Title: Entropic Inequality Constraints from e-separation Relations in Directed Acyclic Graphs with Hidden Variables

Speakers: Beata Zjawin

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 17, 2023 - 4:30 PM

URL: https://pirsa.org/23040109

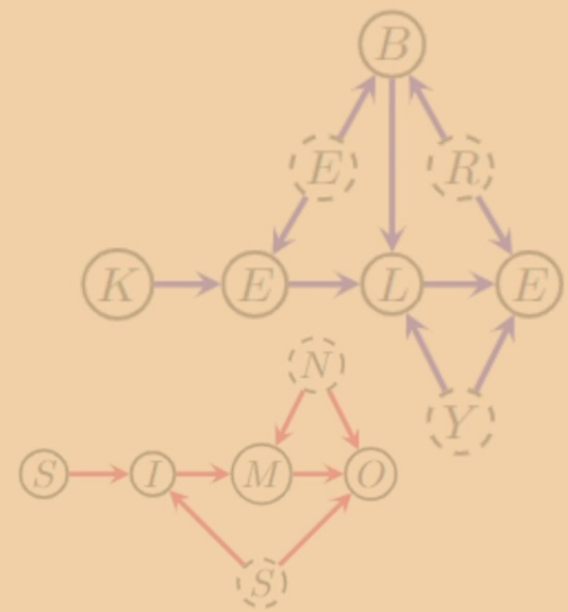
Abstract: When some variables in a directed acyclic graph (DAG) are hidden, a notoriously complicated set of constraints on the distribution of observed variables is implied. In this talk, we present inequality constraints implied by graphical criteria in hidden variable DAGs. The constraints can intuitively be understood to follow from the fact that the capacity of variables along a causal pathway to convey information is restricted by their entropy. For DAGs that exhibit e-separation relations, we present entropic inequality constraints and we show how they can be used to learn about the true causal model from an observed data distribution (arXiv:2107.07087).

Entropic Inequality Constraints from e-separation in Directed Acyclic Graphs with Hidden Variables

Beata Zjawin[1]

Noam Finkelstein[2], Elie Wolfe[3], Ilya Shpitser[2], Robert Spekkens[3]

[1] ICTQT, University of Gdańsk
[2] Johns Hopkins University, Department of Computer Science
[3] Perimeter Institute for Theoretical Physics



Equality constraints

Equality constraints implied by DAGs

- $\mathcal{G} \longrightarrow P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(V \mid pa_{\mathcal{G}}(V))$

d-separation

Fork

$x \leftarrow z \rightarrow y$

Collider

$x \rightarrow z \leftarrow y$

Chain

$x \rightarrow z \rightarrow y$

d-separation

Fork	Collider	Chain
$x \longleftarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$	$x \longrightarrow z \longrightarrow y$

- Forks and chains are said to be open if we do not condition on Z, and closed otherwise;
- Colliders are said to be open if we do condition on Z or its descendants, and closed otherwise;
- A path is open under a conditioning set Z if all contiguous triples along that path are open under that conditioning set.

d-separation

Fork	Collider	Chain
$x \longleftarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$	$x \longrightarrow z \longrightarrow y$

- Forks and chains are said to be open if we do not condition on Z, and closed otherwise;
- Colliders are said to be open if we do condition on Z or its descendants, and closed otherwise;
- A path is open under a conditioning set Z if all contiguous triples along that path are open under that conditioning set.

Let A, B and C be sets of variables in a DAG. A and B are said to be d-separated by C if all paths between A and B are closed after conditioning on C.

