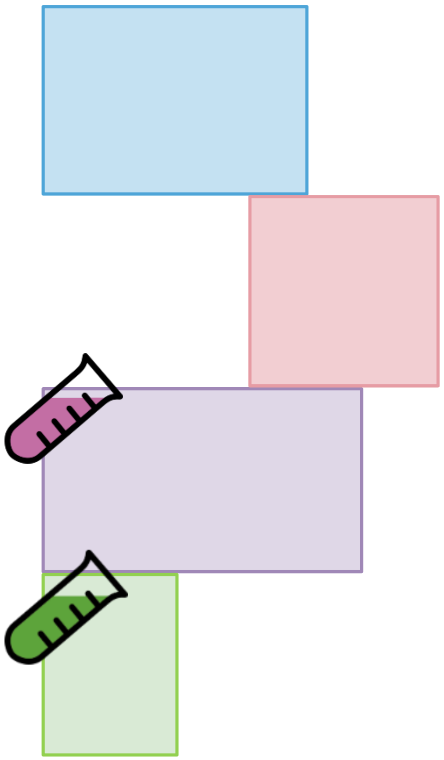
They come from **different experimental/sampling conditions (different distributions).**



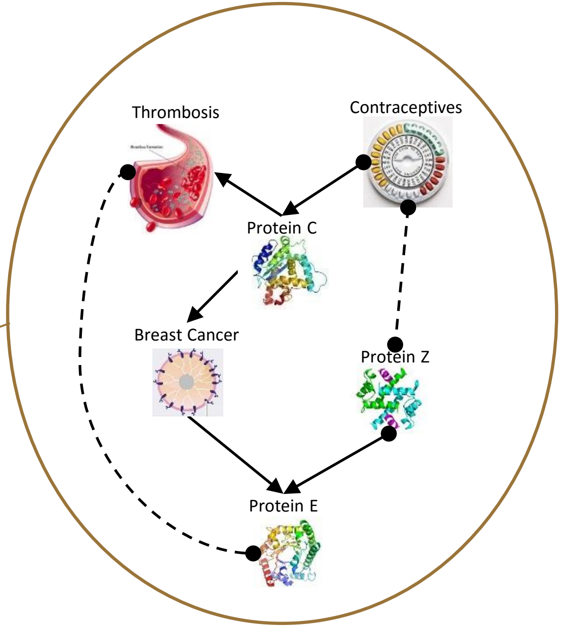
Data come from the **same causal mechanism**.



INTEGRATIVE CAUSAL ANALYSIS



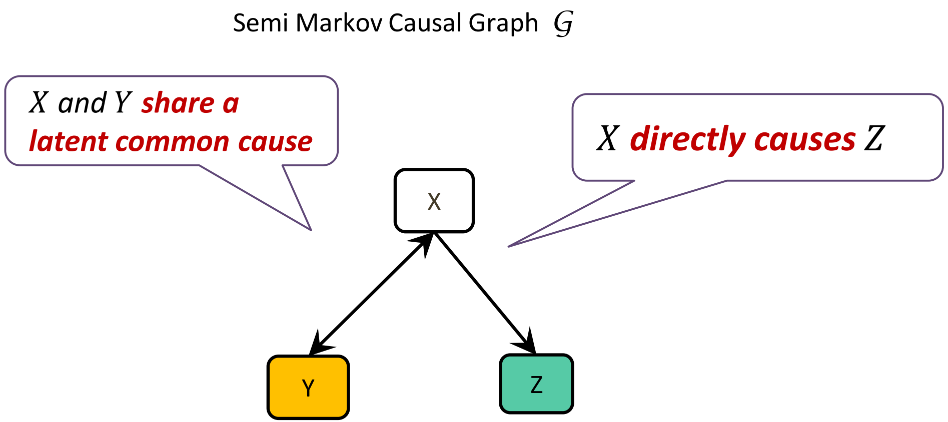
Identify the **causal graphs** that **simultaneously fit** all data.



OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. Causal models.
   iii. m-separation.
   iv. Reverse engineering causal models (single data set).
   v. Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.
   vi. Modeling interventions/selection.
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

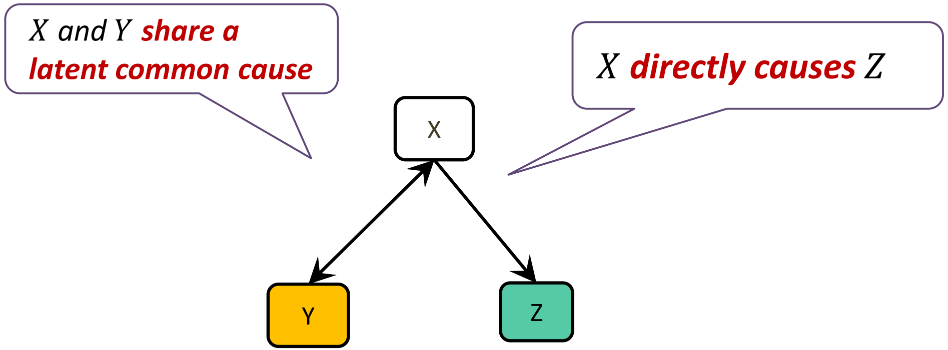
SEMI MARKOV CAUSAL GRAPHS



- Directed edges represent direct causal relationships.
- Bi-directed edges represent confounding (**latent confounders)**.
- Both types of edges allowed for a single pair of variables.
- No directed cycles (no causal feedback).

SEMI MARKOV CAUSAL GRAPHS

Semi Markov Causal Graph $\mathcal{G}$



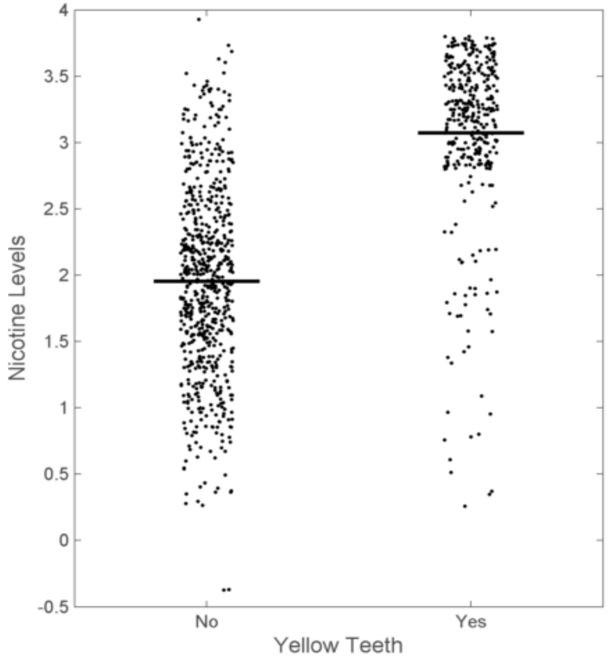
Joint Probability Distribution $\mathcal{P}$

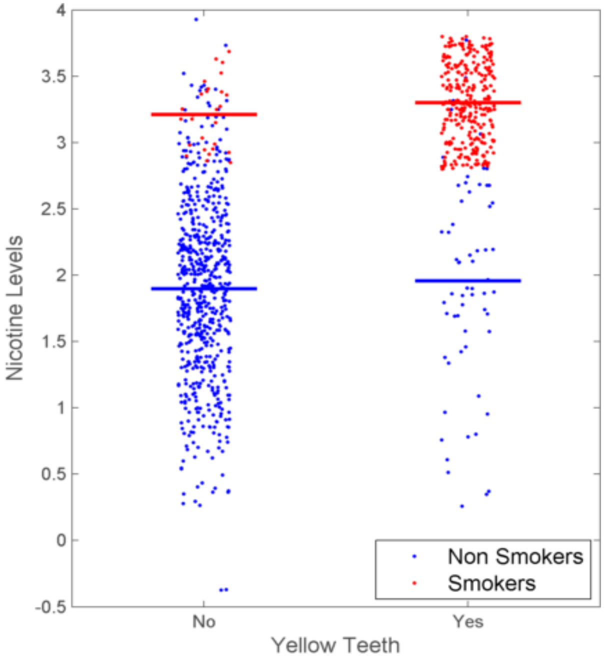
		Z	
X	Y	Yes	No
Yes	Yes	0,01	0,04
Yes	No	0,01	0,04
No	Yes	0,000045	0,044955
No	No	0,000855	0,854145

- Directed edges represent direct causal relationships.
- Bi-directed edges represent confounding (**latent confounders**).
- Both types of edges allowed for a single pair of variables.
- No directed cycles (no causal feedback).

- Joint probability distribution entails conditional (in) dependencies.
- $Ind(X, Y|\boldsymbol{Z}): P(X|Y, \boldsymbol{Z}) = P(X|\boldsymbol{Z})$
- $Dep(X, Y|\boldsymbol{Z}): P(X|Y, \boldsymbol{Z}) \neq P(X|\boldsymbol{Z})$

EXAMPLE OF CONDITIONAL (IN) DEPENDENCE





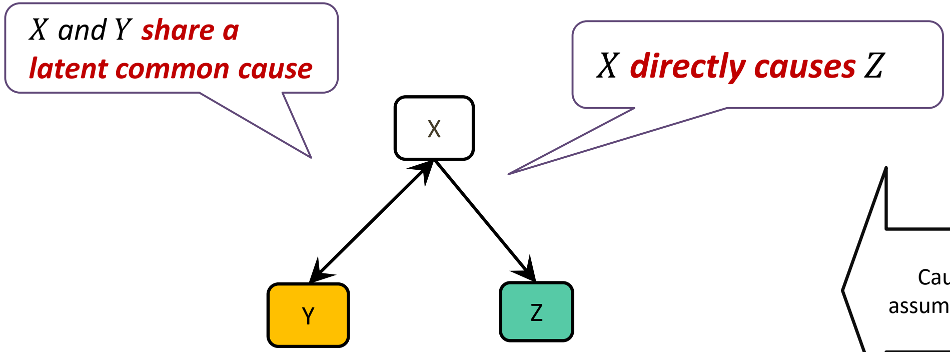
Data measuring:
Smoking,
Yellow Teeth,
Nicotine Levels.

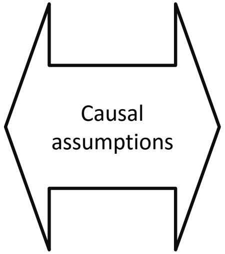
P(Nicotine Levels | Yellow Teeth) ≠ P(Nicotine Levels)

P(Nicotine Levels | Smoking, Yellow Teeth) = P(Nicotine Levels|Smoking)

SEMI MARKOV CAUSAL GRAPHS

Semi Markov Causal Graph $\mathcal{G}$





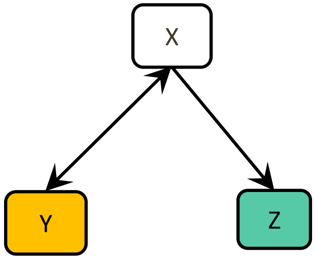
Joint Probability Distribution $\mathcal{P}$

		Z	
X	Y	Yes	No
Yes	Yes	0,01	0,04
Yes	No	0,01	0,04
No	Yes	0,000045	0,044955
No	No	0,000855	0,854145

- Directed edges represent direct causal relationships.
- Bi-directed edges represent confounding (**latent confounders**).
- Both types of edges allowed for a single pair of variables.
- No directed cycles (no causal feedback).

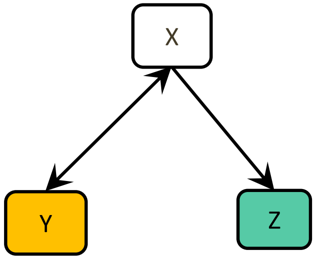
- Joint probability distribution entails conditional (in) dependencies.
- $Ind(X,Y|\boldsymbol{Z}): P(X|Y,\boldsymbol{Z}) = P(X|\boldsymbol{Z})$
- $Dep(X,Y|\boldsymbol{Z}): P(X|Y,\boldsymbol{Z}) \neq P(X|\boldsymbol{Z})$

CAUSAL ASSUMPTIONS



Causal Markov Assumption:
Every variable is independent of its **non-effects** given its **direct causes**.

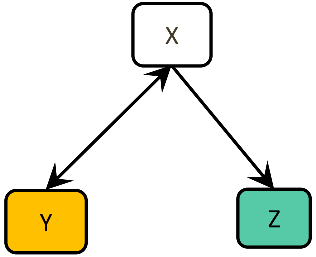
CAUSAL ASSUMPTIONS



$$Ind(Y, Z \mid X)$$

Causal Markov Assumption:
Every variable is independent of its **non-effects** given its **direct causes**.

CAUSAL ASSUMPTIONS

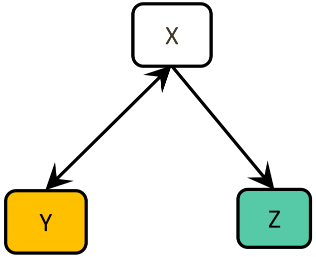


$$Ind(Y, Z \mid X)$$

Causal Markov Assumption:
Every variable is independent of its **non-effects** given its **direct causes.**

Causal Faithfulness Assumption:
Independences stem **only** from the causal structure, **not the parameterization** of the distribution.

CAUSAL ASSUMPTIONS



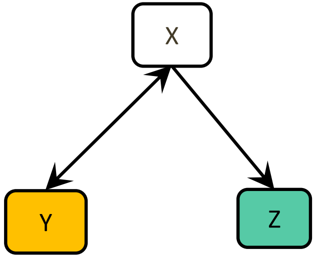
$Ind(Y, Z \mid X)$

$Dep(Y, Z \mid \emptyset)$
$Dep(X, Z \mid \emptyset)$
$Dep(X, Z \mid Y)$
$Dep(Y, X \mid \emptyset)$
$Dep(Y, X \mid Z)$

Causal Markov Assumption:
Every variable is independent of its **non-effects** given its **direct causes**.

Causal Faithfulness Assumption:
Independences stem **only** from the causal structure, **not the parameterization** of the distribution.

CAUSAL ASSUMPTIONS



$Ind(Y, Z \mid X)$

$Dep(Y, Z \mid \emptyset)$
$Dep(X, Z \mid \emptyset)$
$Dep(X, Z \mid Y)$
$Dep(Y, X \mid \emptyset)$
$Dep(Y, X \mid Z)$

Causal Markov Assumption:
Every variable is independent of its **non-effects** given its **direct causes.**

Causal Faithfulness Assumption:
Independences stem **only** from the causal structure, **not the parameterization** of the distribution.

All independencies in the joint probability distribution can be identified in G using the graphical criterion of **m-separation.**

OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. **m-separation.**
   iv. Reverse engineering causal models (single data set).
   v. Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.
   vi. Modeling interventions/selection.
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

m-SEPARATION

A path $X_1, \ldots, X_n$ between X_1 and X_n is **m-connecting given $\boldsymbol{V}$** if for every triple $\langle X_{i-1}, X_i, X_{i+1} \rangle$ on the path:

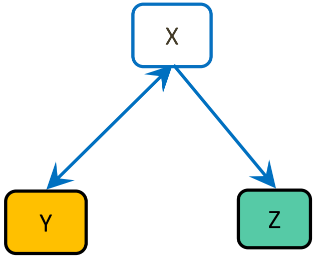
- If $\mathrm{X}_{\mathrm{i}-1} \ast\!\!\rightarrow X_i \leftarrow\!\!\ast X_{i+1}$ (colliding triplet),
  X_i or one of its descendants $\in \boldsymbol{V}$
- Otherwise, $X_i \notin \boldsymbol{V}$

m-connecting path **=>** information flow **=> dependence**

No **m-connecting** path **=>** no information flow **=> independence** (m-separation)

Colliders $\mathrm{X}_{\mathrm{i}-1} \ast\!\!\rightarrow X_i \leftarrow\!\!\ast X_{i+1}$ are **special** and create an asymmetry that will allow us to orient causal direction.

m-SEPARATION

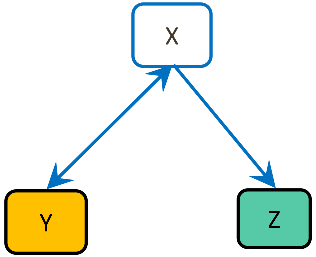


$$Y \leftrightarrow X \longrightarrow Z$$

is m-connecting given $\emptyset$

$$\Leftrightarrow Dep(Y, Z|\emptyset)$$

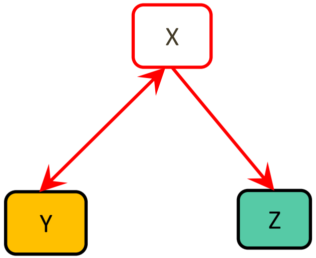
m-SEPARATION



Y $\leftrightarrow$ X $\longrightarrow$ Z

is m-connecting given $\emptyset$

$\Leftrightarrow Dep(Y, Z|\emptyset)$

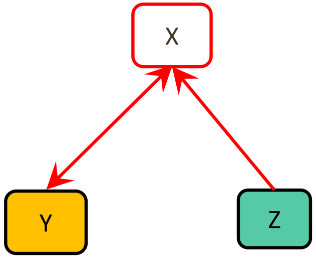


Y $\leftrightarrow$ X $\longrightarrow$ Z

is NOT m-connecting given X

$\Leftrightarrow Ind(Y, Z|X)$

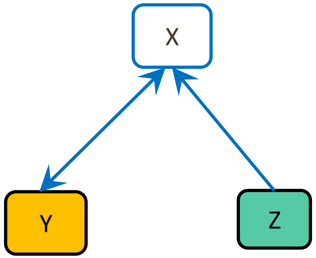
m-SEPARATION



Y $\leftrightarrow$ X $\leftarrow$ Z

is NOT m-connecting given $\emptyset$

$\Leftrightarrow Ind(Y, Z|\emptyset)$

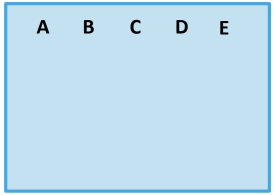


Y $\leftrightarrow$ X $\leftarrow$ Z

is m-connecting given X

$\Leftrightarrow Dep(Y, Z|X)$

CAUSAL MODELLING

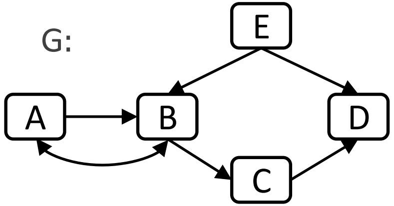


Data set D measuring a set of variables

$A, B \mid E, C$	Ind
$A, B \mid \emptyset$	Dep
...	...
$E, C \mid A, B, C$	Dep

Conditional (in)dependencies (expected) in the joint probability distribution

G:

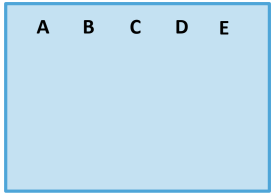


Paths (m-separations/connections) in the causal graph

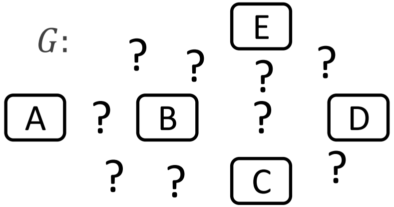
OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. **Reverse engineering causal models (single data set).**
   v. Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.
   vi. Modeling interventions/selection.
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

REVERSE ENGINEERING

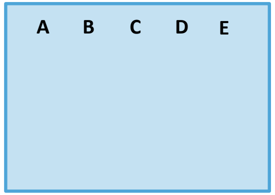


Data set D
measuring a
set of variables



causal graph?

