Entropic Causal Inference: Graph Identifiability

Spencer Compton, Kristjan Greenewald, Dmitriy Katz, Murat Kocaoglu

MIT-IBM Watson AI Lab

PURDUE UNIVERSITY

Motivation

- Learning underlying causal relationships
- Using only observational data: no experiments

Entropic Causality

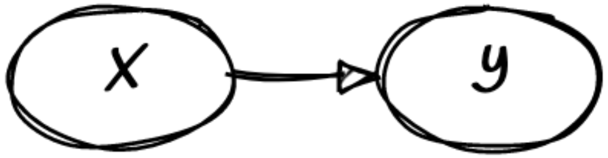
Setting: Categorical variables. No latent confounders.

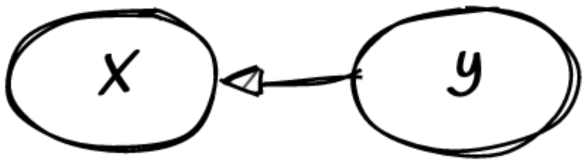
X	Y
0	1
0	1
0	0
1	1
....	....

	$Y=0$	$Y=1$
$X=0$	30/100	20/100
$X=1$	10/100	40/100

Joint Dist. $p(X,Y)$

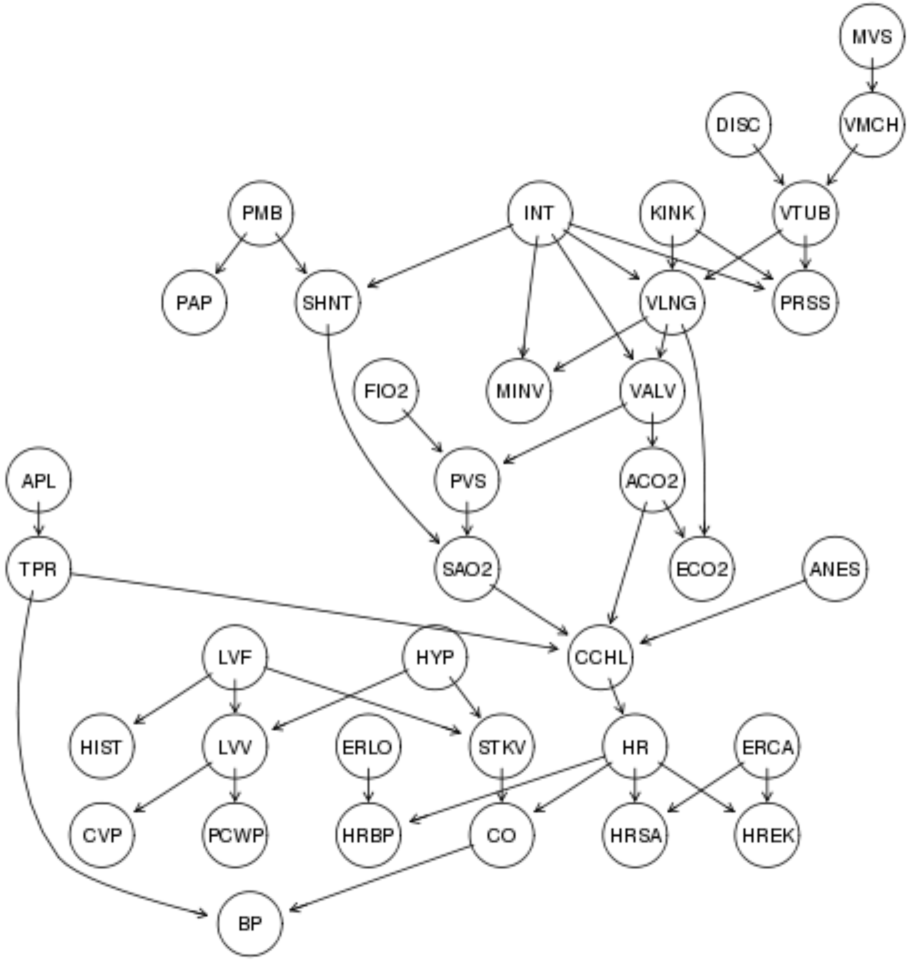
Goal: Determine the causal direction.