$$(\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C})$$

Equality constraints implied by DAGs

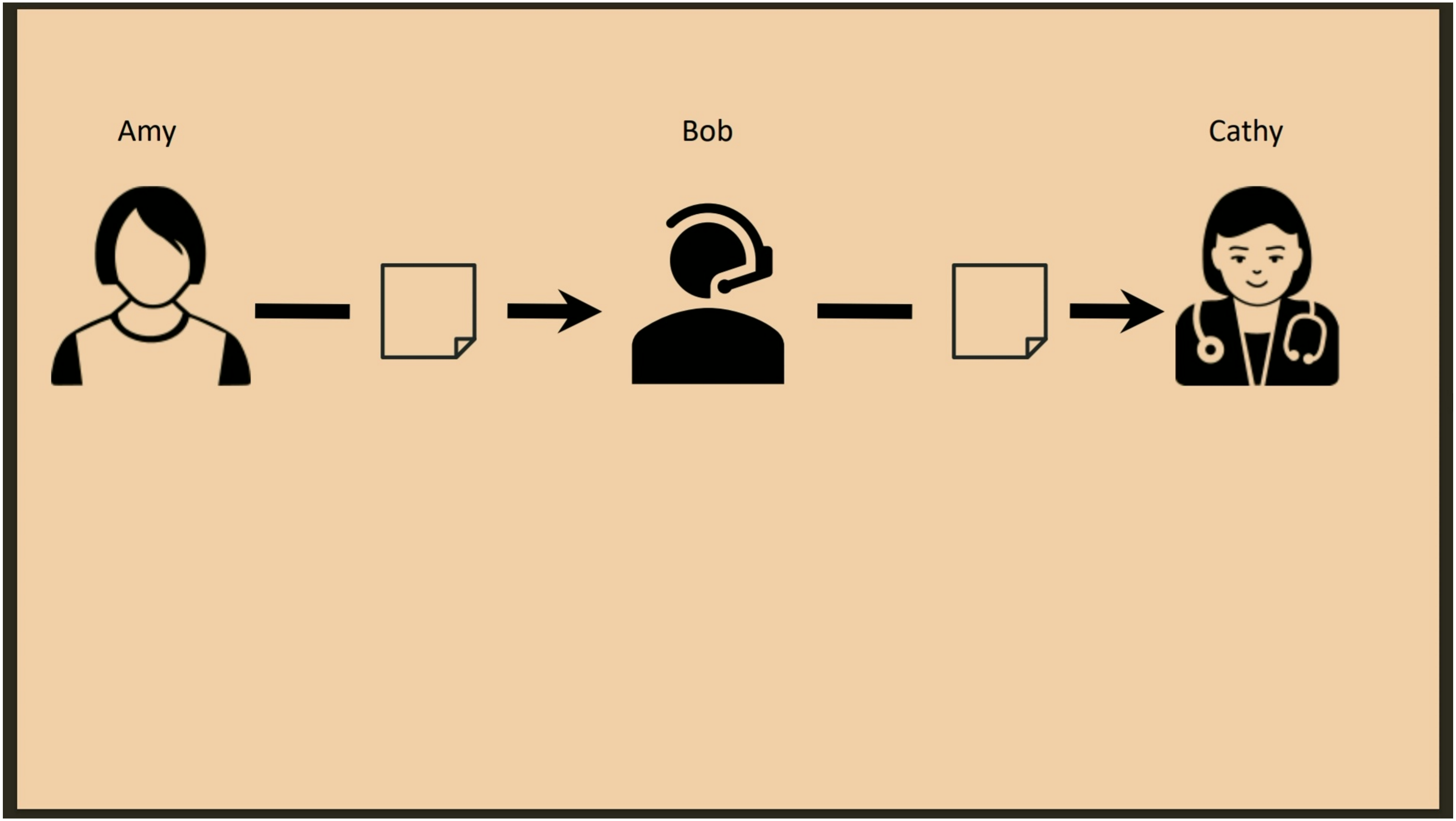
- $\mathcal{G} \longrightarrow P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(V \mid pa_{\mathcal{G}}(V))$
- $(\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C}) \longrightarrow \mathbf{A} \perp \mathbf{B} \mid \mathbf{C}$

