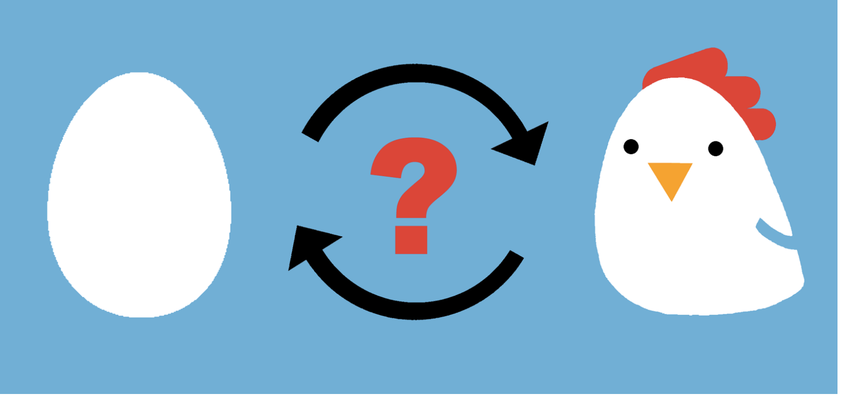


CBMS Conference -- Foundations of Causal Graphical Models and Structure Discovery

Lecture 9

Causal Representation Learning: Recovery of the Hidden World

Instructor: Kun Zhang

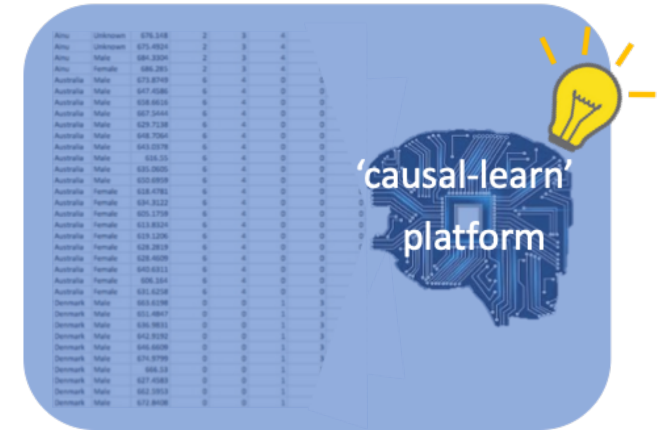
Carnegie Mellon University

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

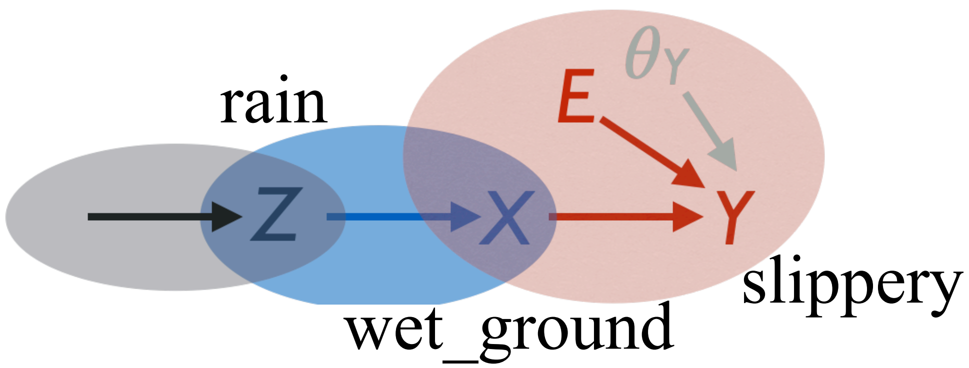
Outline

- Why causal/disentangled representations ?
- How?
  - IID case
    - Linear-Gaussian case
    - Linear, non-Gaussian case
    - Nonlinear case
  - From multiple distributions
  - With temporal information

Uncover Causality from Observational Data?



- Causal system has "irrelevant" modules (Pearl, 2000; Spirtes et al., 1993)



- conditional independence among variables;
- independent noise condition;
- minimal (and independent) changes…

Footprint of causality in data

- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees

- Three dimensions of the problem:

i.i.d. data?	**Parametric constraints?**	**Latent confounders?**
Yes	No	No
No	Yes	Yes

Causal Representation Learning: A Summary

<table>
<tr><th>i.i.d. data?</th><th>Parametric constraints?</th><th>Latent confounders?</th><th>What can we get?</th></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td rowspan="2">(Different types of) equivalence class</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td rowspan="2">Unique identifiability (under structural conditions)</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Non-I, but I.D.</td><td rowspan="2">No/Yes</td><td>No</td><td>(Extended) regression</td></tr>
<tr><td>Yes</td><td>Latent temporal causal processes identifiable!</td></tr>
<tr><td rowspan="4">I., but non-I.D.</td><td>No</td><td rowspan="2">No</td><td>More informative than MEC (CD-NOD)</td></tr>
<tr><td>Yes</td><td>May have unique identifiability</td></tr>
<tr><td>No</td><td rowspan="2">Yes</td><td>Changing subspace identifiable</td></tr>
<tr><td>Yes</td><td>Variables in changing relations identifiable</td></tr>
</table>

A Problem in Psychology: Finding Underlying Mental Conditions?

- 50 questions for big 5 personality test

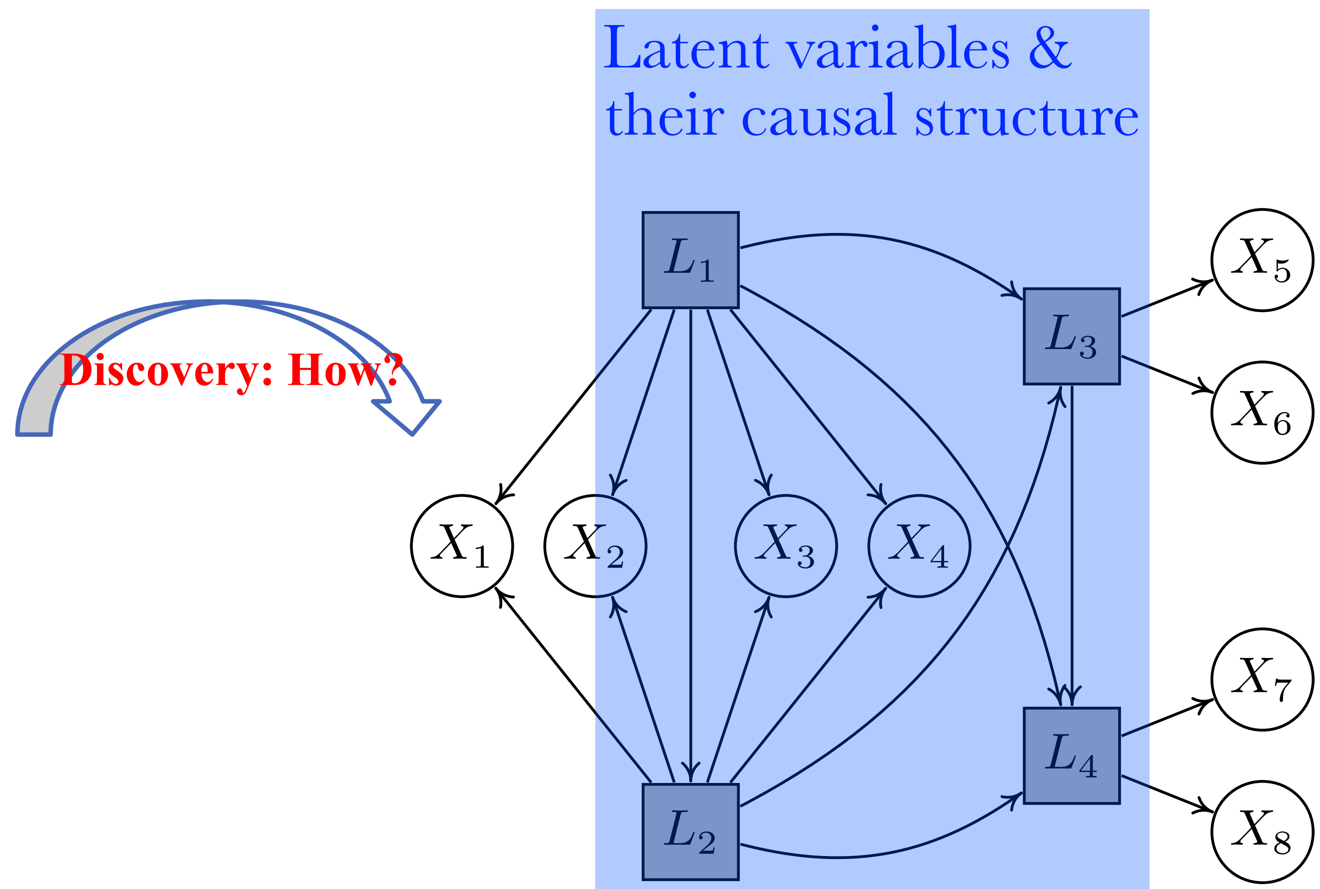
race	age	engnat	gender	hand	source	country	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	A1	A2	A3	A4	A5
3	53	1	1	1	1	US	4	2	5	2	5	1	4	3	5	1	1	5	2	5	1	1	1	1	1	1	1	5	1	5	2
13	46	1	2	1	1	US	2	2	3	3	3	3	1	5	1	5	2	3	4	2	3	4	3	2	2	4	1	3	3	4	4
1	14	2	2	1	1	PK	5	1	1	4	5	1	1	5	5	1	5	1	5	5	5	5	5	5	5	5	5	1	5	5	1
3	19	2	2	1	1	RO	2	5	2	4	3	4	3	4	4	5	5	4	4	2	4	5	5	5	4	5	2	5	4	4	3
11	25	2	2	1	2	US	3	1	3	3	3	1	3	1	3	5	3	3	3	4	3	3	3	3	3	4	5	5	3	5	1
13	31	1	2	1	2	US	1	5	2	4	1	3	2	4	1	5	1	5	4	5	1	4	4	1	5	2	2	2	3	4	3
5	20	1	2	1	5	US	5	1	5	1	5	1	5	4	4	1	2	4	2	4	2	2	3	2	2	2	5	5	1	5	1
4	23	2	1	1	2	IN	4	3	5	3	5	1	4	3	4	3	1	4	4	4	1	1	1	1	1	1	2	5	1	4	3
5	39	1	2	3	4	US	3	1	5	1	5	1	5	2	5	3	2	4	5	3	3	5	5	4	3	3	1	5	1	5	1
3	18	1	2	1	5	US	1	4	2	5	2	4	1	4	1	5	5	2	5	2	3	4	3	2	3	4	2	3	1	4	2
3	17	2	2	1	1	IT	1	5	2	5	1	4	1	4	1	5	5	3	5	3	2	5	3	3	4	3	2	4	2	4	1
13	15	2	1	1	1	IN	3	3	5	3	3	3	2	4	3	3	1	5	3	3	2	3	2	3	2	4	4	4	2	2	5
13	22	1	2	1	2	US	3	3	4	2	4	2	2	3	4	3	3	3	3	3	2	2	4	4	2	3	1	4	1	5	1
3	21	1	2	1	5	US	1	3	2	5	1	1	1	5	1	5	5	3	5	2	5	5	3	2	5	3	1	1	1	4	2
3	28	2	2	1	2	US	3	3	3	4	3	2	2	4	3	5	2	4	4	4	4	4	2	2	3	2	1	4	2	4	2
3	21	1	1	1	5	US	2	3	2	3	3	1	1	3	4	4	2	4	2	4	1	2	2	2	2	2	4	2	4	2	5
13	19	1	2	1	2	FR	1	3	2	4	2	4	1	4	3	4	4	2	3	2	1	3	1	2	2	3	4	2	3	1	4
3	21	1	2	1	5	US	4	1	5	2	5	1	5	3	5	1	5	2	5	2	3	3	3	3	4	2	1	5	2	5	2
3	26	1	2	3	5	GB	2	3	4	3	1	4	1	4	1	5	4	2	5	2	1	4	2	2	2	2	2	2	2	2	2
3	26	1	2	1	1	US	2	2	3	3	3	3	1	3	3	3	4	4	3	1	3	2	2	2	4	4	1	3	2	4	3
13	19	2	2	1	1	IT	1	4	2	5	2	4	2	4	2	2	4	4	4	4	4	4	5	5	4	2	4	5	1	5	2

Learning Hidden Variables & Their Relations

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Measured variables (e.g., answer scores in psychometric questionnaires) were generated by causally related latent variables

X1	X2	X3	X4	X5	X6	X7	X8
4.2	3.6	6.5	6.8	9.6	7.6	2.7	4.8
3.8	1.9	6.5	7.3	8.9	6.9	1.1	4.6
4.2	3.4	6.5	6.9	9.5	7.4	2.5	4.6
4.2	2.2	6.2	6.9	9.6	7.2	1.9	4.8
3.9	1.9	6.5	6.8	9.0	6.8	1.7	4.4
4.0	2.0	6.4	7.2	9.1	7.0	1.0	4.6
3.8	1.7	6.4	7.3	9.0	6.7	0.8	4.3
4.1	2.8	6.5	6.9	9.3	6.7	2.7	4.6
...	...	...	...	...	...	...	...



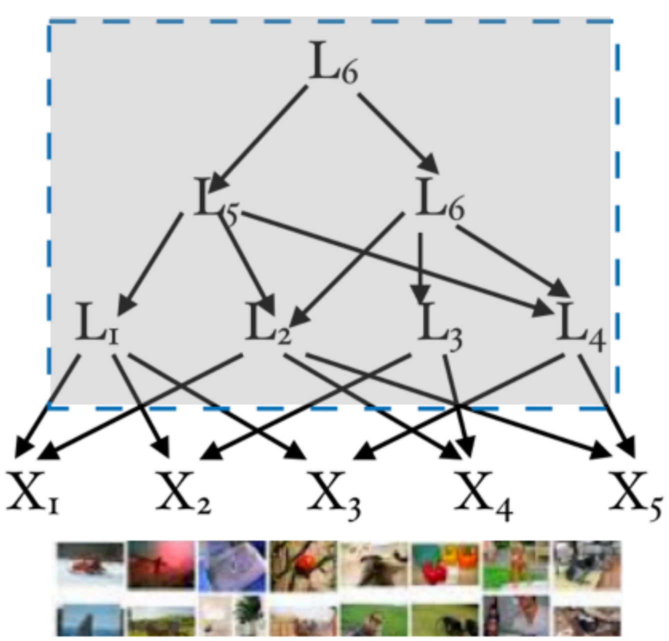
- Find latent variables L_i and their causal relations ?
- Rank deficiency or GIN helps solve the problem

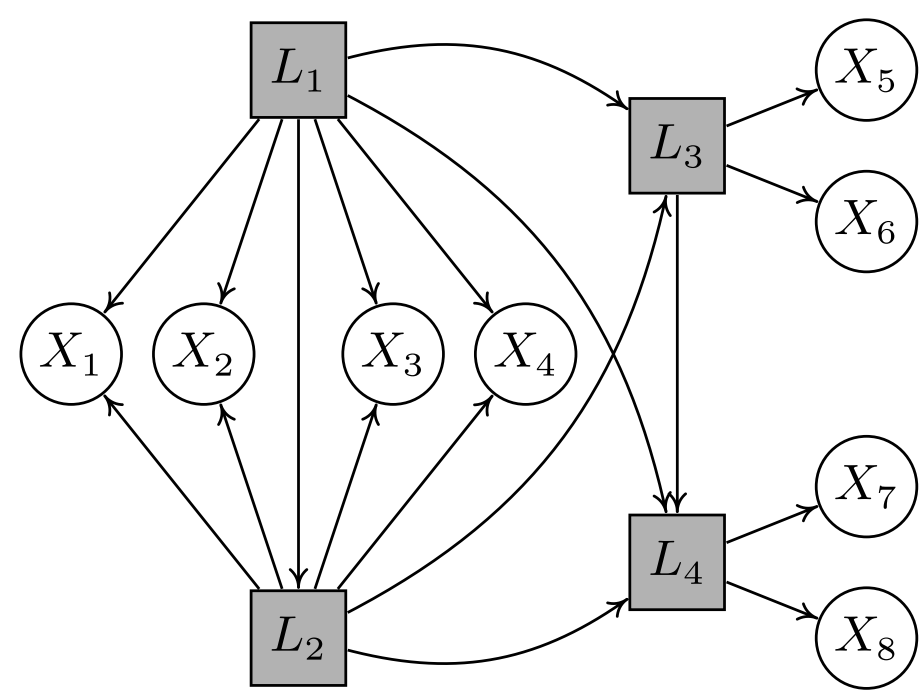
Outline

- Why causal/disentangled representations ?
- How?
  - IID case
    - **Linear-Gaussian case**
    - Linear, non-Gaussian case
    - Nonlinear case
  - From multiple distributions
  - With temporal information

Identifying Latent Causal Model in Linear-Gaussian Cases

Biwei Huang



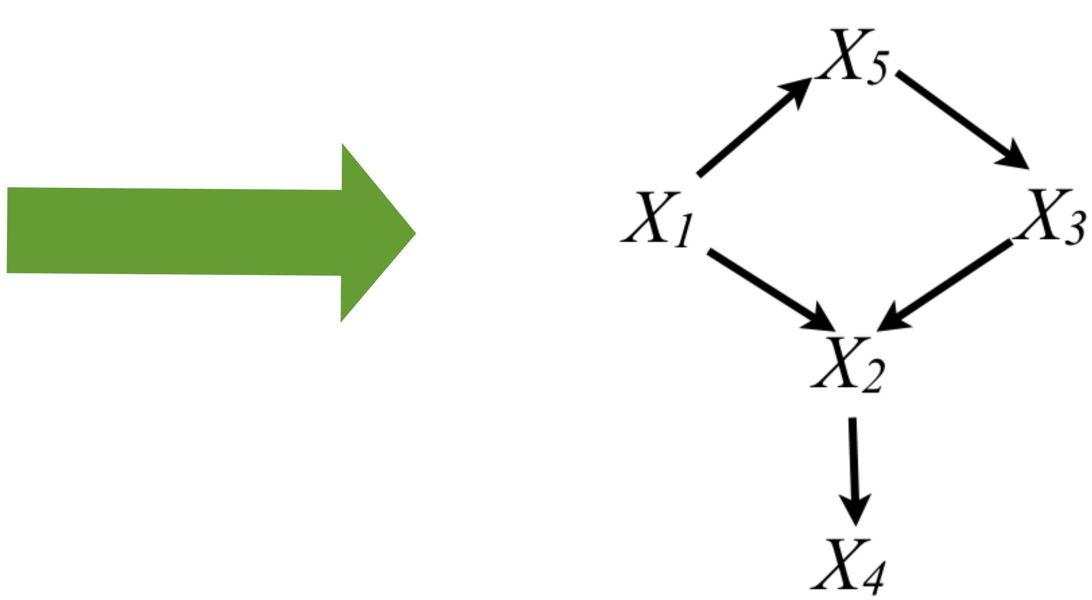


Motivation

Causal discovery:

- Aims to find causal relationships from observational data, without doing interventions
- Traditionally, assumes no latent confounders and only considers causal structure among observed variables
- However, in some cases, measured variables may not be causal variables, e.g., variables with measurement error, image pixels

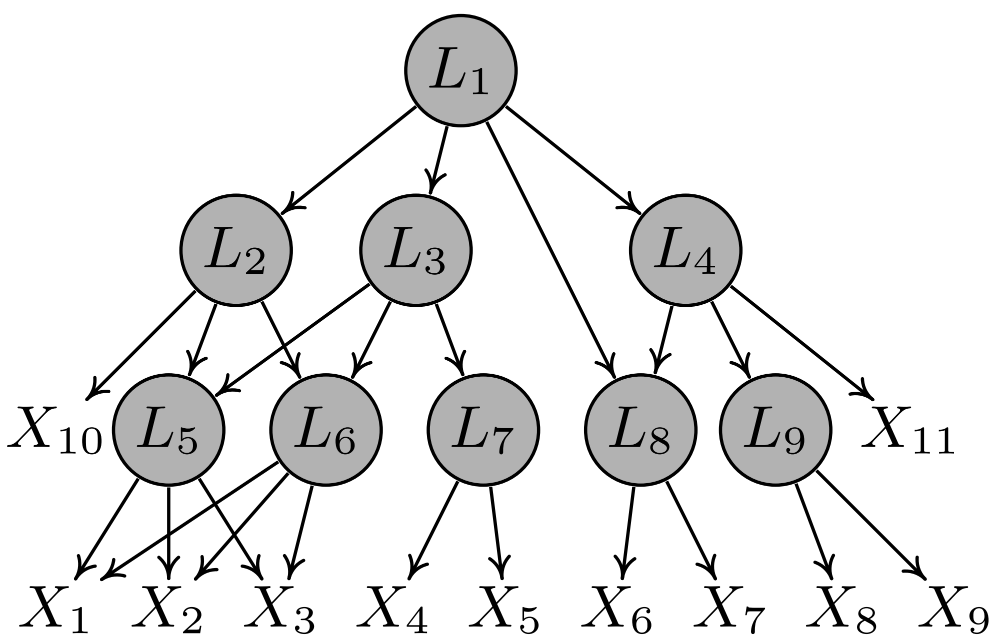
X_1	X_2	X_3	X_4	X_5
1.2	0.7	2.1	1.5	0.9
0.5	0.1	2.2	1.2	1.9
2.9	1.9	3.1	2.2	0.9
2.5	3.5	1.2	1.9	1.4
1.3	1.7	2.3	3.1	2.9

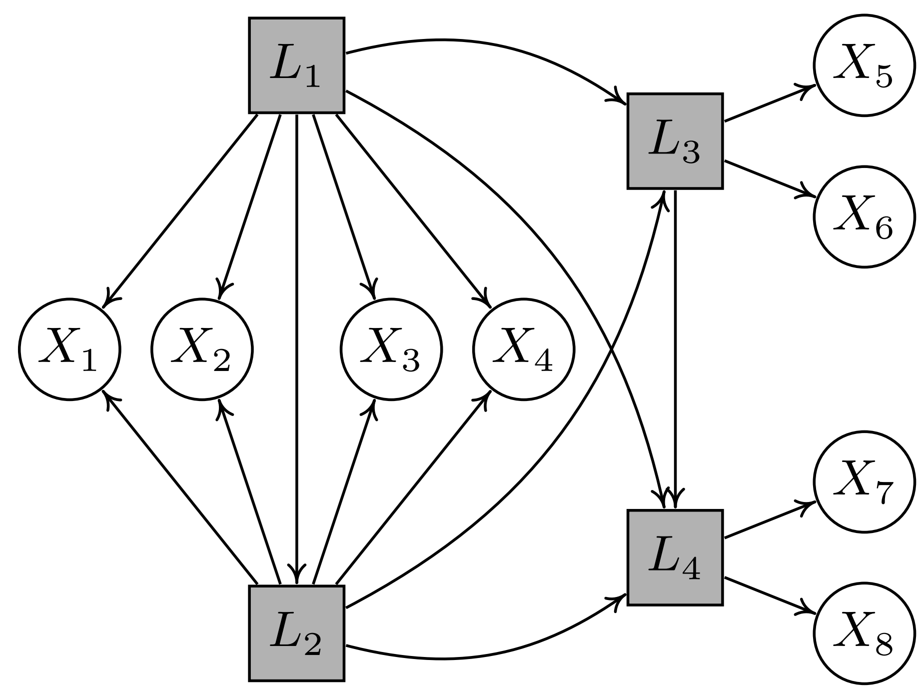


Identifying Latent Causal Graphs

Questions to answer:

- Locate (hierarchical) latent variables (i.e., cluster the lower-level variables)
- Identify the causal structure among all variables

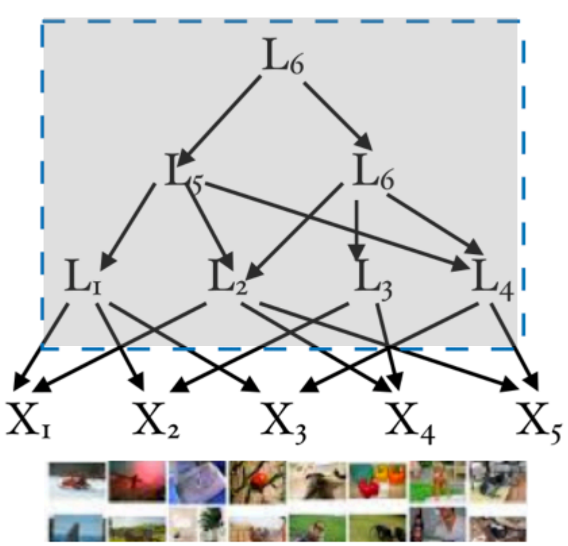


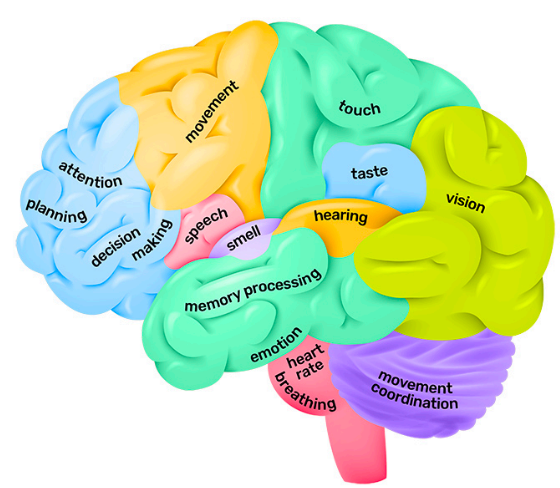


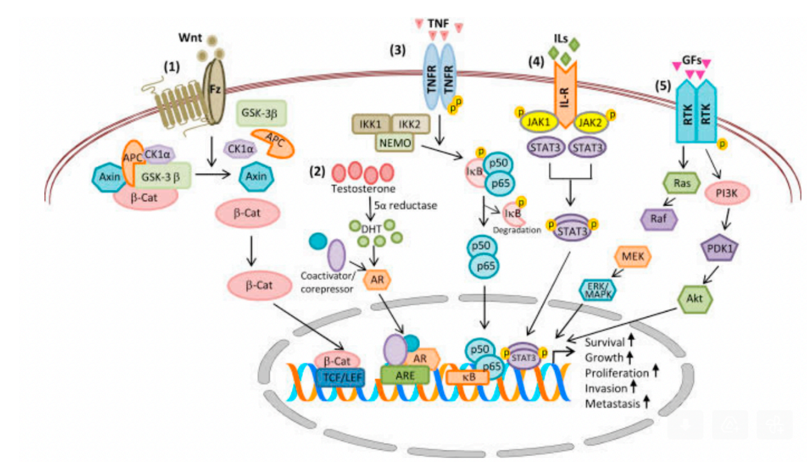
X_i: measured variables
L_i: latent variables

Motivation of Latent Variable Discovery in AI/ Scientific Discovery

- Usually not possible to measure all task-related variables
  - Causal discovery in the presence of latent variables
  - Image/Video/Language understanding
  - Automatically identify and hierarchically cluster the underlying functional brain areas and discover the information flow, from measured voxel data
  - Identify gene regulation process from gene expression data

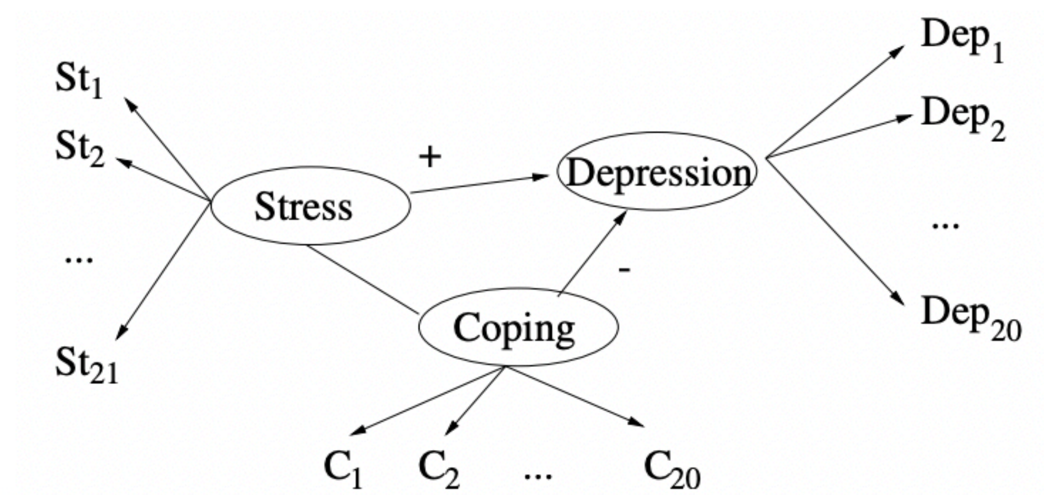


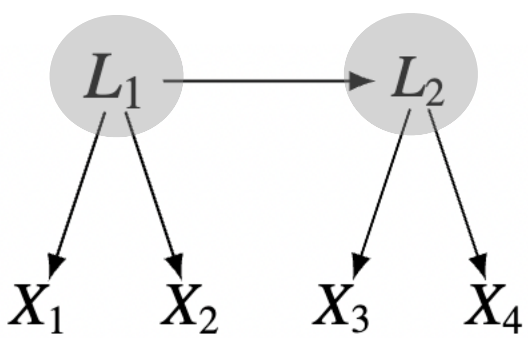




An illustration of Tetrad Conditions

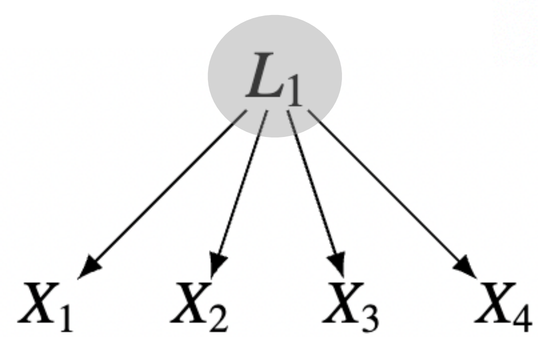
- Tetrad condition [Spearman 1904, Anderson & Rubin 1956]





(a)

$$\rho_{12}\rho_{34} \neq \rho_{23}\rho_{14}$$
$$\rho_{13}\rho_{24} = \rho_{23}\rho_{14}$$
$$\rho_{12}\rho_{34} \neq \rho_{13}\rho_{24}$$



(b)

$$\rho_{12}\rho_{34} = \rho_{23}\rho_{14}$$
$$\rho_{13}\rho_{24} = \rho_{23}\rho_{14}$$
$$\rho_{12}\rho_{34} = \rho_{13}\rho_{24}$$

X_i: measured variables
L_i: latent variables

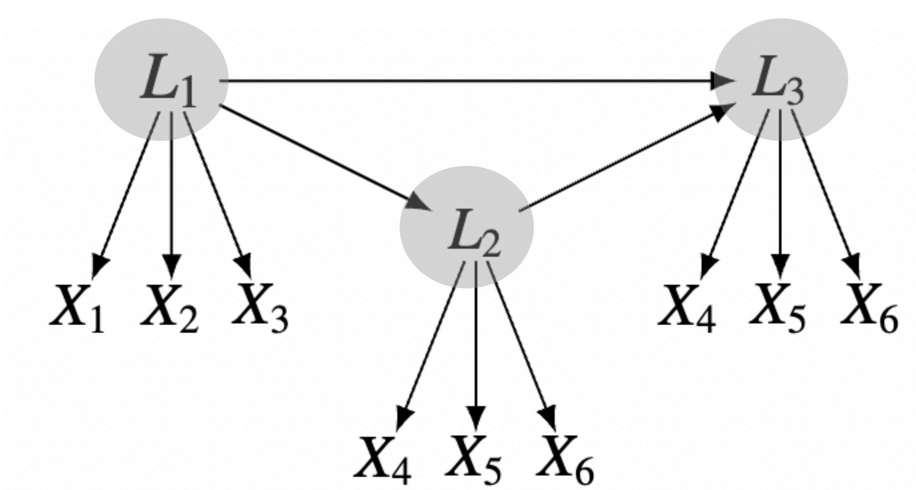
Tetrad condition:
Indicates rank deficiency of 2 x 2 off-diagonal covariance matrices

ρ_{ij} denotes the correlation coefficient between x_i and x_j

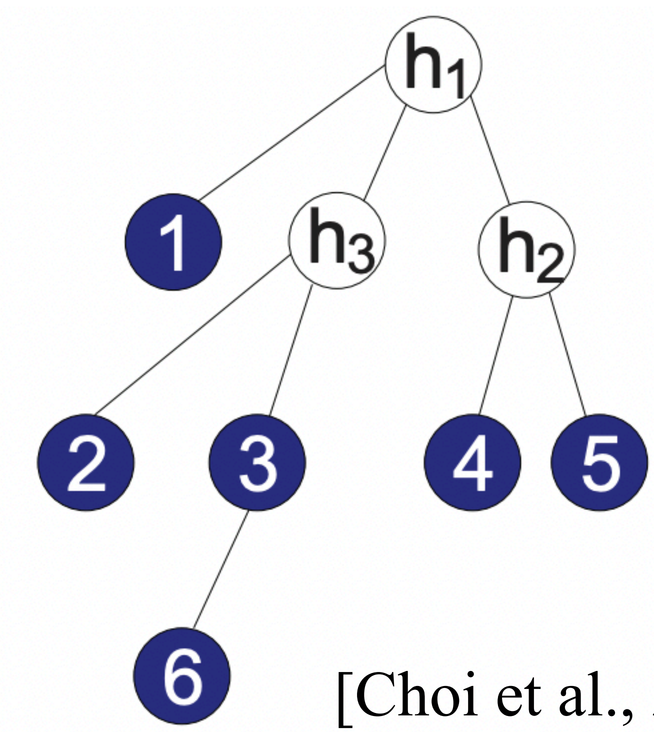
Applications of Tetrad Conditions

- One-factor measurement model [Silva et al., 2006, Kummerfeld et al., 2016]
- Tree structure [Pearl, 1988, Choi et al., 2011]

One-factor measurement model



Tree



[Choi et al., 2011]

Identifying more general latent structures:

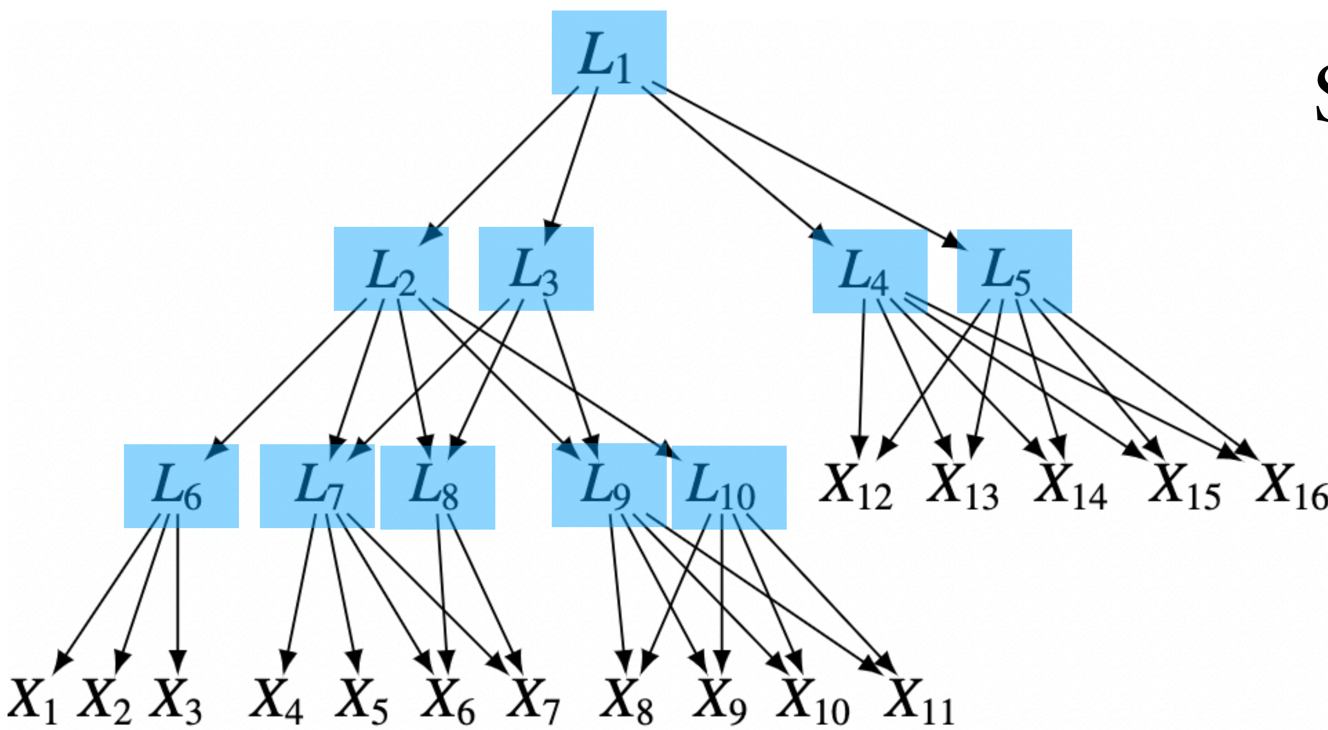
LATENT HIERARCHICAL CAUSAL STRUCTURE DISCOVERY WITH RANK CONSTRAINTS

Biwei Huang *[1] **Charles Low** *[1], **Feng Xie**[3], **Clark Glymour**[1], **Kun Zhang**[1,2]
[1] Carnegie Mellon University
[2] Mohamed bin Zayed University of Artificial Intelligence
[3] Beijing Technology and Business University, China
{bwei.huang, charleslow88, xiefeng009}@gmail.com,
cg09@andrew.cmu.edu, kunz1@cmu.edu

ABSTRACT

Most causal discovery procedures assume that there are no latent confounders in the system, which is often violated in real-world problems. In this paper, we consider a challenging scenario for causal structure identification, where some variables are latent and they form a hierarchical graph structure to generate the measured variables; the children of latent variables may still be latent and only leaf nodes are measured, and moreover, there can be multiple paths between every pair of variables (i.e., it is beyond tree structure). We propose an estimation procedure that can efficiently locate latent variables, determine their cardinalities, and identify the latent hierarchical structure, by leveraging rank deficiency constraints over the measured variables. We show that the proposed algorithm can find the correct Markov equivalence class of the whole graph asymptotically under proper restrictions on the graph structure.

Latent Hierarchical Causal Structure Discovery with Rank Constraints



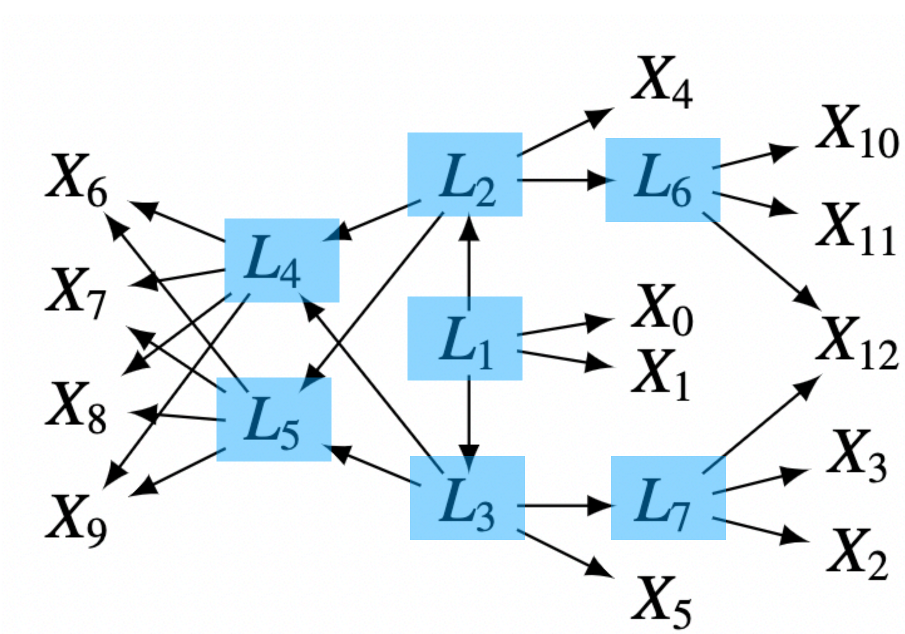
SEM with linear causal relations:

$$X_i = \sum_{L_j \in Pa(X_i)} b_{ij} L_j + \varepsilon_{X_i},$$

$$L_j = \sum_{L_k \in Pa(L_j)} c_{jk} L_k + \varepsilon_{L_j}$$

X_i: measured variables

L_i: latent variables



Questions to answer:

- Locate (hierarchical) latent variables (i.e., cluster the lower-level variables)
- Identify the causal structure among latent variables

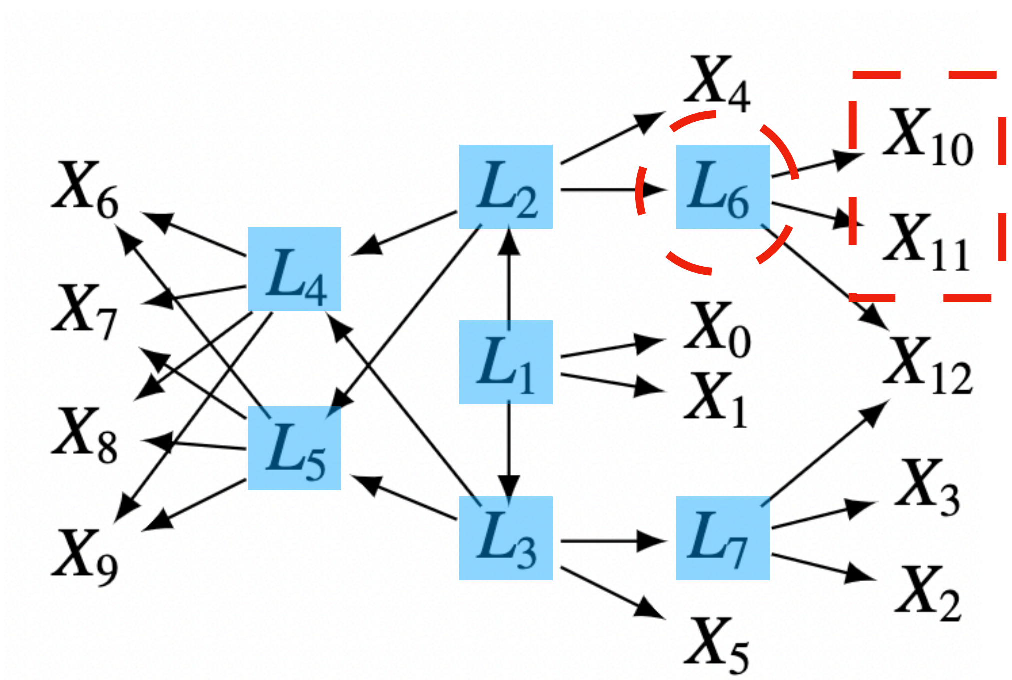
Latent Hierarchical Causal Structure Discovery with Rank Constraints

Basic idea:

- Rank-deficiency constraints over measured variables
  + Specific search procedure

foundation of this method

- $\mathrm{rank}(\Sigma_{X_A,X_B})$, which is deficient, indicates the smallest number of variables that t-separate X_A from X_B



Exp:

Let $X_A = \{X_{10}, X_{11}\}$ and $X_B = \mathbf{X} \setminus X_A$
$\mathrm{rank}(\Sigma_{X_A,X_B}) = 1$ which is rank deficient, because L_6 d-separates X_A from X_B.