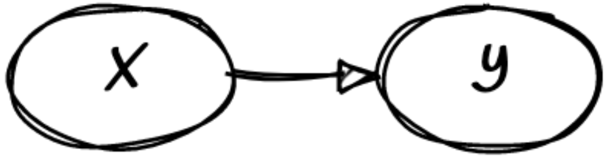
Entropic Causality

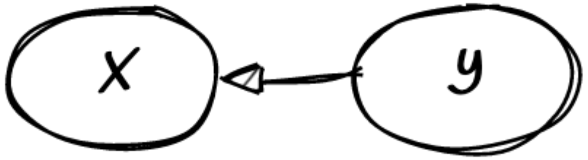
Goal: Learning causal systems with many variables.



Entropic Causality

Joint Dist. $p(X, Y)$





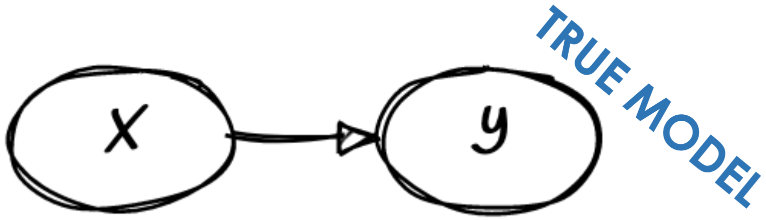
$$\underbrace{p(X)}_{\text{Cause}}\underbrace{p(Y|X)}_{\text{Mechanism}}$$

$$Y = f(X, E), X \perp\!\!\!\perp E$$

$$\underbrace{p(X|Y)}_{\text{Mechanism}}\underbrace{p(Y)}_{\text{Cause}}$$

$$X = g(Y, \tilde{E}), Y \perp\!\!\!\perp \tilde{E}$$

Entropic Causality



$$Y = f(X, E), X \perp\!\!\!\perp E$$

Assumption: True causal mechanism is "simple".

≡ Rényi entropy $H_\alpha(E)$ is small.