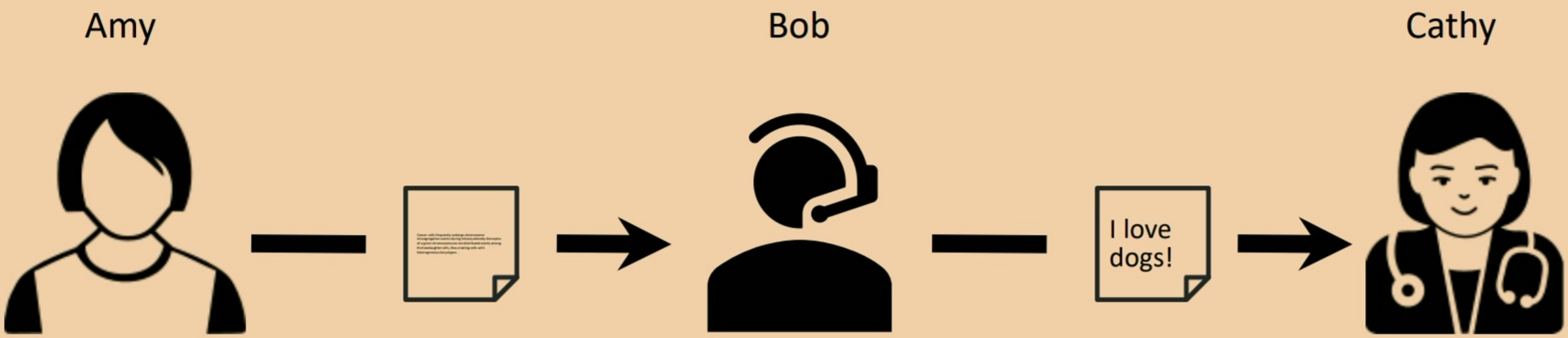
How to study causal models with

hidden variables?