REVERSE ENGINEERING

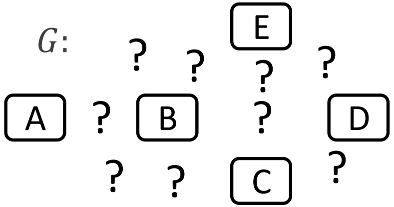


Data set D
measuring a
set of variables

$A, B \mid E, C$	Ind
$A, B \mid \emptyset$	Dep
…	…
$E, C \mid A, B, C$	Dep

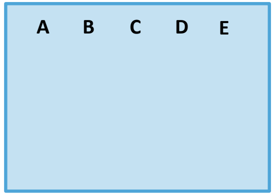
Find the (in)dependencies
using statistical tests.

??



causal graph?

REVERSE ENGINEERING

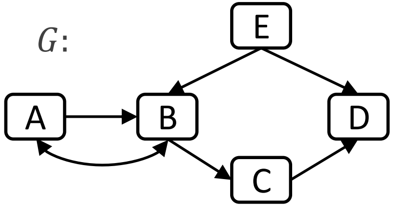


Data set D measuring a set of variables

$A, B \mid E, C$	Ind
$A, B \mid \emptyset$	Dep
…	…
$E, C \mid A, B, C$	Dep

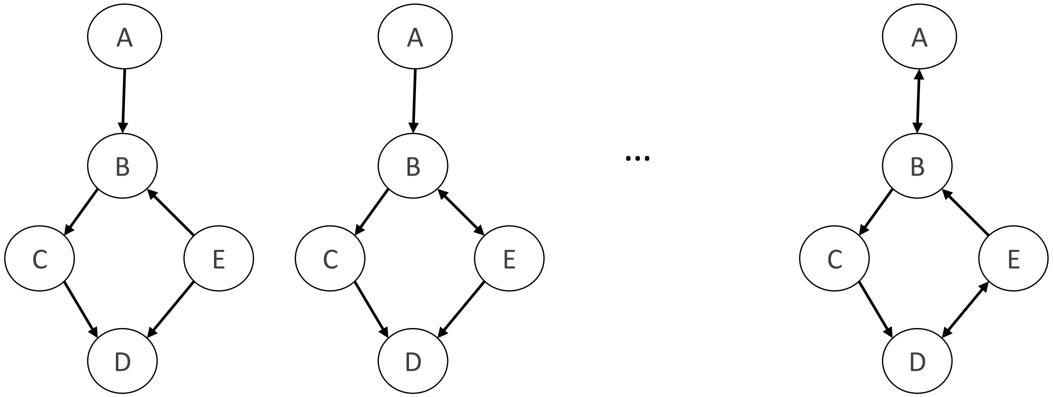
Find the (in)dependencies using statistical tests.

G:



Find a graph that satisfies the implied m-connections/separations.

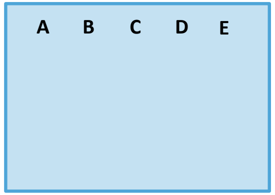
MARKOV EQUIVALENCE



$A, B \mid E, C$	Ind
$A, B \mid \emptyset$	Dep
...	...
$E, C \mid A, B, C$	Dep

- More than one graphs entail the same set of conditional independencies.
- The graphs have some common features (edges/orientations).
- For some types of causal graphs, Markov equivalence classes share the same skeleton.
    - not semi-Markov causal graphs

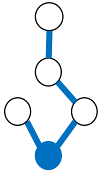
CAUSAL DISCOVERY



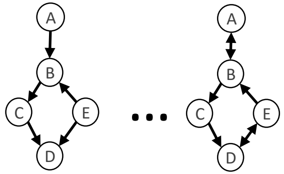
Data

$A,B\|E,C$	Ind
$A,B\|\emptyset$	Dep
...	...
$E,C\|A,B,C$	Dep

(In)dependencies



paths



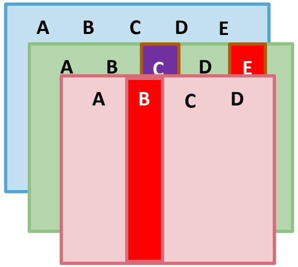
Causal graph(s)

Sound and complete algorithms take as input a data set and output a summary of all the graphs that satisfy all identified conditional independencies.

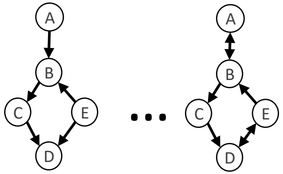
OUTLINE

1. Integrative causal discovery
   - ~~i. Motivation.~~
   - ~~ii. Causal models.~~
   - ~~iii. m-separation.~~
   - ~~iv. Reverse engineering causal models (single data set).~~
   - **v. Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.**
   - vi. Modeling interventions/selection.
2. Logic-based causal discovery
   - i. Converting path constraints to logic formulae.
   - ii. Problem complexity.
   - iii. Conflict resolution.
   - iv. Existing algorithms.
   - v. Reasoning with logic based causal discovery.
   - vi. Non-trivial inferences-validation.

INTEGRATIVE CAUSAL DISCOVERY

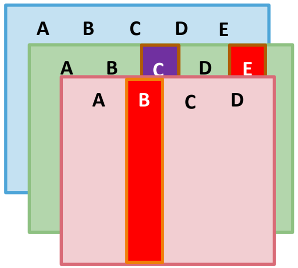


Data sets measuring overlapping variable sets under intervention/selection.

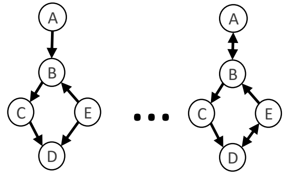


Causal graph(s) that simultaneously fit all data.

INTEGRATIVE CAUSAL DISCOVERY



Data sets measuring overlapping variable sets under intervention/selection.



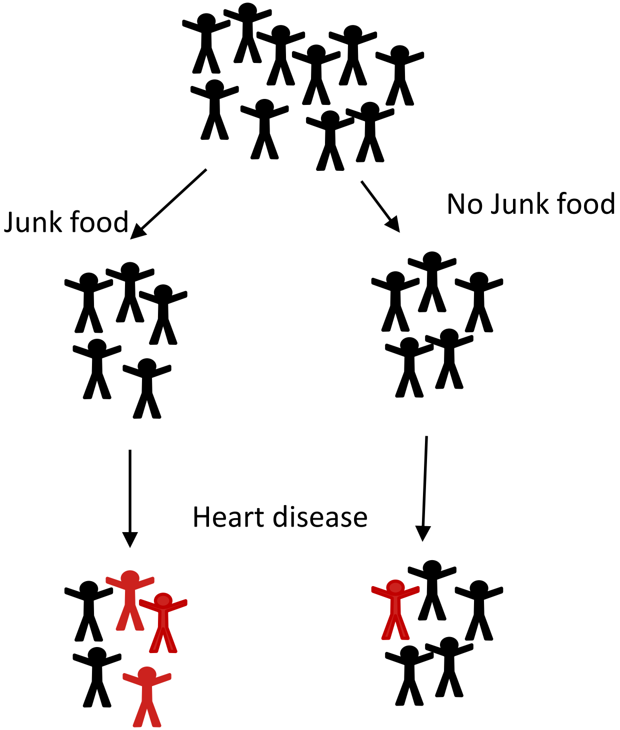
Causal graph(s) that simultaneously fit all data.

- Every data set imposes some constraints.
- Observational data impose m-separation/m-connection constraints on the candidate graph.
- Experimental data?
- Data sampled under selection?

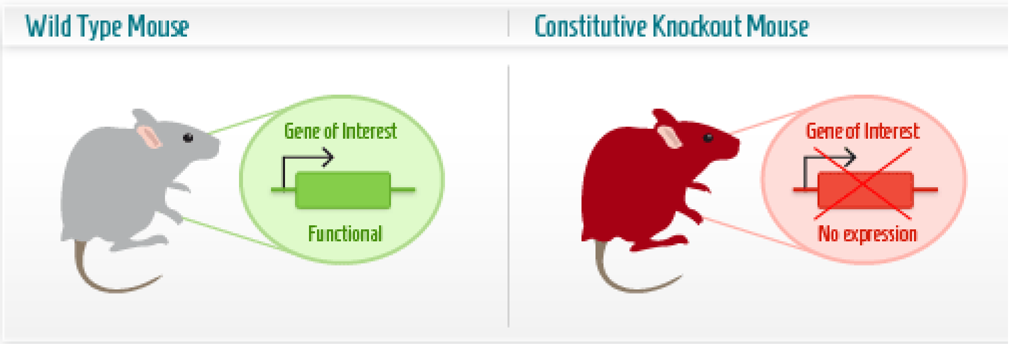
OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. Modeling interventions/selection.
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

INTERVENTIONS (MANIPULATIONS)

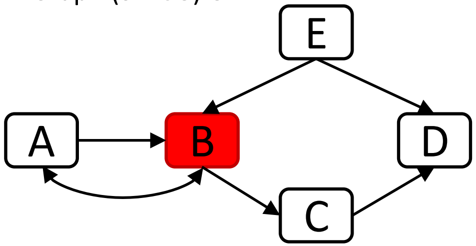


Values of the manipulated variable are **set** solely **by the intervention procedure**
e.g. value of a knockout gene in a knockout mice is set to zero; randomized variable in a randomized control trial.

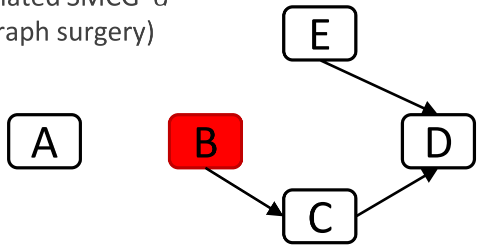


INTERVENTIONS

Graph (SMCG) *G*

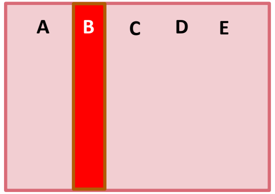


Manipulated SMCG G^B
(after graph surgery)



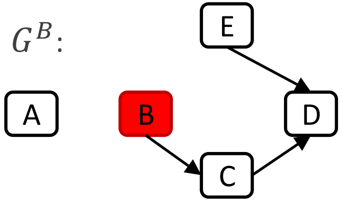
- If you know the causal model, you can model interventions.
- Values of B are **set** solely **by the intervention procedure:** If you know direct causal relations, **remove all edges into the manipulated variable.**
- This procedure is called graph surgery.
  - The resulting graph is called the **manipulated graph** (symb. G^B)

CAUSAL DISCOVERY WITH INTERVENTIONS



$A, B \mid E, C$	Ind
$A, B \mid \emptyset$	Dep
...	...
$E, C \mid A, B, C$	Dep

G^B:



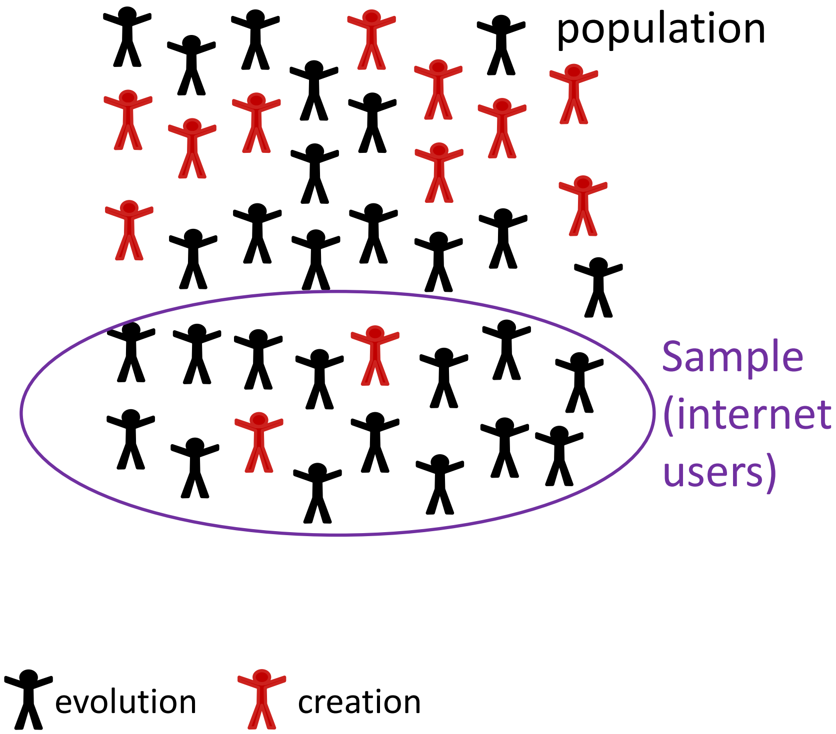
$\nexists$ *m-connecting path from A to D given $\emptyset$ in G^B*
$\nexists$ *m-connecting path from A to D given B in G^B*
$\vdots$
$\nexists$ *m-connecting path from A to D given B, C in G^B*
$\vdots$
$\exists$ *m-connecting path from B to C given $\emptyset$ in G^B*

Dataset D_i measuring a subset of variables, some of which are manipulated

Conditional independencies in D_i

Path constraints on the causal graph after **manipulation**

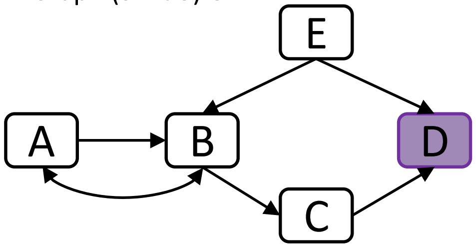
SELECTION BIAS



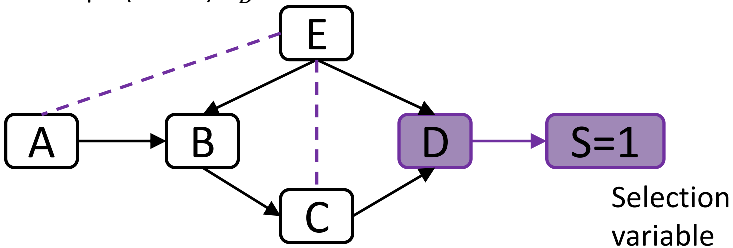
- Samples are selected based on the value of one of your variables.
- e.g. you perform your study in a specific region/on the internet; case-control study for a rare disease.

SELECTION BIAS IN CAUSAL MODELS

Graph (SMCG) G

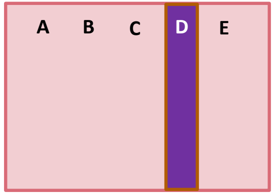


Selected Graph (SMCG) G_D

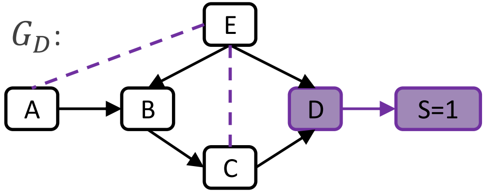


- If you know the causal model, you can model selection bias.
- Samples are selected based on the value of D; The value of D directly affects the probability of being selected.
- S is a child of D, S=1 for all your samples.
- Selected graph, symb. G_D

CAUSAL DISCOVERY WITH SELECTION BIAS



$A, B \mid E, C, S{=}1$	**Ind**
$A, B \mid S{=}1$	**Dep**
…	…
$E, C \mid A, B, D, S{=}1$	**Dep**



$\nexists$ *m-connecting path from A to D given* $\emptyset$ *in* G_D

$\nexists$ *m-connecting path from A to D given B in* G_D

$\vdots$

$\nexists$ *m-connecting path from A to D given* B, *C in* G_D

$\vdots$

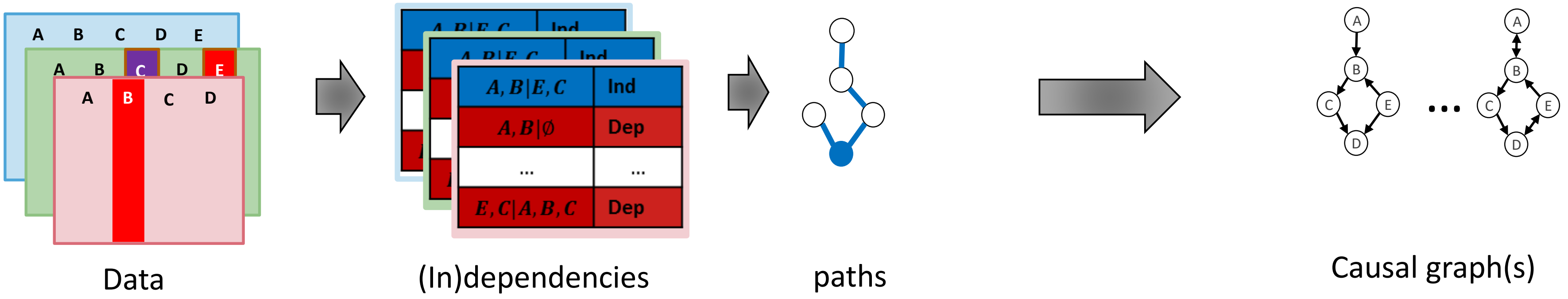
$\exists$ *m-connecting path from B to C given* $\emptyset$ *in* G_D

Dataset D_i measuring a subset of variables, some of which are selected upon

Conditional independencies in D_i

Path constraints on the underlying causal graph after **selection**

INTEGRATIVE CAUSAL DISCOVERY

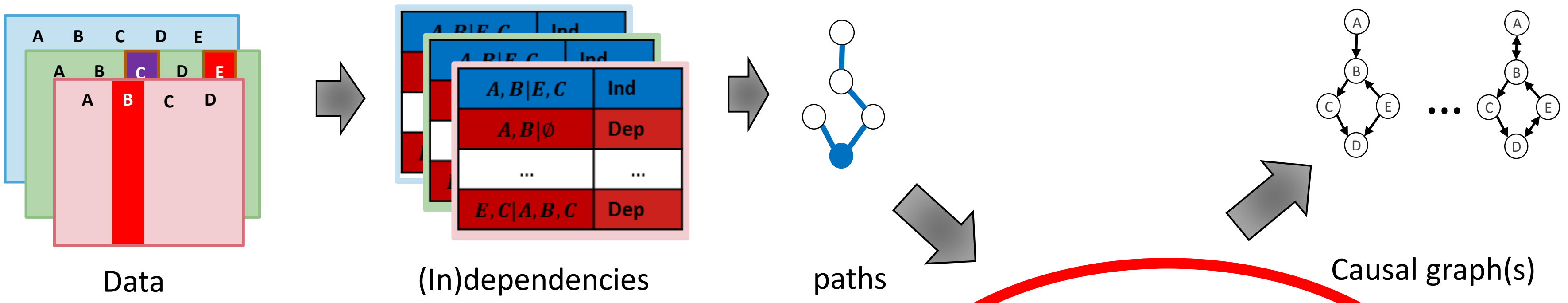


- Every data set imposes some constraints.
- Observational data impose path constraints on the candidate graph.
- Experimental data impose path constraints on the candidate graph after manipulation.
- Data sampled under selection impose path constraints on the candidate graph after selection.
- Easily handles overlapping variable sets
  - Each study imposes constraints on the observed variables.

OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   i. Converting path constraints to logic formulae.
   ii. Problem complexity.
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY



Convert to logic formula!

Variables of the formula correspond to graph features (edges, orientations).

Truth setting assignments encode graphs that satisfy all path constraints after manipulation/selection.

Logic encoding Φ of path constraints in the causal graph

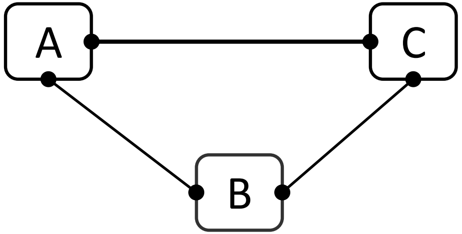
$$[E_{A\to D} \vee [E_{A\to B} \wedge E_{B\to D}] \vee [E_{A\to C} \wedge E_{C\to D}] \vee \cdots]$$

$$\vdots$$

$$[E_{A\to C} \vee [E_{A\to B} \wedge E_{B\to C}] \vee [E_{A\leftrightarrow C} \wedge E_{C\to D}] \vee \ldots]$$

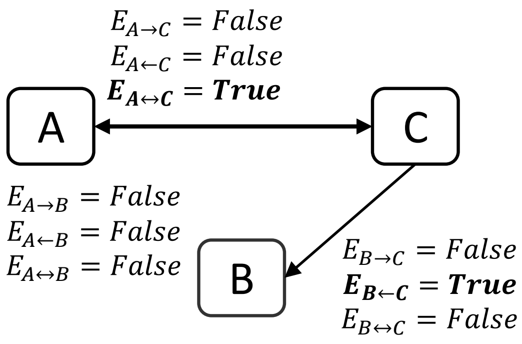
CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



CONVERSION TO LOGIC FORMULA: EXAMPLE

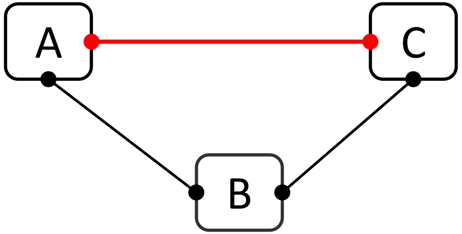
- Edges of the graph as Boolean variables
  - $E_{A \to B} = true$ if $A \to B$ in G, $false$ otherwise.
  - $E_{A \leftarrow B} = true$ if $A \to B$ in G, $false$ otherwise.
  - $E_{A \leftrightarrow B} = true$ if $A \leftrightarrow B$ in G, $false$ otherwise.
- $E_{A \to B}$ and $E_{A \leftarrow B}$ are mutually exclusive: $\neg E_{A \to B} \vee \neg E_{A \leftarrow B}$.



Assignments to Boolean variables correspond to graphs.

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.

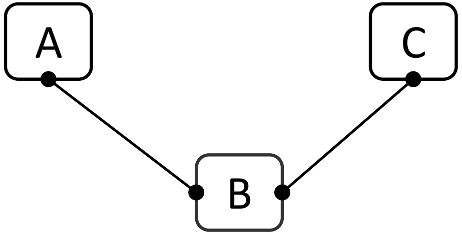


A-C does not exist

$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



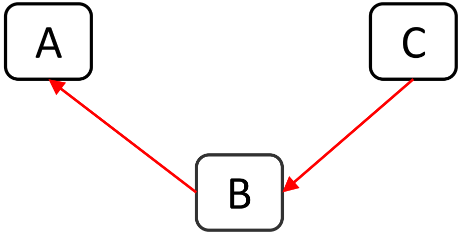
A-C does not exist

$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

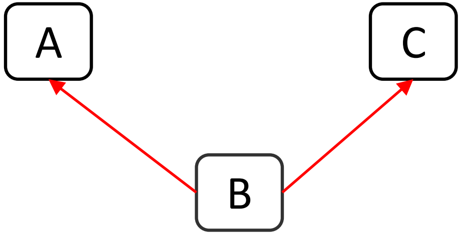
$\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

$\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

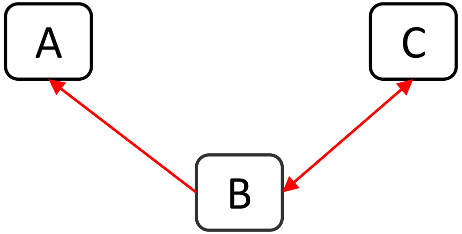
A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$

$\neg(E_{A\leftarrow B} \wedge E_{B\to C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

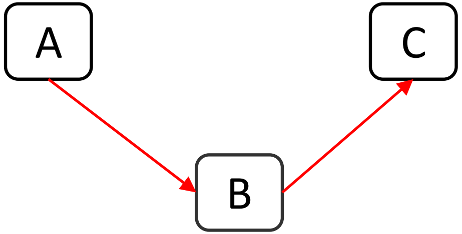
$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

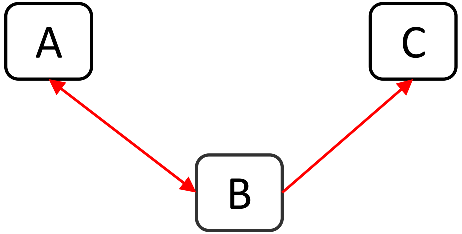
$$\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$$

A-B-C is not m-connecting

$$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$$
$$\neg(E_{A\leftarrow B} \wedge E_{B\to C})$$
$$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$$
$$\neg(E_{A\to B} \wedge E_{B\to C})$$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

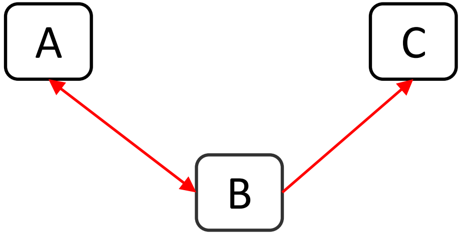
$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$
$\neg(E_{A\rightarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\rightarrow C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE

- Suppose you know nothing about the causal structure G of A, B, C.
- In an observational data set, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G.



A-C does not exist

$$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$$

A-B-C is not m-connecting

$$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$$
$$\neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C})$$
$$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$$
$$\neg(E_{A\rightarrow B} \wedge E_{B\rightarrow C})$$
$$\neg(E_{A\leftrightarrow B} \wedge E_{B\rightarrow C})$$

Logic formula:

$$\begin{gathered}(\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}) \wedge \\ \neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C}) \wedge \\ \neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C}) \wedge \\ \neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C}) \wedge \\ \neg(E_{A\rightarrow B} \wedge E_{B\rightarrow C}) \wedge \\ \neg(E_{A\leftrightarrow B} \wedge E_{B\rightarrow C})\end{gathered}$$

CONVERSION TO LOGIC FORMULA: EXAMPLE

Logic formula: TRUE

$$(\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}) \wedge$$

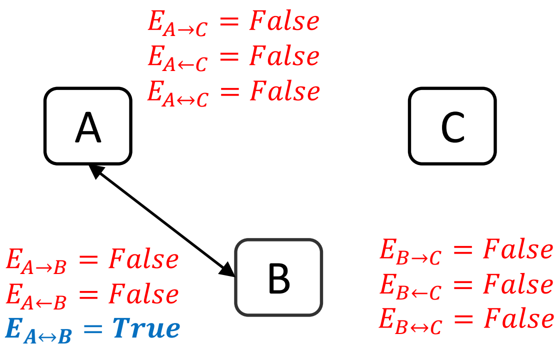
$$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C}) \wedge$$

$$\neg(E_{A\leftarrow B} \wedge E_{B\to C}) \wedge$$

$$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C}) \wedge$$

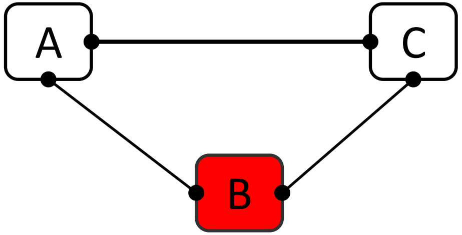
$$\neg(E_{A\to B} \wedge E_{B\to C}) \wedge$$

$$\neg(E_{A\leftrightarrow B} \wedge E_{B\to C})$$



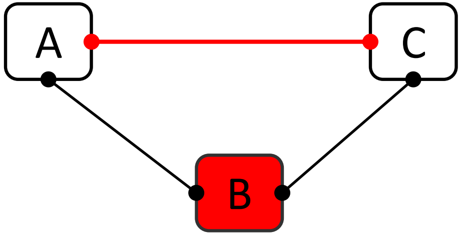
CONVERSION TO LOGIC FORMULA: EXAMPLE (INTERVENTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set where B is manipulated, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G^B.



CONVERSION TO LOGIC FORMULA: EXAMPLE (INTERVENTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set where B is manipulated, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G^B.

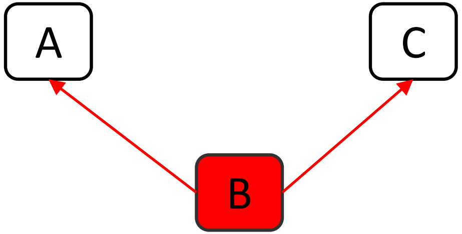


A-C does not exist

$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

CONVERSION TO LOGIC FORMULA: EXAMPLE (INTERVENTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set where B is manipulated, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G^B.



A-C does not exist

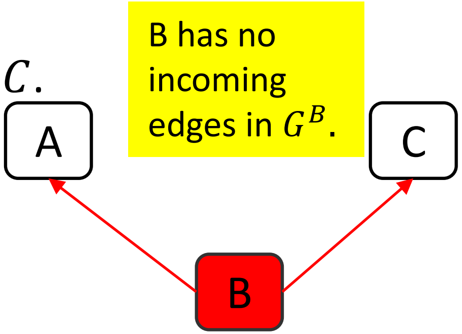
$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{B\rightarrow A} \wedge E_{B\rightarrow C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE (INTERVENTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set where B is manipulated, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms: $\nexists$ m-connecting path between A and C given $\emptyset$ in G^B.



A-C does not exist

$\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{B\to A} \wedge E_{B\to C})$

Logic formula:

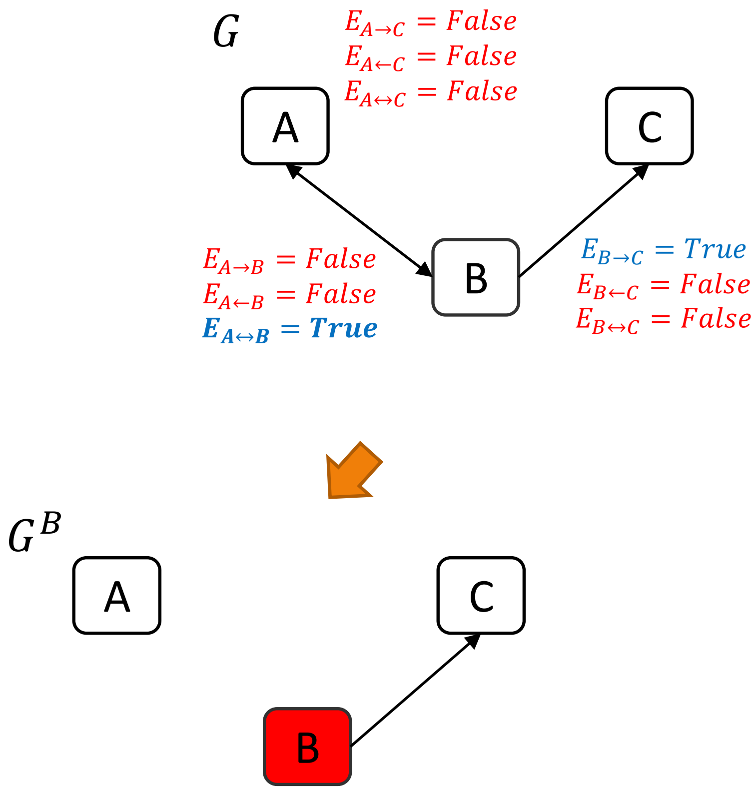
$$(\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}) \wedge$$
$$\neg(E_{A\leftarrow B} \wedge E_{B\to C})$$

CONVERSION TO LOGIC FORMULA: EXAMPLE

Logic formula: TRUE

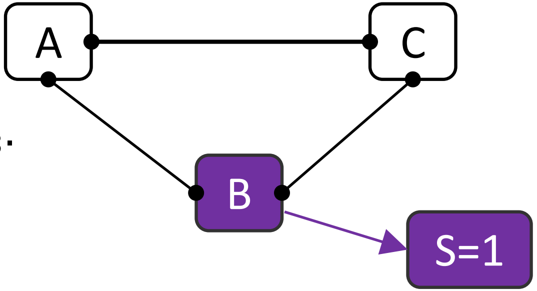
$$(\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}) \wedge$$

$$\neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C}) \wedge$$



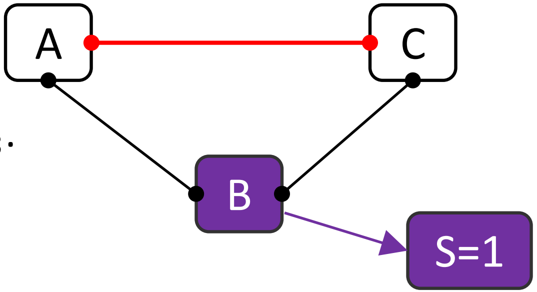
CONVERSION TO LOGIC FORMULA: EXAMPLE (SELECTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set selected based on B, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|S = 1)$
- In path terms: $\nexists$ m-connecting path between A and C given $S = 1$ in G_{B}.



CONVERSION TO LOGIC FORMULA: EXAMPLE (SELECTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set selected based on B, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|S = 1)$
- In path terms: $\nexists$ m-connecting path between A and C given $S = 1$ in G_{B}.

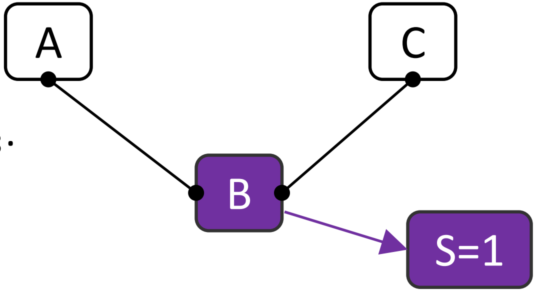


A-C does not exist

$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

CONVERSION TO LOGIC FORMULA: EXAMPLE (SELECTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set selected based on B, $\mathrm{Ind}(\mathrm{A}, \mathrm{C} | S = 1)$
- In path terms: $\nexists$ m-connecting path between A and C given $S = 1$ in G_{B}.



A-C does not exist

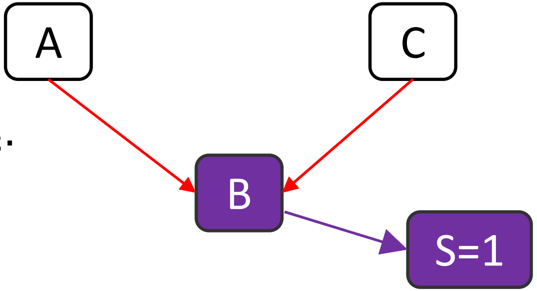
$\neg E_{A \to C} \land \neg E_{A \leftarrow C} \land \neg E_{A \leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A \leftarrow B} \land E_{B \leftarrow C})$
$\neg(E_{A \leftarrow B} \land E_{B \to C})$
$\neg(E_{A \leftarrow B} \land E_{B \leftrightarrow C})$
$\neg(E_{A \to B} \land E_{B \to C})$
$\neg(E_{A \leftrightarrow B} \land E_{B \to C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE (SELECTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set selected based on B, $\mathrm{Ind}(\mathrm{A}, \mathrm{C} | S = 1)$
- In path terms: $\nexists$ m-connecting path between A and C given $S = 1$ in G_{B}.



A-C does not exist

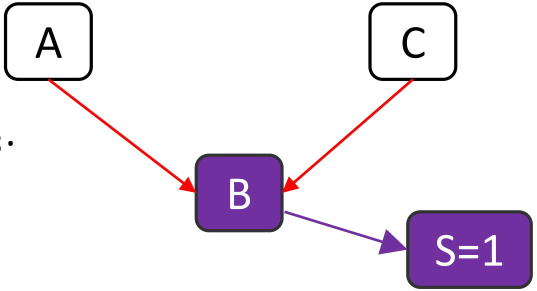
$\neg E_{A\rightarrow C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$
$\neg(E_{A\rightarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\rightarrow C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\leftarrow C})$

CONVERSION TO LOGIC FORMULA: EXAMPLE (SELECTION)

- Suppose you know nothing about the causal structure G of A, B, C.
- In a data set selected based on B, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|S = 1)$
- In path terms: $\nexists$ m-connecting path between A and C given $S = 1$ in G_{B}.