Kocaoglu, Dimakis, Vishwanath, Hassibi '17, AAAI

$H_0(.)$: Support Size

Compton, Kocaoglu, Greenewald, Katz '20, NeurIPS

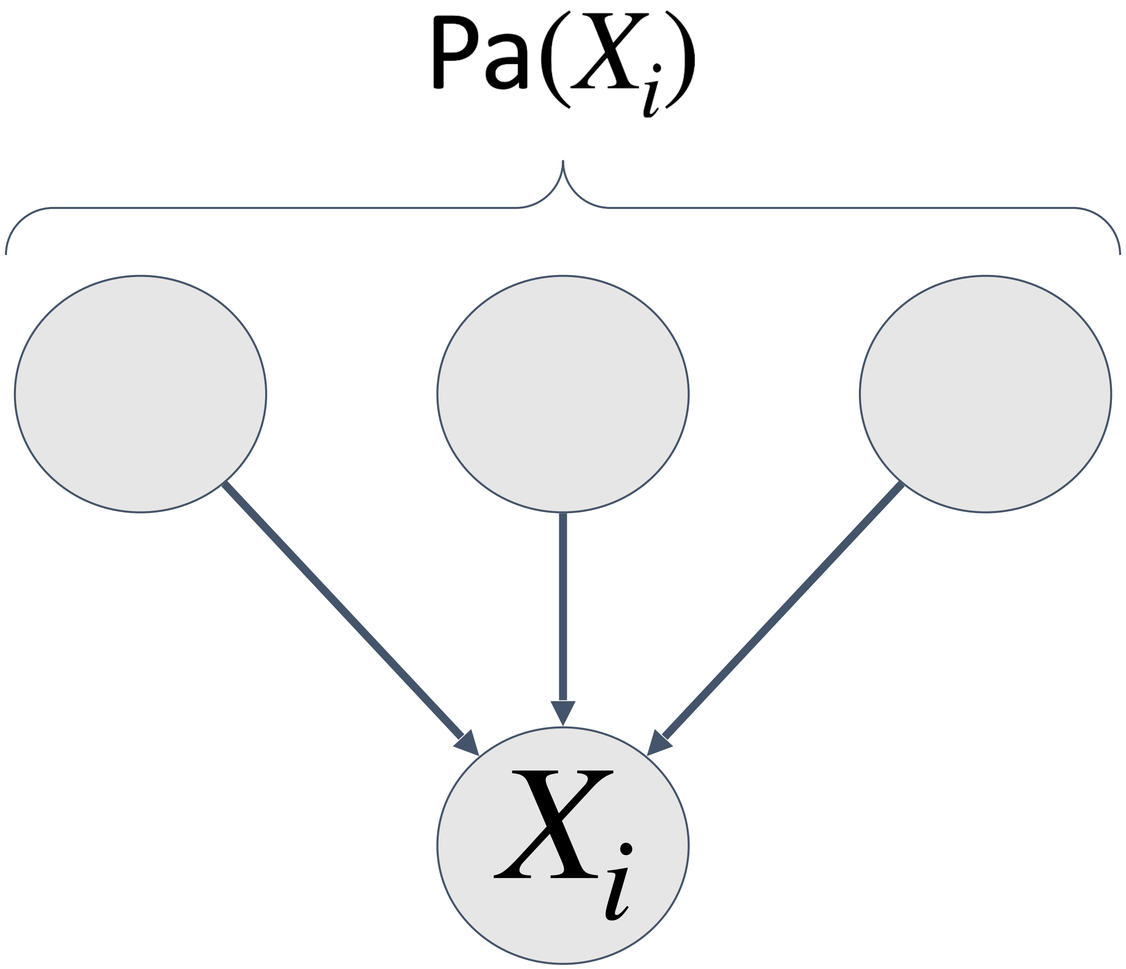
$H_1(.)$: Shannon Entropy

Is it true that the Rényi entropy of **any model** in the **wrong** causal direction will be **large**?

Our Results

- Relaxing assumptions for pairs
- Extend entropic causality to larger graphs with a provably correct peeling algorithm

Graph Setting



$$X_i = f(\text{Pa}(X_i), E_i), \text{Pa}(X_i) \perp\!\!\!\perp E_i$$

Graph Assumptions

Graph Assumptions: Consider an SCM where $X_i = f_i(Pa_i, E_i), Pa_i \perp\!\!\!\perp E_i$, $X_i \in [n], E_i \in [m]$. We assume $H(E_i) = o(\log(\log(n)))$,

E_i has $\Omega(n)$ states with $\Omega\left(\frac{1}{n\log(n)}\right)$ mass and f_i is uniformly random.

Graph Assumptions

Graph Assumptions: Consider an SCM where $X_i = f_i(Pa_i, E_i), Pa_i \perp\!\!\!\perp E_i$, $X_i \in [n], E_i \in [m]$. We assume $H(E_i) = o(\log(\log(n)))$,

E_i has $\Omega(n)$ states with $\Omega\left(\frac{1}{n\log(n)}\right)$ mass and f_i is uniformly random.

In summary, assuming all E_i have non-negligible support and low entropy, while all functions are randomly chosen.

Source-Pathwise Comparisons

Theorem: Given our graph assumptions, consider any pair of nodes X, Y where X is a source and there is a directed path from X to Y. The pairwise minimum-entropy comparison orients $X \rightarrow Y$ with high probability.

Source-Pathwise Comparisons

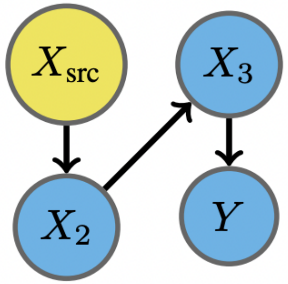
Theorem: Given our graph assumptions, consider any pair of nodes X, Y where X is a source and there is a directed path from X to Y. The pairwise minimum-entropy comparison orients $X \to Y$ with high probability.