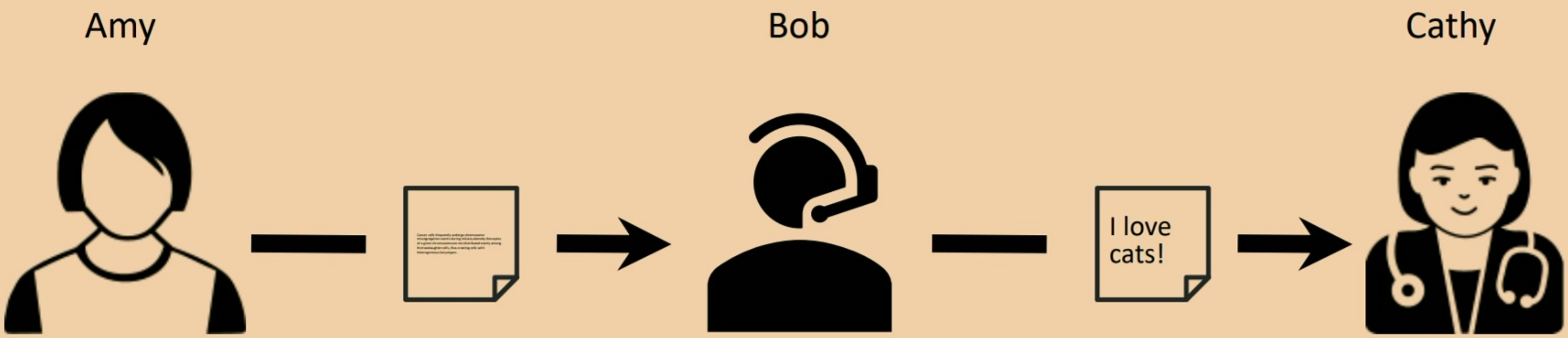
Constraints in hidden variable models

- Quantifier elimination algorithms

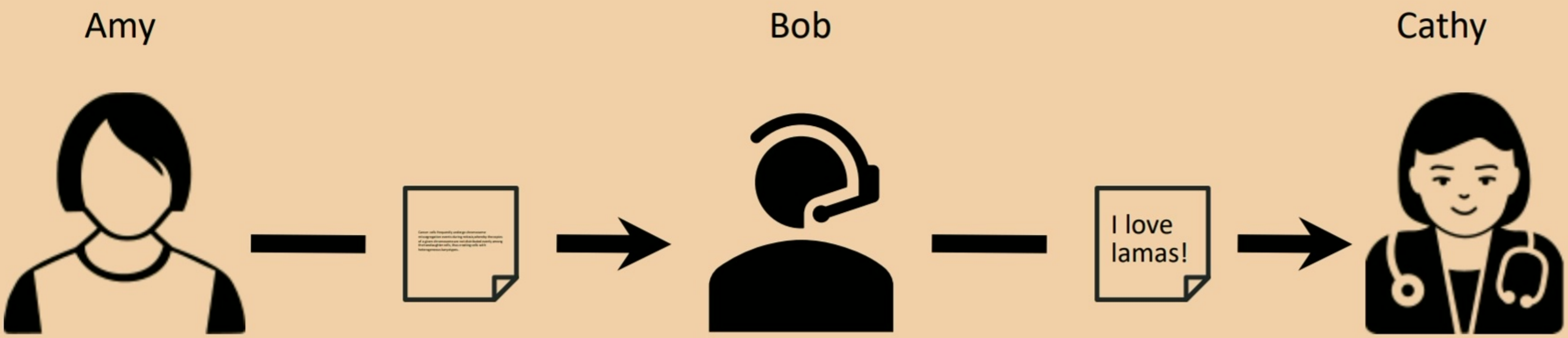

Amy
Bob
Cathy



- If Bob only ever sends the same one message, regardless of what he gets from Amy, Cathy can't find anything out about Amy's note from Bob.
- Bob's notes have zero **Entropy** $H(X) \equiv -\sum_{x \in \mathcal{X}} P(x) \log_2 P(x)$



- As $H(B)$ increases — there is variety in Bob's notes — the *potential* for Cathy to learn about Amy's note from Bob's note increases.
- But there are no guarantees — Bob may be sending Cathy nonsense.



- The **information** shared between Amy and Cathy is bounded from above by the entropy of Bob's notes:

$$I(X:Y) \equiv H(X) + H(Y) - H(X,Y)$$

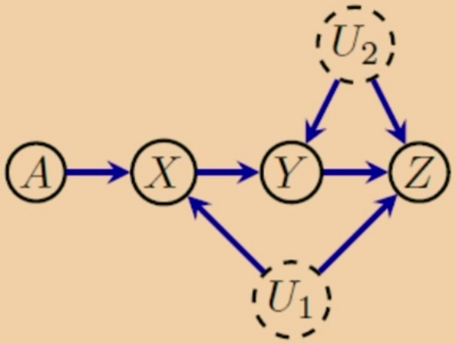
$$I(A:C) \leq H(B)$$

Intuition about passing information

- The amount of “fluid”(information) that can get through a bottleneck cannot exceed its “size” (entropy).

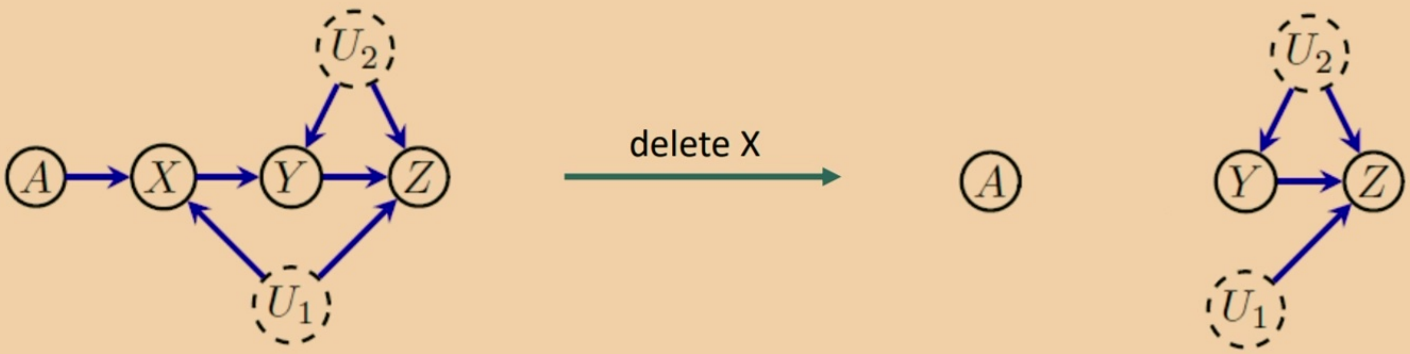
What is a bottleneck?

- Bottleneck variables (between A and Z) - variables that are between A and Z along some path



What is a bottleneck?

- Bottleneck variables (between A and Z) - variables that are between A and Z along some path



e-separation

- A node can be deleted from a graph by removing the node and all of its incoming and outgoing edges.