A-C does not exist

$\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}$

A-B-C is not m-connecting

$\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\to C})$
$\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C})$
$\neg(E_{A\to B} \wedge E_{B\to C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\to C})$
$\neg(E_{A\to B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\leftarrow C})$
$\neg(E_{A\to B} \wedge E_{B\leftrightarrow C})$
$\neg(E_{A\leftrightarrow B} \wedge E_{B\leftrightarrow C})$

Logic formula:

$$
\begin{gathered}
(\neg E_{A\to C} \wedge \neg E_{A\leftarrow C} \wedge \neg E_{A\leftrightarrow C}) \wedge \\
\neg(E_{A\leftarrow B} \wedge E_{B\leftarrow C}) \wedge \\
\neg(E_{A\leftarrow B} \wedge E_{B\to C}) \wedge \\
\neg(E_{A\leftarrow B} \wedge E_{B\leftrightarrow C}) \wedge \\
\neg(E_{A\to B} \wedge E_{B\to C}) \wedge \\
\neg(E_{A\leftrightarrow B} \wedge E_{B\to C}) \wedge \\
\neg(E_{A\to B} \wedge E_{B\leftarrow C}) \wedge \\
\neg(E_{A\leftrightarrow B} \wedge E_{B\leftarrow C}) \wedge \\
\neg(E_{A\to B} \wedge E_{B\leftrightarrow C}) \wedge \\
\neg(E_{A\leftrightarrow B} \wedge E_{B\leftrightarrow C})
\end{gathered}
$$

CONVERSION TO LOGIC FORMULA: INPUT CONSTRAINTS

Path constraints corresponding to (conditional) dependencies and independencies from multiple datasets.

Information about the datasets

- Whether your samples were selected based on some variables.
- Variables that were manipulated in your data set.

Many more ways to encode constraints into logic

- Different variable choices (e.g. edge *-*, orientations).
- Different constraint choices depending on the problem at hand.
  - Ancestral paths
  - Inducing paths.
  - Colliders/non-colliders.

CONVERSION TO LOGIC FORMULA: VERSATILITY

Logic-based causal discovery **trivially** and **collectively** handles cases for which no algorithm existed!

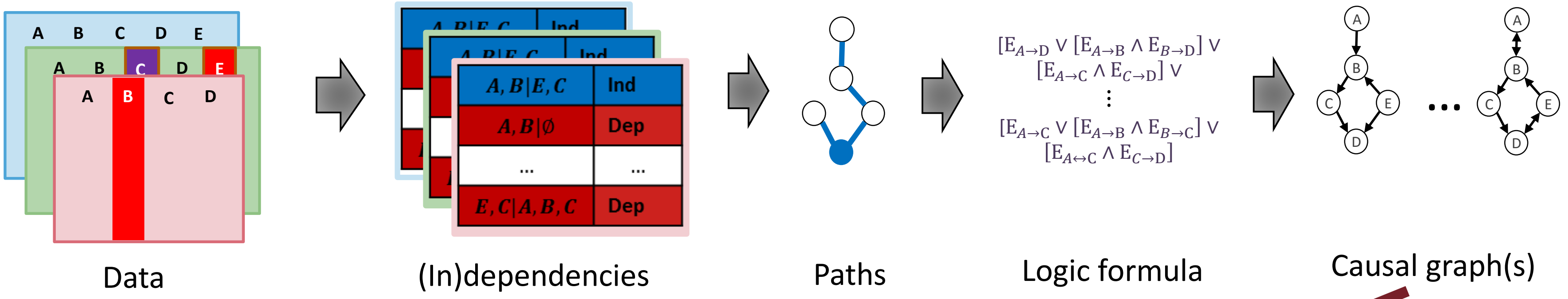
- Incorporating prior knowledge.
  - Algorithms for learning Bayesian networks can only enforce the presence/absence of direct edges.
  - Easily impose presence/absence of direct edges, directed paths or m-connections (associations).
  - root/leaf nodes.
- Learning semi-Markov causal graphs.
  - no learning algorithm until logic-based causal discovery.
- Combining heterogeneous data sets.
- Soft interventions.

- Sound and complete algorithms with incomplete knowledge (e.g. can not perform some tests of independence).

OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   i. ~~Converting path constraints to logic formulae.~~
   ii. **Problem complexity.**
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY



Exponential number of
1.Independencies
2.Paths
3. Solutions

PROBLEM COMPLEXITY: EXAMPLE

For a data set with 10 variables:

$2^8 = 256$ different conditioning sets

For each conditioning set, you need to consider all possible paths with up to 9 edges:

$\sum_{k=2}^{10} \frac{8}{10-k} = 1435$ paths per pair of variables.

In total: $\binom{10}{2} = 45\ variable\ pairs \times 256\ cond\ sets \times 1435\ paths = 16531200\ path\ constraints.$

For a network of 10 variables:

135 possible edges.

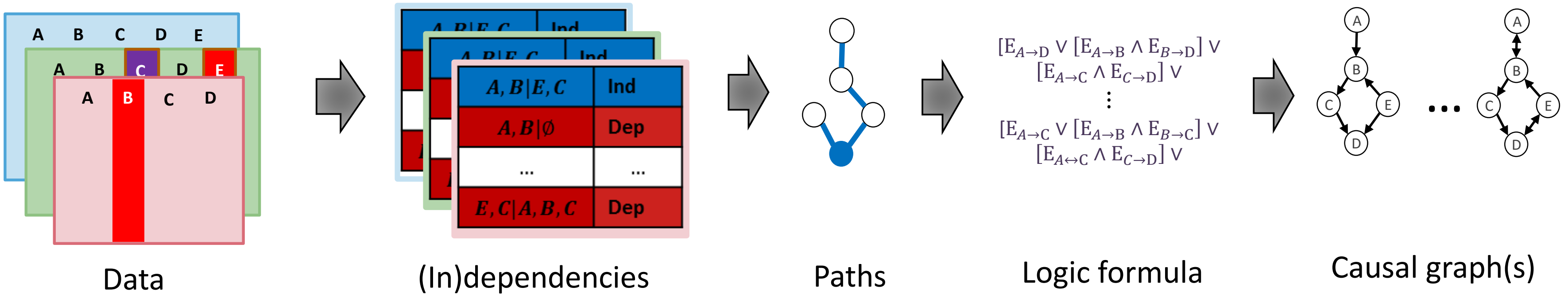
$2^{135} \sim 10^{40}$ different graphs.

Brute force approach only works for ~10 variables regardless of encoding.

Several heuristics for scaling up (depending on the algorithm).

You can take into account <u>all dependencies and independencies, even for a small number of variables.</u>

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY

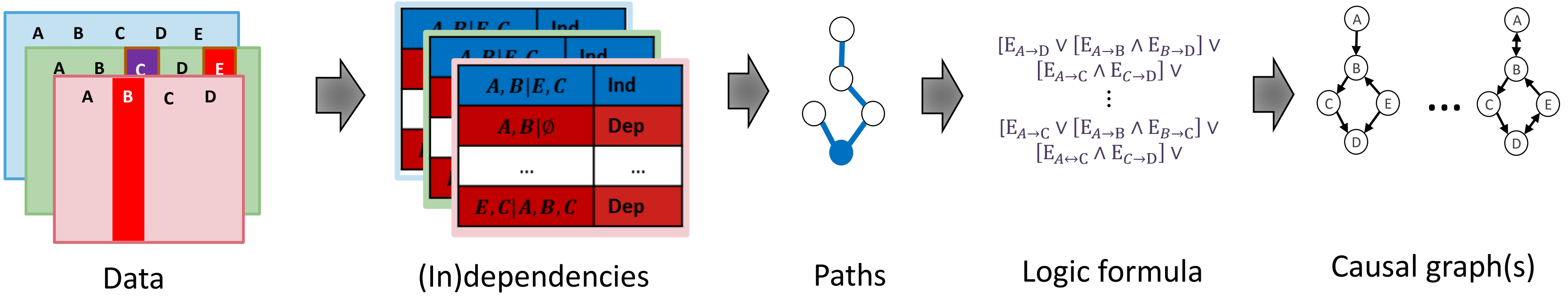


Reduce the number of independencies:

Run FCI and use only the tests performed by FCI.

Limit max conditioning set size.

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY

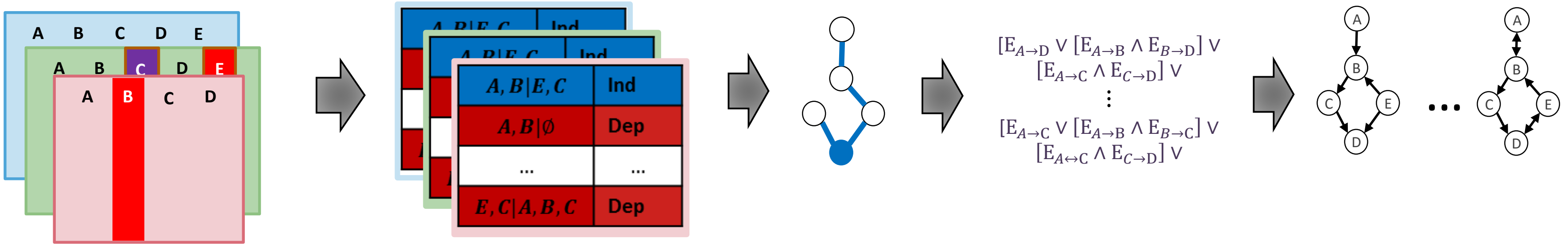


Reduce the number of paths:

Use inducing paths that connect paths on the graph to ∃ of independence (given any set).

Limit the maximum path length.

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY



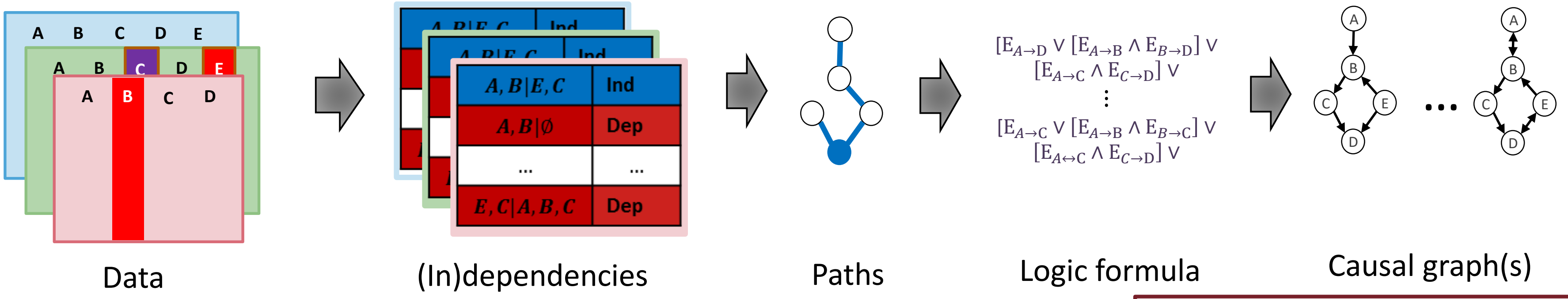
Data (In)dependencies Paths Logic formula Causal graph(s)

Need a clever way to encode constraints!

e.g. recursively encode paths.

Convert to CNF for most SAT solvers.

LOGIC-BASED INTEGRATIVE CAUSAL DISCOVERY

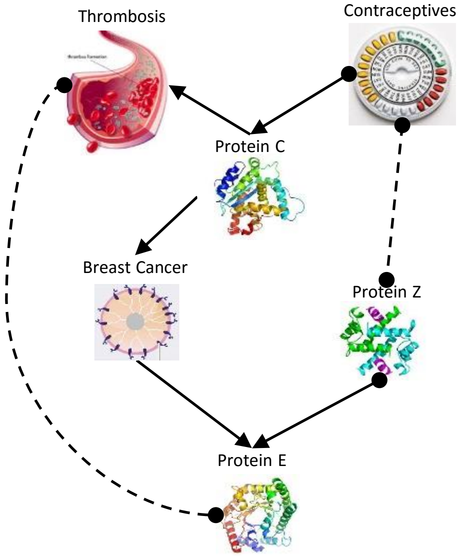


No need to enumerate all solutions!

Query the formula for
- A single causal graph.
- A causal graph with specific features.
- Features that are invariant in all possible causal graphs.

SUMMARIZING PAIRWISE RELATIONS

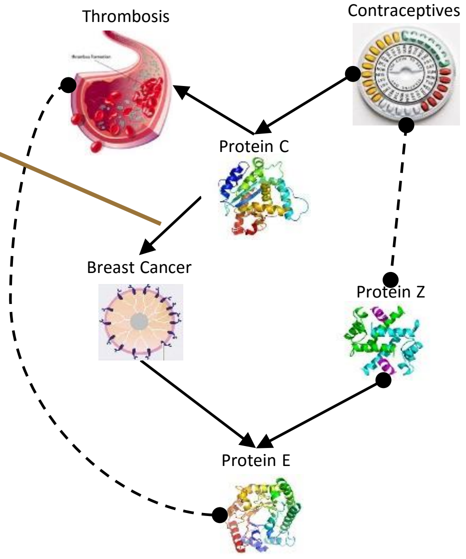
Absent edges:
Absent in **all**
solutions



SUMMARIZING PAIRWISE RELATIONS

Absent edges:
Absent in **all**
solutions

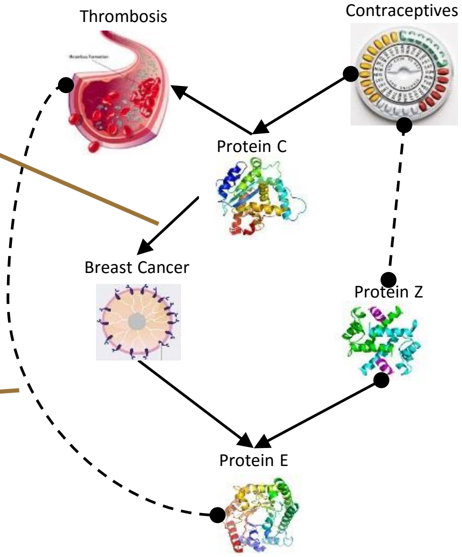
solid edges:
present in **all**
solutions



SUMMARIZING PAIRWISE RELATIONS

Absent edges:
Absent in **all**
solutions

solid edges:
present in **all**
solutions



dashed edges:
present in **some**
solutions

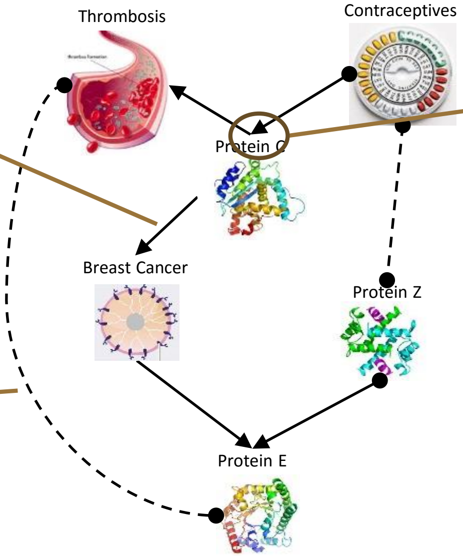
SUMMARIZING PAIRWISE RELATIONS

Absent edges: Absent in **all** solutions

solid edges: present in **all** solutions

solid endpoints: same orientation in **all** solutions

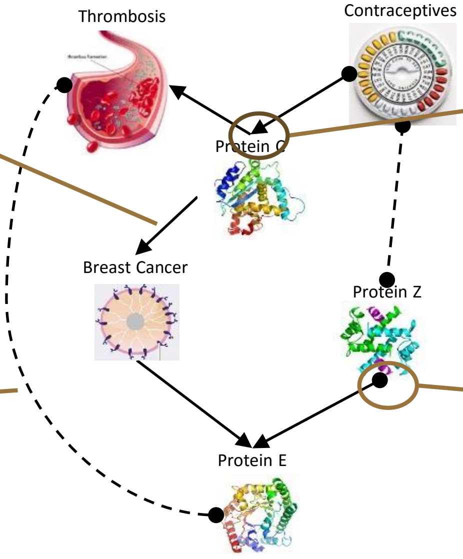
dashed edges: present in **some** solutions



SUMMARIZING PAIRWISE RELATIONS

Absent edges: Absent in **all** solutions

solid edges: present in **all** solutions



solid endpoints: same orientation in **all** solutions

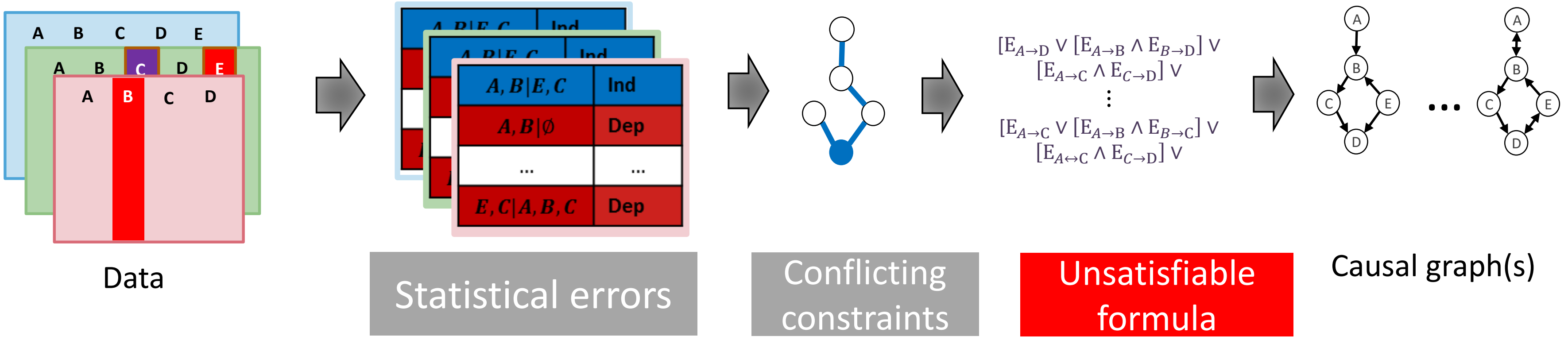
dashed edges: present in **some** solutions