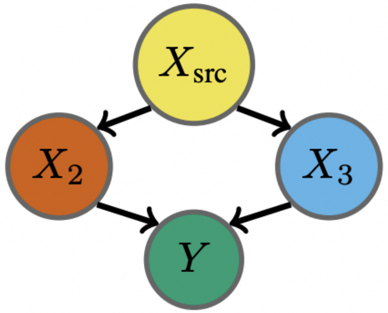
These comparisons enable us to find the sources!

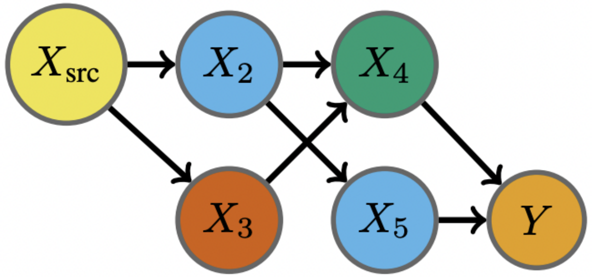
Source-Pathwise Comparisons

Theorem: Given our graph assumptions, consider any pair of nodes X, Y where X is a source and there is a directed path from X to Y. The pairwise minimum-entropy comparison orients $X \to Y$ with high probability.

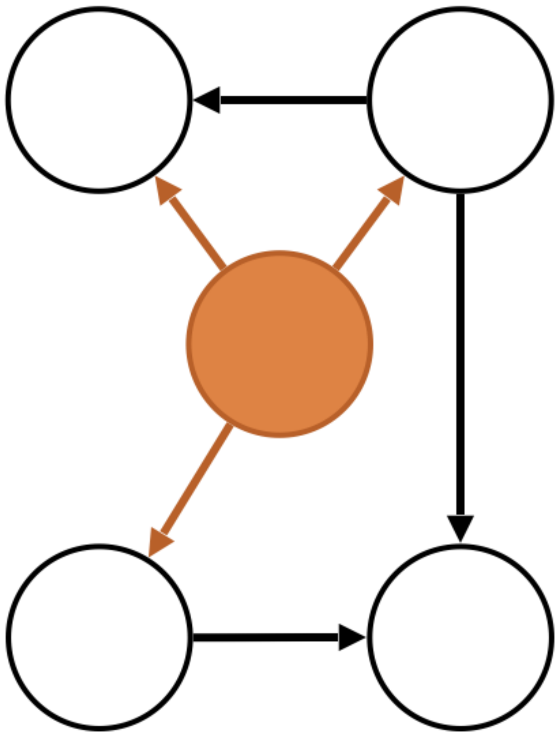
These comparisons enable us to find the sources!