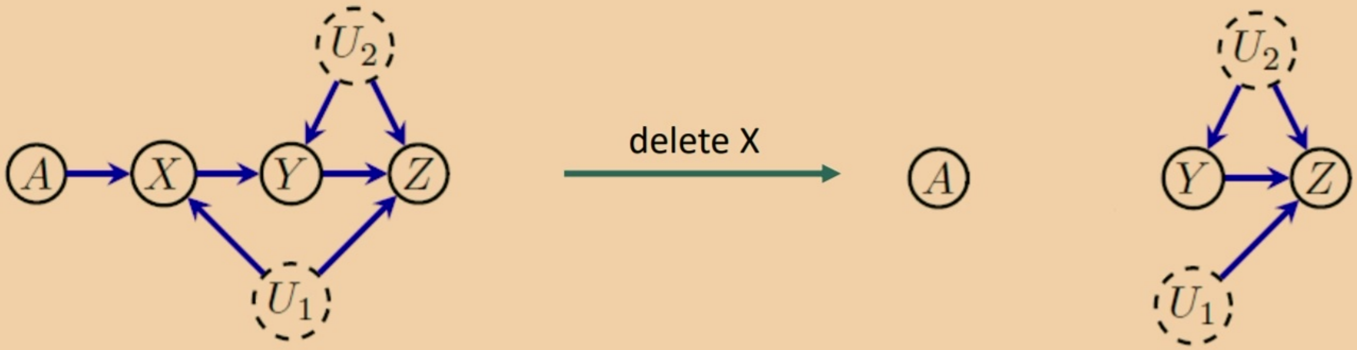
Let A, B, C and D be sets of variables in a DAG. A and B are said to be e-separated by C after deletion of D if A and B are d-separated by C after deletion of every variable in D.

$$(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C} \; \textit{upon} \; \neg\mathbf{D})$$

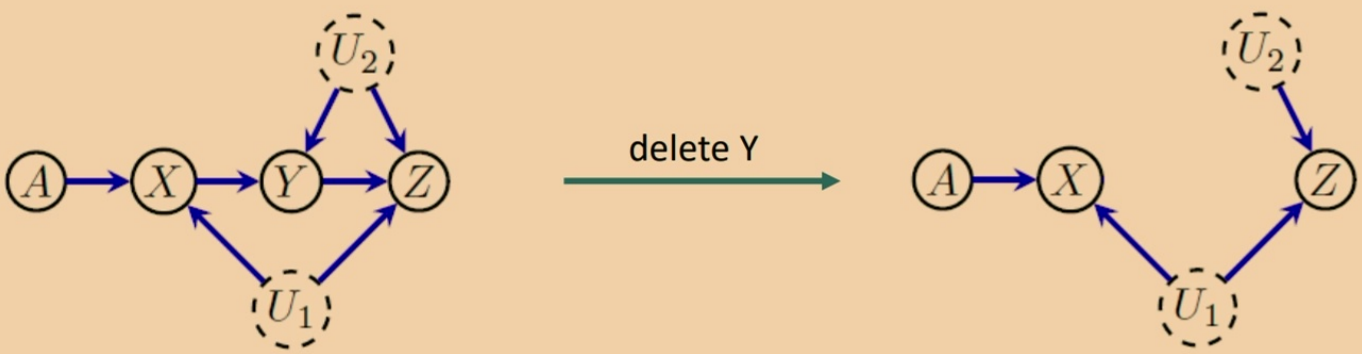
D is a bottleneck for A and B conditional on C

- All information shared between A and B must flow through D

e-separation



$A \perp_e YZ$ *upon* $\neg X$



$A \perp_e Z$ *upon* $\neg Y$

The amount of information that can get through a bottleneck cannot exceed its entropy.

We can formalize bottlenecks using e-separation.

Entropic constraints from e-separation

Theorem. *Suppose observed variables are discrete. If* $(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C}\ \mathit{upon}\ \neg\mathbf{D})$ *and no element of C is a descendant of any in D, then for any value c in the domain of C, the following constraints hold:*

$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}{=}\mathbf{c}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

If in addition, no element of A is a descendant of any in D, then for any value c in the domain of C, the following stronger constraints hold:

$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}{=}\mathbf{c}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

Entropic Inequality Constraints from e-separation Relations in Directed Acyclic Graphs with Hidden Variables
Noam Finkelstein, Beata Zjawin, Elie Wolfe, Ilya Shpitser, Robert Spekkens (UAI 2021) arxiv: 2107.07087