Circle endpoints: orientation varies in **different** solutions

OUTLINE

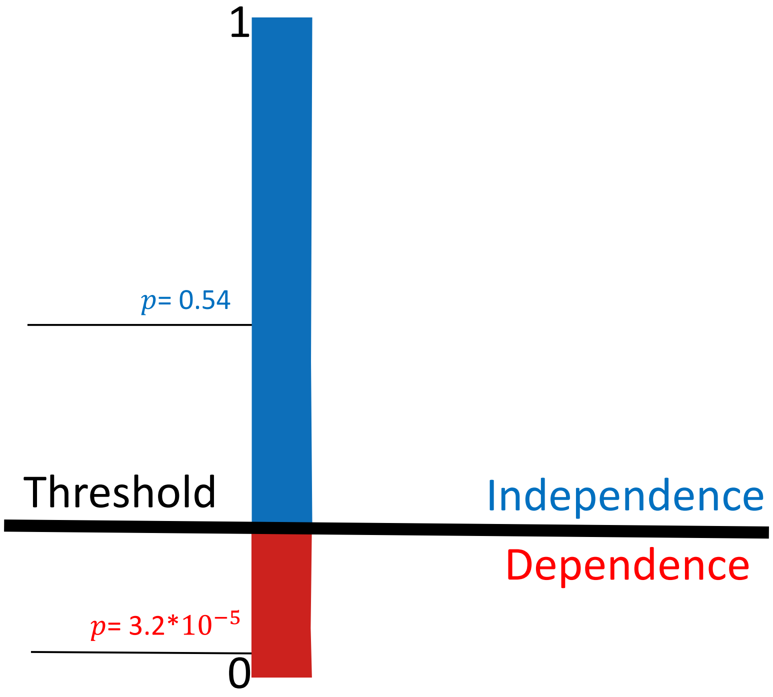
1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   i. ~~Converting path constraints to logic formulae.~~
   ii. ~~Problem complexity.~~
   iii. Conflict resolution.
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

STATISTICAL ERRORS RESULT IN CONFLICTING INPUTS

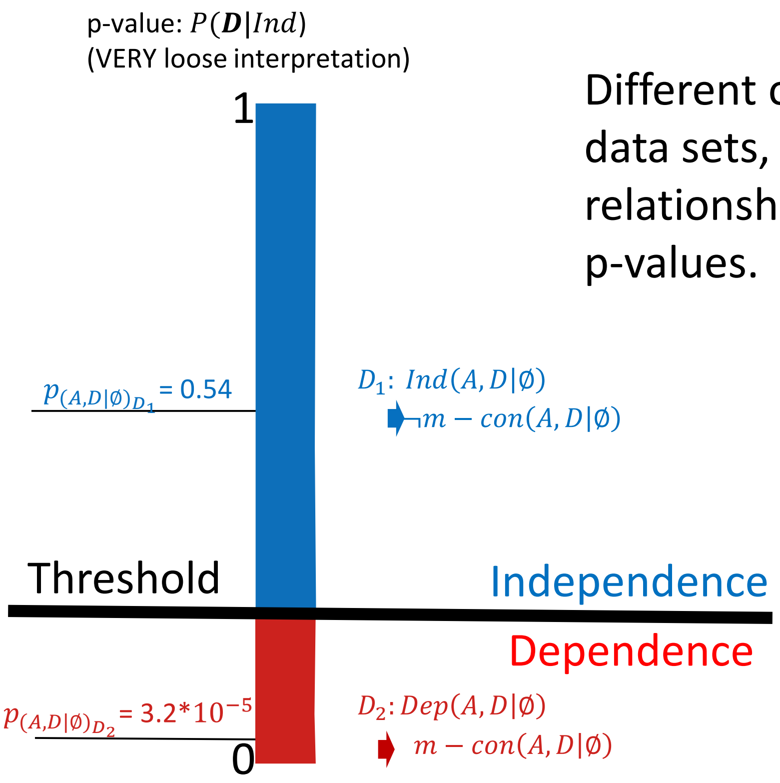


TESTING (CONDITIONAL) INDEPENDENCE

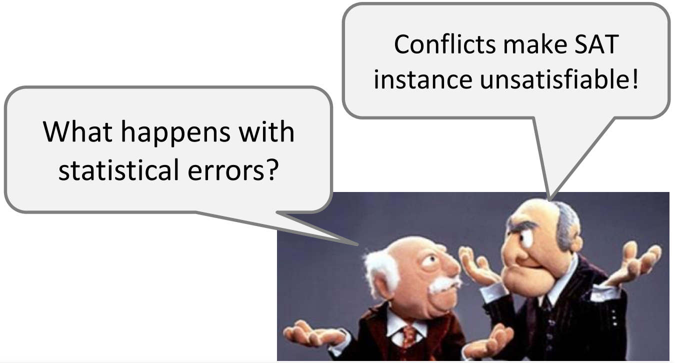
p-value: $P(\boldsymbol{D}|Ind)$
(VERY loose interpretation)



TESTING (CONDITIONAL) INDEPENDENCE

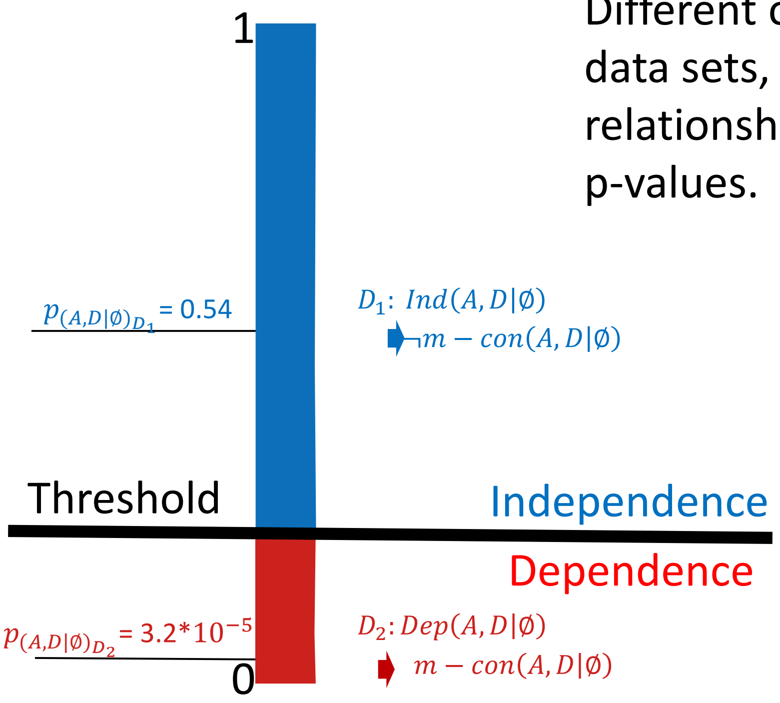


Different observational data sets, same relationship, different p-values.



TESTING (CONDITIONAL) INDEPENDENCE

p-value: $P(\boldsymbol{D}|Ind)$
(VERY loose interpretation)

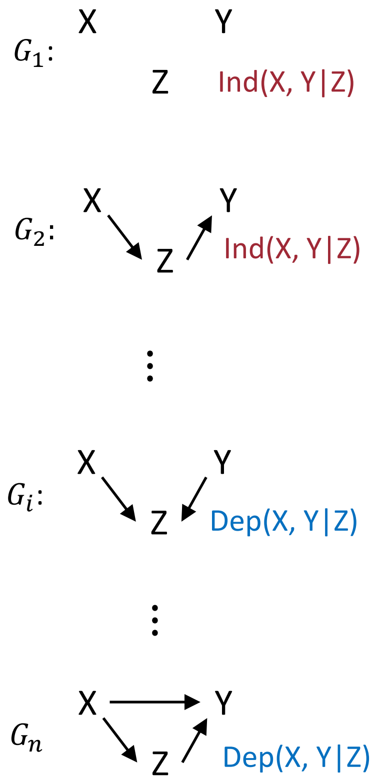


Different observational data sets, same relationship, different p-values.

How can you decide if Independence is more probable than dependence?

ESTIMATING $P(Ind|\boldsymbol{D})$ USING BAYESIAN SCORING (1)

- You want to estimate $P(Ind(X,Y|Z)|\boldsymbol{D})$
- Score every possible DAG over X, Y, Z: $P(\boldsymbol{D}|G)$.
    - You can use BDE, BGE to compute $P(\boldsymbol{D}|G)$.
- Some of these DAGs entail dependence (m-connection, some independence (m-separation).
    - Define a prior over graphs.
- Take the weighted average:
- $P(Ind(X,Y|\boldsymbol{Z})|\boldsymbol{D}) \propto \sum_{G:G\ entails\ Ind(X,Y|\boldsymbol{Z})} P(\boldsymbol{D}|G) \times P(G)$
- Exponential number of DAGs.
- Use one graph per Markov equivalence class (still exponential).
- Still not possible for more than 5-6 variables.



[BCCD, Claassen and Heskes, UAI 2012]

ESTIMATING $P(Ind|\boldsymbol{D})$ USING BAYESIAN SCORING (2)

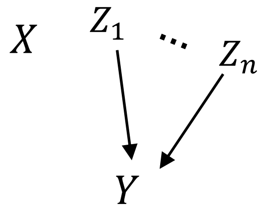
- You want to estimate $P(Ind(X,Y|\boldsymbol{Z})|\boldsymbol{D})$

- Independence $Ind(X,Y|\boldsymbol{Z})$: $P(X,Y|\boldsymbol{Z}) = P(X|\boldsymbol{Z})P(Y|\boldsymbol{Z})$
- Dependence $Dep(X,Y|\boldsymbol{Z})$: $P(X,Y|\boldsymbol{Z}) = P(X|\boldsymbol{Z})P(Y|X,\boldsymbol{Z})$

- $P(Ind(X,Y|\boldsymbol{Z})|\boldsymbol{D}) = \dfrac{P(Y|\boldsymbol{Z})\,\pi_0}{P(Y|\boldsymbol{Z})\pi_0 + P(Y|X,\boldsymbol{Z})(1-\pi_0)}$.

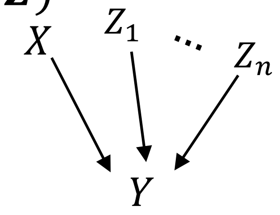
- Use BDE, BGE to estimate $P(Y|\boldsymbol{Z}), P(Y|X,\boldsymbol{Z})$.
- π_0: Prior for independence is an input parameter.

$P(Y|\boldsymbol{Z})$



vs.

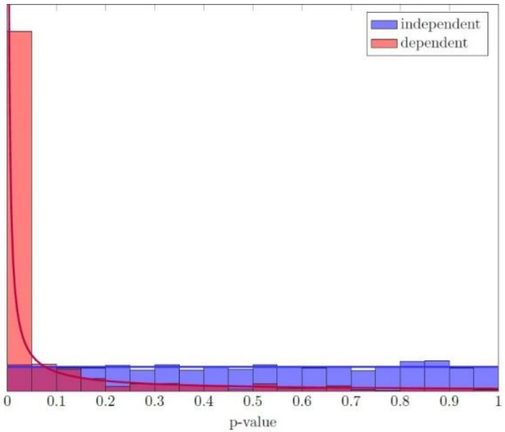
$P(Y|X,\boldsymbol{Z})$



[M&B, Margaritis and Bromberg, CI 2009]

ESTIMATING $P(Ind|\boldsymbol{D})$ FROM P-VALUES

- p-values coming from independence follow a $Beta(1,1)$ distribution
- p-values coming from dependence follow a distribution in (0, 1) with declining density
    - Can be modeled with a $Beta(\xi, 1), \xi \in (0,1)$ distribution.

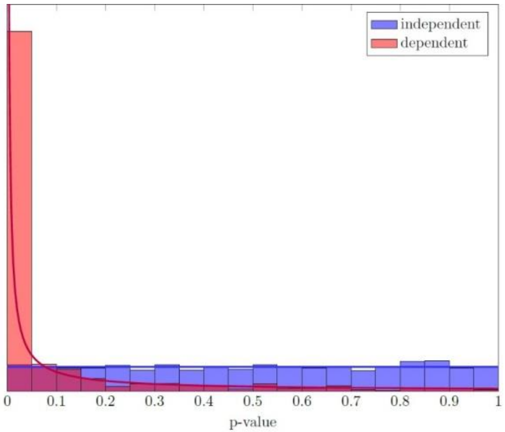


[PROPER, Triantafillou et al, PGM 2014]

ESTIMATING $P(Ind|\boldsymbol{D})$ FROM P-VALUES

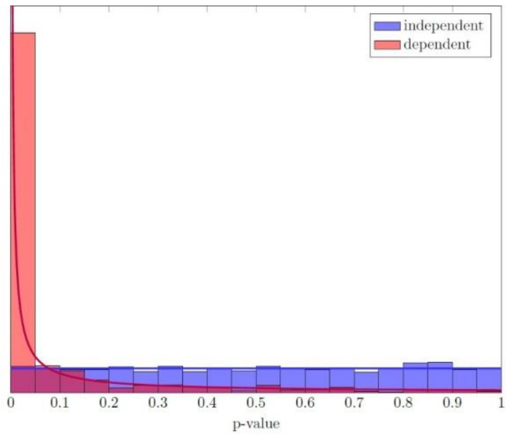
- p-values coming from independence follow a $Beta(1,1)$ distribution
- p-values coming from dependence follow a distribution in (0, 1) with declining density
  - Can be modeled with a $Beta(\xi, 1), \xi \in (0,1)$ distribution.

- Let π_0 be the proportion of independencies.
- $f(p|\pi_o, \xi) = \pi_0 + (1-\pi_0)\xi p^{\xi-1}$.



[PROPER, Triantafillou et al, PGM 2014]

ESTIMATING $P(Ind|\boldsymbol{D})$ FROM P-VALUES



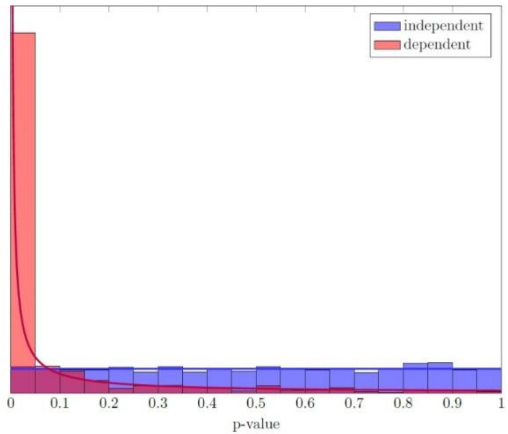
- p-values coming from independence follow a $Beta(1, 1)$ distribution
- p-values coming from dependence follow a distribution in (0, 1) with declining density
  - Can be modeled with a $Beta(\xi, 1), \xi \in (0, 1)$ distribution.

- Let π_0 be the proportion of independencies.
- $f(p|\pi_o, \xi) = \pi_0 + (1 - \pi_0)\xi p^{\xi - 1}$.

- You can find estimate π_0, ξ from the empirical distribution of your p-values
  - Find $\widehat{\pi_0}$ using [Storey and Tibshirani, 2003] (assumes i.i.d. p-values)
  - Find $\hat{\xi}$ by minimizing negative log likelihood

[PROPER, Triantafillou et al, PGM 2014]

ESTIMATING $P(Ind|\boldsymbol{D})$ FROM P-VALUES



- p-values coming from independence follow a $Beta(1,1)$ distribution
- p-values coming from dependence follow a distribution in (0, 1) with declining density
  - Can be modeled with a $Beta(\xi,1), \xi \in (0,1)$ distribution.

- Let π_0 be the proportion of independencies.
- $f(p|\pi_o,\xi) = \pi_0 + (1-\pi_0)\xi p^{\xi-1}$.

- You can find estimate π_0, ξ from the empirical distribution of your p-values
  - Find $\widehat{\pi_0}$ using [Storey and Tibshirani, 2003] (assumes i.i.d. p-values)
  - Find $\hat{\xi}$ by minimizing negative log likelihood

$$P(Ind|p) = \frac{\dfrac{\widehat{\pi_0}}{(1-\widehat{\pi_0})\hat{\xi}p^{(1-\hat{\xi})}}}{1+\dfrac{\widehat{\pi_0}}{(1-\widehat{\pi_0})\hat{\xi}p^{(1-\hat{\xi})}}}$$

[PROPER, Triantafillou et al, PGM 2014]

ESTIMATING $P(Ind|\boldsymbol{D})$

- Bayesian methods
  - Use the data directly.
  - No problem if you have data sets with different sample sizes etc.
  - Computationally expensive.
  - Choose a prior for π_0.

- PROPER (based on p-values)
  - Scalable, no computational overhead, benefits from larger p-value populations (more tests).
  - Estimate π_0 from the data.
  - p-values are not i.i.d.

CONFLICT RESOLUTION STRATEGIES

P(constraint)	Ind/Dep	path constraint
0.999	Dep	∃ m-connecting path from A to D given Ø in S^{I_n}
0.998	Ind	∄ m-connecting path from A to D given Ø in S^{I_1}
⋮	⋮	⋮
0.510	Dep	∃ m-connecting path from A to B given Ø in S^{I_1}

- Assign weights according to P(constraint), maximize the sum of weights.
- Rank by probability, greedily satisfy constraints.

Maximizing sum of weights is the best strategy
Use greedy to scale up.

OUTLINE

1. Integrative causal discovery
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   i. ~~Converting path constraints to logic formulae.~~
   ii. ~~Problem complexity.~~
   iii. ~~Conflict resolution.~~
   iv. Existing algorithms.
   v. Reasoning with logic based causal discovery.
   vi. Non-trivial inferences-validation.

EXISTING ALGORITHMS

Vary in:

- Type of constraints:
  - different types of paths (m-connecting, inducing, ancestral).
  - translation to logic formula.
- Types of heterogeneity:
  - Soft/hard interventions, selection.
- Preprocessing:
  - Heuristics to limit number of constraints / paths.
- Conflict Resolution
  - Method for calculating probabilities.
  - Conflict resolution strategy (greedy/ max SAT / weighted max SAT).
- CS solver
  - Initially SAT solvers, more recently ASP.
- Scalability
  - Depends on choices above. Be exact/ focus on scalability.
  - Difficult to determine
    - huge variance depending on the problem.

Implementations vary heuristics are typically easy to incorporate in any algorithm.

-maximum conditioning set size/ path length.

-greedy /weighted max SAT.