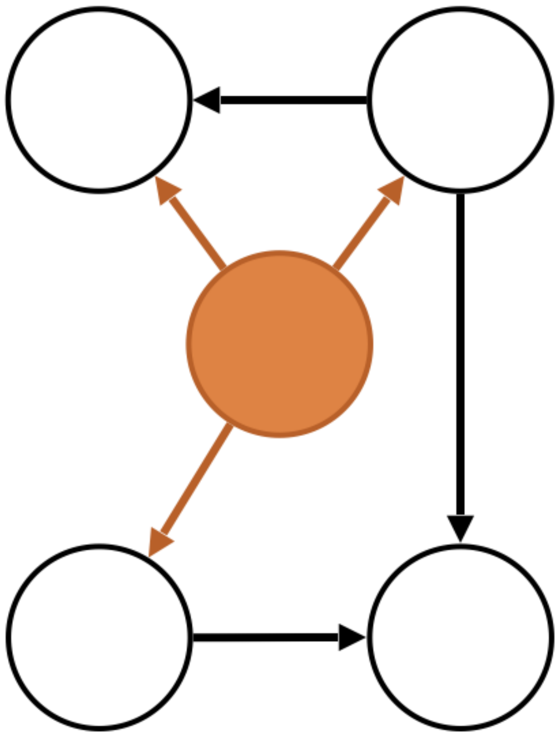


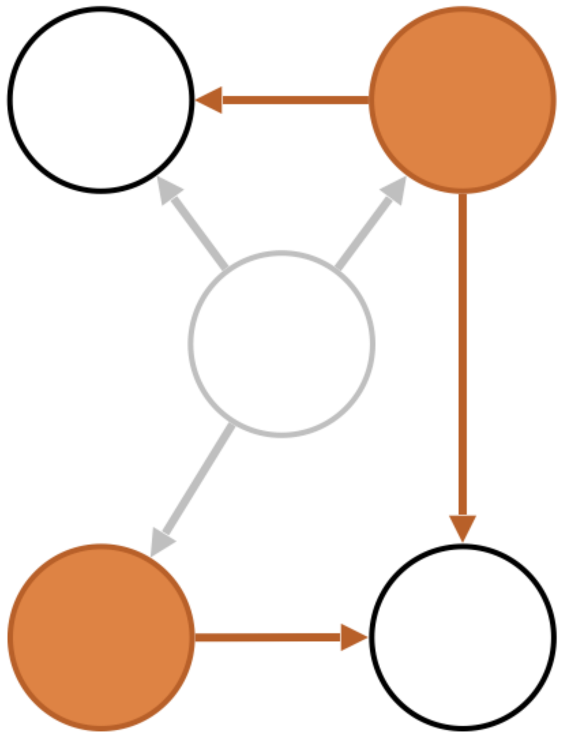


Source-Finding Peeling Algorithm

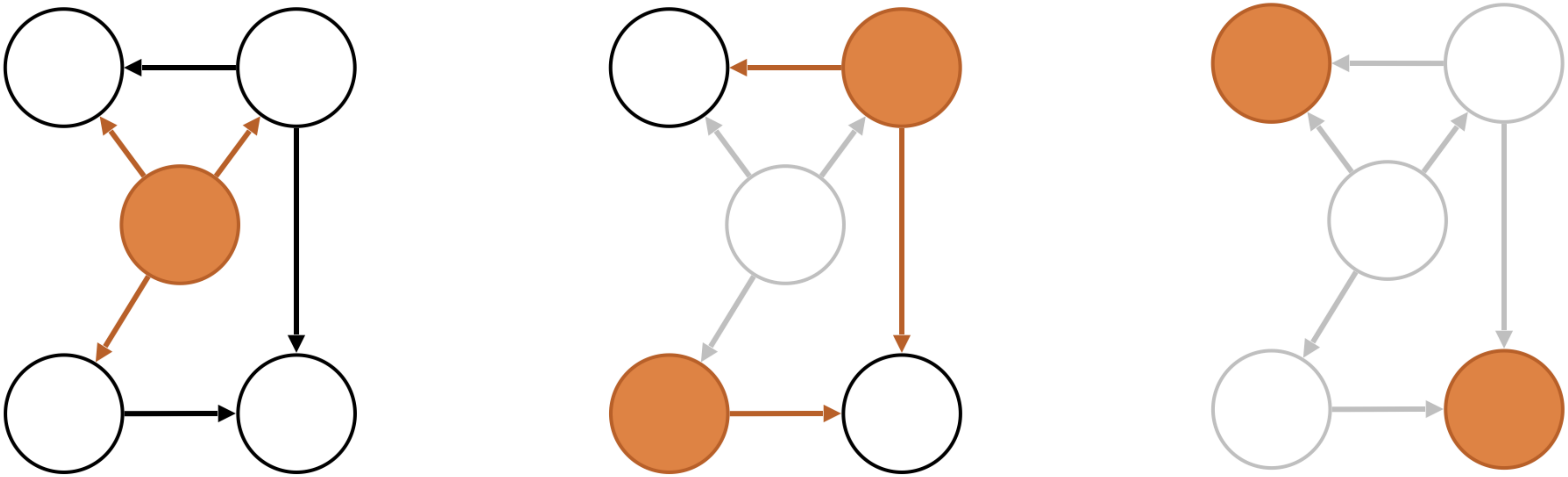