However, we cannot directly know the location of these latent variables in the graph

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

Input : Date from a set of measured variables $\mathbf{X}_{\mathcal{G}}$
Output : Markov equivalence class $\mathcal{G}'$

1. Find clusters and assign latent covers greedily

$$\mathcal{G}' = \text{findCausalClusters}\ (\mathbf{X}_{\mathcal{G}})\ ;$$

2. Refine incorrect clusters and covers from greedy search

$$\mathcal{G}' = \text{refineClusters}\ (\mathcal{G}')\ ;$$

3. Refine edges and find v structures

$$\mathcal{G}' = \text{refineEdges}\ (\mathcal{G}')\ ;$$

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

Input : Date from a set of measured variables $\mathbf{X}_{\mathcal{G}}$
Output : Markov equivalence class $\mathcal{G}'$

1. Find clusters and assign latent covers greedily

$$\mathcal{G}' = \text{findCausalClusters}\ (\mathbf{X}_{\mathcal{G}})\ ;$$

2. Refine incorrect clusters and covers from greedy search

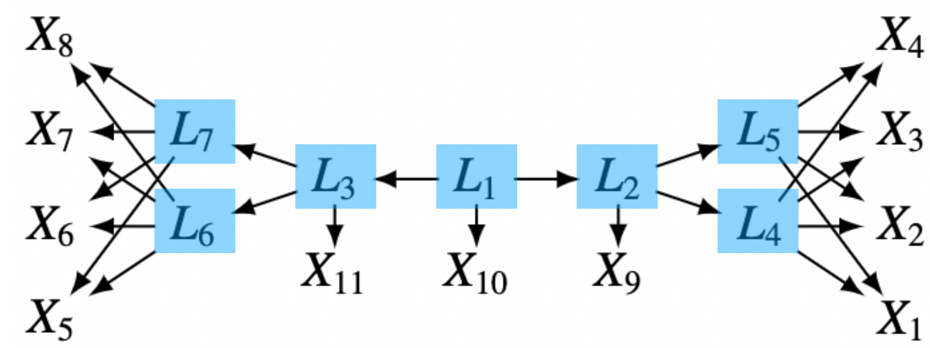
$$\mathcal{G}' = \text{refineClusters}\ (\mathcal{G}')\ ;$$

3. Refine edges and find v structures

$$\mathcal{G}' = \text{refineEdges}\ (\mathcal{G}')\ ;$$

Latent Hierarchical Causal Structure Discovery with Rank Constraints

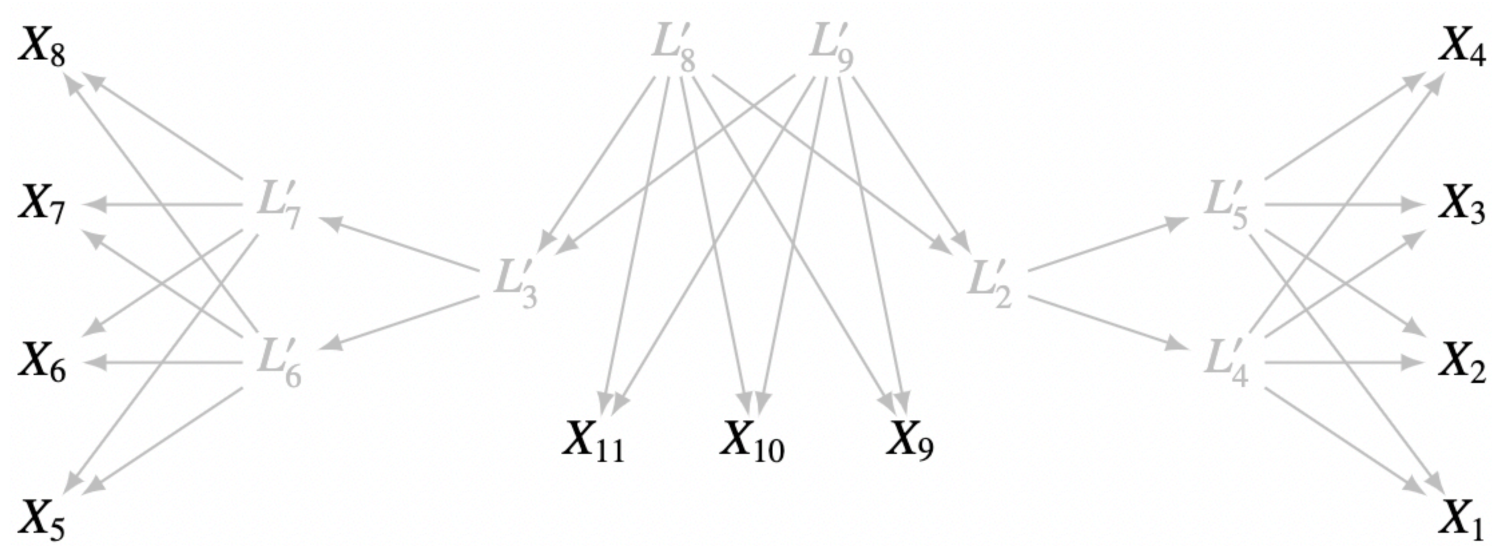
Search procedure:



1. Find clusters and assign latent covers greedily *(findCausalClusters)*

Rule 1:

If $\mathbf{A}$ is a rank-deficiency set with $\mathrm{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = k$, assign a latent variable $\mathbf{L}$ with size k as the parents of each A_i

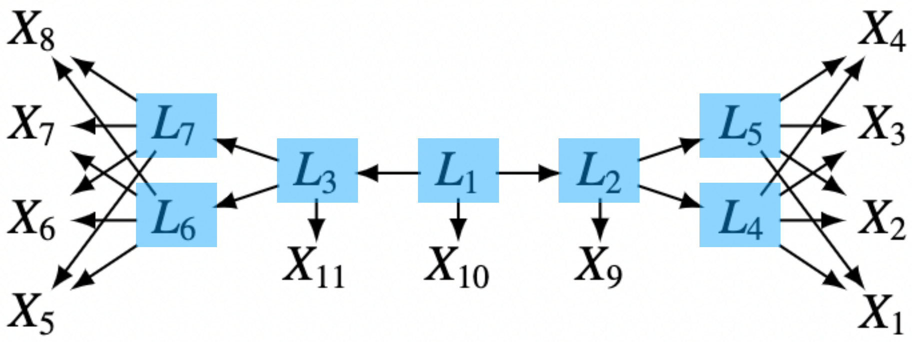


Input the data from measured variables

Latent Hierarchical Causal Structure Discovery with Rank Constraints

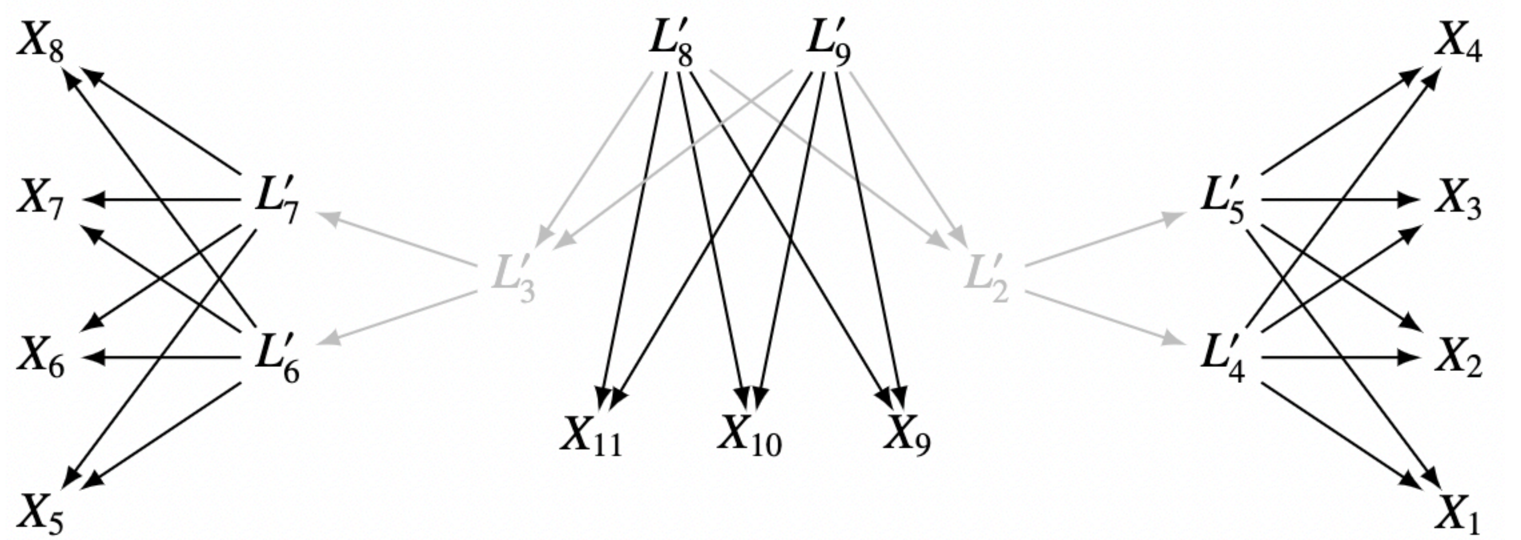
Search procedure:

1. Find clusters and assign latent covers greedily *(findCausalClusters)*



Rule 1:

If $\mathbf{A}$ is a rank-deficiency set with $\text{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = k$, assign a latent variable $\mathbf{L}$ with size k as the parents of each A_i



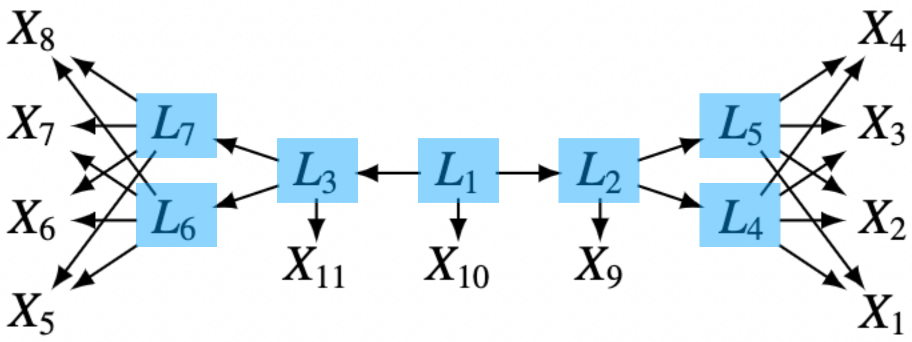
Find latent covers $\{L'_6, L'_7\}$, $\{L'_4, L'_5\}$, and $\{L'_8, L'_9\}$.

Exp:
Let $\mathbf{X}_A = \{X_5, X_6, X_7\}$ and $\mathbf{X}_B = \mathbf{X} \backslash \mathbf{X}_A$.
$\texttt{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = 2$ being rank deficient.
Add a latent cover $\{L'_6, L'_7\}$ with size 2 above $\mathbf{X}_A$.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

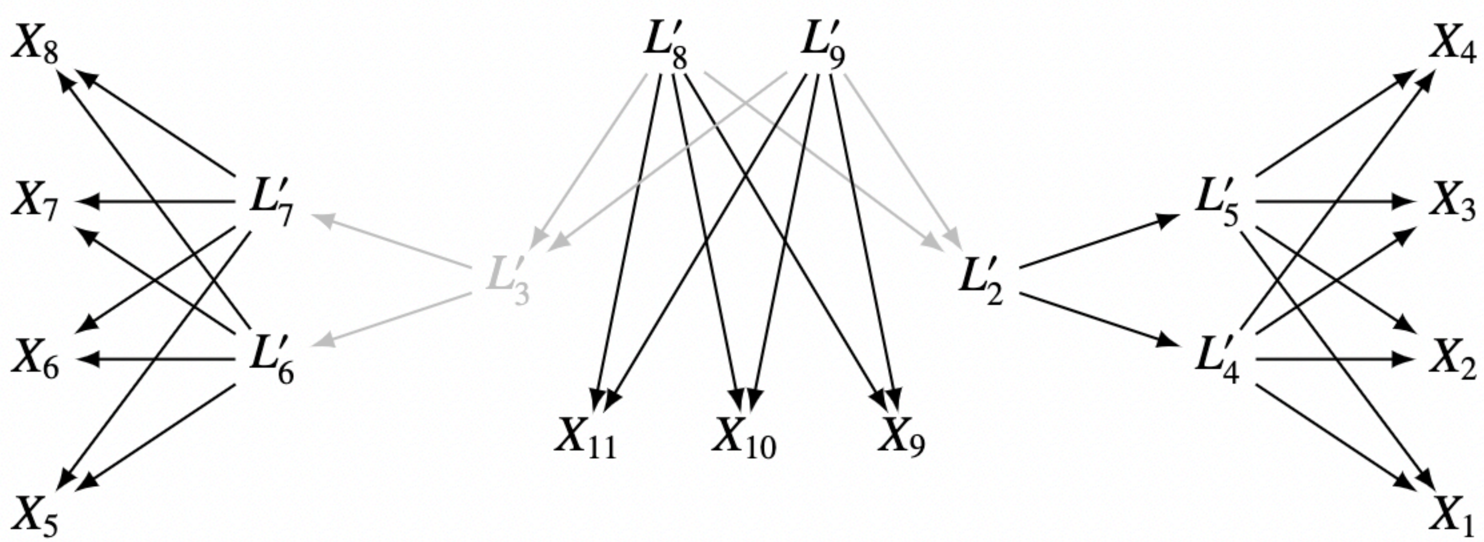
Search procedure:

1. Find clusters and assign latent covers greedily *(findCausalClusters)*



Rule 1:

If $\mathbf{A}$ is a rank-deficiency set with $\text{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = k$, assign a latent variable $\mathbf{L}$ with size k as the parents of each A_i



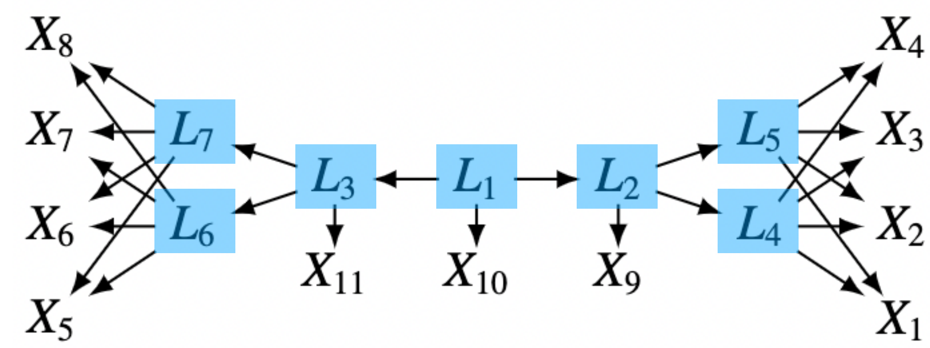
Find a latent cover $\{L'_2\}$.

Let $\mathbf{A} = \{L'_4, L'_5\}$.
Use its measured descendants $\mathbf{X}_A = \{X_1, X_2, X_3, X_4\}$ as a surrogate for rank test.
$\text{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = 1$ being rank deficient.
Add a latent cover $\{L'_2\}$ with size 1 above $\mathbf{A}$.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

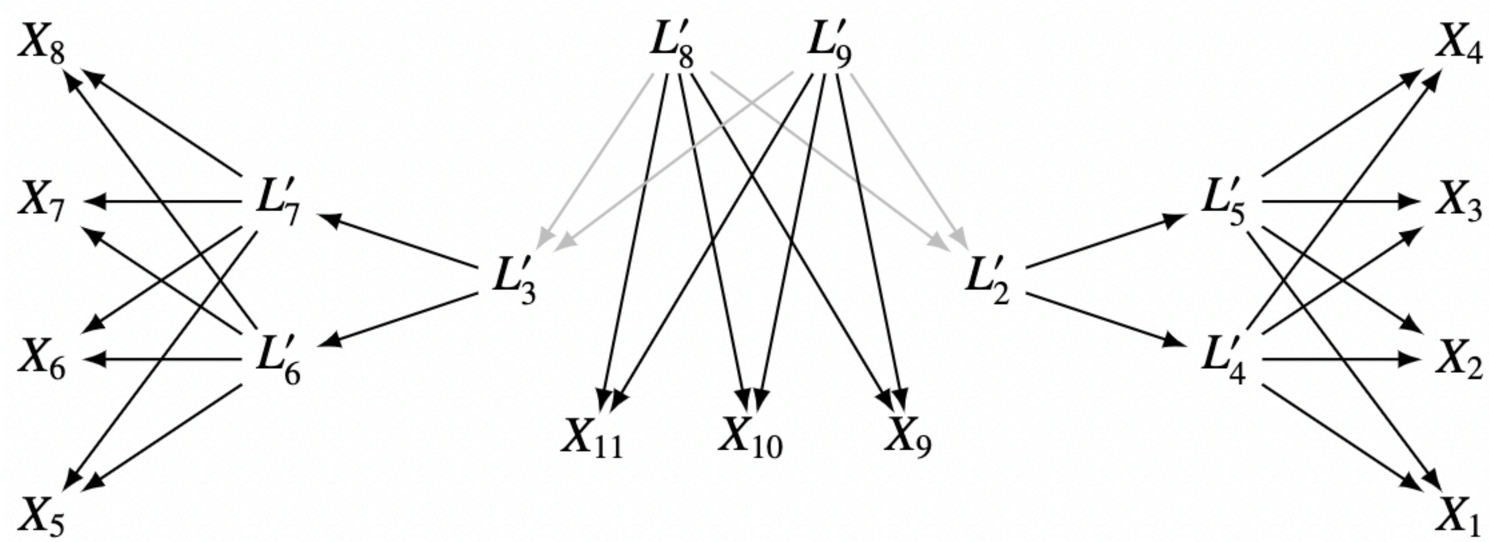
Search procedure:

1. Find clusters and assign latent covers greedily
   (findCausalClusters)



Rule 1:

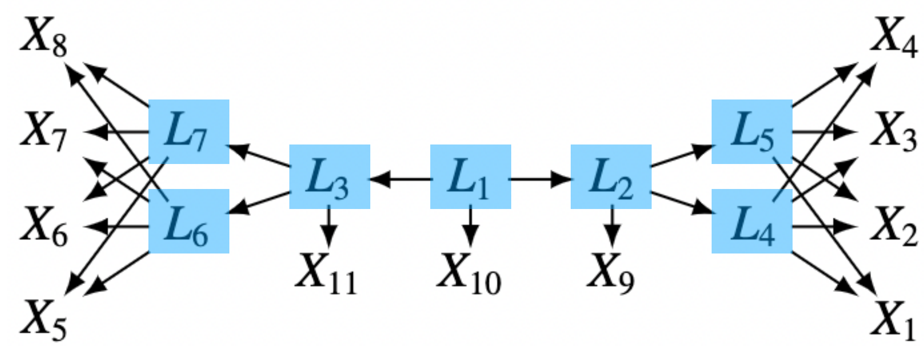
If $\mathbf{A}$ is a rank-deficiency set with $\mathrm{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = k$, assign a latent variable $\mathbf{L}$ with size k as the parents of each A_i



Find a latent cover $\{L'_3\}$.

Let $\mathbf{A} = \{L'_6, L'_7\}$.
Use its measured descendants $\mathbf{X}_A = \{X_5, X_6, X_7, X_8\}$ as a surrogate for rank test.
$\mathtt{rank}(\Sigma_{\mathbf{X}_A,\mathbf{X}_B}) = 1$ being rank deficient.
Add a latent cover $\{L'_3\}$ with size 1 above $\mathbf{A}$.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

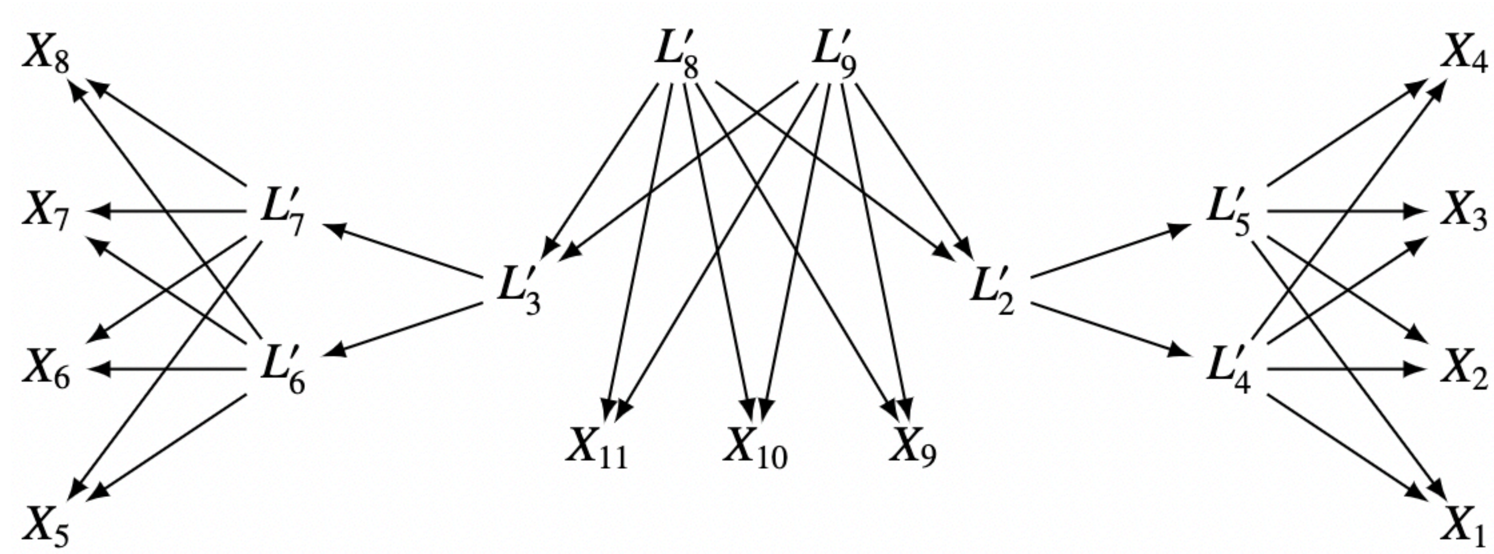


Search procedure:

1. Find clusters and assign latent covers greedily *(findCausalClusters)*

Rule 1:

If $\mathbf{A}$ is a rank-deficiency set with $\text{rank}(\Sigma_{\mathbf{X}_A, \mathbf{X}_B}) = k$, assign a latent variable $\mathbf{L}$ with size k as the parents of each A_i



Connect $\{L'_8, L'_9\}$ to $\{L'_2\}$ and to $\{L'_3\}$, and Step 1 ends.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

Input : Date from a set of measured variables $\mathbf{X}_{\mathcal{G}}$
Output : Markov equivalence class $\mathcal{G}'$

1. Find clusters and assign latent covers greedily

$$\mathcal{G}' = \text{findCausalClusters}\ (\mathbf{X}_{\mathcal{G}})\ ;$$

2. Refine incorrect clusters and covers from greedy search

$$\mathcal{G}' = \text{refineClusters}\ (\mathcal{G}')\ ;$$

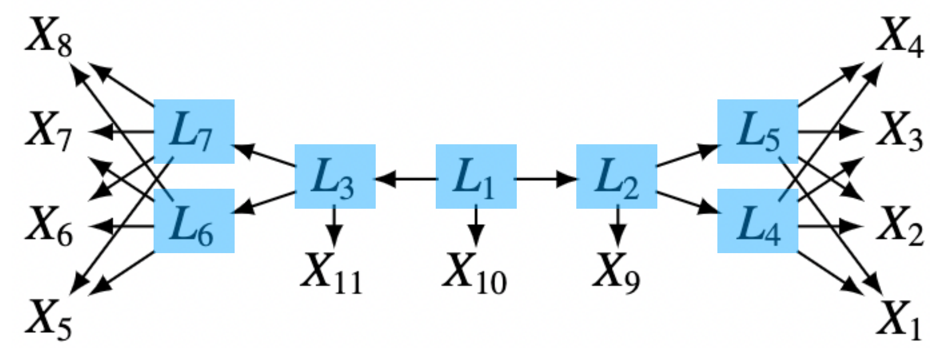
3. Refine edges and find v structures

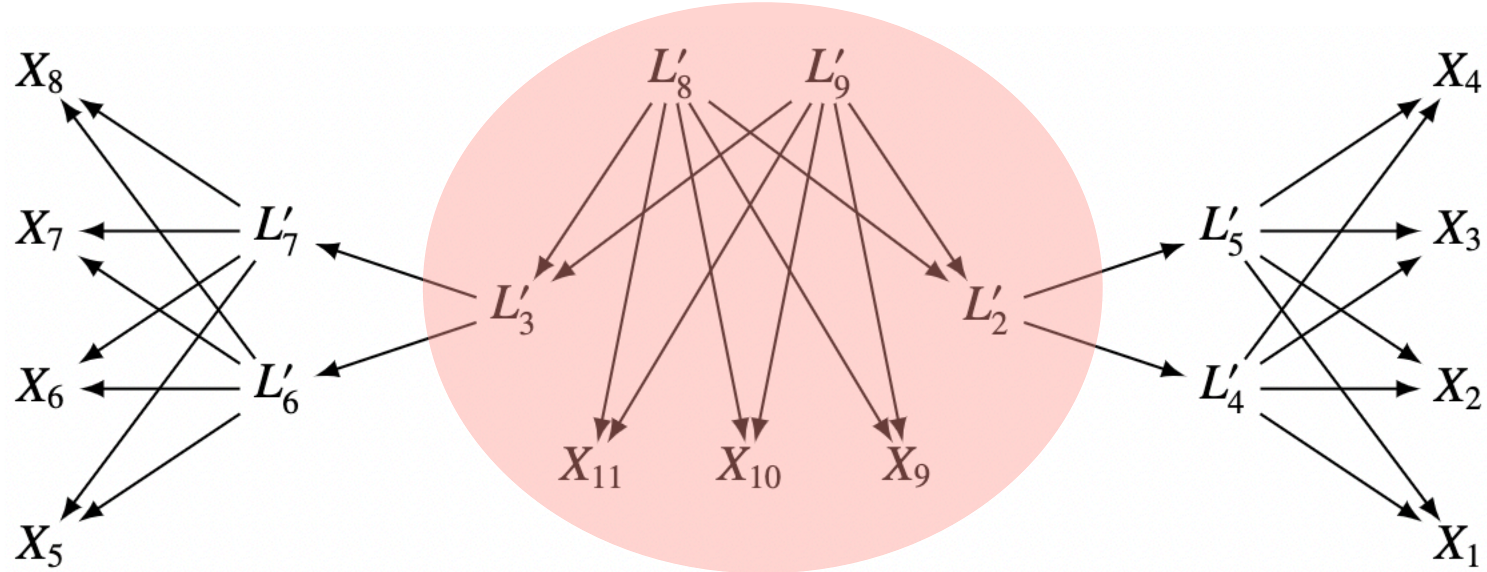
$$\mathcal{G}' = \text{refineEdges}\ (\mathcal{G}')\ ;$$