CSAT+

Type of Constraints	m-connecting paths, inducing paths.
Type of Heterogeneity	Overlapping variables.
Preprocessing	Runs FCI on multiple data sets. Additional preprocessing rules for additional edge removals/orientations.
Conflict resolution	None (oracle only)
CSP solver	MINISAT
Scalability	~37 variables (ALARM network)

[Triantafillou, Tsamardinos and Tollis, AISTATS 2010]

LOCI

Type of Constraints	ancestral paths. Converts [minimal] conditional independencies to ancestral relations: $Ind(X, Y \mid [\boldsymbol{Z}]) \Rightarrow \boldsymbol{Z} \rightarrow \cdots \rightarrow X \vee \boldsymbol{Z} \rightarrow \cdots \rightarrow Y$
Type of Heterogeneity	None (substitutes FCI orientation steps).
Preprocessing	FCI skeleton step.
Conflict resolution	None (single data set, runs similar to FCI orientation rules)
CSP solver	custom set of rules
Scalability	unknown (probably similar to FCI).

[Claassen and Heskes, UAI 2011]

SAT-BASED CAUSAL DISCOVERY

Type of Constraints	m-connecting paths
Type of Heterogeneity	Overlapping variables, interventions also allows cycles.
Preprocessing	None. Can use a subset of (in) dependencies depending on assumptions (e.g. FCI tests only)
Conflict resolution	None (oracle only)
CSP solver	MINISAT
Scalability	8-12 variables

[Hyttinen, Hoyer, Eberhardt and Järvisalo, UAI 2013]

CONSTRAINT-BASED CAUSAL DISCOVERY

Type of Constraints	m-connecting paths. encoded in ASP based on marginalization and conditioning.
Type of Heterogeneity	Overlapping variables, interventions allows cycles
Preprocessing	none
Conflict resolution	Default: based on M&B, maximize sum of weights (find global optimum), also tried maximizing the number of independencies/ number of constraints
CSP solver	ASP
Scalability	7 variables

[Hyttinen, Eberhardt and Järvisalo, UAI 2014]

COMBINE

Type of Constraints	inducing paths Drastically reduces the number of constraints ($\exists, \nexists$ path) to 1 per variable pair & data set (compared to 2^n)
Type of Heterogeneity	Overlapping variables, interventions
Preprocessing	FCI on each data set.
Conflict resolution	Default: based on PROPER, greedy search. also implemented: BCCD , weighted maxSAT.
CSP solver	MINISAT
Scalability	100 variables (additionally limits maximum path length)

[Triantafillou and Tsamardinos, JMLR 2015]

ETIO

Type of Constraints	m-connecting paths. encoded in ASP based on extension of the Bayes-Ball algorithm (used to determine m-connections/m-separations in graphs) for SMCGs with selection.
Type of Heterogeneity	Overlapping variables, interventions, selection.
Preprocessing	none
Conflict resolution	based on PROPER/M&B, greedy
CSP solver	ASP
Scalability	10-15 variables

[Borboudakis and Tsamardinos, KDD 2016]

ACI

Type of Constraints	m-connections, ancestry relations
Type of Heterogeneity	Overlapping variables, various types of interventions
Preprocessing	none
Conflict resolution	based on M&B, weighted maxSAT
CSP solver	ASP
Scalability	10-15 variables

[S. Magliacane, T. Claassen, J.M. Mooij, *arXiv*]

MORE

- Using conversion to logic to incorporate prior knowledge in maximal ancestral graphs.
  - [Borboudakis, Triantafillou and Tsamardinos, ESANN 2011].
- Using conversion to logic for causal discovery from time-course data
  - Causal Discovery from Subsampled Time Series Data by Constraint Optimization, [Hyttinen, Plis, Järvisalo, Eberhardt and Danks, arXiv, 2016]
- Using conversion to logic for identifying chain graphs.
  - Learning Optimal Chain Graphs with Answer Set Programming[Sonntag, Järvisalo, Penã, Hyttinen, UAI 2015]
- Using conversion to logic to identify semi-Markov causal graphs.
  - [Penã, UAI 2016]

OVERVIEW

Different data distributions, same causal mechanism: use causal modeling to connect.

Algorithms can handle datasets of different variable sets, different experimental conditions, prior causal knowledge.

Identify the set of causal graphs that simultaneously fit all datasets .

Convert problem to SAT or ASP.

Logic formula encodes a set of causal models that simultaneously fit all the data sets.

QUESTIONS

-How can you reason with this set of models?

-Is it useful? Do you make additional inferences than analyzing each data set in isolation?

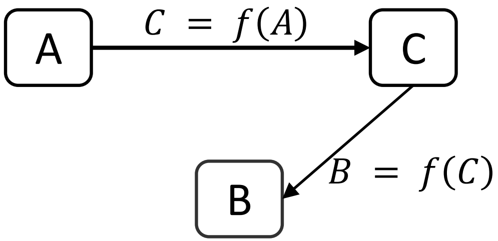
OUTLINE

1. Integrative causal discovery
   1. ~~Motivation.~~
   2. ~~Causal models.~~
   3. ~~m-separation.~~
   4. ~~Reverse engineering causal models (single data set).~~
   5. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   6. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   1. ~~Converting path constraints to logic formulae.~~
   2. ~~Problem complexity.~~
   3. ~~Conflict resolution.~~
   4. ~~Existing algorithms.~~
   5. **Reasoning with logic based causal discovery.**
   6. Non-trivial inferences-validation.

ESTIMATING CAUSAL EFFECTS

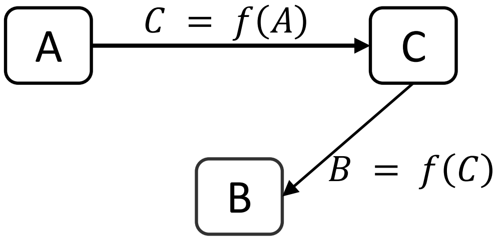
You are interested in computing $P(B|do(A = a))$
In general, $P(B|do(A = a)) \neq P(B|A)$

If you know the causal graph, you can use the rules of do-calculus to transform **post-intervention** probabilities to **pre-intervention** probabilities.



ESTIMATING CAUSAL EFFECTS

You are interested in computing $P(B|do(A=a))$
In general, $P(B|do(A=a)) \neq P(B|A)$



If you know the causal graph, you can use the rules of do-calculus to transform **post-intervention** probabilities to **pre-intervention** probabilities.

[Rule 1] $Ind(Y,Z\|X,W)_{G^{X}} \Rightarrow P(y\|do(x),z,w) = P(y\|do(x),w)$.	Insert/delete observations
[Rule 2] $Ind(Y,I_Z\|X,Z,W)_{G^{X}} \Rightarrow P(y\|do(x),do(z),w) = P(y\|do(x),z,w)$.	Exchange action/observation
[Rule 3] $Ind(Y,I_Z\|X,W)_{G^{X}} \Rightarrow P(y\|do(x),do(z),w) = P(y\|do(x),w)$.	Insert/delete action

Check m-separations ⇒

Apply rules until you have a formula with pre-intervention probabilities

[Shpitser and Pearl (2006): Return a formula if identifiable]

DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*

Causal effect $P(B|do(A))$

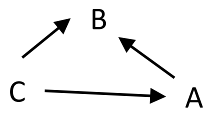
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*

Find a graph consistent with Φ



Causal effect $P(B|do(A))$

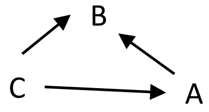
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*

Find a graph consistent with Φ



Shpitser and Pearl (2006)

Causal effect $P(B|do(A))$

$$F_1 = \sum_c P(b|a,c)P(c)$$

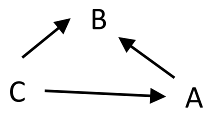
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

<u>Constraints in logic formula Φ</u>

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given* A, C)

<u>Find a graph consistent with Φ</u>



Shpitser and Pearl (2006)

<u>Causal effect $P(B|do(A))$</u>

$$F_1 = \sum_c P(b|a,c)P(c)$$

Identify and negate the m-separations used to derive the formula

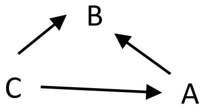
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

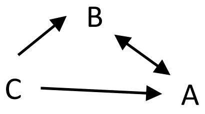
DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given* A, C)

Find a graph consistent with Φ





Shpitser and Pearl (2006)

Causal effect $P(B|do(A))$

$$F_1 = \sum_c P(b|a,c)P(c)$$

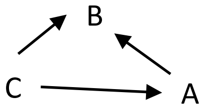
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

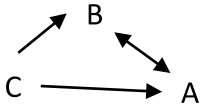
DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given* A, C)

Find a graph consistent with Φ





Shpitser and Pearl (2006)

Causal effect $P(B|do(A))$

$$F_1 = \sum_c P(b|a,c)P(c)$$

$$F_2 = P(b)$$

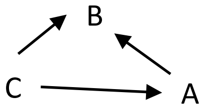
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

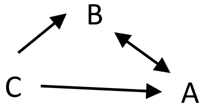
DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

<u>Constraints in logic formula Φ</u>

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given* A, C)
$\exists$ *m-connecting path from* I_A *to B given* $\emptyset$)

<u>Find a graph consistent with Φ</u>





Shpitser and Pearl (2006)

<u>Causal effect $P(B|do(A))$</u>

$$F_1 = \sum_c P(b|a,c)P(c)$$

$$F_2 = P(b)$$

Identify and negate the m-separations used to derive the formula

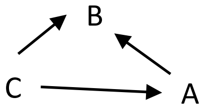
[Hyttinen, Eberhardt and Jarvislao, UAI 2015]

DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given* C
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given* B
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given* A
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given* A, C)
$\exists$ *m-connecting path from* I_A *to B given* $\emptyset$)

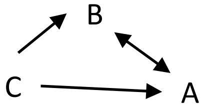
Find a graph consistent with Φ



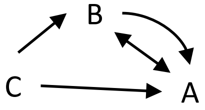
Shpitser and Pearl (2006)

Causal effect $P(B|do(A))$

$$F_1 = \sum_c P(b|a,c)P(c)$$



$$F_2 = P(b)$$



$$NA$$

.
.
.

UNSAT

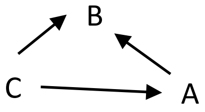
[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

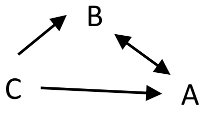
DO-CALCULUS WHEN THE GRAPH IS UNKNOWN

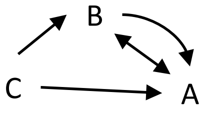
Constraints in logic formula Φ

$\exists$ *m-connecting path from A to B given* $\emptyset$
$\exists$ *m-connecting path from A to B given C*
$\exists$ *m-connecting path from A to C given* $\emptyset$
$\exists$ *m-connecting path from A to C given B*
$\exists$ *m-connecting path from B to C given* $\emptyset$
$\exists$ *m-connecting path from B to C given A*
$\nexists$ *directed path from A to C*
$\nexists$ *directed path from A to B*
$\nexists$ *directed path from B to C*
($\exists$ *m-connecting path from* I_A *to C given* $\emptyset \vee$
$\exists$ *m-connecting path from* I_A *to B given A, C*)
$\exists$ *m-connecting path from* I_A *to B given* $\emptyset$)

Find a graph consistent with Φ







.
.
.

UNSAT

Shpitser and Pearl (2006)

Causal effect $P(B|do(A))$

$$F_1 = \sum_c P(b|a,c)P(c)$$

$$F_2 = P(b)$$

$$NA$$

[Hyttinen, Eberhardt and Järvisalo, UAI 2015]

QUESTIONS

-How can you reason with this set of models?

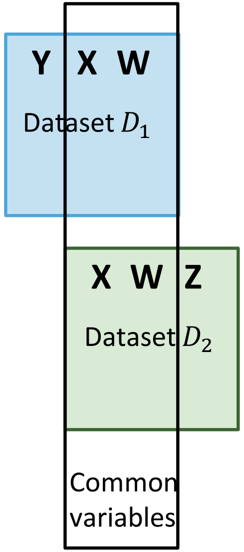
You can use do-calculus and estimate (a population of) causal effects.

-Is it useful? Do you make additional inferences than analyzing each data set in isolation?

OUTLINE

1. ~~Integrative Causal Discovery~~
   - i. ~~Motivation.~~
   - ii. ~~Causal models.~~
   - iii. ~~m-separation.~~
   - iv. ~~Reverse engineering causal models (single data set).~~
   - v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   - vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   - i. ~~Converting path constraints to logic formulae.~~
   - ii. ~~Problem complexity.~~
   - iii. ~~Conflict resolution.~~
   - iv. ~~Existing algorithms.~~
   - v. ~~Reasoning with logic based causal discovery.~~
   - **vi. Non-trivial inferences-validation.**

EXAMPLE INPUT- OUTPUT



CI pattern C_1

$X, W \mid Y$	Ind
$X, W \mid \emptyset$	Dep
$X, Y \mid \emptyset$	Dep
$Y, W \mid \emptyset$	Dep
$Y, X \mid W$	Dep
$Y, W \mid X$	Dep

CI pattern C_2

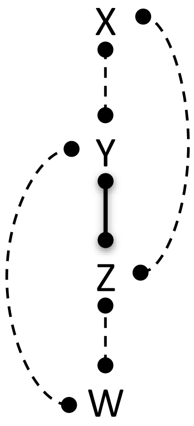
$X, W \mid Z$	Ind
$X, W \mid \emptyset$	Dep
$X, Z \mid \emptyset$	Dep
$Z, W \mid \emptyset$	Dep
$Z, W \mid X$	Dep
$Z, X \mid W$	Dep

Formula Φ

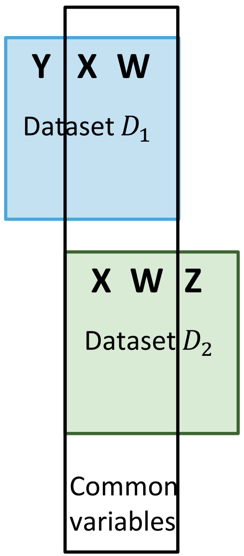
$\nexists m.\,connecting\ path\ \ from\ X\ to\ W\ given\ Y \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ Y\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ Y\ to\ W\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ Y\ given\ W \wedge$
$\exists m.\,connecting\ path\ \ from\ Y\ to\ W\ given\ X \wedge$

$\nexists m.\,connecting\ path\ \ from\ X\ to\ W\ given\ Z \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ Z\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ Zto\ W\ given\ \emptyset \wedge$
$\exists m.\,connecting\ path\ \ from\ X\ to\ Z\ given\ W \wedge$
$\exists m.\,connecting\ path\ \ from\ Z\ to\ W\ given\ X$

Summary of solutions



EXAMPLE INPUT- OUTPUT



CI pattern C_1

$X,W\|Y$	Ind
$X,W\|\emptyset$	Dep
$X,Y\|\emptyset$	Dep
$Y,W\|\emptyset$	Dep
$Y,X\|W$	Dep
$Y,W\|X$	Dep

CI pattern C_2

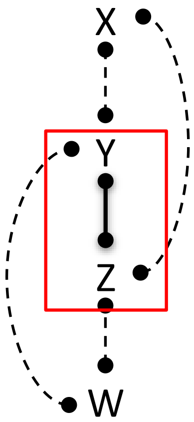
$X,W\|Z$	Ind
$X,W\|\emptyset$	Dep
$X,Z\|\emptyset$	Dep
$Z,W\|\emptyset$	Dep
$Z,W\|X$	Dep
$Z,X\|W$	Dep

Formula Φ

$\nexists m.\, connecting\ path\ \ from\ X\ to\ W\ given\ Y \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ Y\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ Y\ to\ W\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ Y\ given\ W \wedge$
$\exists m.\, connecting\ path\ \ from\ Y\ to\ W\ given\ X \wedge$

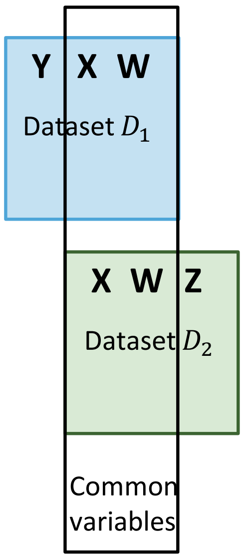
$\nexists m.\, connecting\ path\ \ from\ X\ to\ W\ given\ Z \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ Z\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ Z to\ W\ given\ \emptyset \wedge$
$\exists m.\, connecting\ path\ \ from\ X\ to\ Z\ given\ W \wedge$
$\exists m.\, connecting\ path\ \ from\ Z\ to\ W\ given\ X$

Summary of solutions



Predict that Y and Z are associated even though they are not measured in the same data set.

TEST IF IT WORKS IN REAL DATA.



CI pattern C_1

$X, W \| Y$	Ind
$X, W \| \emptyset$	Dep
$X, Y \| \emptyset$	Dep
$Y, W \| \emptyset$	Dep
$Y, X \| W$	Dep
$Y, W \| X$	Dep

CI pattern C_2

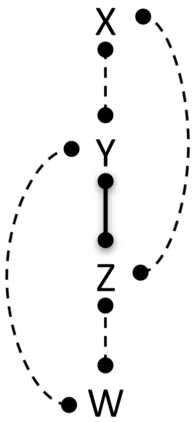
$X, W \| Z$	Ind
$X, W \| \emptyset$	Dep
$X, Z \| \emptyset$	Dep
$Z, W \| \emptyset$	Dep
$Z, W \| X$	Dep
$Z, X \| W$	Dep

Formula Φ

$\nexists m.\mathit{connecting\ path\ \ from\ X\ to\ W\ given\ Y} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ Y\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ Y\ to\ W\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ Y\ given\ W} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ Y\ to\ W\ given\ X} \wedge$

$\nexists m.\mathit{connecting\ path\ \ from\ X\ to\ W\ given\ Z} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ W\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ Z\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ Zto\ W\ given\ \emptyset} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ X\ to\ Z\ given\ W} \wedge$
$\exists m.\mathit{connecting\ path\ \ from\ Z\ to\ W\ given\ X}$

Summary of solutions



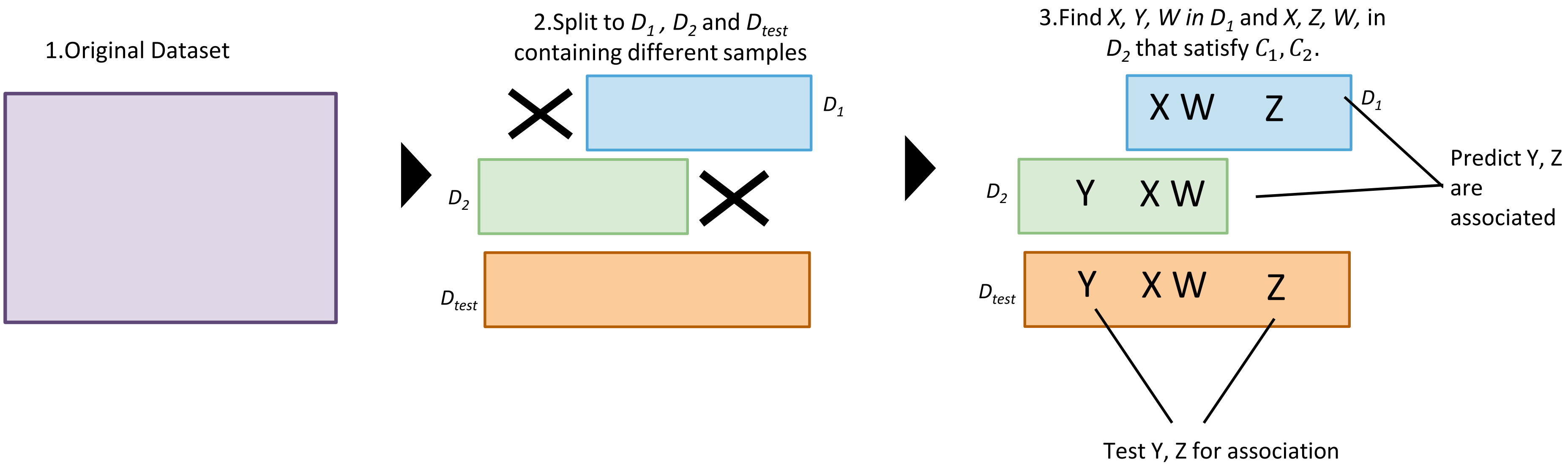
Find data sets D_1, D_2 measuring overlapping variables

Look for patterns C_1, C_2.