Source-Finding Peeling Algorithm



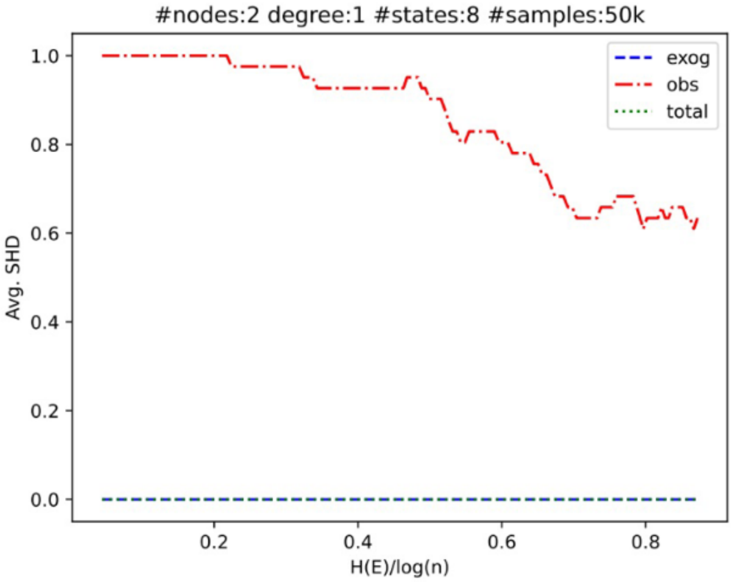


Source-Finding Peeling Algorithm

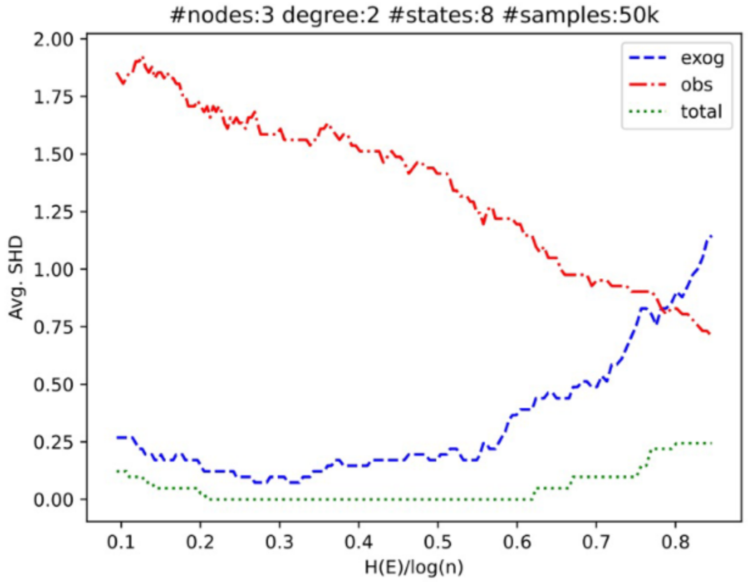


Provably learns general graphs!