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Pink area: incorrect clusters due to the greedy search in step 1

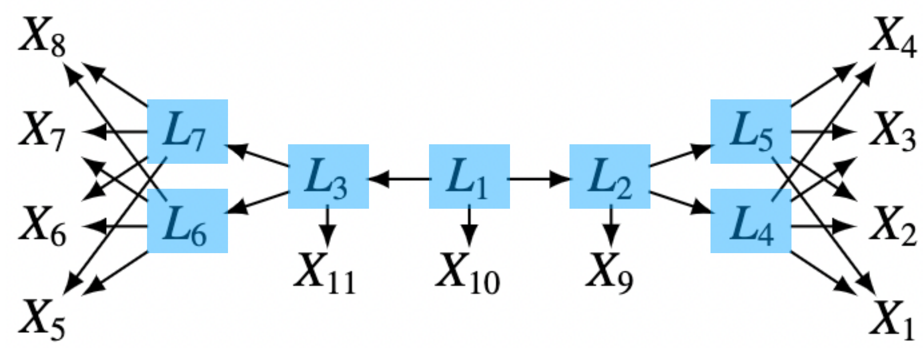
Rule 2:

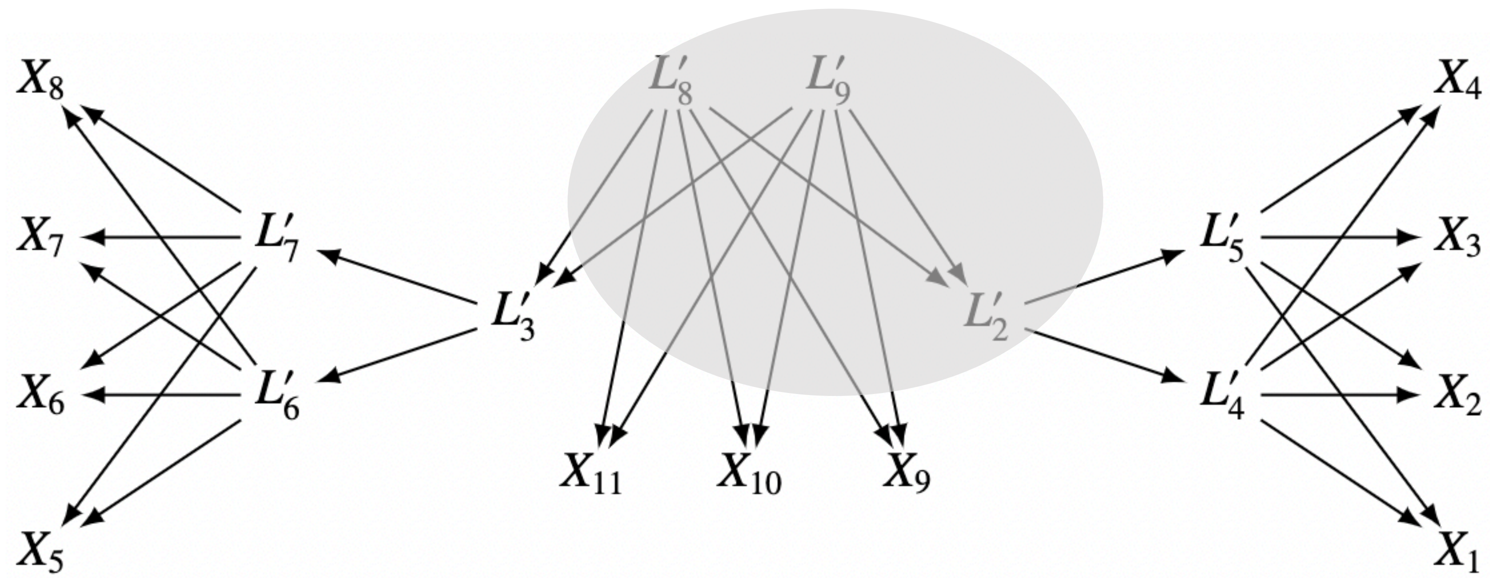
For each discovered latent cover $\mathbf{L}$,
let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$
and apply findCausalClusters to
$\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Refine $\{L'_2\}$ by first removing $\{L'_2\}$ and its parents $\{L'_8, L'_9\}$.

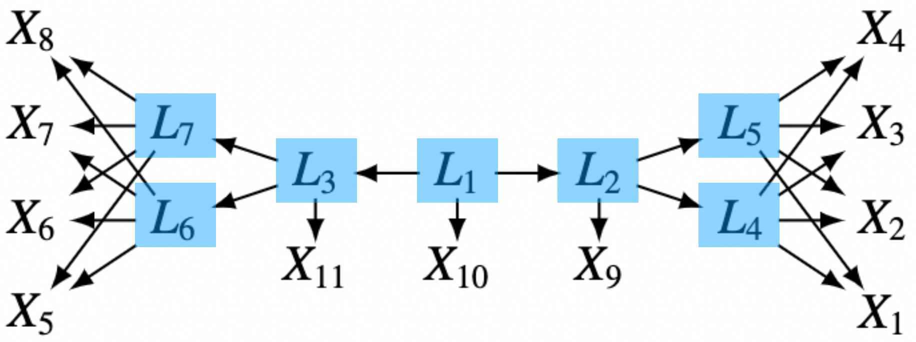
Rule 2:

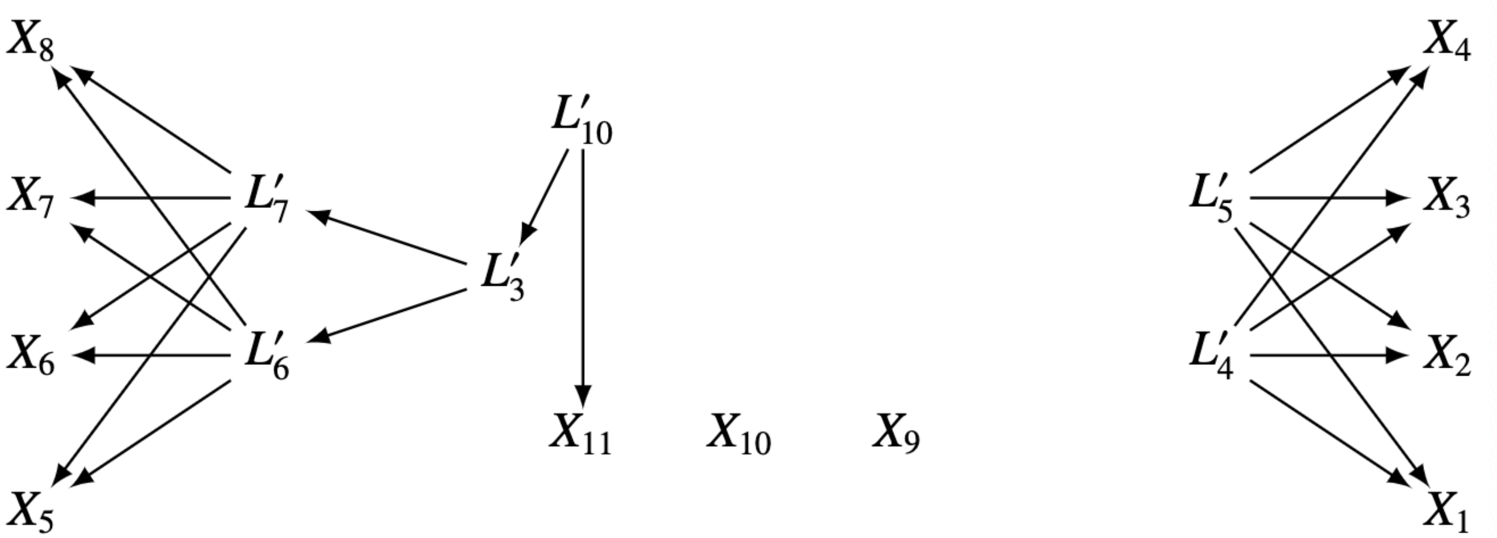
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Rule 2:

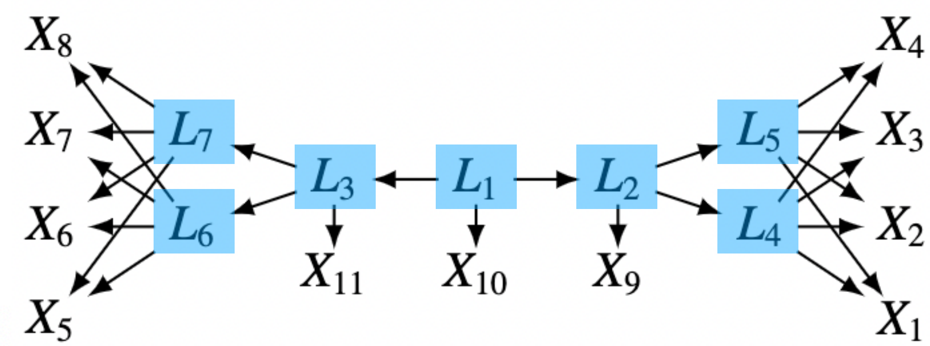
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

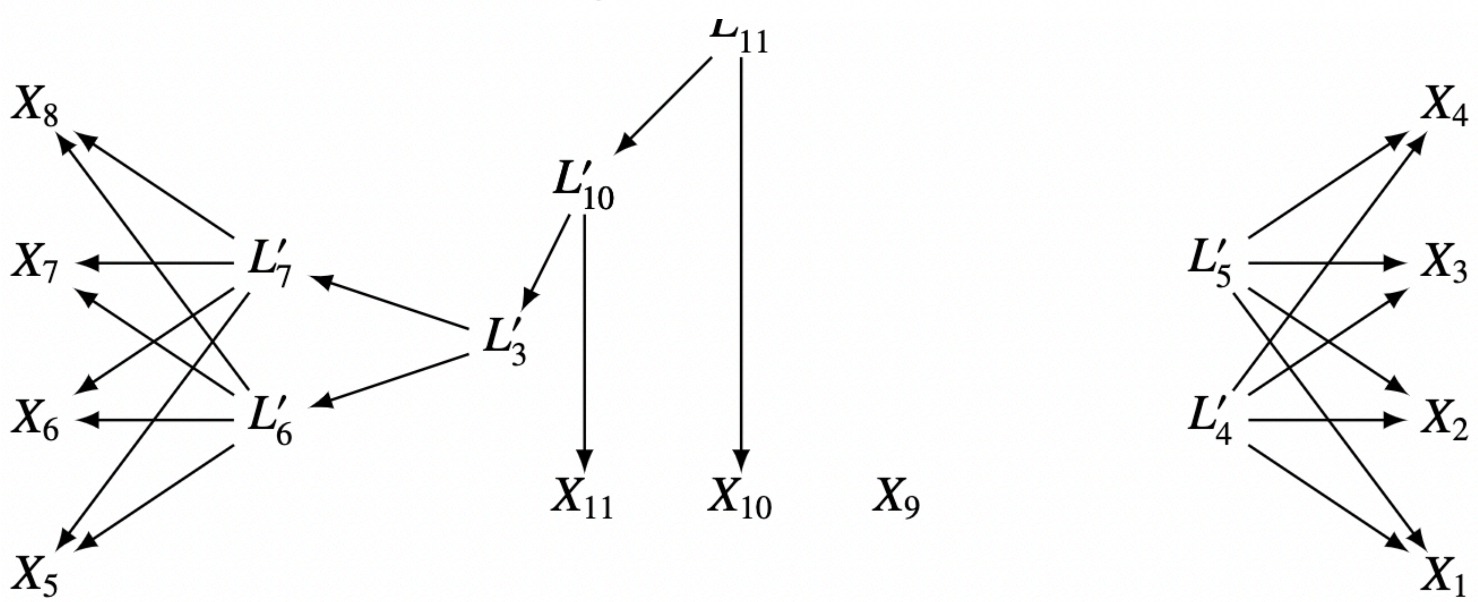
Perform *findCausalClusters* over $\{L'_3, L'_4, L'_5, X_9, X_{10}, X_{11}\}$, and then we can find a latent cover $\{L'_{10}\}$.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Next perform *findCausalClusters* over $\{L'_{10}, L'_4, L'_5, X_9, X_{10}\}$, and we can find a latent cover $\{L'_{11}\}$.

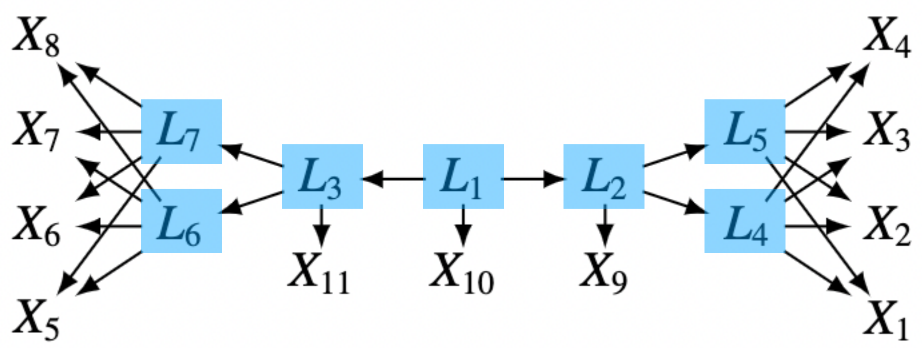
Rule 2:

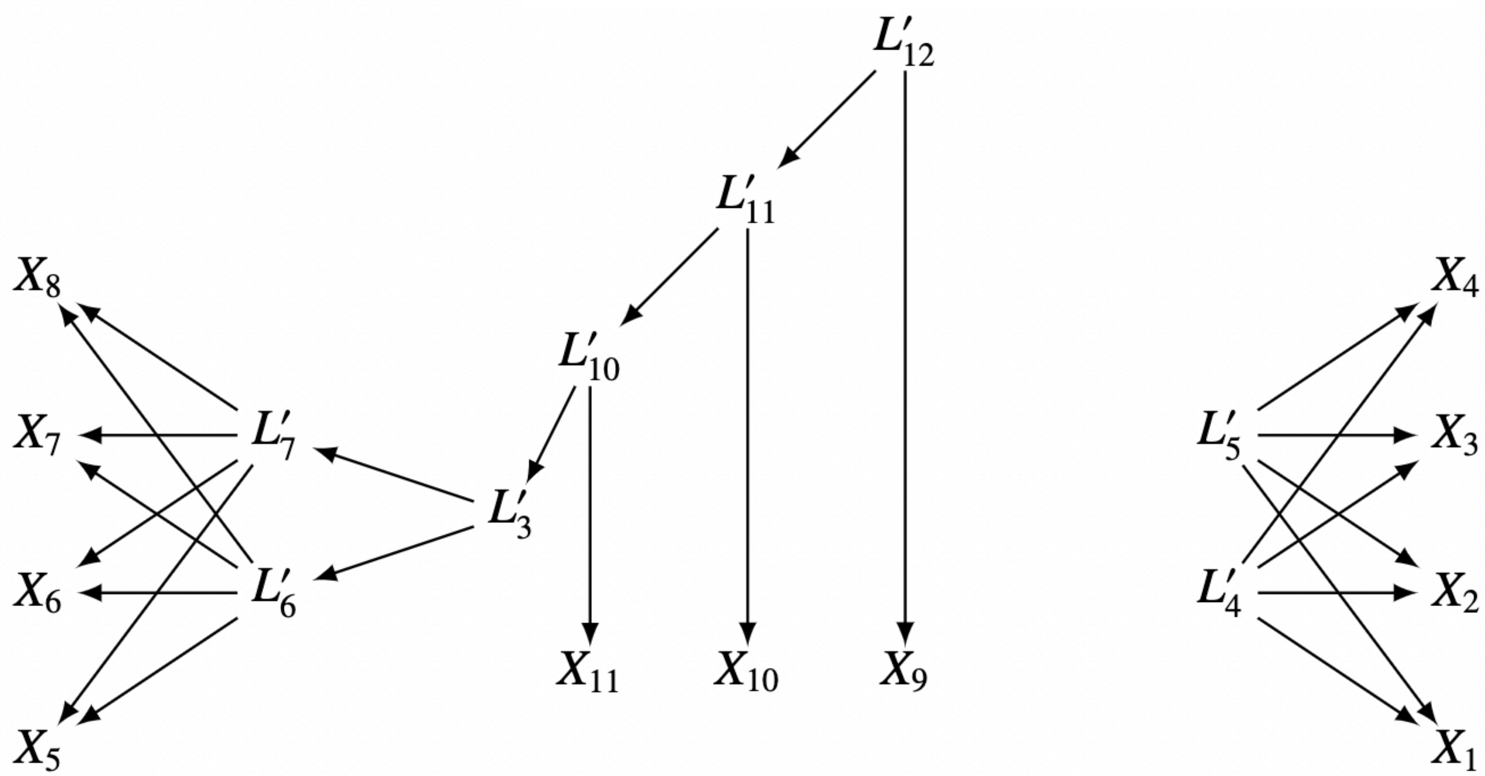
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Next perform *findCausalClusters* over $\{L'_{11}, L'_4, L'_5, X_9\}$, and we can find a latent cover L'_{12}.

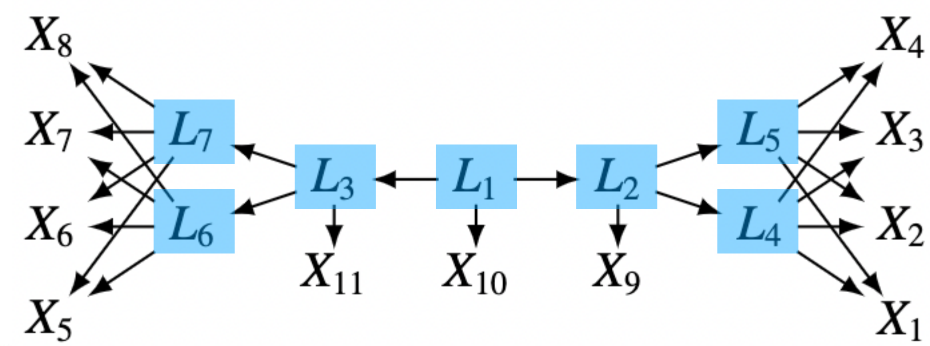
Rule 2:

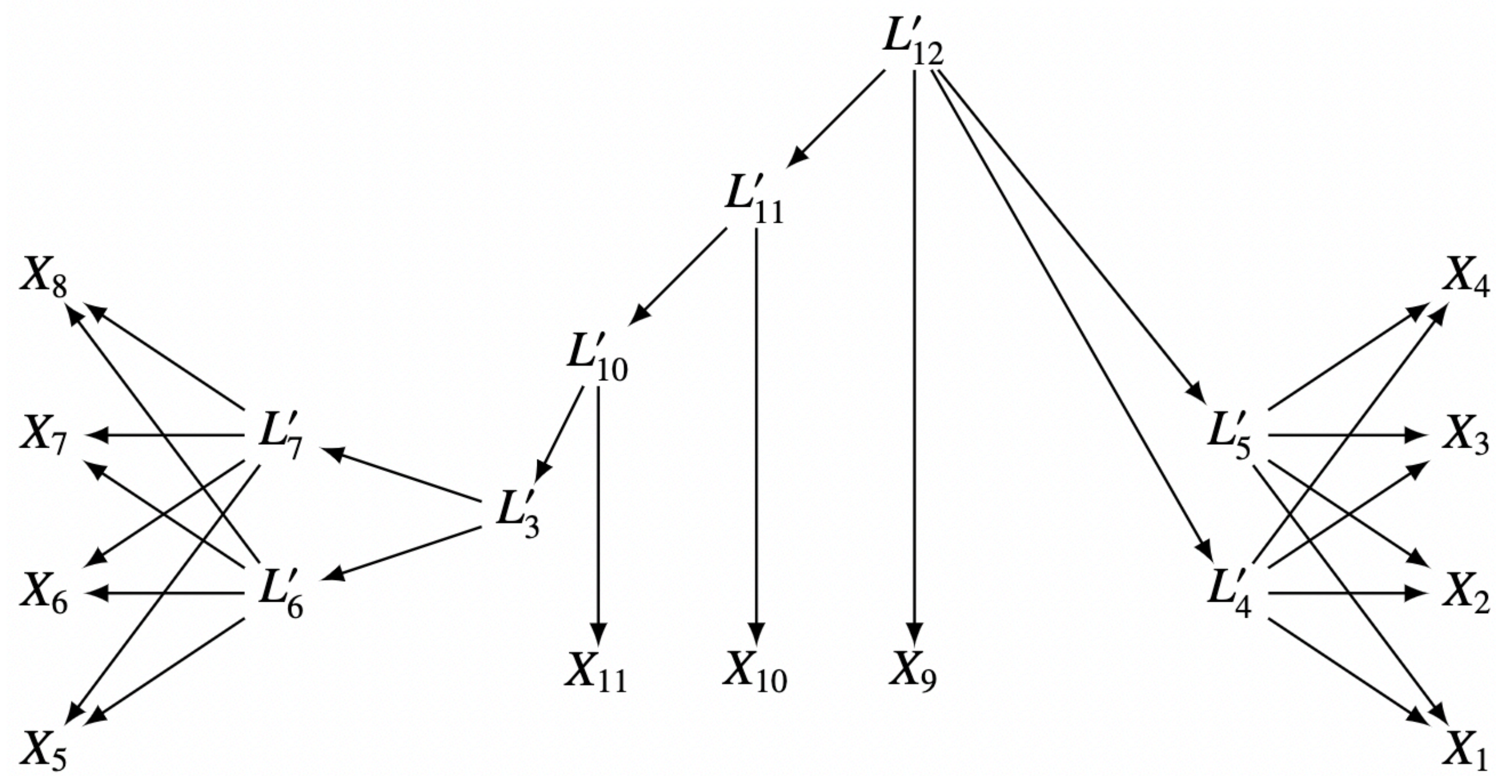
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Connect $\{L'_{12}\}$ to $\{L'_4, L'_5\}$.