Predict that Y, Z are associated.

Find a third data set measuring both Y, Z and test if they are associated.

TEST IF IT WORKS ON REAL DATA (SIMULATE SCENARIO)



Restrict inferences only to cases where the probability of errors is small, i.e. *p*-values are extreme.

$p_{XY.Z} < 0.05$ accept $Dep(X, Y|\mathbf{Z})$
$p_{XY.Z} > 0.3$ accept $Ind(X, Y|\mathbf{Z})$
Else, undecided (forgo making any inferences)

[Tsamardinos, Triantafillou and Lagani, JMLR 2012]

DATASETS

Name	# instances	# variables	Group Size	Variables type	Scientific domain
Covtype	581012	55	55	Nominal/Ordinal	Agricultural
Read	681	26	26	Nominal/Continuous/Ordinal	Business
Infant-mortality	5337	83	83	Nominal	Clinical study
Compactiv	8192	22	22	Continuous	Computer science
Gisette	7000	5000	50	Continuous	Digit recognition
Hiva	4229	1617	50	Nominal	Drug discovering
Breast-Cancer	286	17816	50	Continuous	Gene expression
Lymphoma	237	7399	50	Continuous	Gene expression
Wine	4898	12	12	Continuous	Industrial
Insurance-C	9000	84	84	Nominal/Ordinal	Insurance
Insurance-N	9000	86	86	Nominal/Ordinal	Insurance
p53	16772	5408	50	Continuous	Protein activity
Ovarian	216	2190	50	Continuous	Proteomics
C&C	1994	128	128	Continuous	Social science
ACPJ	15779	28228	50	Continuous	Text mining
Bibtex	7395	1995	50	Nominal	Text mining
Delicious	16105	1483	50	Nominal	Text mining
Dexter	600	11035	50	Nominal	Text mining
Nova	1929	12709	50	Nominal	Text mining
Ohsumed	5000	14373	50	Nominal	Text mining

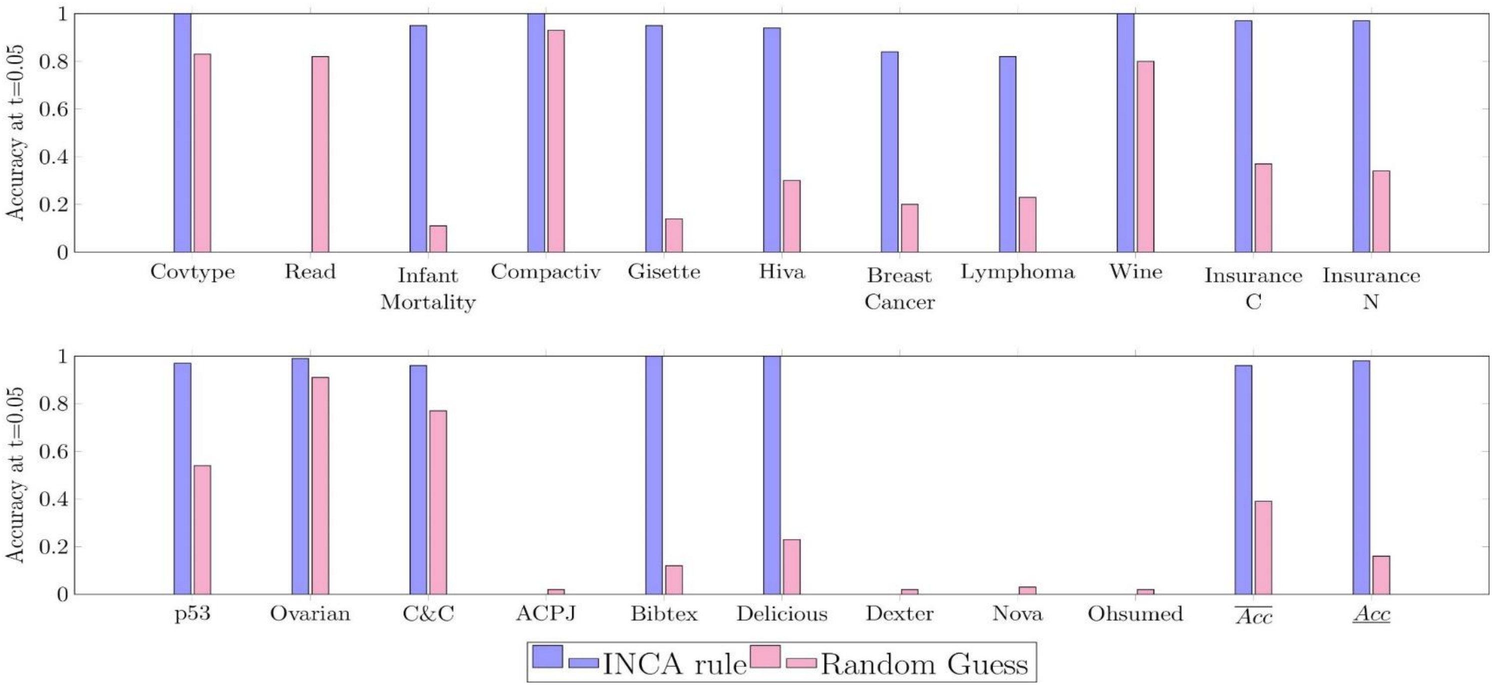
[Tsamardinos, Triantafillou and Lagani, JMLR 2012]

DATASETS

Name	# instances	# variables	Group Size	Variables type	Scientific domain	# predictions
Covtype	581012	55	55	Nominal/Ordinal	Agricultural	222
Read	681	26	26	Nominal/Continuous/Ordinal	Business	0
Infant-mortality	5337	83	83	Nominal	Clinical study	22
Compactiv	8192	22	22	Continuous	Computer science	135
Gisette	7000	5000	50	Continuous	Digit recognition	423
Hiva	4229	1617	50	Nominal	Drug discovering	554
Breast-Cancer	286	17816	50	Continuous	Gene expression	1833
Lymphoma	237	7399	50	Continuous	Gene expression	7712
Wine	4898	12	12	Continuous	Industrial	4
Insurance-C	9000	84	84	Nominal/Ordinal	Insurance	1839
Insurance-N	9000	86	86	Nominal/Ordinal	Insurance	226
p53	16772	5408	50	Continuous	Protein activity	46647
Ovarian	216	2190	50	Continuous	Proteomics	539165
C&C	1994	128	128	Continuous	Social science	99241
ACPJ	15779	28228	50	Continuous	Text mining	0
Bibtex	7395	1995	50	Nominal	Text mining	1
Delicious	16105	1483	50	Nominal	Text mining	856
Dexter	600	11035	50	Nominal	Text mining	0
Nova	1929	12709	50	Nominal	Text mining	0
Ohsumed	5000	14373	50	Nominal	Text mining	0

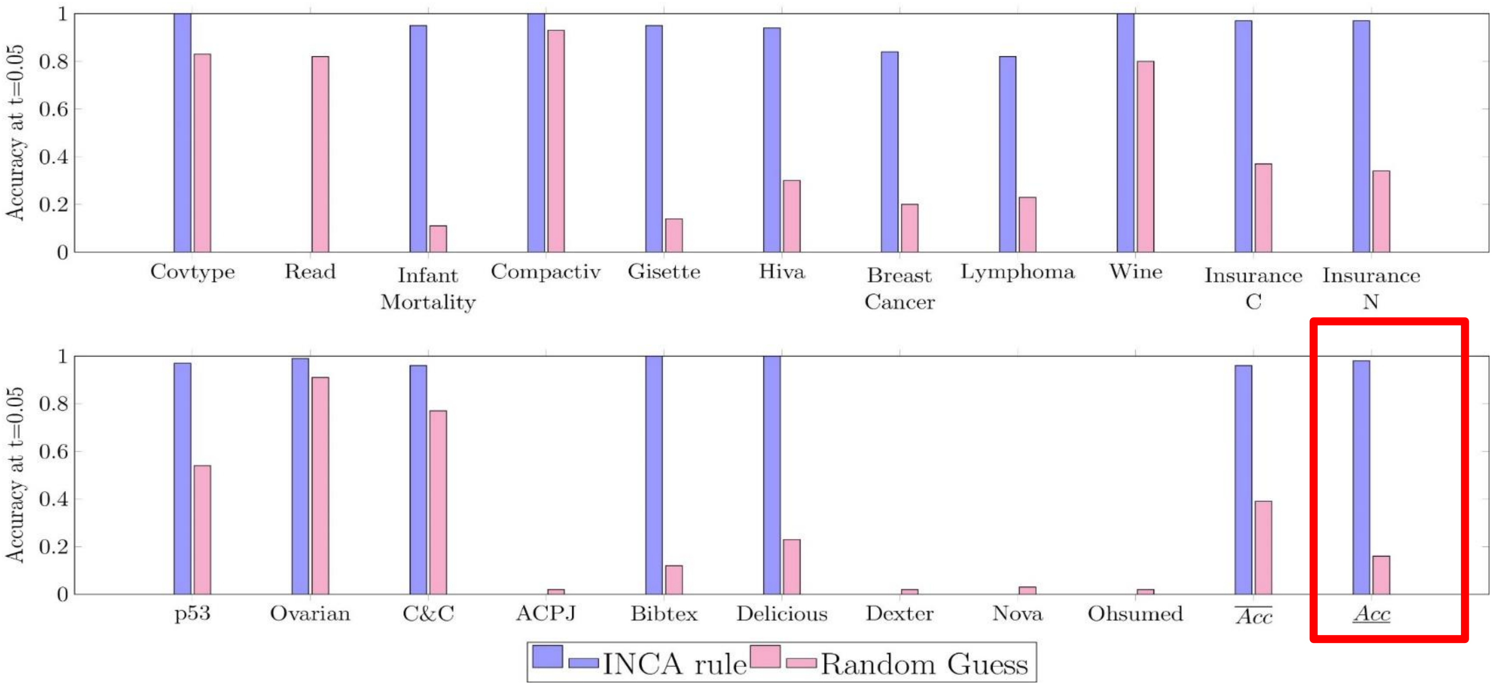
[Tsamardinos, Triantafillou and Lagani, JMLR 2012]

HOW DID WE DO?



- About 700000 predictions in 20 datasets.
- Accuracy: The percentage of p-values < 0.05.
    - May include false positives and exclude false negatives.

HOW DID WE DO?

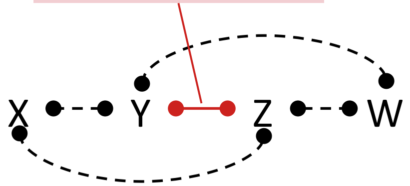


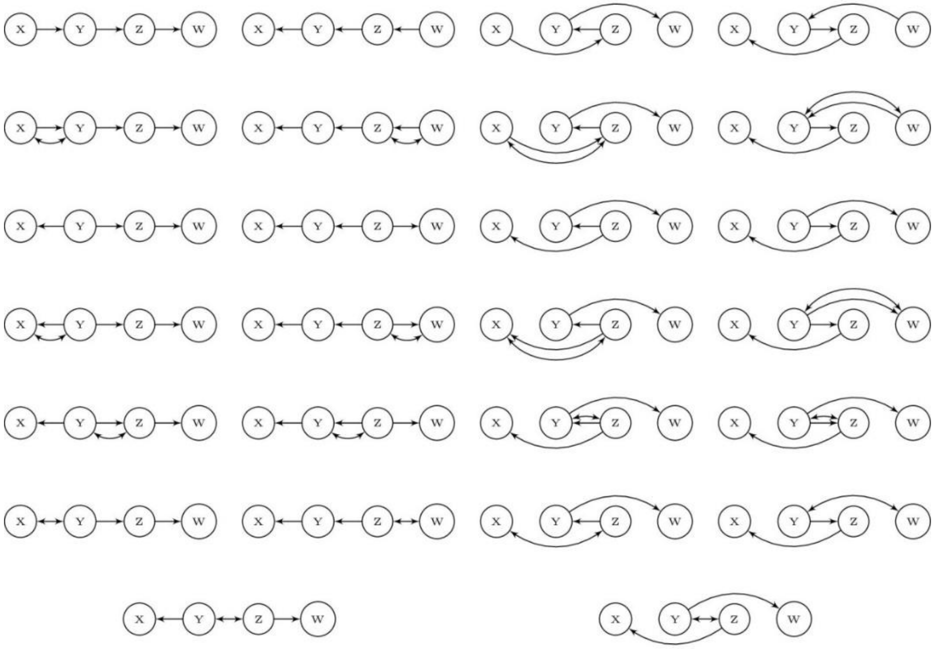
- About 700000 predictions in 20 datasets.
- Accuracy: The percentage of p-values < 0.05.
  - May include false positives and exclude false negatives.

98% accuracy vs.
16% of random guessing

PREDICT CORRELATION STRENGTH ρ_{YZ}

How strong is the correlation of Y and Z?

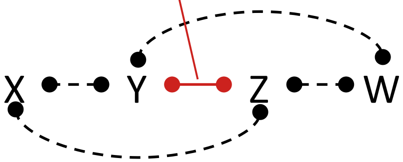




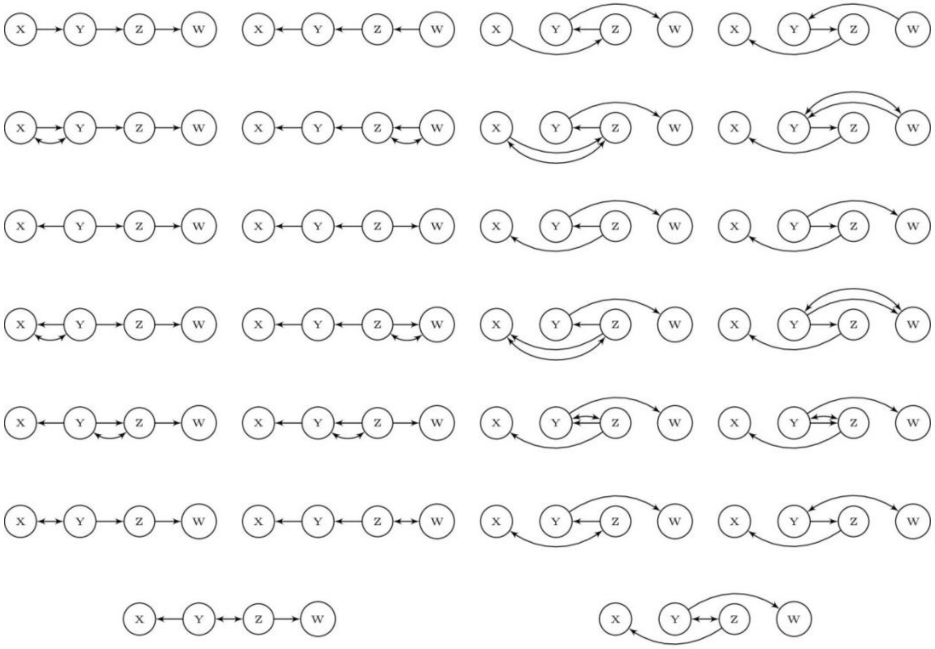
26 possible SMCGs.

PREDICT CORRELATION STRENGTH ρ_{YZ}

How strong is the correlation of Y and Z?



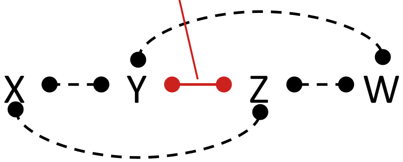
- Assume multivariate normality and interpret SMCG as path diagram.
- Use the (measured) sample correlations
  - r_{YX}, r_{YW}, r_{XW} (D_1)
  - r_{ZX}, r_{ZW}, r_{XW} (D_2)
- Use rules of path analysis to predict $\widehat{r_{YZ}}$.



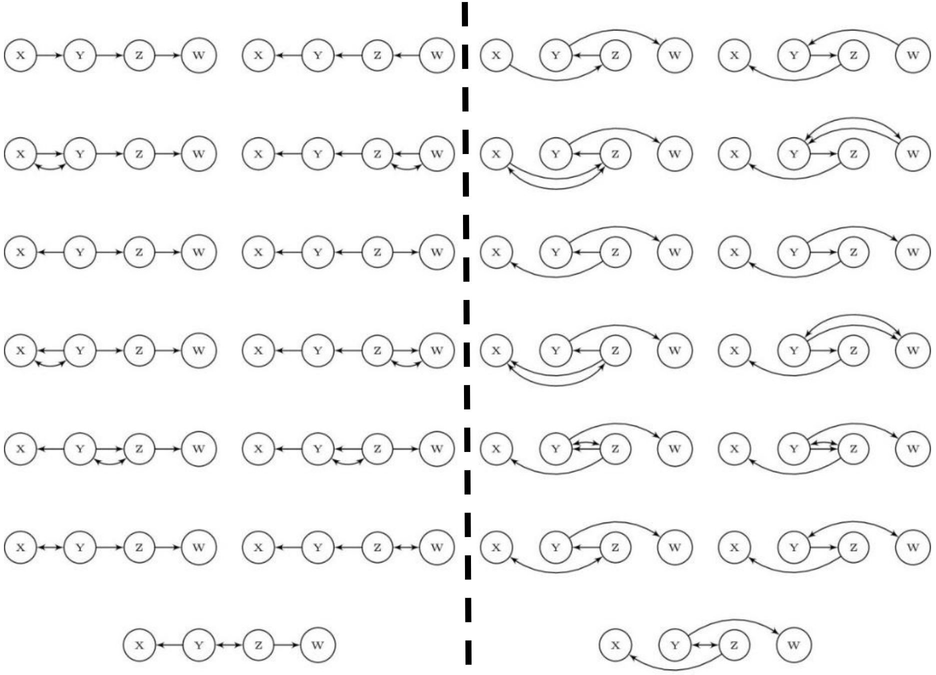
26 possible SMCGs.