Entropic constraints from e-separation

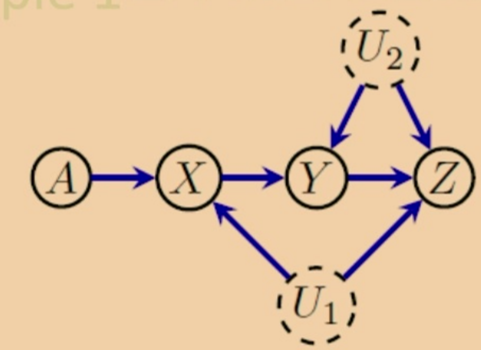
Theorem. *Suppose observed variables are discrete. If* $(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C}\ upon\ \neg\mathbf{D})$ *and no element of C is a descendant of any in D, then for any value c in the domain of C, the following constraints hold:*

$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}{=}\mathbf{c}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

If in addition, no element of A is a descendant of any in D, then for any value c in the domain of C, the following stronger constraints hold:

$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}{=}\mathbf{c}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

Example 1



$$I(A : XYZ) \leq H(X)$$
$$I(A : YZ) \leq H(Y)$$

Entropic constraints from e-separation

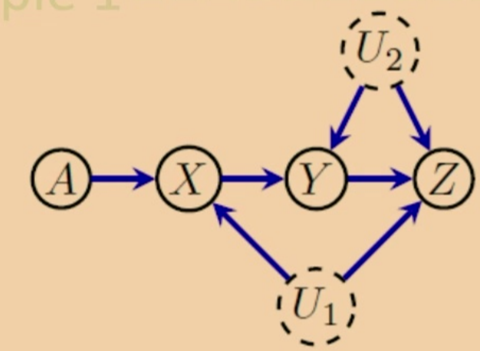
Theorem. *Suppose observed variables are discrete. If* $(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C}\ upon\ \neg\mathbf{D})$ *and no element of C is a descendant of any in D, then for any value c in the domain of C, the following constraints hold:*

$$I(\mathbf{A}:\mathbf{B} \mid \mathbf{C}{=}\mathbf{c}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A}:\mathbf{B} \mid \mathbf{C}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

If in addition, no element of A is a descendant of any in D, then for any value c in the domain of C, the following stronger constraints hold:

$$I(\mathbf{A}:\mathbf{B}, \mathbf{D} \mid \mathbf{C}{=}\mathbf{c}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A}:\mathbf{B}, \mathbf{D} \mid \mathbf{C}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

Example 1



$$I(A:XYZ) \leq H(X)$$
$$I(A:YZ) \leq H(Y)$$

We recover all Shannon-type entropic inequality constraints implied by the graph

Entropic constraints from e-separation

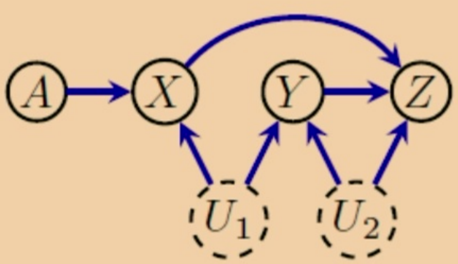
Theorem. *Suppose observed variables are discrete. If* $(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C}\ upon\ \neg\mathbf{D})$ *and no element of C is a descendant of any in D, then for any value c in the domain of C, the following constraints hold:*

$$I(\mathbf{A}:\mathbf{B} \mid \mathbf{C}{=}\mathbf{c}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A}:\mathbf{B} \mid \mathbf{C}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

If in addition, no element of A is a descendant of any in D, then for any value c in the domain of C, the following stronger constraints hold:

$$I(\mathbf{A}:\mathbf{B},\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A}:\mathbf{B},\mathbf{D} \mid \mathbf{C}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

Example 2



$$I(A:XYZ) \leq H(X)$$
$$I(A:YZ|X) \leq H(Y|X)$$

Entropic constraints from e-separation

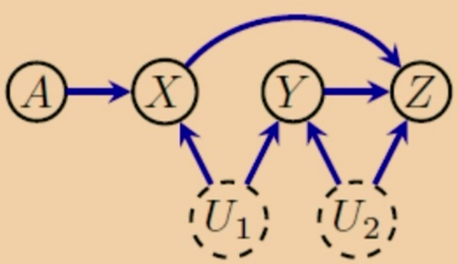
Theorem. *Suppose observed variables are discrete. If* $(\mathbf{A} \perp_e \mathbf{B} \mid \mathbf{C} \textit{ upon } \neg\mathbf{D})$ *and no element of C is a descendant of any in D, then for any value c in the domain of C, the following constraints hold:*