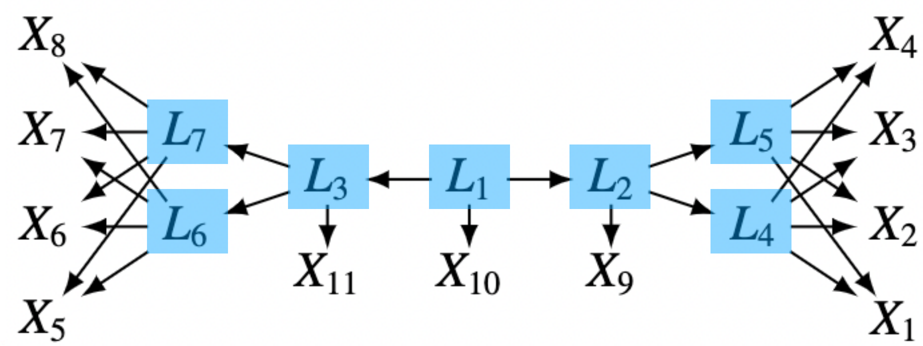
Rule 2:

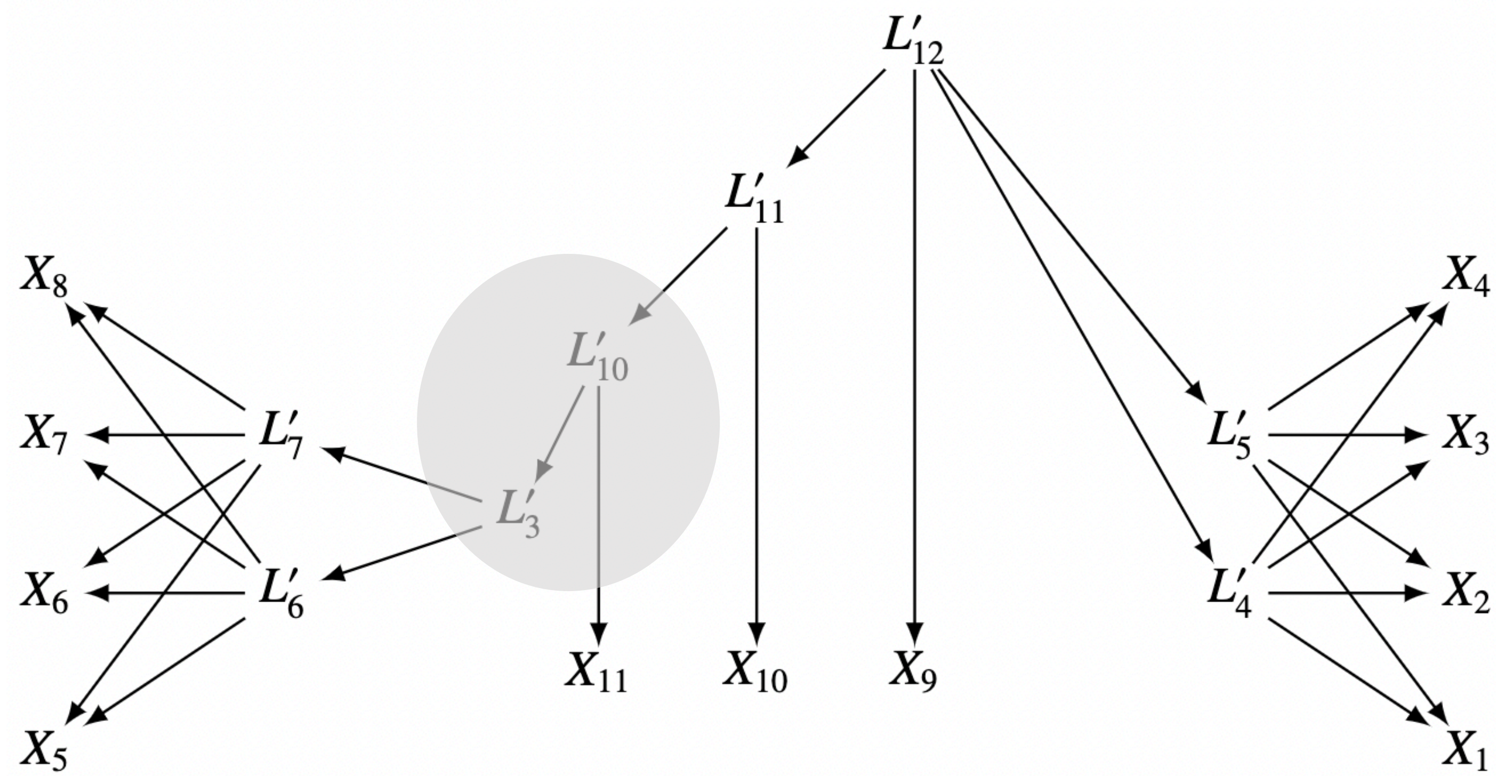
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*



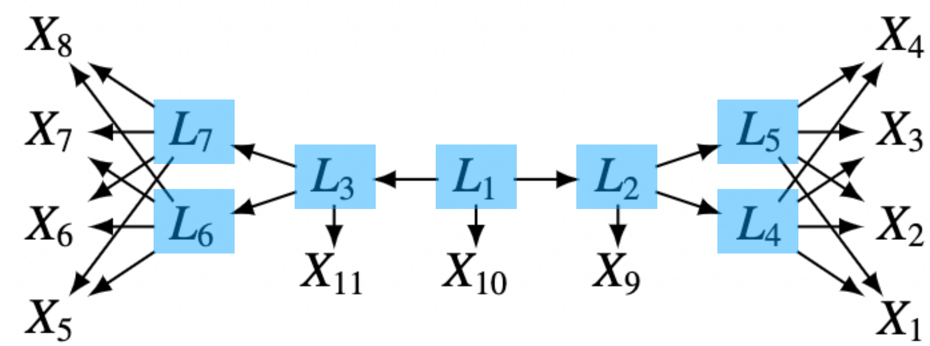


Refine $\{L'_3\}$ by first removing $\{L'_3\}$ and its parents $\{L'_{10}\}$.

Rule 2:

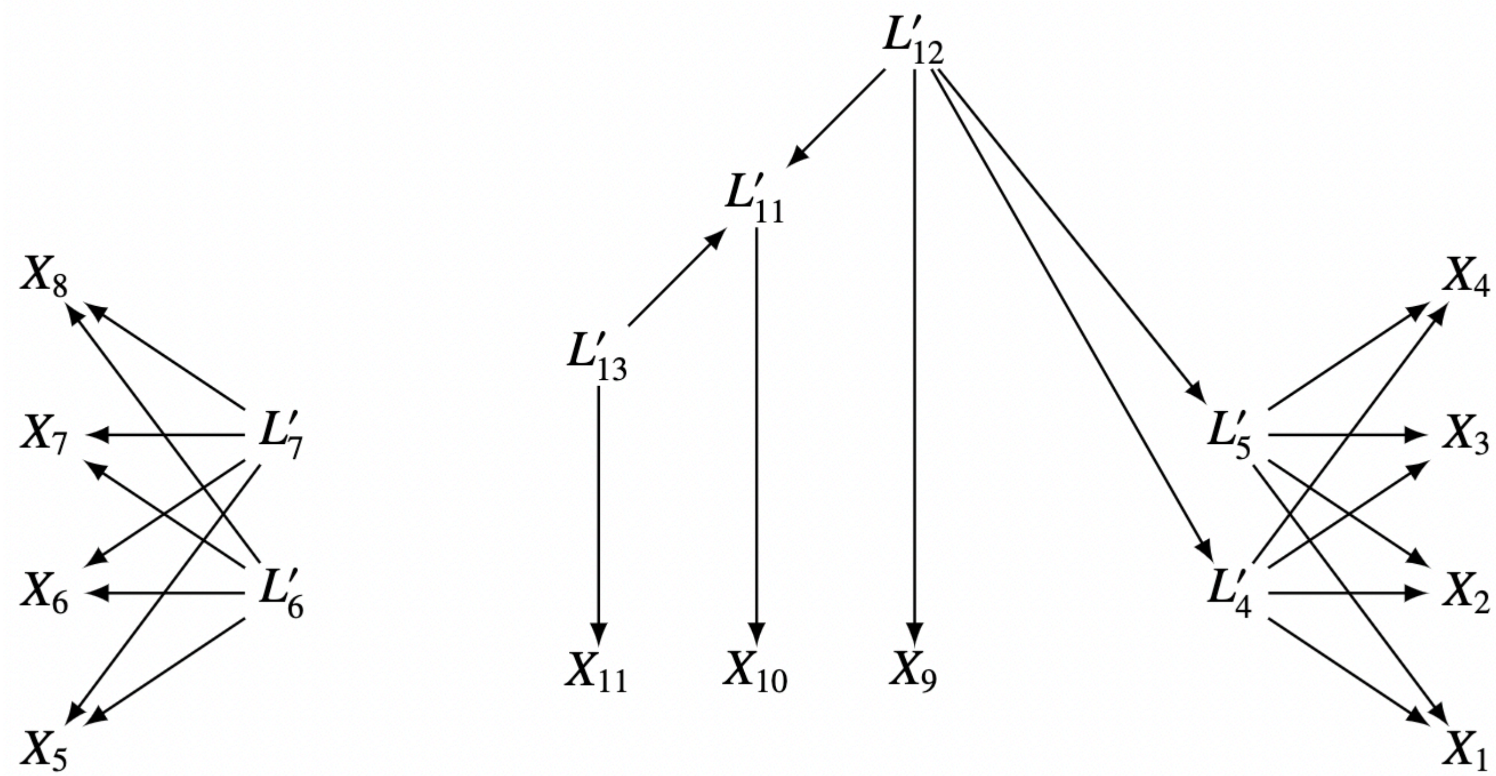
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints



Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*



Perform *findCausalClusters* over $\{L'_6, L'_7, L'_{11}, X_{11}\}$, and then we can find a latent cover $\{L'_{13}\}$.

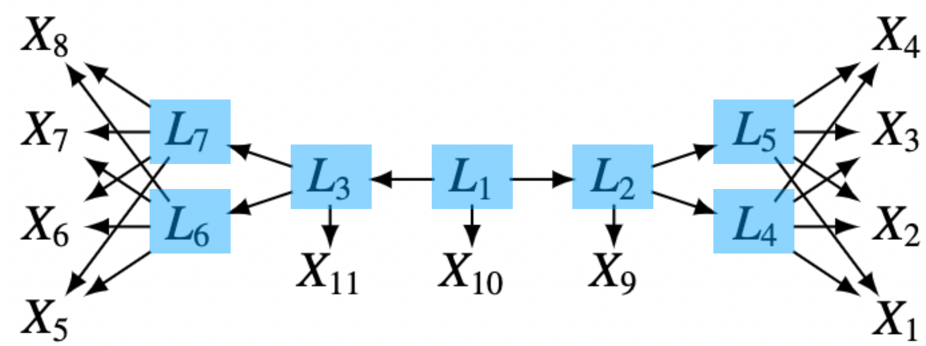
Rule 2:

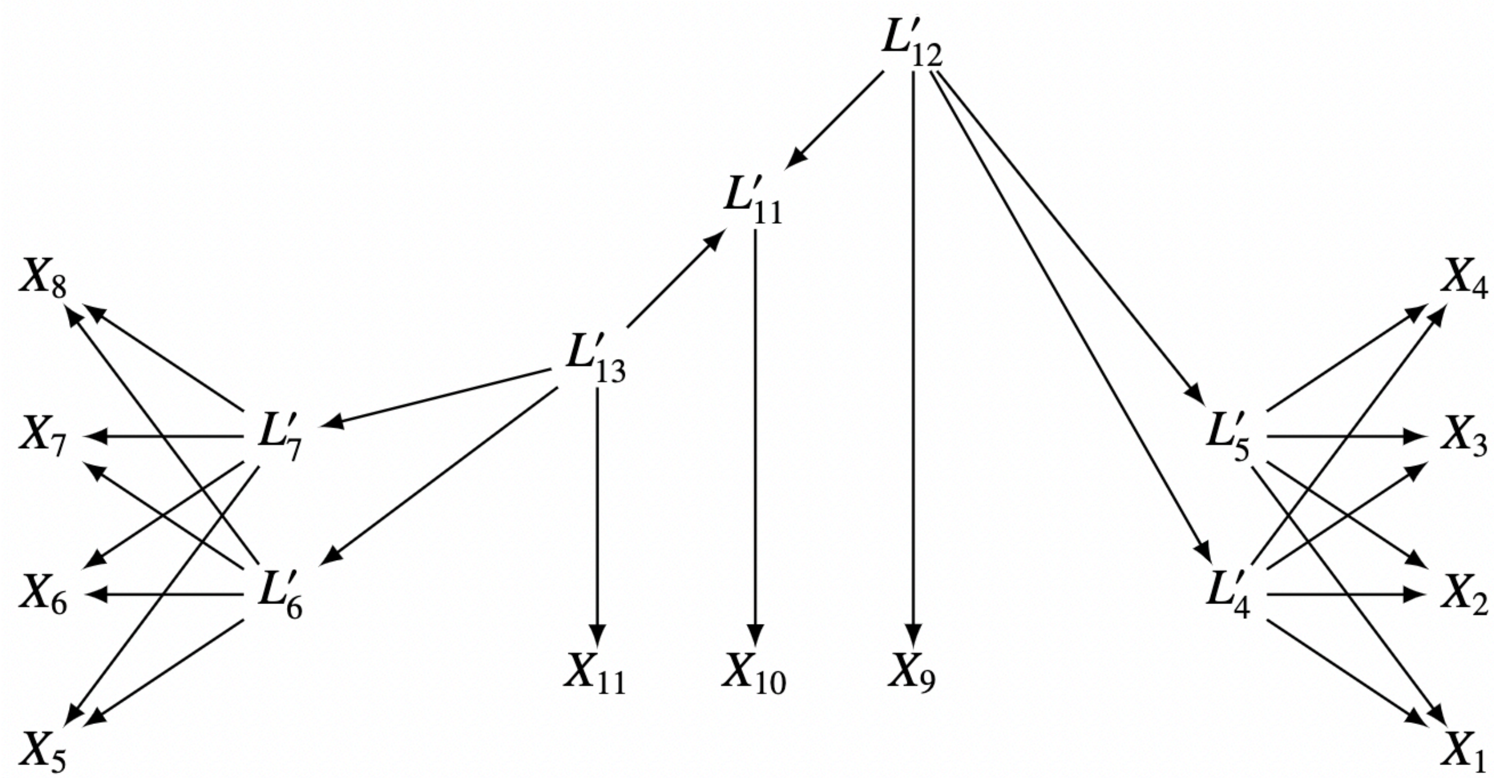
For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

2. Refine incorrect clusters and covers from greedy search *(refineClusters)*





Connect $\{L'_{13}\}$ to $\{L'_6, L'_7\}$.

Rule 2:

For each discovered latent cover $\mathbf{L}$, let $\mathbf{V} = Gp_{G'}(\mathbf{L}) \cup Sib_{G'}(\mathbf{L}) \cup Ch_{G'}(\mathbf{L})$ and apply findCausalClusters to $\mathbf{V}$ to refine the clusters

Latent Hierarchical Causal Structure Discovery with Rank Constraints

Search procedure:

Input : Date from a set of measured variables $\mathbf{X}_{\mathcal{G}}$
Output : Markov equivalence class $\mathcal{G}'$

1. Find clusters and assign latent covers greedily

$$\mathcal{G}' = \text{findCausalClusters}\ (\mathbf{X}_{\mathcal{G}})\ ;$$

2. Refine incorrect clusters and covers from greedy search

$$\mathcal{G}' = \text{refineClusters}\ (\mathcal{G}')\ ;$$

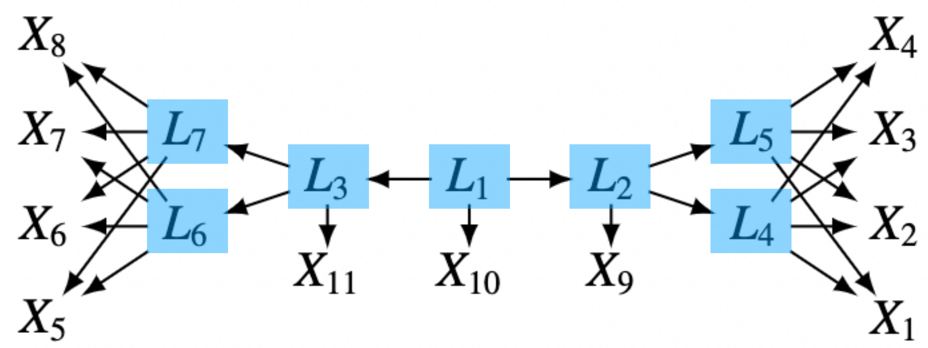
3. **Refine edges and find v structures**

$$\mathcal{G}' = \text{refineEdges}\ (\mathcal{G}')\ ;$$

Latent Hierarchical Causal Structure Discovery with Rank Constraints

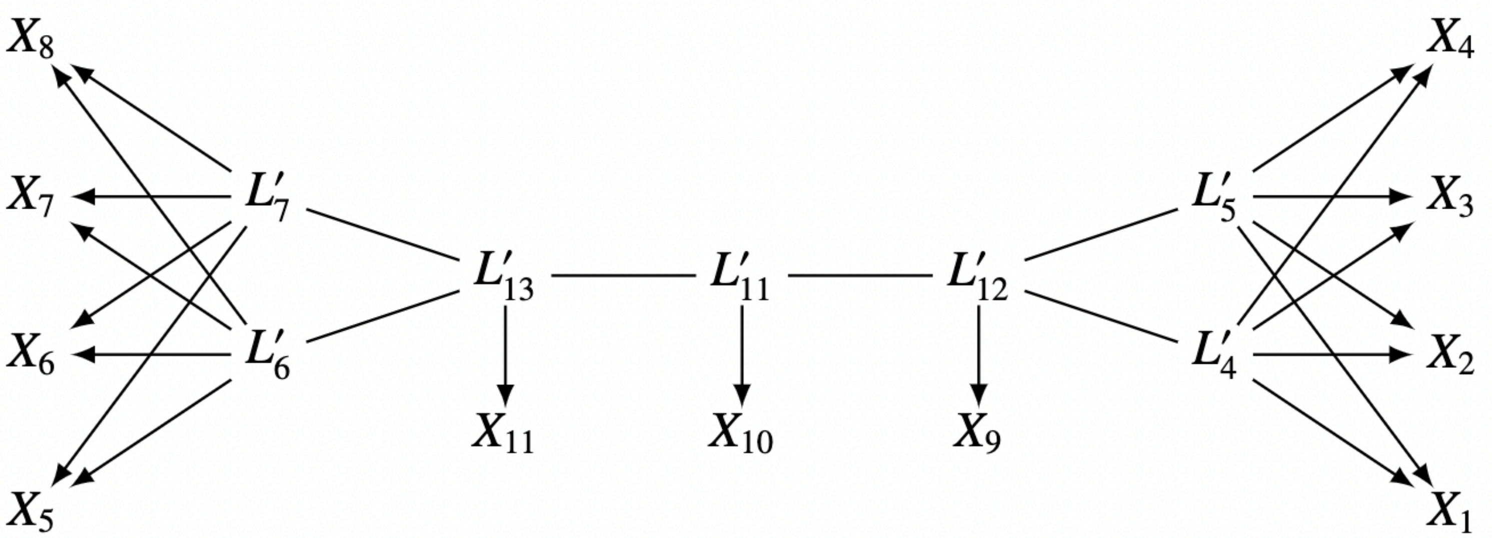
Search procedure:

3. Refine edges and find v structures
   (refineEdges)



Rule 3:

For a pair $(\mathbf{L}_A, \mathbf{L}_B)$,
let $\mathcal{A} \leftarrow \{\mathbf{L}_A, \mathbf{C}_1^A, \mathbf{C}_2^A, ...\}$ and
$\mathcal{B} \leftarrow \{\mathbf{L}_B, \mathbf{C}_1^B, \mathbf{C}_2^B, ...\}$.
If there exists such $\mathcal{A}, \mathcal{B}$ such that
$\texttt{rank}_{\mathcal{G}'}(\Sigma_{\mathcal{A},\mathcal{B}})$ is rank deficient,
remove all edges between $\mathbf{L}_A, \mathbf{L}_B$ in $\mathcal{G}'$



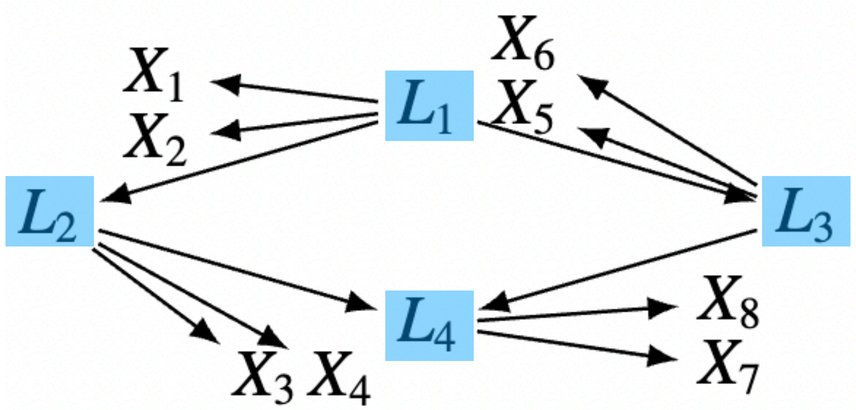
Perform *refineEdges* to refine the edges and output the Markov equivalence class.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

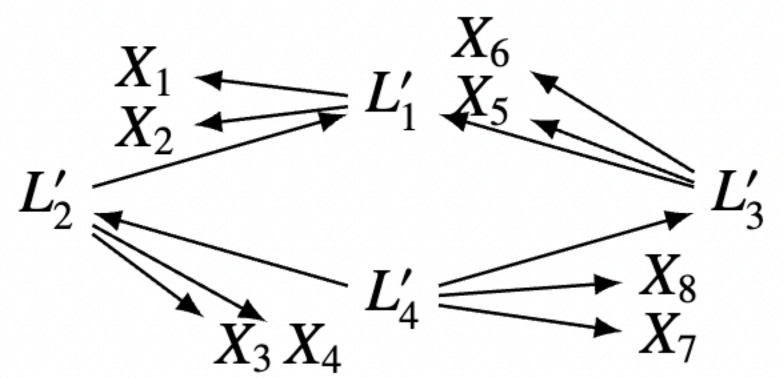
Search procedure:

3. Refine edges and find v structures
(refineEdges)

Another example that contains the v structure:



A graph with v structure.

