



Dependency Patterns For Latent Variable Discovery

By Xuhui Zhang, Kevin Korb, Ann Nicholson and Steven Mascaró



Presentation Layout

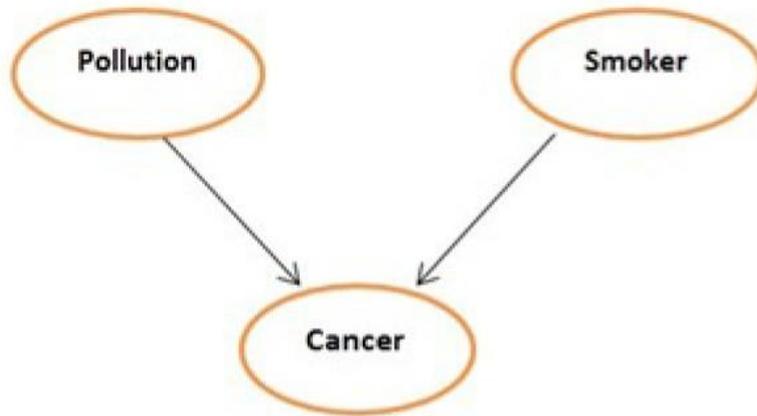
1. Background
2. Dependency Patterns (Triggers) Discovery for latent variable
3. Applying Triggers in causal discovery
4. Analysis & Future Work



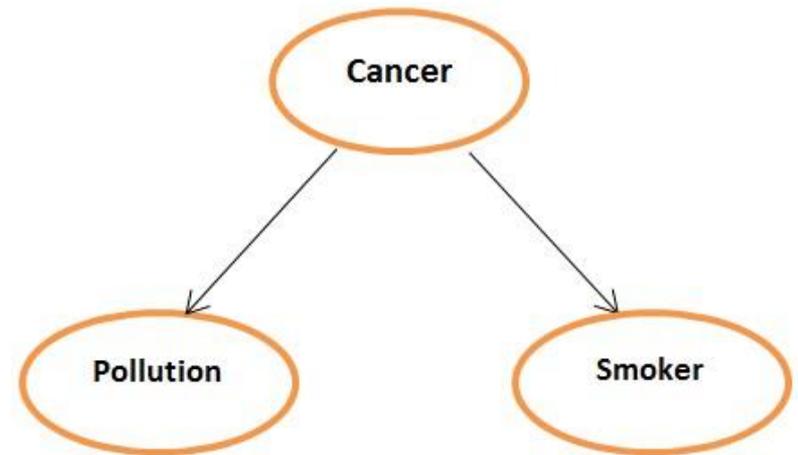
Background

Causal Model & Non-Causal Model

Causal model

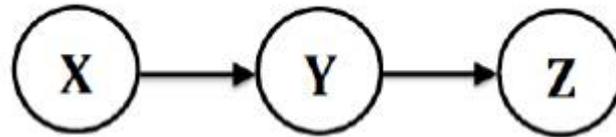


Naive Bayes (anti-causal) model

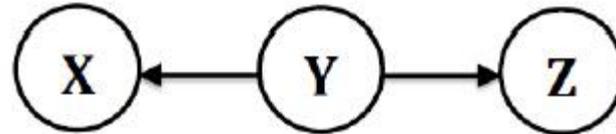


D-Separation

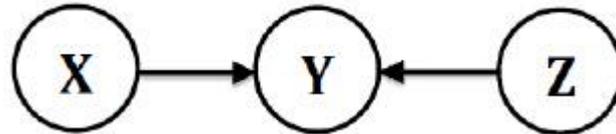
Chain:



Common Cause:



Common Effect:



Latent Variables

Latent variables are those we cannot measure them or those do not know they exist.

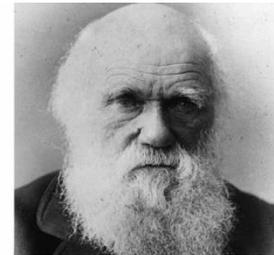
Newton explained motion via **gravity** (a latent variable).

Darwin proposed evolutionary theory but could only guess about the role of **genes** (a latent variable).

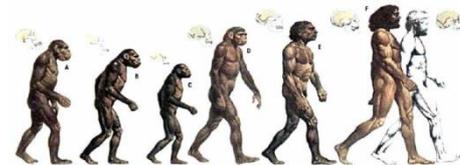


Gravity.

It's not just a good idea.
It's the Law.