$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}{=}\mathbf{c}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B} \mid \mathbf{C}, \mathbf{D}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

If in addition, no element of A is a descendant of any in D, then for any value c in the domain of C, the following stronger constraints hold:

$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}{=}\mathbf{c}) \leq H(\mathbf{D} \mid \mathbf{C}{=}\mathbf{c}),$$
$$I(\mathbf{A} : \mathbf{B}, \mathbf{D} \mid \mathbf{C}) \leq H(\mathbf{D} \mid \mathbf{C}).$$

Example 2



$$I(A : XYZ) \leq H(X)$$
$$I(A : YZ|X) \leq H(Y|X)$$

Can be strengthen:

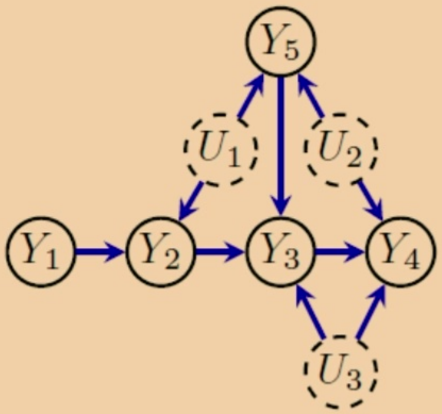
$$I(A : XYZ) \leq H(X|Y)$$

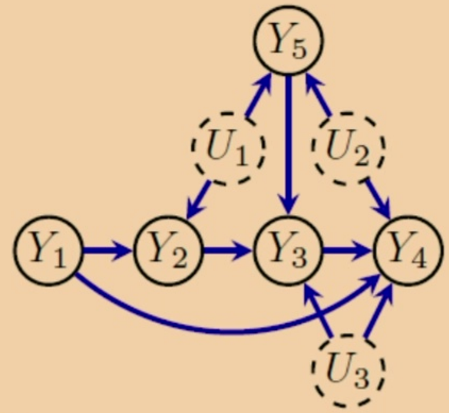
Related results

- It is possible to relate our inequality constraints to equality constraints (to the d-separation-based conditional independence and Verma constraints (in identified post-intervention distributions)).

Proposition. *If A is d-separated from B by {C,D}, then A is also e-separated from B by C upon deleting D.*

Causal discovery





$$Y_1 \perp Y_3 \mid Y_2 Y_5$$

$$Y_1 \perp Y_5$$

$$(Y_1 \perp_e Y_3 Y_4 \mid Y_2 \text{ upon } \neg Y_5)$$

$$(Y_1 Y_2 \perp_e Y_4 \mid \text{ upon } \neg Y_3)$$

$$(Y_2 \perp_e Y_4 \mid Y_1 \text{ upon } \neg Y_3)$$

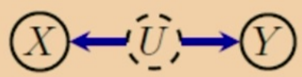
$$(Y_1 \perp_e Y_3 \mid Y_2 \text{ upon } \neg Y_5)$$

$$(Y_2 \perp_e Y_4 \mid Y_1 \text{ upon } \neg Y_3)$$

Latent variables

Observed data:

		Y			
		0	1	2	3
X	0	0.002	0.001	0.400	0.001
	1	0.003	0.005	0.005	0.066
	2	0.224	0.003	0.003	0.001
	3	0.002	0.281	0.001	0.002

Task: Decide between 

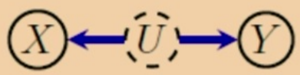
and 

Assumption: |U|=3

Latent variables

Observed data:

		Y			
		0	1	2	3
X	0	0.002	0.001	0.400	0.001
	1	0.003	0.005	0.005	0.066
	2	0.224	0.003	0.003	0.001
	3	0.002	0.281	0.001	0.002

Task: Decide between 

and 

Assumption: |U|=3

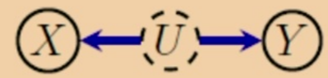
Proposition. *If* $(\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C}, \mathbf{U})$, *then* $|\mathbf{U}| \geq 2^{I(\mathbf{A}:\mathbf{B}|\mathbf{C})}$

$$2^{I(X:Y)} \approx 2^{1.594} \approx 3.018$$

Latent variables

Observed data:

		Y: 0	1	2	3
	0	0.002	0.001	0.400	0.001
X	1	0.003	0.005	0.005	0.066
	2	0.224	0.003	0.003	0.001
	3	0.002	0.281	0.001	0.002



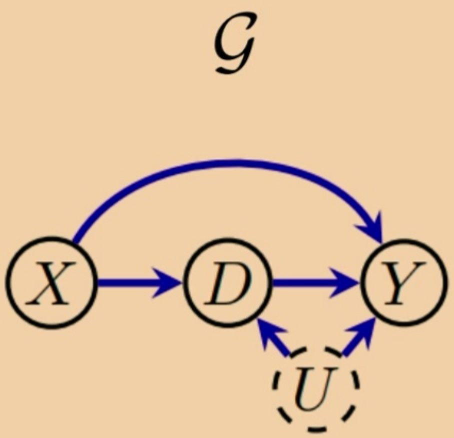
Task: Bound |U|

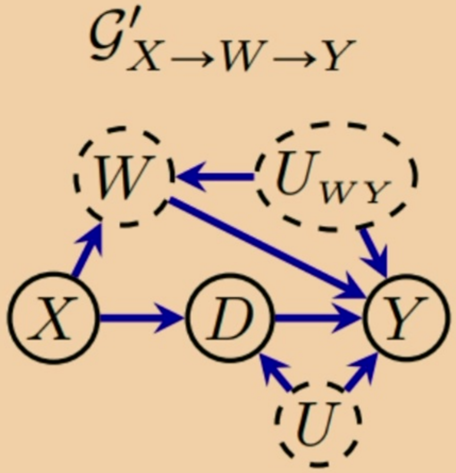
Quantifying causal influence

- Traditional approach: Average Causal Effect defined as $E[Y(X=x) - Y(X=x')]$

Quantifying causal influence

Minimal Mediary Entropy (MME) for direct causal effect:

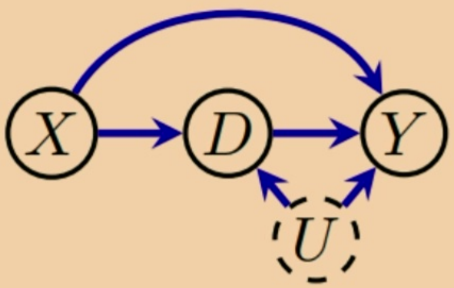




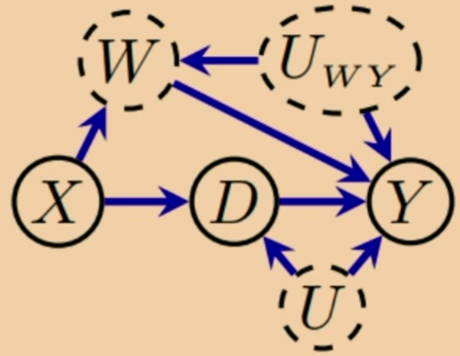
MME (X→Y) is the smallest entropy H(W) over all structural equation models reproducing the observed data distribution over the modified DAG in which W has finite cardinality.

Quantifying causal influence

$\mathcal{G}$



$\mathcal{G}'_{X\to W\to Y}$



MME (X→Y) is the smallest entropy H(W) over all structural equation models reproducing the observed data distribution over the modified DAG in which W has finite cardinality.

$$(\boldsymbol{A} \perp_e \boldsymbol{B} \,|\, \boldsymbol{C} \; upon \; \neg\{\boldsymbol{D}, W\}) \longrightarrow \text{MME}_{X\to Y}$$

$\boldsymbol{A} \subset \{X\} \cup an(X)$

$\boldsymbol{B} \subset \{Y\} \cup desc(Y)$

$$\geqslant \max_{\boldsymbol{c}} I(\boldsymbol{A}:\boldsymbol{B}\,|\,\boldsymbol{C}{=}\boldsymbol{c},\boldsymbol{D}) - H(\boldsymbol{D}\,|\,\boldsymbol{C}{=}\boldsymbol{c})$$

$$\geqslant I(\boldsymbol{A}:\boldsymbol{B}\,|\,\boldsymbol{C},\boldsymbol{D}) - H(\boldsymbol{D}\,|\,\boldsymbol{C}).$$