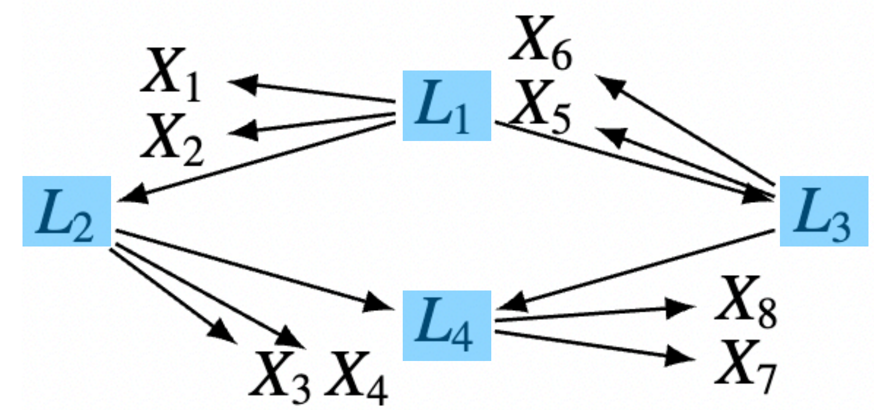


Possible Output from Phase II.

Latent Hierarchical Causal Structure Discovery with Rank Constraints

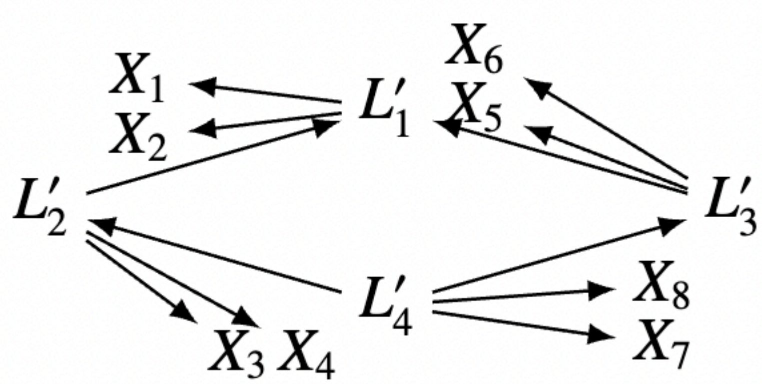
Search procedure:

3. Refine edges and find v structures
   (refineEdges)

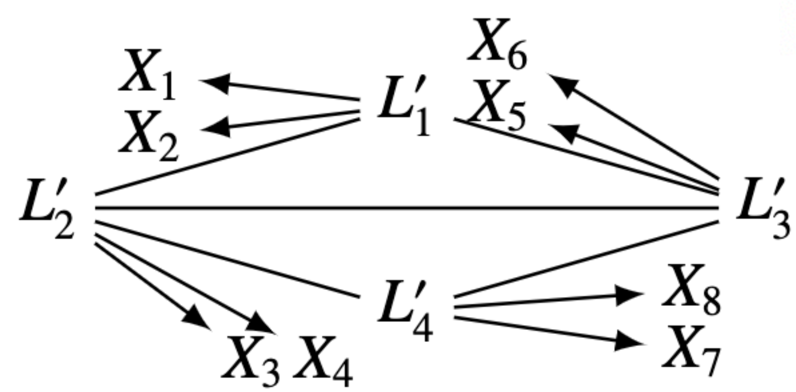


A graph with v structure.

Another example that contains the v structure:



Possible Output from Phase II.



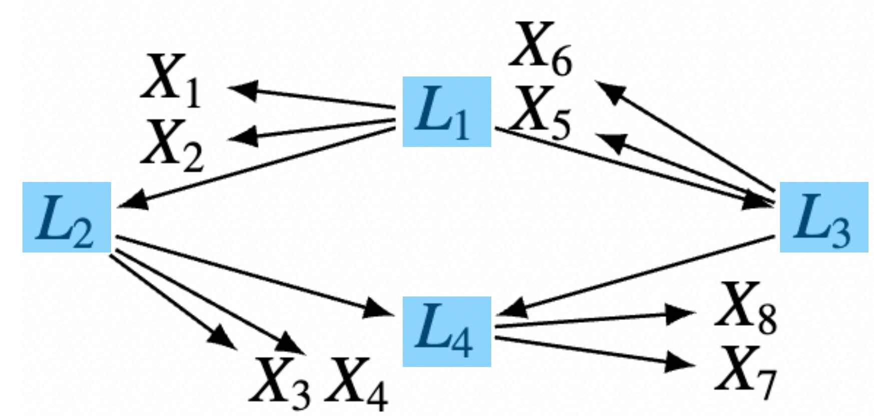
Connect L'_2 and L'_3

Latent Hierarchical Causal Structure Discovery with Rank Constraints

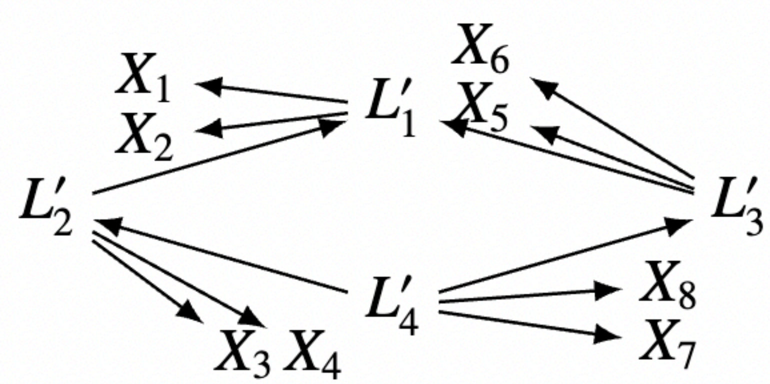
Search procedure:

3. Refine edges and find v structures
   (refineEdges)

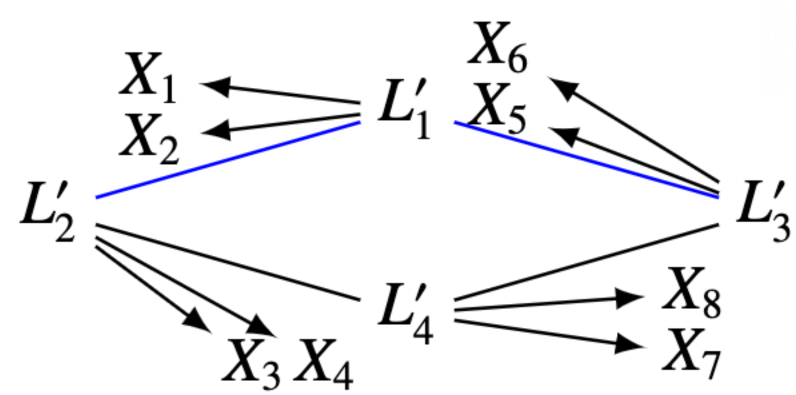
Another example that contains the v structure:



A graph with v structure.



Possible Output from Phase II.



Remove the edge between L'_2 and L'_3

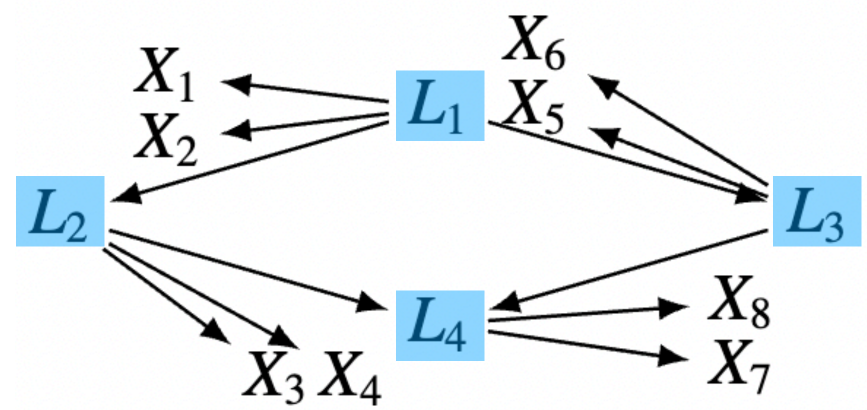
Let $\mathcal{A} = \{L'_2, X_1\}$ and $\mathcal{B} = \{L'_3, X_2\}$
$\texttt{rank}(\Sigma_{\mathcal{A},\mathcal{B}}) = 1$ being rank deficient
It means that L'_1 d-separates L'_2 from L'_3
Remove the edge between L'_2 and L'_3

Latent Hierarchical Causal Structure Discovery with Rank Constraints

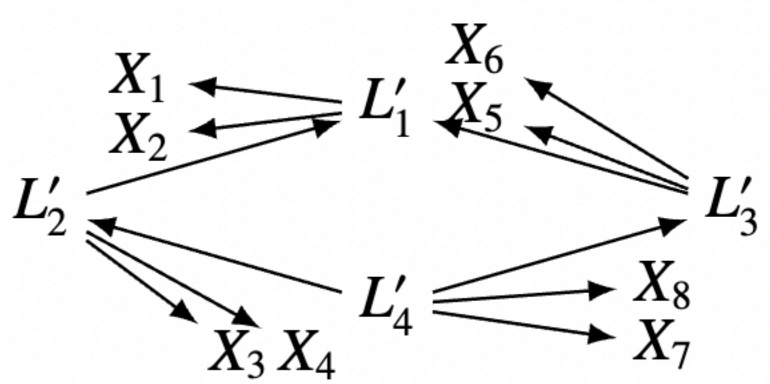
Search procedure:

3. Refine edges and find v structures *(refineEdges)*

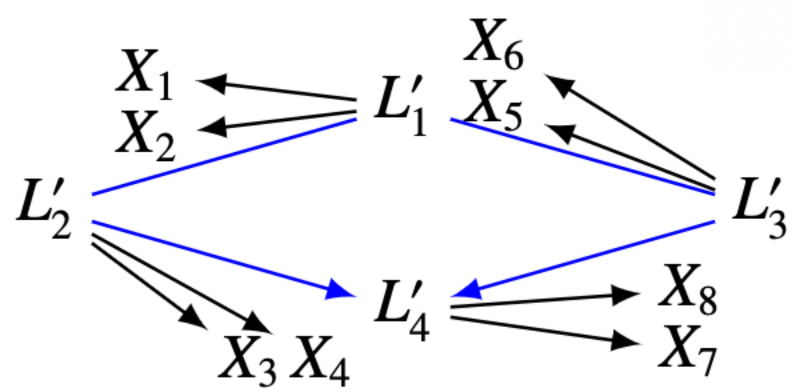
Another example that contains the v structure:



A graph with v structure.



Possible Output from Phase II.



Form a V structure: $L'_2 \rightarrow L'_4 \leftarrow L'_3$.

Let $\mathcal{A}^1 = \{L'_2, L'_4, X_1\}$ and $\mathcal{B}^1 = \{L'_3, X_2\}$

$\mathrm{rank}(\Sigma_{\mathcal{A}^1,\mathcal{B}^1}) = 2 > 1 = \mathrm{rank}(\Sigma_{\mathcal{A},\mathcal{B}})$

Let $\mathcal{A}^2 = \{L'_2, X_1\}$ and $\mathcal{B}^2 = \{L'_3, L'_4, X_2\}$

$\mathrm{rank}(\Sigma_{\mathcal{A}^2,\mathcal{B}^2}) = 2 > 1 = \mathrm{rank}(\Sigma_{\mathcal{A},\mathcal{B}})$

so $L'_2 \rightarrow L'_4 \leftarrow L'_3$

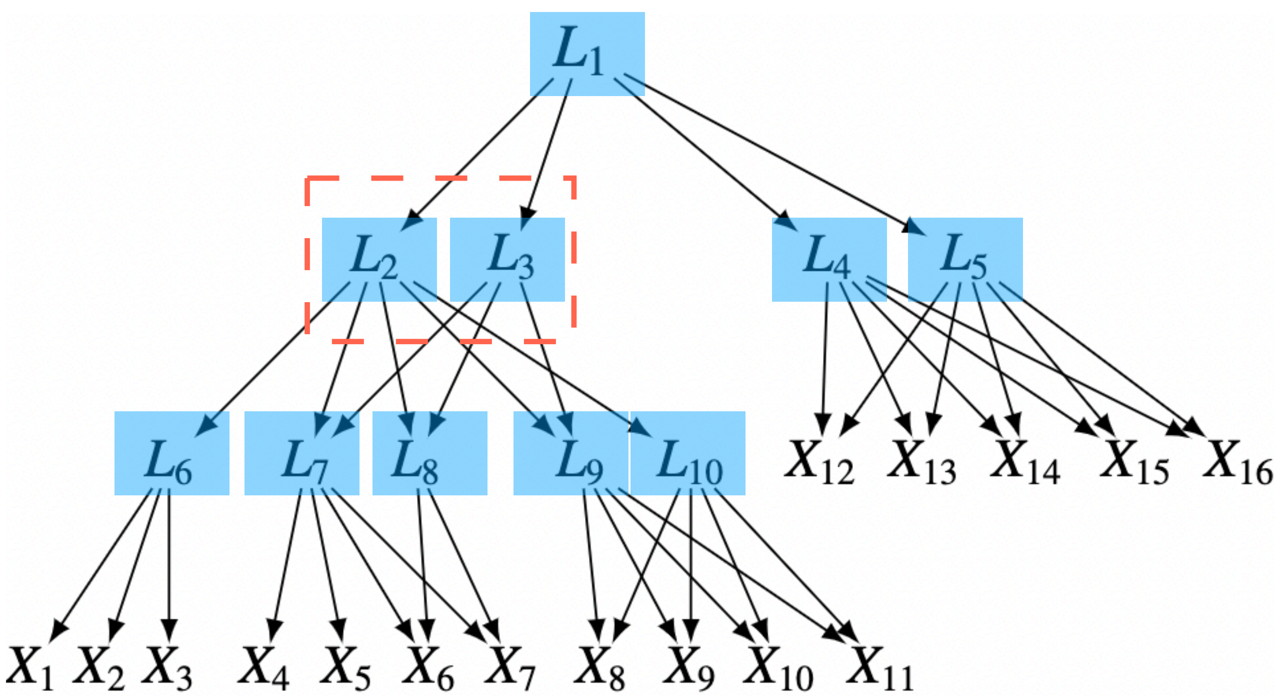
Latent Hierarchical Causal Structure Discovery with Rank Constraints

Main identifiability condition:

For each latent variable set **L** with size k, it has

- at least k+1 pure children (can be either latent or measured) and
- another k+1 neighbors

Pure children: if **V** are pure children of latent variables **L**, then **V** do not have other parents besides **L**



Exp:
Let **L**={L_2,L_3} with size 2:
3 pure children: {L_6,L_7,L_8}
3 neighbors: {L_1,L_9,L_{10}}

Latent Hierarchical Causal Structure Discovery with Rank Constraints

The proposed approach works for

✓ Latent hierarchical causal structure with linear causal relations

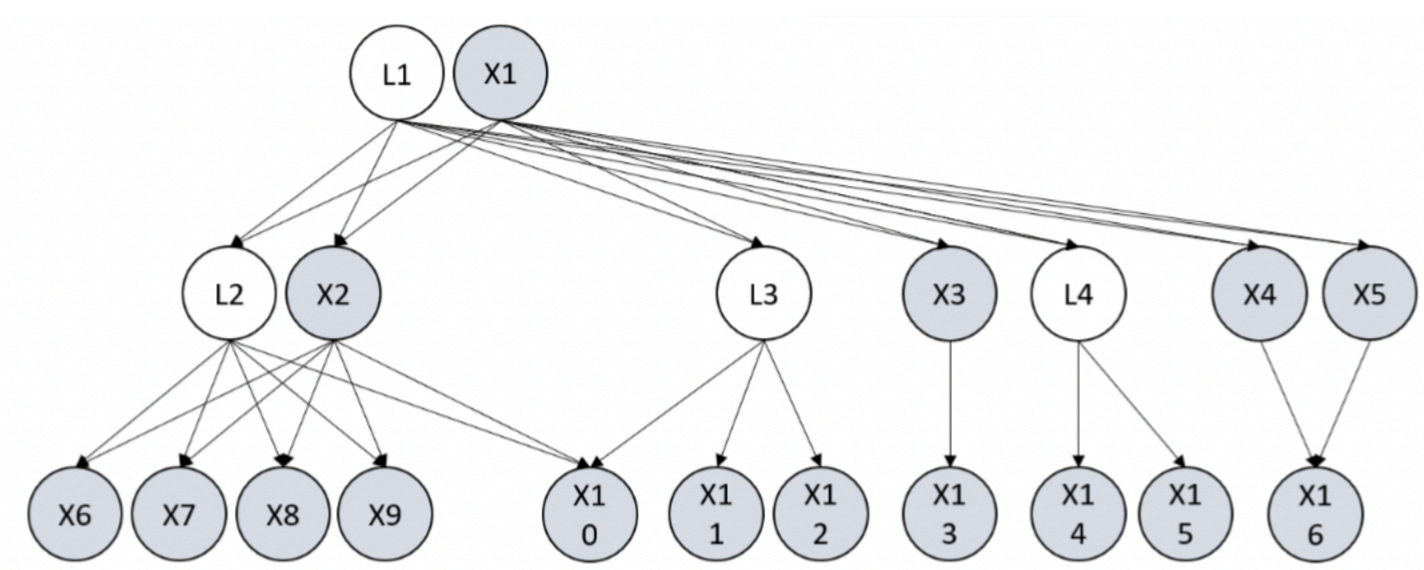
✓ Each latent variable set with size k has at least (k+1) pure children

✓ Multiple latent parents for each (measured or latent) variable

Extension to the general case: to allow observed variables to be causes as well

- Direct causal influences among observed variables
  - In the special case with no latent variables, it returns the same graph as the PC algorithm

- Causal-related latent variables
  - Latent variables may form a hierarchical structure
  - Latent variables can serve as both confounders and intermediate variables for observed variables

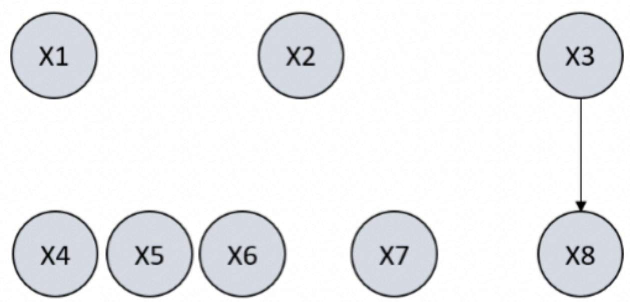


Xinshuai Dong, Biwei Huang, Peter Spirtes, Kun Zhang, et al., ongoing work

Extension to the general case: to allow observed variables to be causes as well

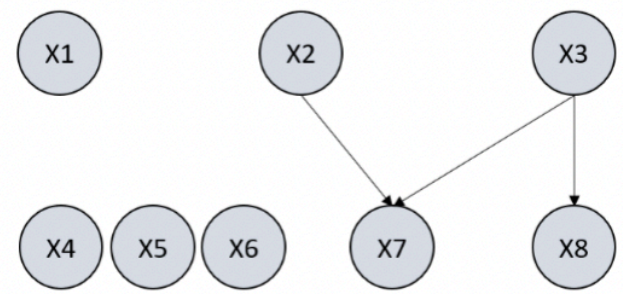
Conditional independence Vs. Rank:

let A, B, and C be disjoint subsets of $[m]$. Then the conditional independence statement $X_A \perp\!\!\!\perp X_B | X_C$ holds for X, if and only if $\Sigma_{A\cup C, B\cup C}$ has rank C.