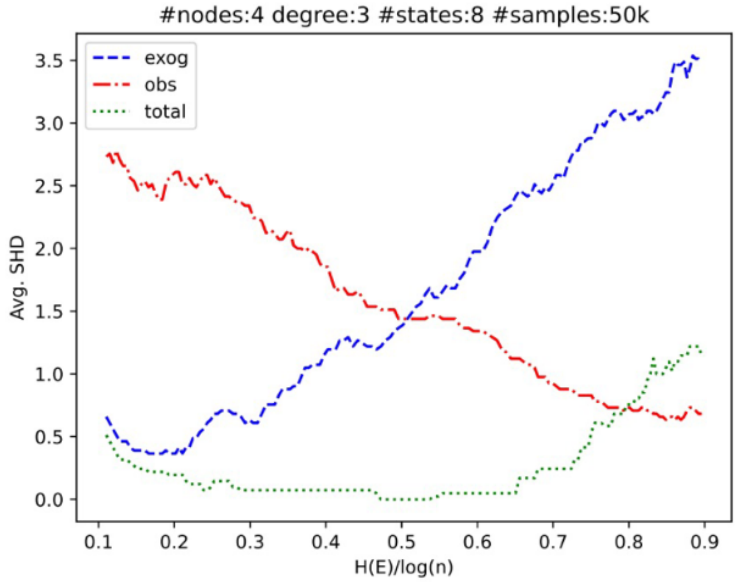
Experiments: Synthetic Graphs



(a) Bivariate

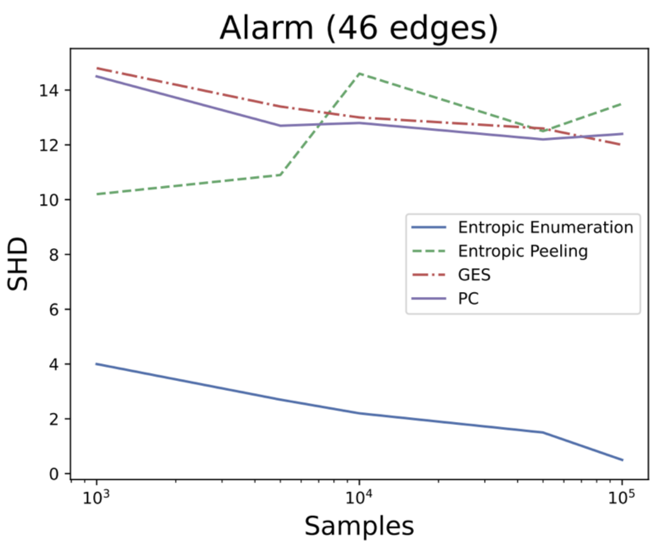


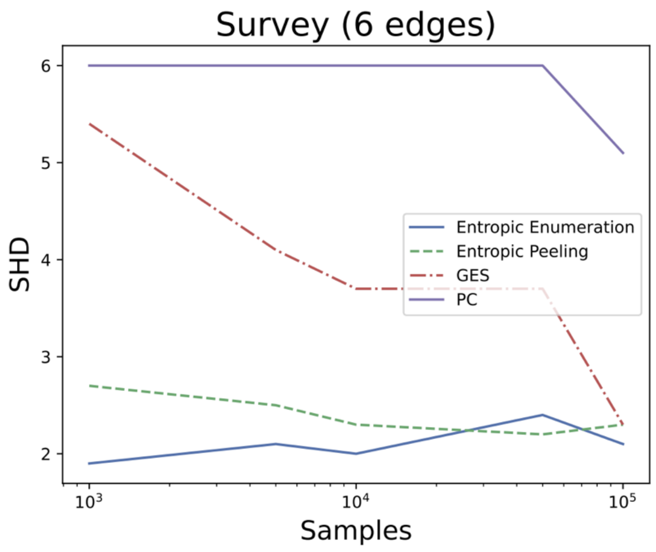
(b) 3-Node Complete Graph

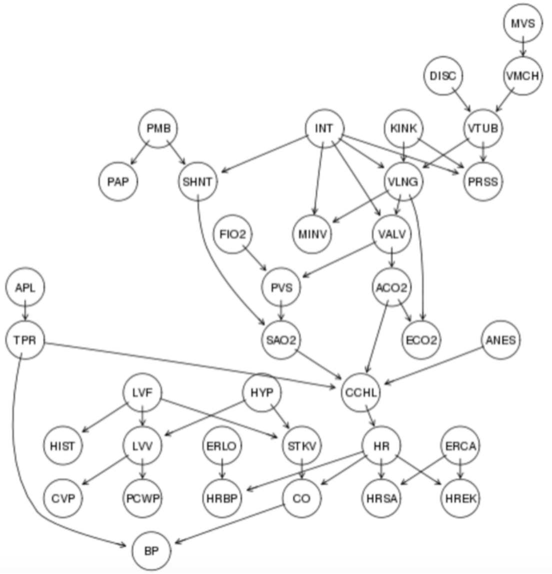


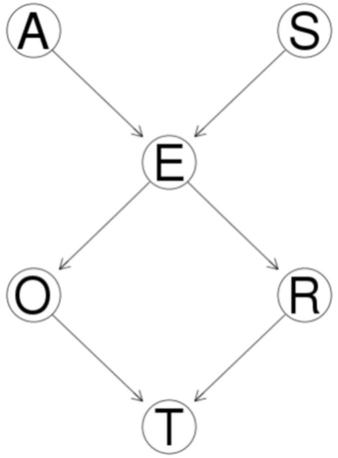
(c) 4-Node Complete Graph

Experiments: Real-World Graphs









THANK YOU