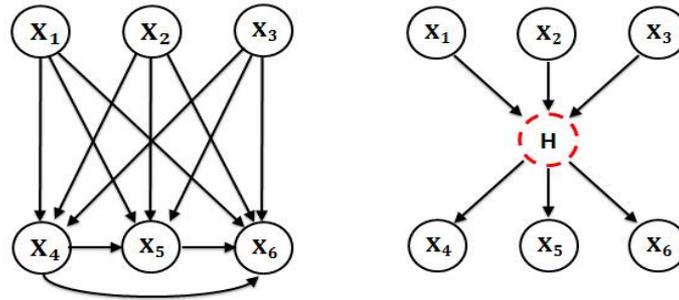


Charles Robert Darwin (12 February 1809 – 19 April 1882)

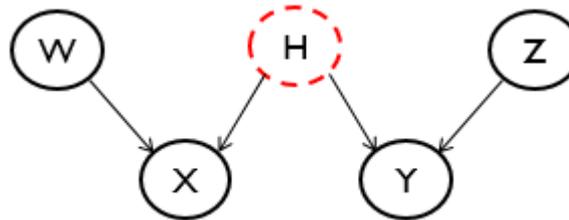


Why learning latent variables is important?

Simplify the network:



Help us to explain the true dependency structure:



Latent variables discovery algorithms

- **Constraint-based algorithms**

- 1) Involve statistical tests for conditional independence.
- 2) Return a statistically equivalent class that contains the true model.

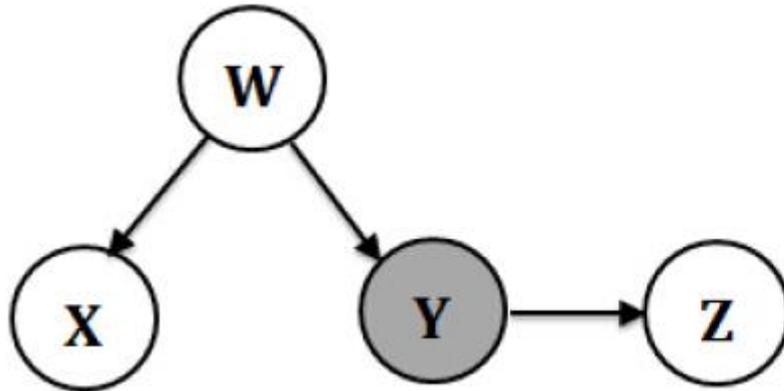
- **Metric-based algorithms**

- 1) Use a scoring metric to evaluate potential models
- 2) Find a network structure with a good score