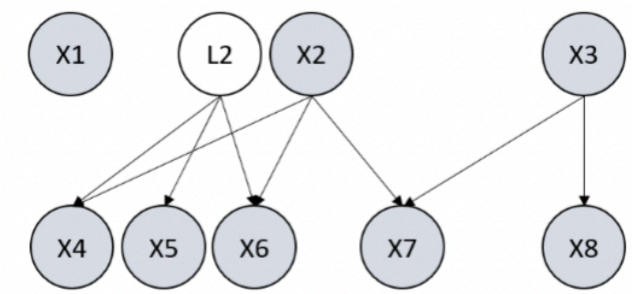
(a) Take $\mathcal{X} = \{\{X_3\}\}$ and $\mathcal{C} = \{\{X_8\}\}$.

$$\boldsymbol{rank}(\Sigma_{\mathcal{C}\cup\mathcal{X},\mathcal{N}\cup\mathcal{X}}) = 1$$



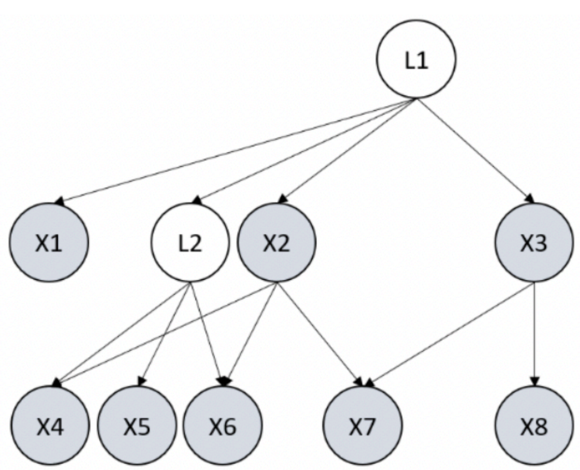
(b) Take $\mathcal{X} = \{\{X_2, X_3\}\}$ and $\mathcal{C} = \{\{X_7\}, \{X_8\}\}$.

$$\boldsymbol{rank}(\Sigma_{\mathcal{C}\cup\mathcal{X},\mathcal{N}\cup\mathcal{X}}) = 2$$



(c) Take $\mathcal{X} = \{\{X_2\}\}$ and $\mathcal{C} = \{\{X_4\}, \{X_5\}, \{X_6\}\}$.

$$\boldsymbol{rank}(\Sigma_{\mathcal{C}\cup\mathcal{X},\mathcal{N}\cup\mathcal{X}}) = 2$$



(d) Take $\mathcal{X} = \{\}$ and $\mathcal{C} = \{\{X_1\}, \{L_2, X_2\}, \{X_3\}\}$

$$\boldsymbol{rank}(\Sigma_{\mathcal{C}\cup\mathcal{X},\mathcal{N}\cup\mathcal{X}}) = 1$$

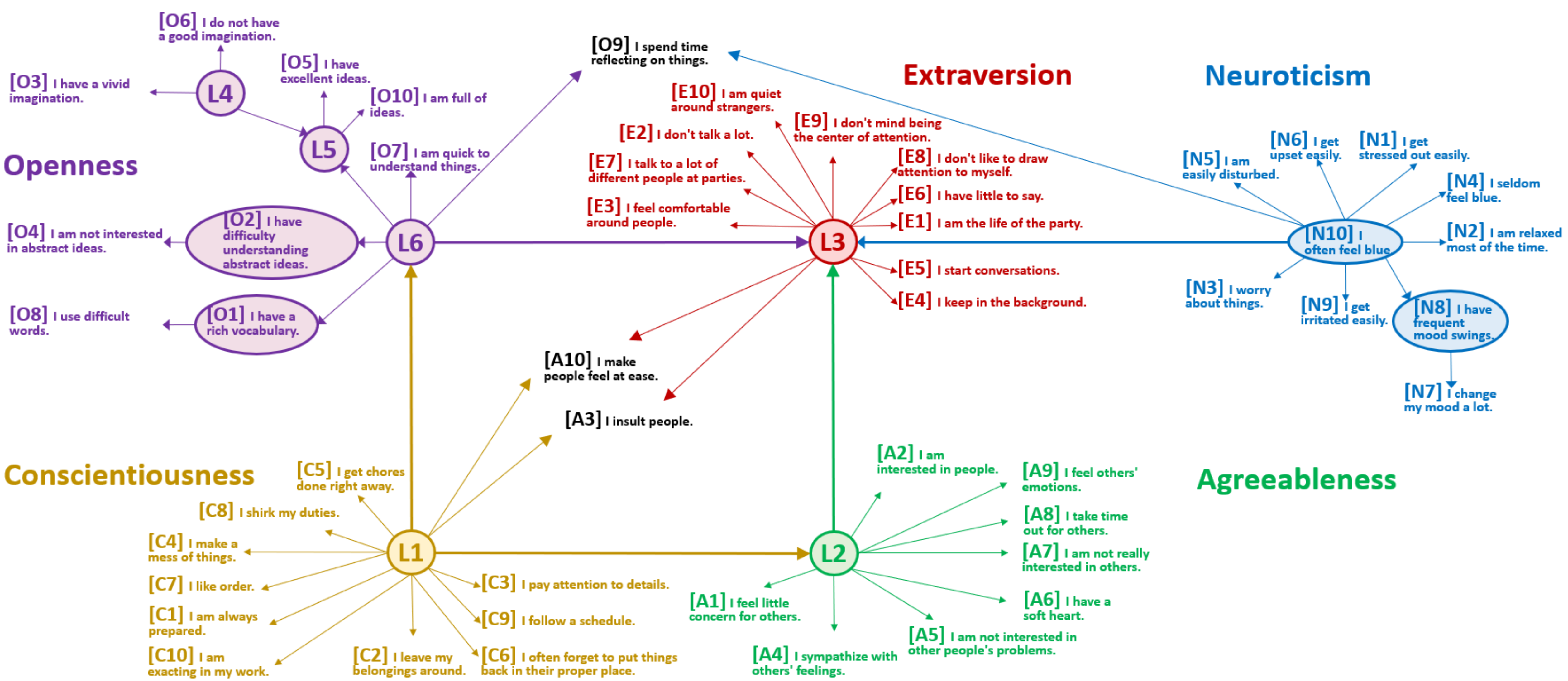
Extension to the general case: To allow observed variables to be causes as well

		F1 score for skeleton among $\mathbf{X}_\mathcal{G}$					
Algorithm		Ours	Hier. rank	PC	FCI	GIN	RCD
Observed only	2k	**0.95** (0.04)	-	0.78 (0.01)	0.07 (0.01)	-	0.79 (0.01)
	5k	**0.97** (0.01)	-	0.81 (0.01)	0.15 (0.07)	-	0.91 (0.01)
	10k	**0.98** (0.02)	-	0.82 (0.01)	0.25 (0.18)	-	0.93 (0.01)
Latent+tree	2k	**0.79** (0.16)	-	0.46 (0.02)	0.00 (0.00)	-	0.30 (0.03)
	5k	**0.86** (0.10)	-	0.44 (0.00)	0.03 (0.04)	-	0.38 (0.01)
	10k	**0.97** (0.04)	-	0.44 (0.00)	0.18 (0.07)	-	0.39 (0.02)
Latent+measm	2k	**0.84** (0.11)	-	0.50 (0.02)	0.00 (0.00)	-	0.30 (0.02)
	5k	**0.93** (0.08)	-	0.49 (0.01)	0.05 (0.03)	-	0.32 (0.02)
	10k	**0.95** (0.05)	-	0.48 (0.02)	0.03 (0.05)	-	0.42 (0.09)
Latent general	2k	**0.68** (0.02)	-	0.44 (0.01)	0.27 (0.09)	-	0.39 (0.06)
	5k	**0.71** (0.03)	-	0.45 (0.01)	0.31 (0.10)	-	0.44 (0.05)
	10k	**0.78** (0.06)	-	0.45 (0.01)	0.32 (0.05)	-	0.44 (0.01)

		F1 score for skeleton among $\mathbf{V}_\mathcal{G}$ (both $\mathbf{X}_\mathcal{G}$ and $\mathbf{L}_\mathcal{G}$)					
Algorithm		Ours	Hier. rank	PC	FCI	GIN	RCD
Latent+tree	2k	**0.84** (0.11)	0.58 (0.01)	0.36 (0.01)	0.00 (0.00)	0.37 (0.03)	0.24 (0.04)
	5k	**0.92** (0.05)	0.60 (0.01)	0.36 (0.00)	0.02 (0.02)	0.41 (0.03)	0.33 (0.00)
	10k	**0.98** (0.02)	0.60 (0.01)	0.36 (0.00)	0.15 (0.08)	0.41 (0.03)	0.33 (0.01)
Latent+measm	2k	**0.81** (0.12)	0.52 (0.05)	0.37 (0.01)	0.00 (0.00)	0.40 (0.02)	0.26 (0.03)
	5k	**0.88** (0.11)	0.52 (0.05)	0.49 (0.01)	0.04 (0.03)	0.46 (0.03)	0.29 (0.01)
	10k	**0.91** (0.09)	0.53 (0.05)	0.49 (0.01)	0.02 (0.03)	0.47 (0.05)	0.34 (0.04)
Latent general	2k	**0.66** (0.01)	0.44 (0.02)	0.31 (0.01)	0.17 (0.06)	0.30 (0.04)	0.32 (0.03)
	5k	**0.72** (0.03)	0.45 (0.03)	0.33 (0.01)	0.21 (0.07)	0.38 (0.04)	0.34 (0.02)
	10k	**0.80** (0.05)	0.45 (0.04)	0.33 (0.01)	0.21 (0.04)	0.35 (0.01)	0.36 (0.01)

Example: Big 5 Questions Are Well Designed but…

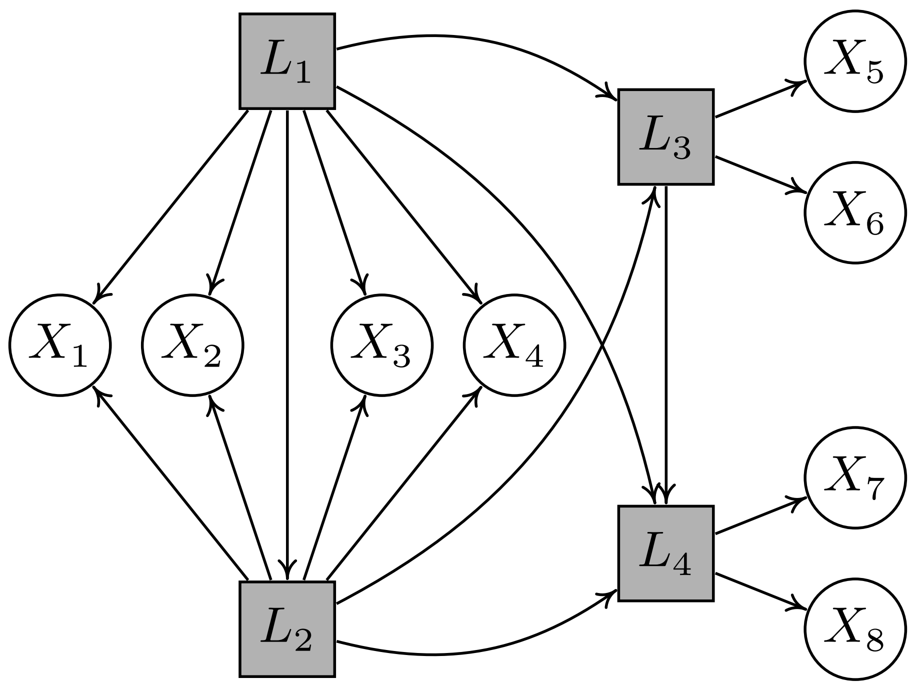
Big 5:
openness; conscientiousness; extraversion; agreeableness; neuroticism

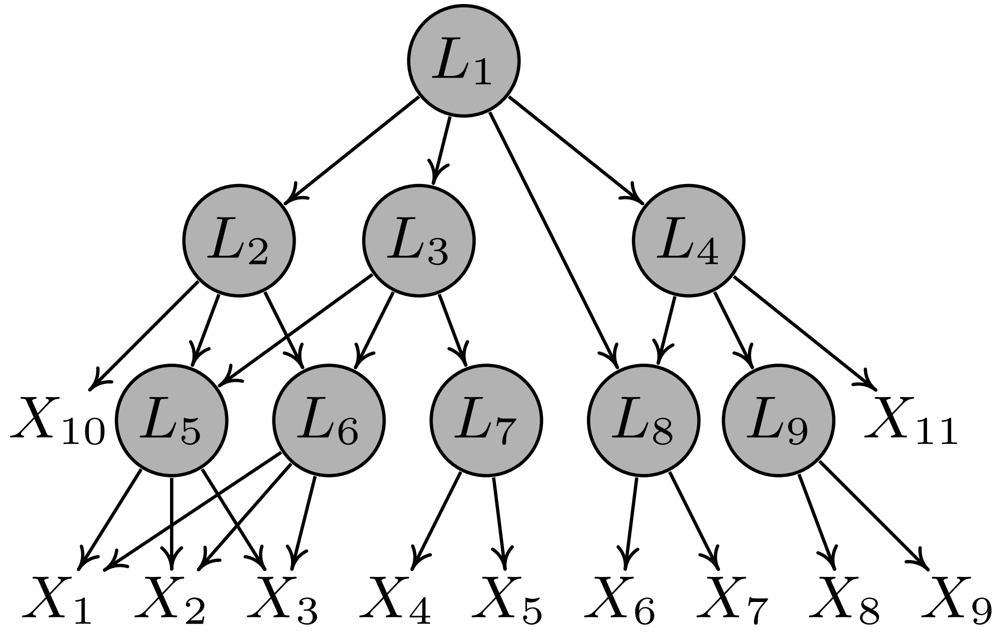


Nice results by Xinshuai Dong

Intermediate Summary: Identifying Latent Causal Graphs

- Further issues…
  - Weaker conditions on the structure?
  - Nonstationary cases?
  - More general or domain-specific cases?





X_i: measured variables
L_i: latent variables

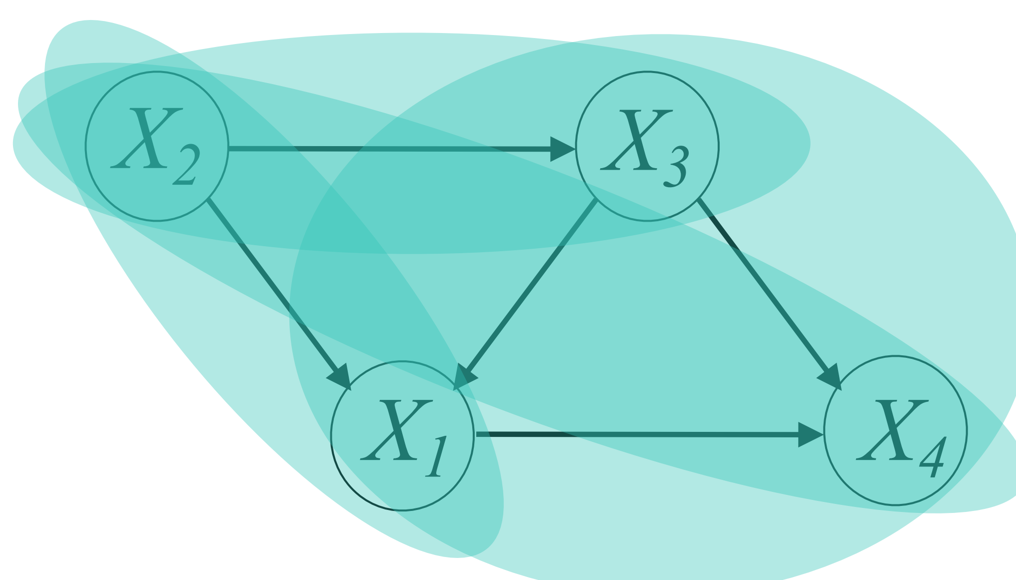
Outline

- Why?
- How?
  - IID case
    - Linear-Gaussian case
    - **Linear, non-Gaussian case**
    - Nonlinear case
  - From multiple distributions
  - With temporal information

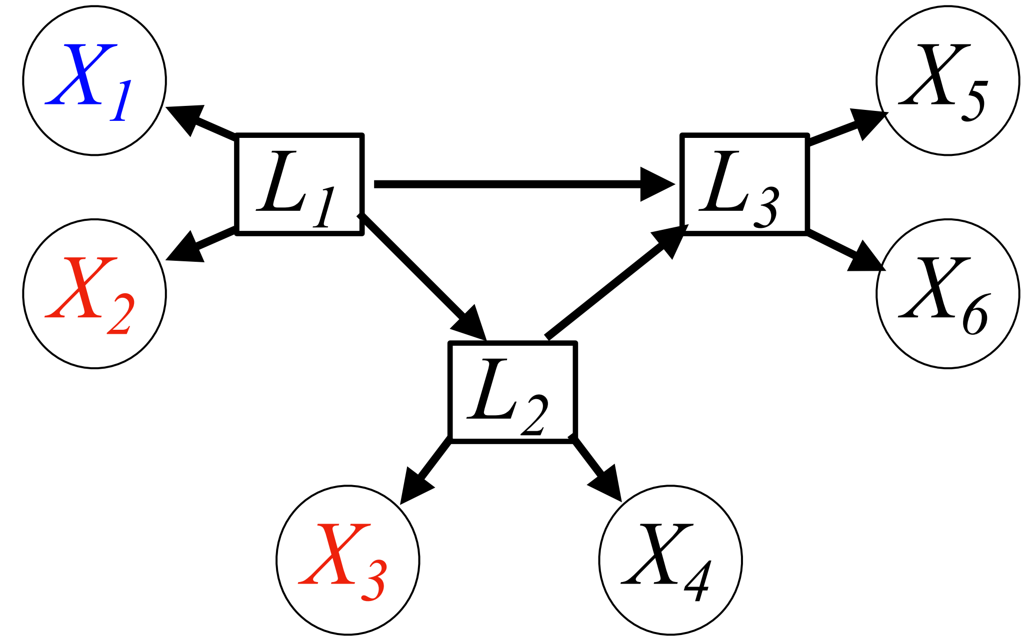
Recap: Independent Noise (IN) Condition

$$\mathbf{Z} \longrightarrow Y$$

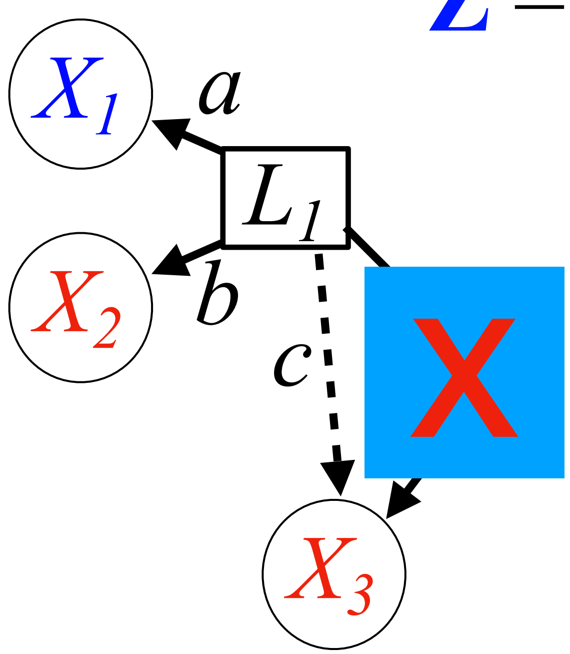
- $(\mathbf{Z}, Y)$ follows the <u>IN condition</u> iff regression residual $Y - \tilde{w}^{\intercal}\mathbf{Z}$ is independent from $\mathbf{Z}$
- Help determine causal orders and estimate the Linear, Non-Gaussian Acyclic Causal model (LiNGAM)

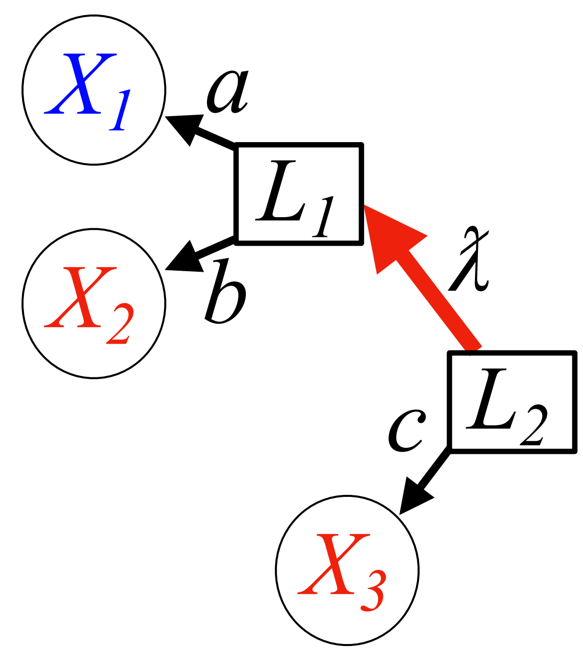


Generalized Independent Noise Condition (GIN)



$\mathbf{Z} = \{X_1\}$ $\mathbf{Y} = \{X_2, X_3\}$





$$
\begin{aligned}
& c \cdot X_2 - b \cdot X_3 \\
=& c(bL1 + E_2) - b(cL_1 + E_3) \\
=& cE_2 - bE_3,
\end{aligned}
$$

independent from L_1 and from X_1,

and we know $\frac{b}{c} = \frac{Cov(X_2, X_3)}{Cov(X_1, X_3)}$

Nontrivial linear combination of X_2 and X_3 will involve the noise term in L_1, hence dependent on X_1

Linear, Non-Gaussian Case: GIN

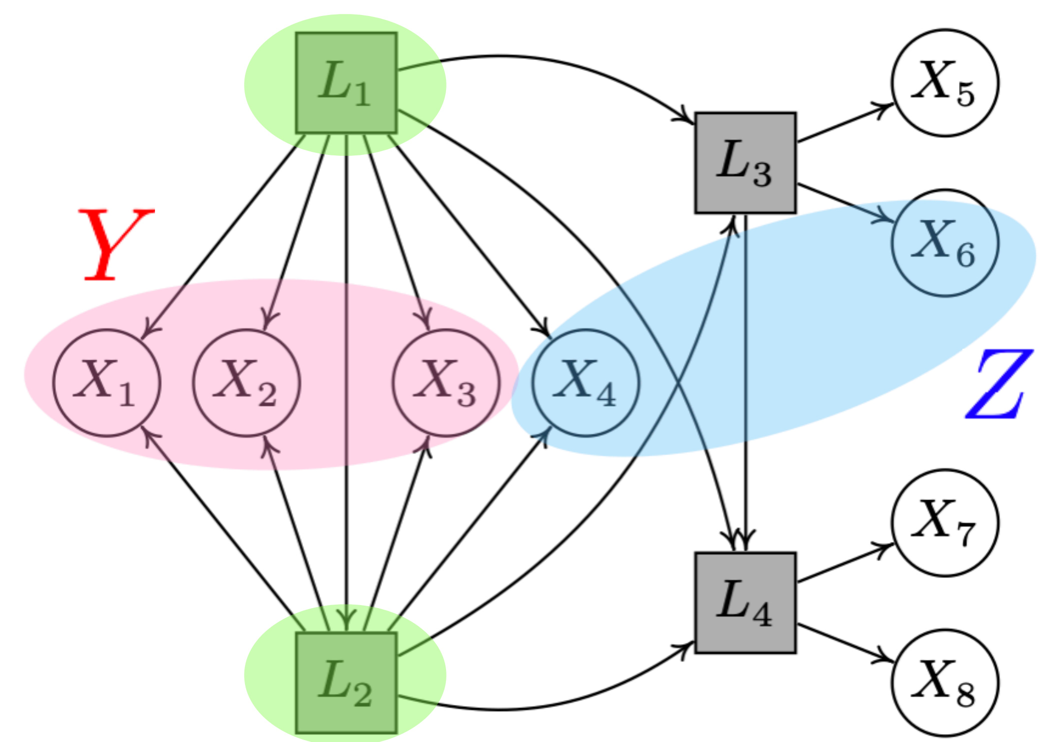
- Generalized Independent Noise (GIN) Condition:

  (Z, Y) follows the GIN condition $\Longleftrightarrow \omega^\top Y \perp\!\!\!\perp Z$,

  where $\omega^\top \mathrm{Cov}(Y, Z) = 0$ and $\omega \neq 0$

- Graphical criterion

  (Z, Y) follows the GIN condition iff there is an exogenous set S of **PA(Y) that blocks all paths between Y and Z**, where 0<=|S|<=min(|Z|, |Y|-1)

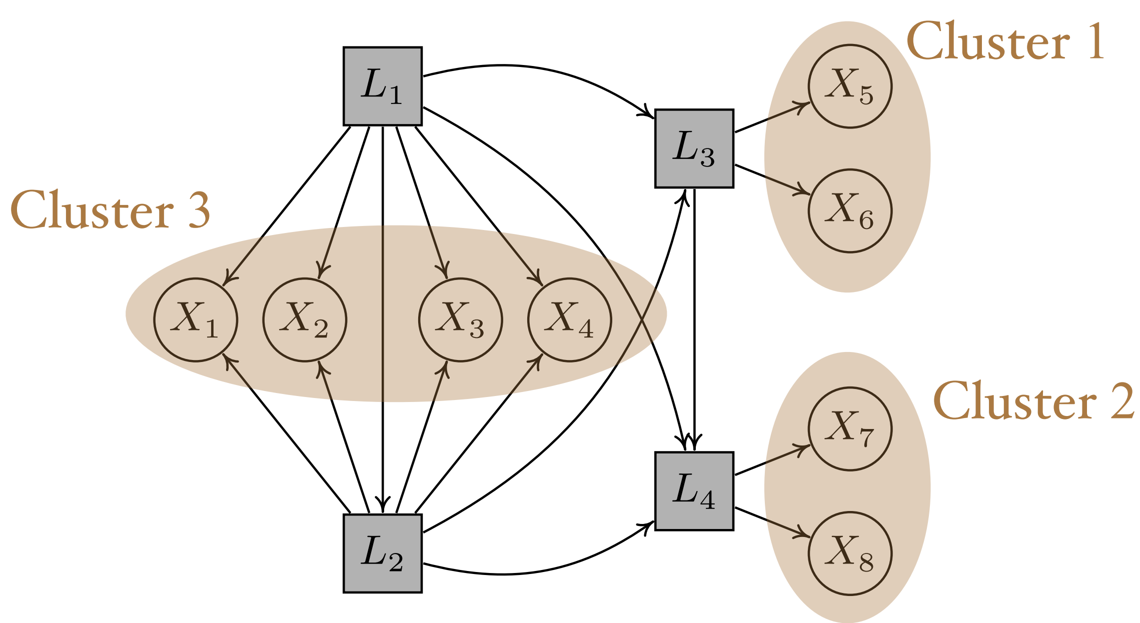


X_i: observed variables
L_i: latent variables

GIN Condition for Estimating Linear Non-Gaussian Latent Graphs

- A two-step algorithm to identify the latent variable graph
  - By testing for GIN conditions over the input $X_1, \cdots, X_8$