PREDICT CORRELATION STRENGTH ρ_{YZ}

How strong is the correlation of Y and Z?



- Assume multivariate normality and interpret SMCG as path diagram.
- Use the (measured) sample correlations
  - r_{YX}, r_{YW}, r_{XW} (D_1)
  - r_{ZX}, r_{ZW}, r_{XW} (D_2)
- Use rules of path analysis to predict $\widehat{r_{YZ}}$.



13 models imply

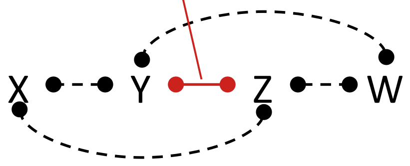
$$\widehat{r_{YZ}}^{1} \approx \frac{1}{2}\left(\frac{r_{XZ}}{r_{XY}} + \frac{r_{YW}}{r_{ZW}}\right)$$

13 models imply

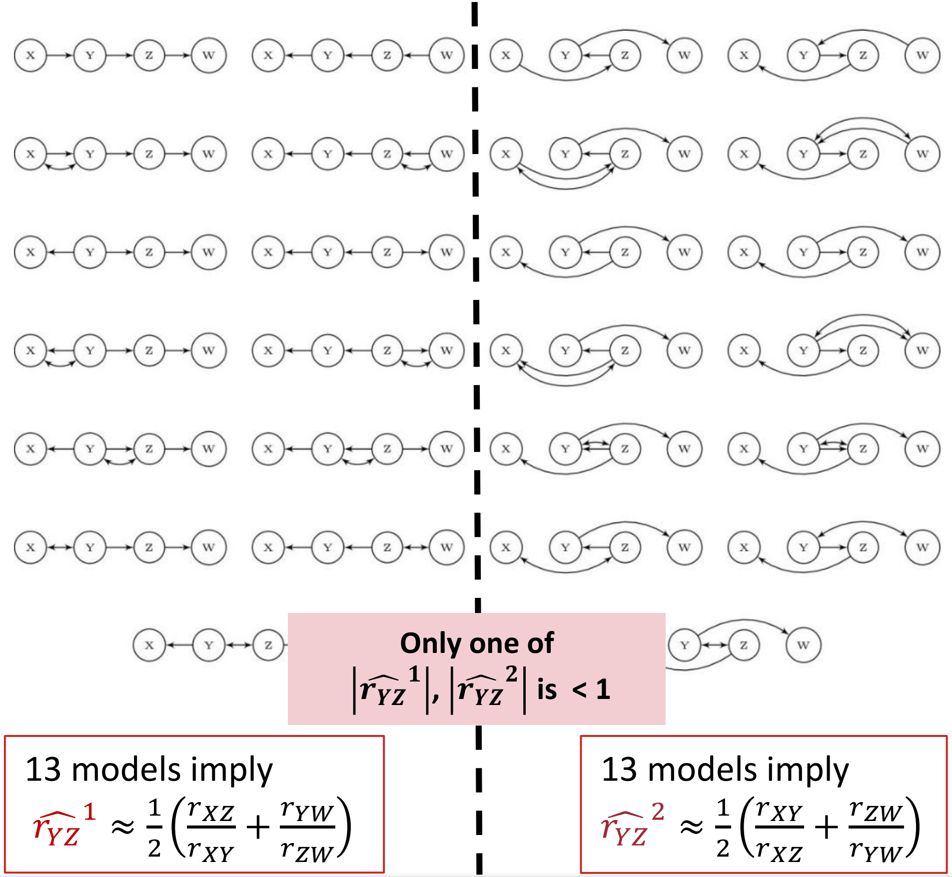
$$\widehat{r_{YZ}}^{2} \approx \frac{1}{2}\left(\frac{r_{XY}}{r_{XZ}} + \frac{r_{ZW}}{r_{YW}}\right)$$

PREDICT CORRELATION STRENGTH ρ_{YZ}

How strong is the correlation of Y and Z?



- Assume multivariate normality and interpret SMCG as path diagram.
- Use the (measured) sample correlations
  - r_{YX}, r_{YW}, r_{XW} (D_1)
  - r_{ZX}, r_{ZW}, r_{XW} (D_2)
- Use rules of path analysis to predict $\widehat{r_{YZ}}$.



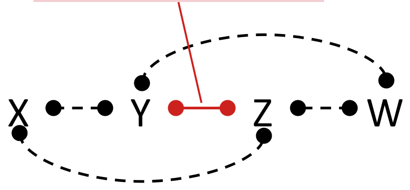
Only one of $\left|\widehat{r_{YZ}}^1\right|, \left|\widehat{r_{YZ}}^2\right|$ **is < 1**

13 models imply
$$\widehat{r_{YZ}}^1 \approx \frac{1}{2}\left(\frac{r_{XZ}}{r_{XY}} + \frac{r_{YW}}{r_{ZW}}\right)$$

13 models imply
$$\widehat{r_{YZ}}^2 \approx \frac{1}{2}\left(\frac{r_{XY}}{r_{XZ}} + \frac{r_{ZW}}{r_{YW}}\right)$$

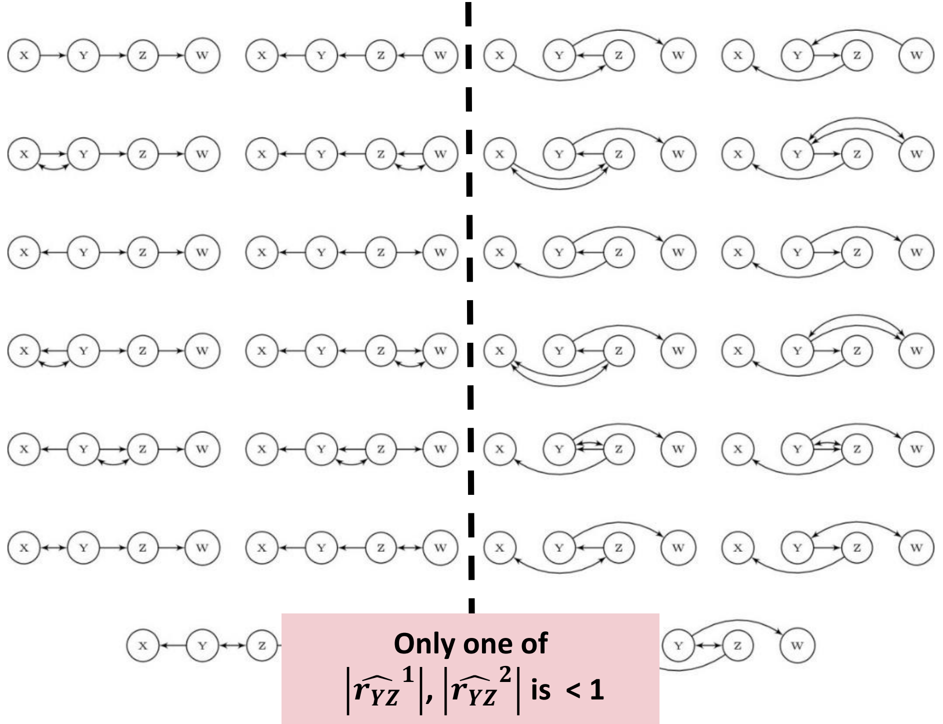
PREDICT CORRELATION STRENGTH ρ_{YZ}

How strong is the correlation of Y and Z?



- Assume multivariate normality and interpret SMCG as path diagram.
- Use the (measured) sample correlations
  - r_{YX}, r_{YW}, r_{XW} (D_1)
  - r_{ZX}, r_{ZW}, r_{XW} (D_2)
- Use rules of path analysis to predict $\widehat{r_{YZ}}$.

You can uniquely identify the skeleton of the graph AND predict the correlation coefficient of Y, Z!

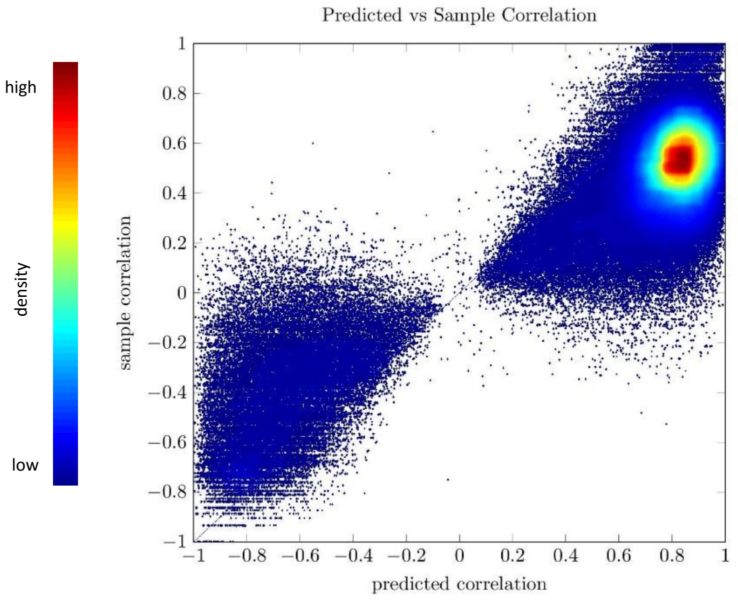


Only one of $\left|\widehat{r_{YZ}}^1\right|, \left|\widehat{r_{YZ}}^2\right|$ is < 1

13 models imply
$\widehat{r_{YZ}}^1 \approx \frac{1}{2}\left(\frac{r_{XZ}}{r_{XY}} + \frac{r_{YW}}{r_{ZW}}\right)$

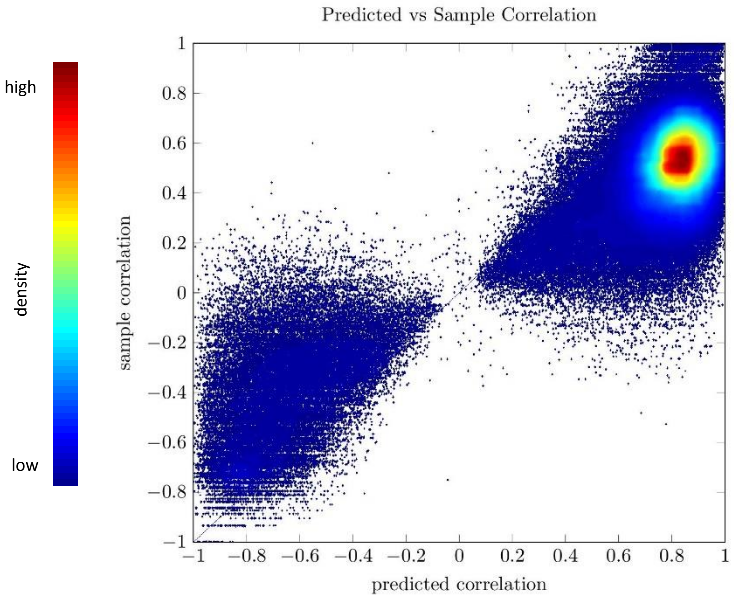
13 models imply
$\widehat{r_{YZ}}^2 \approx \frac{1}{2}\left(\frac{r_{XY}}{r_{XZ}} + \frac{r_{ZW}}{r_{YW}}\right)$

HOW DID WE DO?

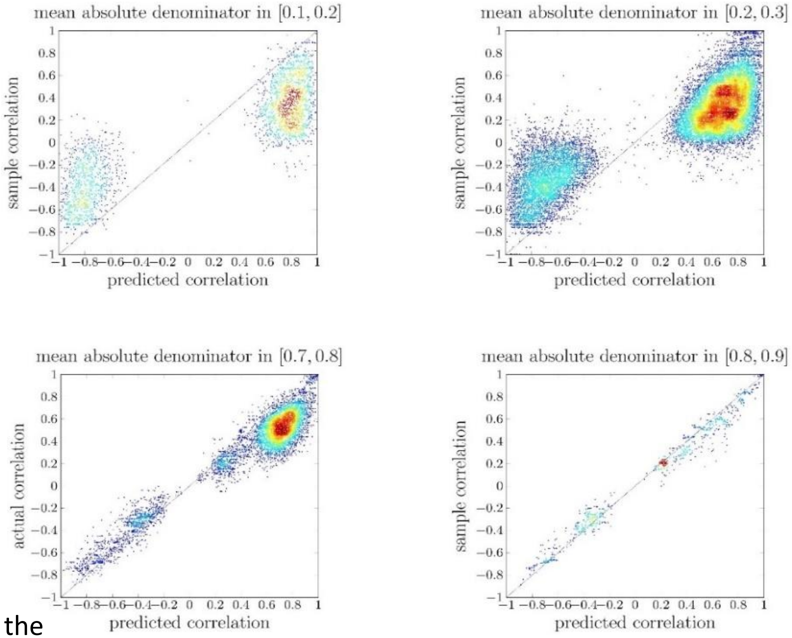


- Clear trend in predicted vs sample correlations.
- Also a systematic bias because the predictions have been selected based on the independence tests.
- Correlation of predicted vs sample correlations is 0.89.
- Predictions based on large correlations have reduced bias.

HOW DID WE DO?



Predicted vs sample correlations over all data sets, grouped by mean absolute value of the denominators used in their computations



- Clear trend in predicted vs sample correlations
- Also a systematic bias because the predictions have been selected based on the independence tests
- Correlation of predicted vs sample correlations is 0.89
- Predictions based on large correlations have reduced bias.

QUESTIONS

-How can you reason with this set of models?

You can use do-calculus and estimate (a population of) causal effects.

-Is it useful? Do you make additional inferences than analyzing each data set in isolation?

You can make non-trivial inferences, quantitative with additional assumptions.

OUTLINE

1. ~~Integrative causal discovery~~
   i. ~~Motivation.~~
   ii. ~~Causal models.~~
   iii. ~~m-separation.~~
   iv. ~~Reverse engineering causal models (single data set).~~
   v. ~~Problem formulation: Reverse engineering causal models from multiple heterogeneous data sets.~~
   vi. ~~Modeling interventions/selection.~~
2. Logic-based causal discovery
   i. ~~Converting path constraints to logic formulae.~~
   ii. ~~Problem complexity.~~
   iii. ~~Conflict resolution.~~
   iv. ~~Existing algorithms.~~
   v. ~~Reasoning with logic based causal discovery.~~
   vi. ~~Non-trivial inferences-validation.~~

KEY-POINTS

Integrative logic-based causal discovery.

Different data distributions, same causal mechanism: use causal modeling to connect.

Can handle datasets of different variable sets, different experimental conditions, prior causal knowledge.

Identify the set of causal graphs that simultaneously fit all datasets and reason with this set.

Convert problem to SAT or ASP; exploit 40 years of SAT-solving technology.

Query-based approach to avoid explosion of possible solutions!

Vision of automatically analyzing a large portion of available datasets in a domain.

WHAT IS NEXT?

Improving scalability.

Improving quality of learning and robustness.

Further removing restrictive assumptions (e.g., Faithfulness).

Making quantitative predictions.

Extensions for temporal data.

Additional constraints (e.g. Verma constraints).

Feature selection from multiple data sets.

Apply it to real problems.

REFERENCES

1. D. Margaritis and F. Bromberg, **Efficient Markov Network Discovery Using Particle Filters**, *Computational Intelligence (2009).*
2. S. Triantafillou, I Tsamardinos and IG Tollis**, Learning Causal Structure from Overlapping Variable Sets**, *AISTATS 2010*.
3. G. Borboudakis, S. Triantafillou, V. Lagani, I. Tsamardinos, **A Constraint-based Approach to Incorporating Prior Knowledge in Causal Models**, *ESANN 2011.*
4. T. Claassen and T. Heskes, **A Logical Characterization of Constraint-based Causal Discovery**. *UAI 2011.*
5. T. Claassen and T. Heskes, **A Bayesian Approach to Constraint-based Causal Inference**, *UAI 2012*.
6. I. Tsamardinos, S. Triantafillou, V. Lagani, **Towards Integrative Causal Analysis of Heterogeneous Data Sets and Studies,** *Journal of Machine Learning Research (2012).*
7. A. Hyttinen, P. O. Hoyer, F. Eberhardt, and M. Järvisalo, **Discovering Cyclic Causal Models with Latent Variables: A General SAT-Based Procedure ,** *UAI 2013.*

REFERENCES(2)

8. A. Hyttinen, F. Eberhardt, and M. Järvisalo, **Constraint-based Causal Discovery: Conflict Resolution with Answer Set Programming,** *UAI 2014.*
9. S. Triantafillou, I. Tsamardinos, A. Roumpelaki, **Learning Neighborhoods of High Confidence in Constraint-Based Causal Discovery**, *PGM 2014.*
10. S. Triantafillou and I. Tsamardinos, **Constraint-based Causal Discovery from Multiple Interventions over Overlapping Variable Sets**, *Journal of Machine Learning Research (2015).*
11. D. Sonntag, M. Järvisalo, Jose M. Pena, A. Hyttinen, **Learning Optimal Chain Graphs with Answer Set Programming,** *UAI 2015.*
12. A. Hyttinen, S. Plis, M. Järvisalo, F. Eberhardt, and D. Danks, **Causal Discovery from Subsampled Time Series Data by Constraint Optimization,** *submitted*.
13. G. Borboudakis and I. Tsamardinos. **Towards Robust and Versatile Causal Discovery for Business Applications**. *KDD 2016*.
14. S. Magliacane, T. Claassen, J.M. Mooij, **Ancestral Causal Inference**, *arXiv:1606.07035*

ACKNOWLEDGEMENTS

Mens x machina group, University of Crete.
Jan Lemeire, Frederick Eberhardt, Antti Hyttinen, Joris Mooij
ERC CAUSALPATH