*Dependency Patterns (Triggers) Discovery
for latent variable*

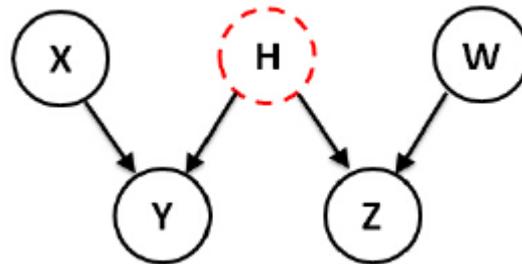
Dependency Matrix



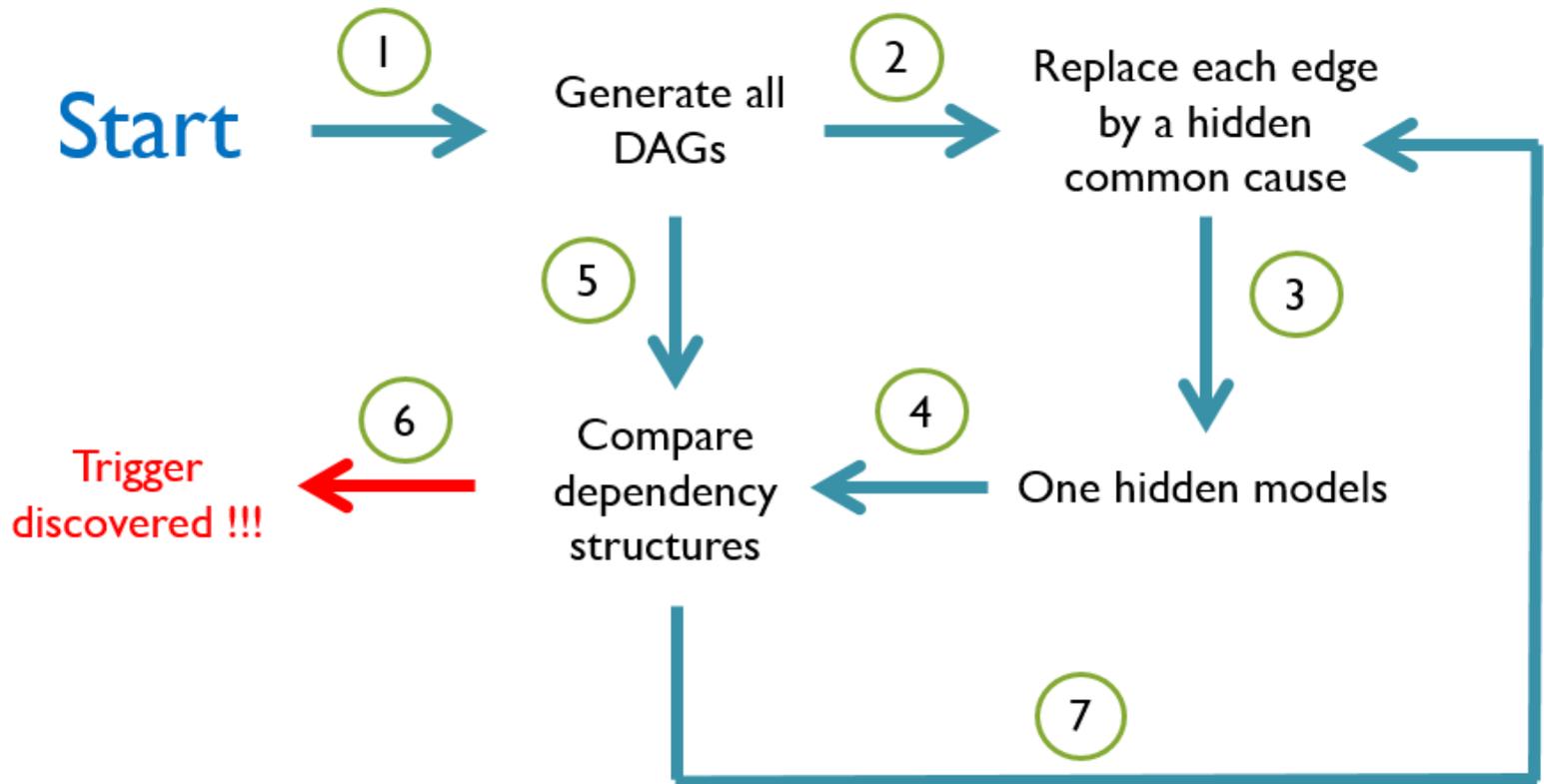
	W	X	Y	Z
W	0	1	0	0
X	1	0	0	0
Y	0	0	0	0
Z	0	0	0	0

What is Trigger?

- Latent variables are typically considered only in scenarios where they are **common causes** (Friedman, 1997).
- The set of dependencies of a latent variables is a trigger if and only if these dependency sets cannot be matched by any fully models.
- Trigger models can better encode the actual dependencies and independencies.



A Systematic Search for Finding Triggers

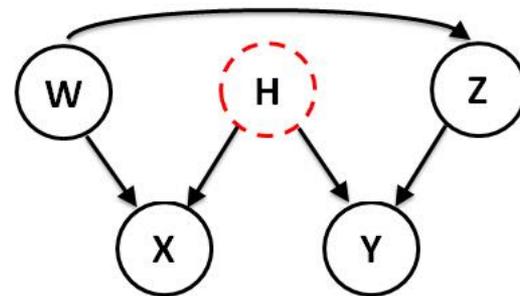
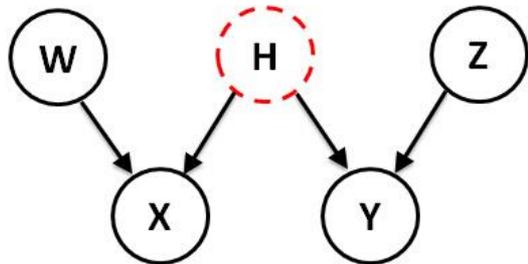


Experiment Results (Triggers)

Triggers:

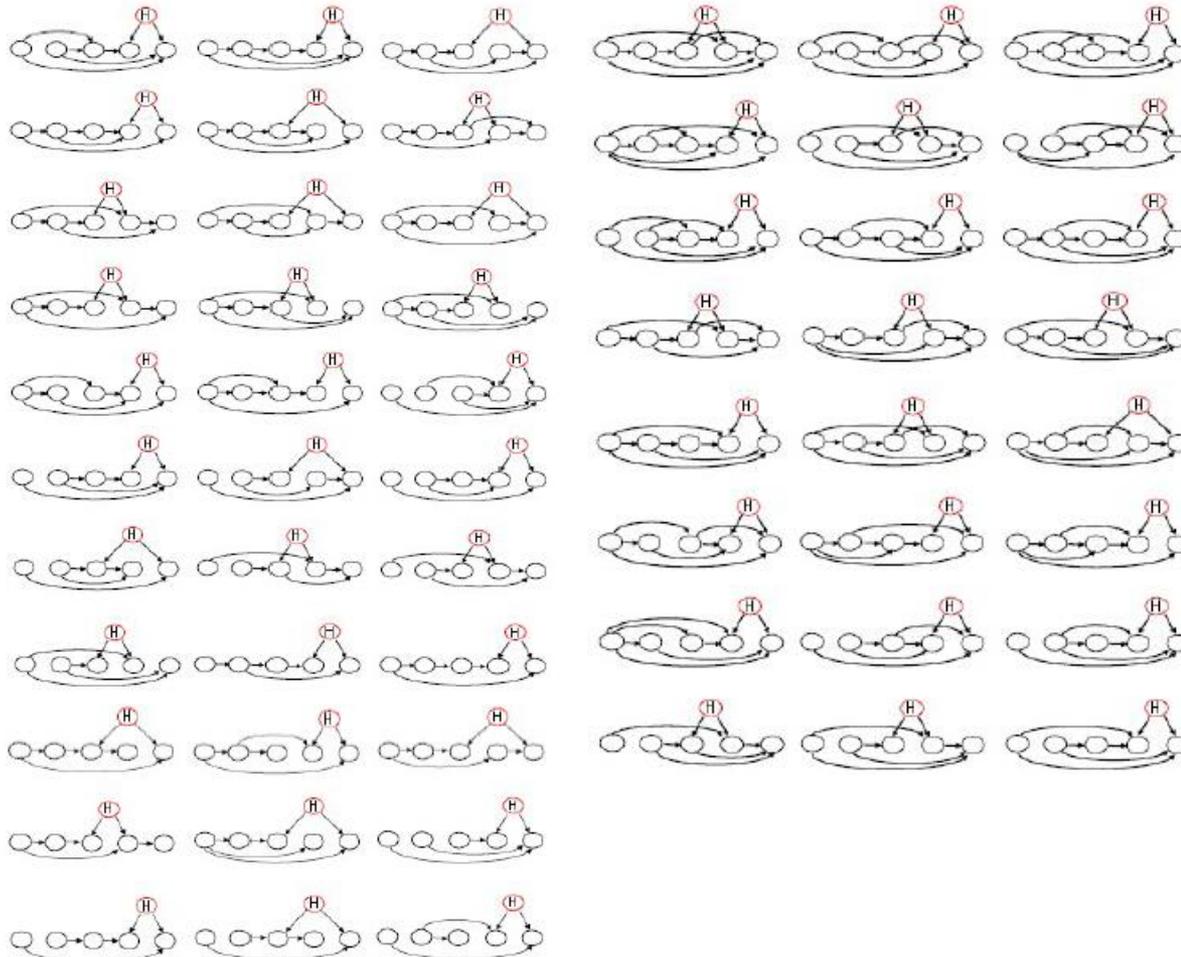
Number of observed variables	DAGs	Triggers
3	6	0
4	31	2
5	302	57

E.g., for four variables, the two triggers are:



Experiment Results (Triggers)

For five variables, all the triggers are:





Applying Triggers in causal discovery

Triggers + Chi-square test

- Pre-compute all triggers.
- Get the full dependency matrices of a given data set by applying conditional chi-square test.
- Check whether the dependencies match any one of the triggers.

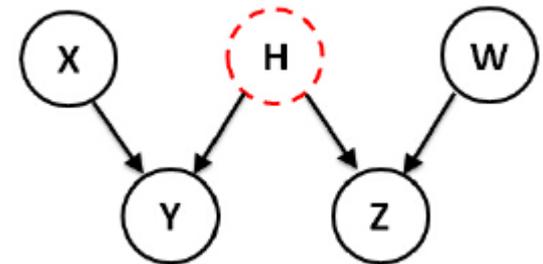
A Simulated Data

```
@ATTRIBUTE B {False, True}  
@ATTRIBUTE C {False, True}  
@ATTRIBUTE A {False, True}  
@ATTRIBUTE D {False, True}
```

```
⏏  
@DATA  
False, False, True, False  
False, True, True, False  
False, False, True, True  
False, False, True, True  
True, True, False, False  
False, True, True, False  
True, True, False, False  
False, False, True, True  
False, False, False, True  
False, False, True, True  
False, False, True, True  
True, False, False, True  
False, True, True, False  
False, True, True, False  
False, False, True, True  
False, True, True, False  
False, True, True, False  
False, True, True, False  
False, False, True, True  
False, False, False, True  
False, False, True, False  
False, True, True, True  
True, False, False, True  
False, False, True, True  
False, False, False, True  
True, False, False, True  
False, True, True, True
```

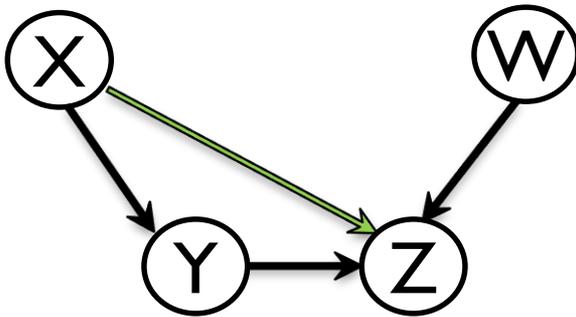
Results

Given ϕ : W & Z, X & Y, Y& Z
Given W: X & Y, Y & Z
Given X: W & Z, Y & Z
Given Y: W & Z, X & Z
Given Z: W & Y, X & Y
Given W and X: Y & Z
Given W and Y: X & Z
Given W and Z: X & Y
Given X and Y: W & Z
Given X and Z: W & Y
Given Y and Z: W & X

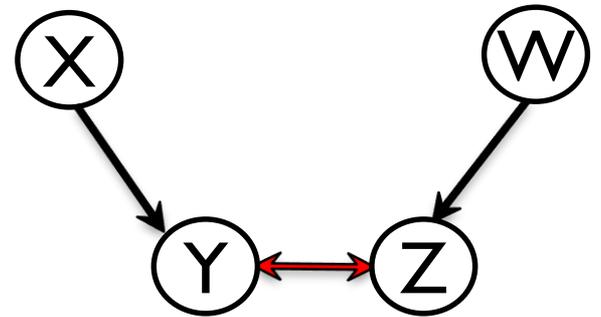


Results

CaMML

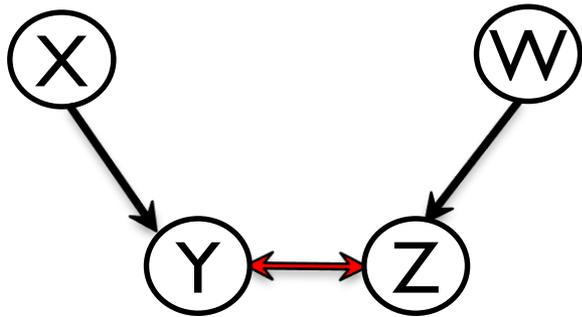


GeNIe (PC)

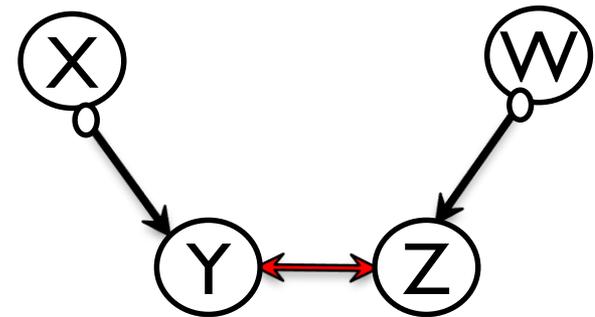


Results

Tetrad (PC)



Tetrad (FCI)





Analysis & Future work

Analysis

- CaMML fails to detect latent variable because there is no latent variable discovery algorithms implemented in CaMML.
- Our simple learner have successfully matched the dependencies in the data with one of the triggers.
- Both PC and FCI use an arc with two arrowheads to imply the existence of a latent variable.

Future work

- Extent the program with more than one latent variables.
- Parameterize latent variable models.
- Try to implement our trigger program into CaMML.

References

- Korb, Kevin B, & Nicholson, Ann E. (2003). *Bayesian artificial intelligence*: CRC press.
- Pearl, Judea. (2000). *Causality: models, reasoning and inference* (Vol. 29): Cambridge Univ Press.
- Spirtes, Peter, Glymour, Clark, & Scheines, Richard. (1993). *Causation, prediction, and search* (Vol. 81): Springer New York.
- Friedman, Nir. (1997). *Learning belief networks in the presence of missing values and hidden variables*. ICML.
- Meek, Christopher. (1997). *Graphical Models: Selecting causal and statistical models*. PhD thesis, Carnegie Mellon University.
- O'Donnell, R, Korb, K, & Allison, Lloyd. (2007). *Causal KL: Evaluating causal discovery*: Citeseer.



Thanks for listening!!!

Any questions?

Xuhui Zhang

Nov 25th, 2015

@ABNMS