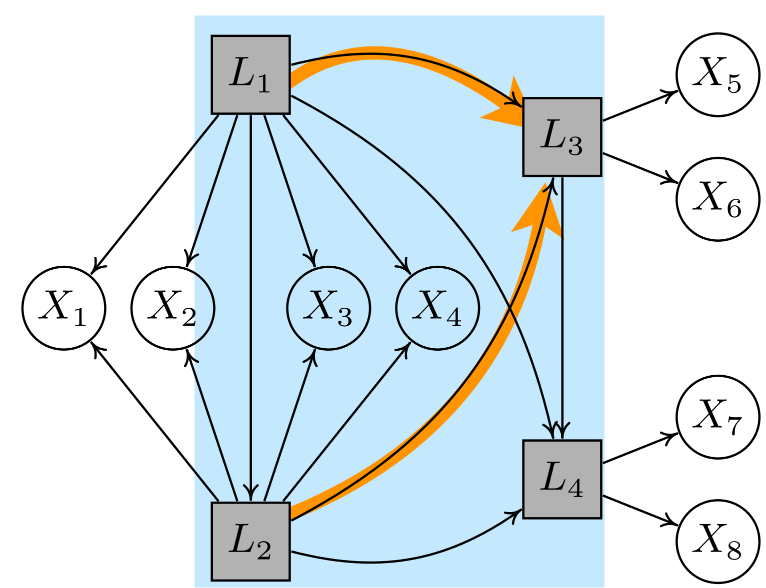
Step 1: find ***causal clusters***



$$(\overbrace{\{X_1,\cdots,X_4,X_7,X_8\}}^{\mathbf{Z}}, \overbrace{\{X_5,X_6\}}^{\mathbf{Y}})$$

satisfies GIN condition

Step 2: determine ***causal structure*** of the latent variables



$$(\overbrace{\{X_3,X_4\}}^{\mathbf{Z}}, \overbrace{\{X_1,X_2,X_5\}}^{\mathbf{Y}})$$

Cluster 3 Cluster 1 & 3

satisfies GIN condition

GIN-Based Method: Application to Teacher's Burnout Data

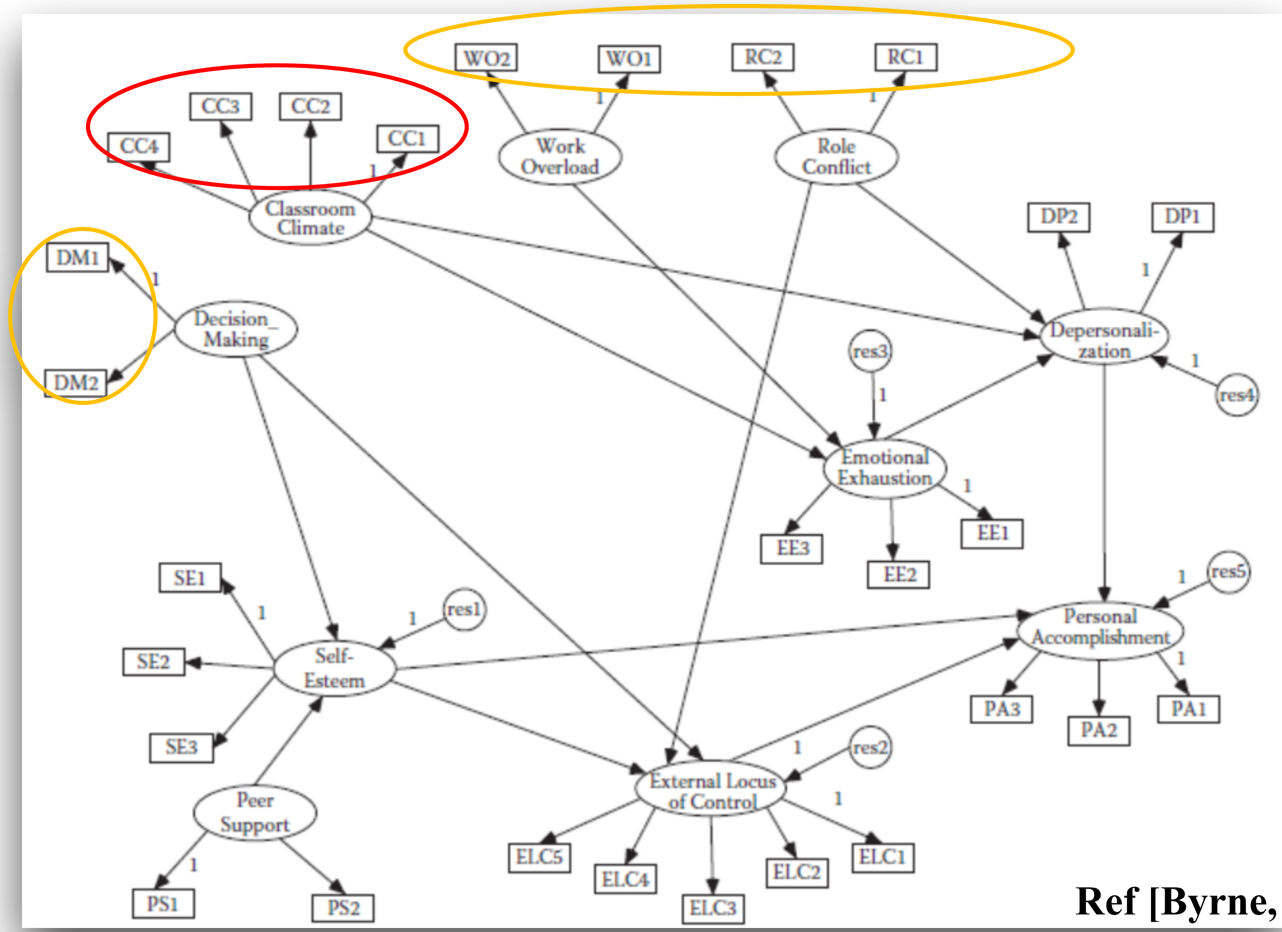
- Contains 28 measured variables
- Discovered clusters and causal order of the latent variables:

Causal Clusters	Observed variables
$\mathcal{S}_1$ (1)	$RC_1, RC_2, WO_1, WO_2, DM_1, DM_2$
$\mathcal{S}_2$ (1)	CC_1, CC_2, CC_3, CC_4
$\mathcal{S}_3$ (1)	PS_1, PS_2
$\mathcal{S}_4$ (1)	$ELC_1, ELC_2, ELC_3, ELC_4, ELC_5$
$\mathcal{S}_5$ (2)	$SE_1, SE_2, SE_3, EE_1, EE_2, EE_3, DP_1, PA_3$
$\mathcal{S}_6$ (3)	DP_2, PA_1, PA_2

$$L(\mathcal{S}_1) \succ L(\mathcal{S}_2) \succ L(\mathcal{S}_3) \succ L(\mathcal{S}_5) \succ L(\mathcal{S}_4) \succ L(\mathcal{S}_6).$$

(from root to leaf)

Hypothesized model by experts



Ref [Byrne, 2010]

- Consistent with the hypothesized model

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

Outline

- Why?
- How?
  - IID case
    - Linear-Gaussian case
    - Linear, non-Gaussian case
    - **Nonlinear case**
  - From multiple distributions
  - With temporal information

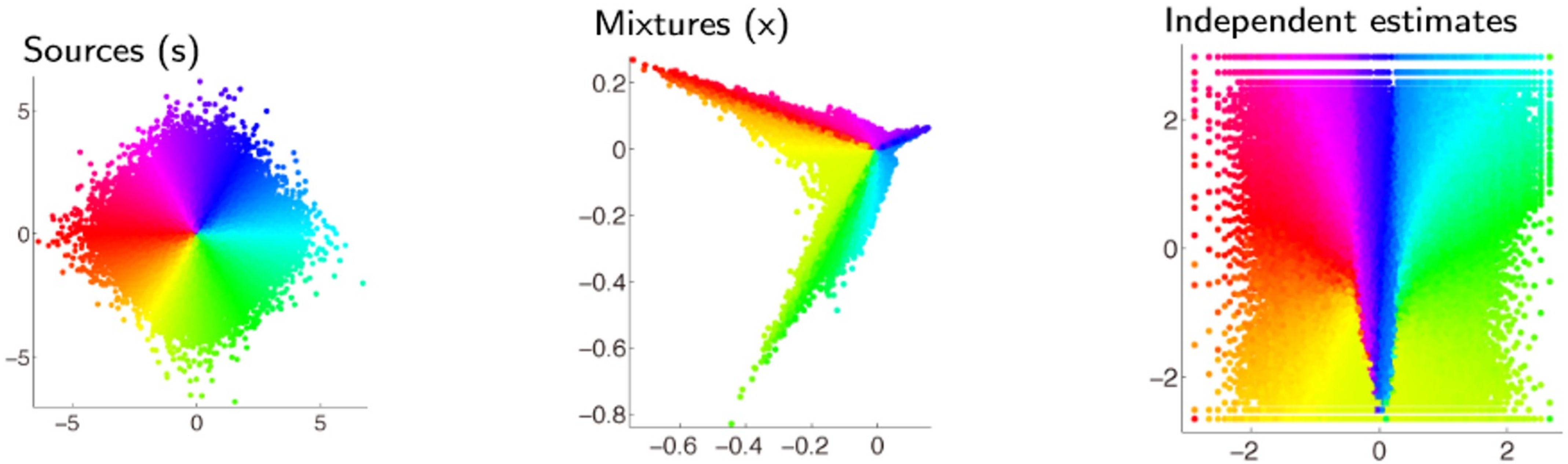
By Yujia Zheng

Identifiability of nonlinear ICA: challenge

Is nonlinear ICA identifiable?

$$x = f(s)$$

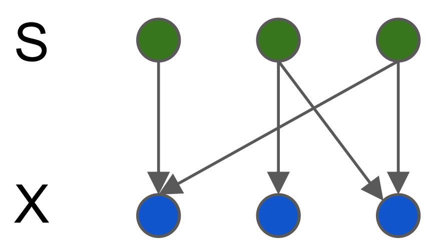
No, it's ill-posed without further assumptions



Identifiability of nonlinear ICA: auxiliary variables

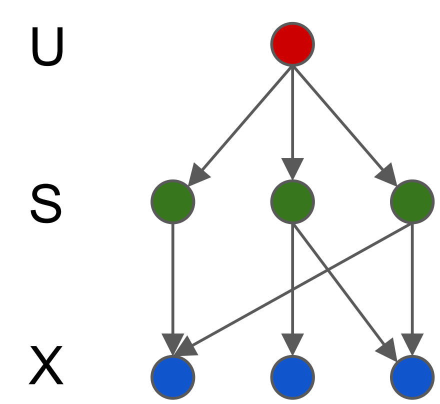
Independence alone is too weak → Conditional independence is strong enough

S1, S2, … , SN are marginally independent



$$x = f(s)$$

S1, S2, … , SN are conditionally independent given an auxiliary variable U (e.g., domain index)

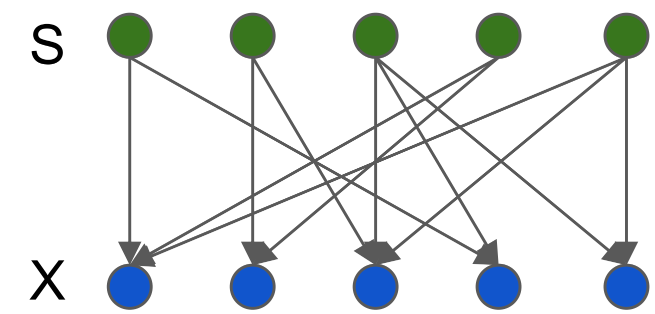


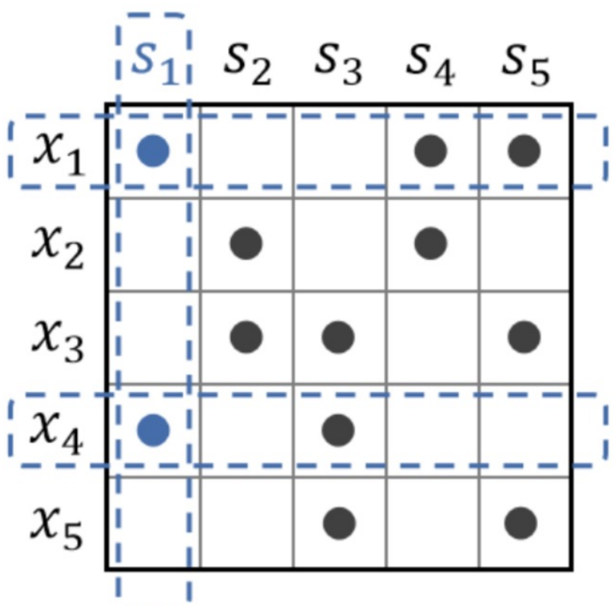
[Hyvarinen et al., Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning, AISTAT 2019]

Identifiability of nonlinear ICA: structural sparsity

Structural Sparsity) For all $k \in \{1, \ldots, n\}$, there exists $\mathcal{C}_k$ such that

$$\bigcap_{i \in \mathcal{C}_k} \text{supp}(\mathbf{J_f}(\mathbf{s})_{i,:}) = \{k\}.$$





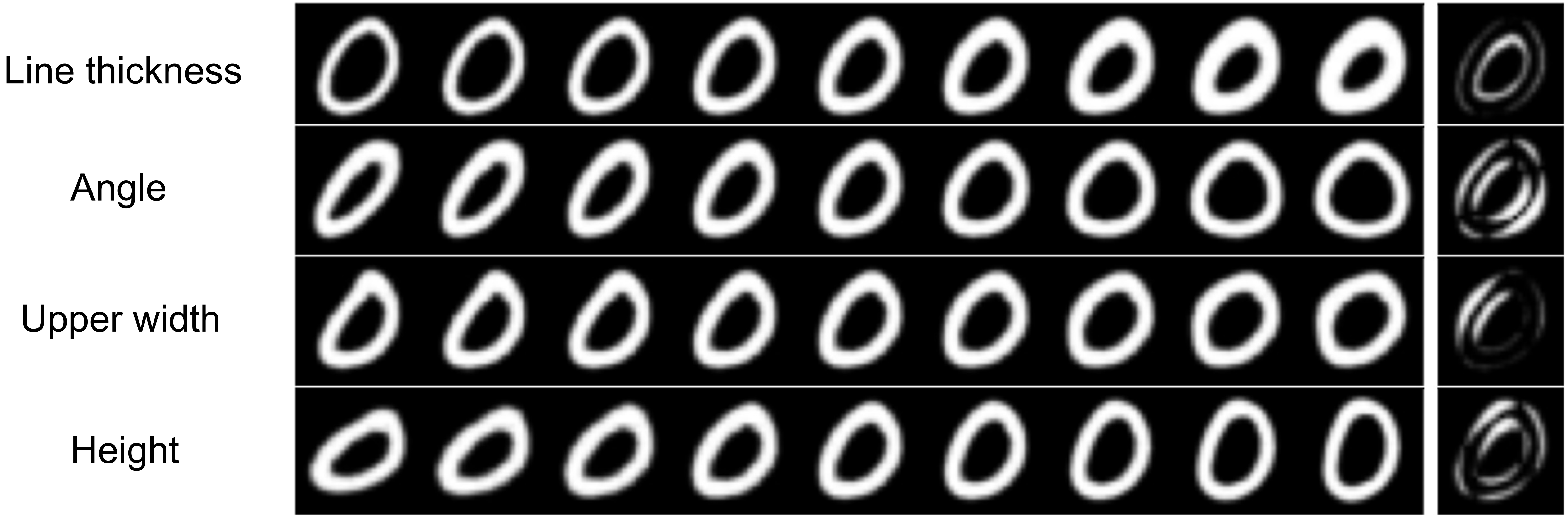
Graphically, for every latent source **s_i**, there exists a set of observed variable(s) such that the intersection of their/its parent(s) is **s_i**

Example: for **s_1**, there exists **x_1** and **x_4** such that the intersection of their parents is **s_1**

Failure: two sources influence the same set of observed variables

[Zheng et al., On the Identifiability of Nonlinear ICA: Sparsity and Beyond, NeurIPS 2022]

Identifiability of nonlinear ICA: real-world images



Identification results on EMNIST

Each row represents an identified source with its value varying

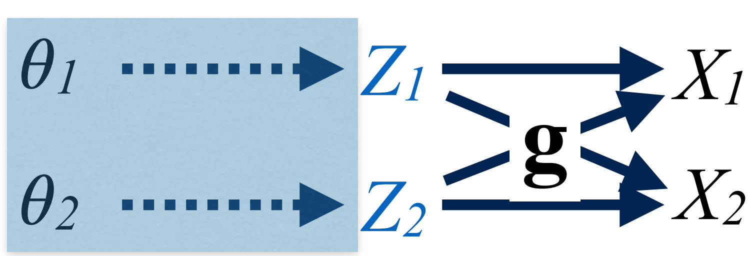
Outline

- Why?
- How?
  - IID case
    - Linear-Gaussian case
    - Linear, non-Gaussian case
    - Nonlinear case
  - **From multiple distributions**
  - With temporal information

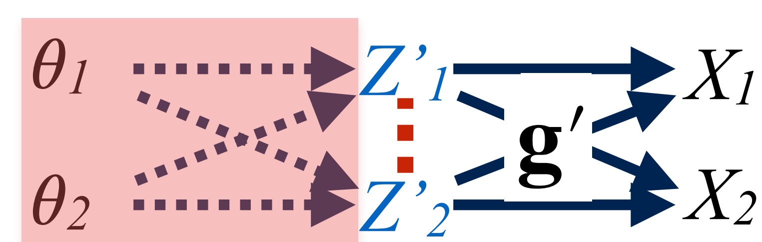
Nonlinear ICA with Multiple Domains

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Nonlinear ICA: observed variables follow $\mathbf{X} = \mathbf{g}(\mathbf{Z})$, in which Z_i are mutually independent
  - Solutions to nonlinear ICA high non-unique
  - If the dstr of each Z_i change across multiple domains, generally their are identifiable (up to component-wise transformations)
  - Why?



for $\mathbf{Z}' = h(\mathbf{Z})$:



- *Hyvärinen, Pajunen, Nonlinear independent component analysis: Existence and uniqueness results. Neural networks, 1999.*
- *Hyvarinen, Sasaki, Turner, "Nonlinear ICA using auxiliary variables and generalized contrastive learning," In The 22nd International Conference on Artificial Intelligence and Statistics, 2019.*

Partial Identifiability for Domain Adaptation

Lingjing Kong[1] **Shaoan Xie**[1] **Weiran Yao**[1] **Yujia Zheng**[1] **Guangyi Chen**[2,1] **Petar Stojanov**[3]
Victor Akinwande[1] **Kun Zhang**[2,1]

Abstract

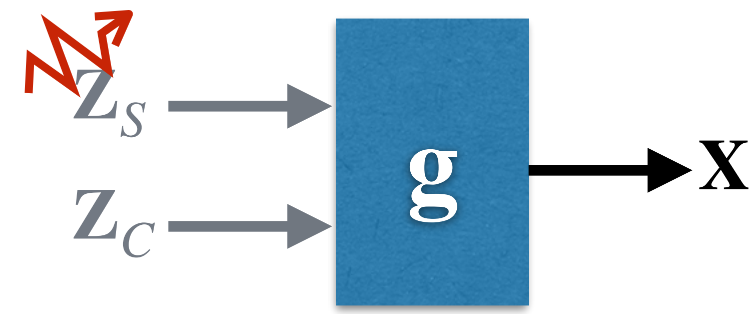
Unsupervised domain adaptation is critical to many real-world applications where label information is unavailable in the target domain. In general, without further assumptions, the joint distribution of the features and the label is not identifiable in the target domain. To address this issue, we rely on a property of minimal changes of causal mechanisms across domains to minimize unnecessary influences of domain shift. To encode this property, we first formulate the data generating process using a latent variable model with two partitioned latent subspaces: invariant components whose distributions stay the same across domains, and sparse changing components that vary across domains. We further constrain the domain shift to have a restrictive influence on the changing components. Under mild conditions, we show that the latent variables are partially identifiable, from

domain indices $\mathbf{u}$, the training (source domain) data follows multiple joint distributions $p_{\mathbf{x},\mathbf{y}|\mathbf{u}_1}$, $p_{\mathbf{x},\mathbf{y}|\mathbf{u}_2}$, ..., $p_{\mathbf{x},\mathbf{y}|\mathbf{u}_M}$,[1] and the test (target domain) data follows the joint distribution $p_{\mathbf{x},\mathbf{y}|\mathbf{u}^{\mathcal{T}}}$, where $p_{\mathbf{x},\mathbf{y}|\mathbf{u}}$ may vary across $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_M$. During training, for each i-th source domain, we are given labeled observations $(\mathbf{x}_k^{(i)}, \mathbf{y}_k^{(i)})_{k=1}^{m_i}$ from $p_{\mathbf{x},\mathbf{y}|\mathbf{u}_i}$, and target domain unlabeled instances $(\mathbf{x}_k^{\mathcal{T}})_{k=1}^{m_{\mathcal{T}}}$ from $p_{\mathbf{x},\mathbf{y}|\mathbf{u}^{\mathcal{T}}}$. The main goal of domain adaptation is to make use of the available observed information, to construct a predictor that will have optimal performance in the target domain.

It is apparent that without further assumptions, this objective is ill-posed. Namely, since the only available observations in the target domain are from the marginal distribution $p_{\mathbf{x}|\mathbf{u}^{\mathcal{T}}}$, the data may correspond to infinitely many joint distributions $p_{\mathbf{x},\mathbf{y}|\mathbf{u}^{\mathcal{T}}}$. This mandates making additional assumptions on the relationship between the source and the target domain distributions, with the hope to be able to reconstruct (identify) the joint distribution in the target domain $p_{\mathbf{x},\mathbf{y}|\mathbf{u}^{\mathcal{T}}}$. Typically, these assumptions entail some measure of sim-

Finding Changing Hidden Variables for Transfer Learning

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



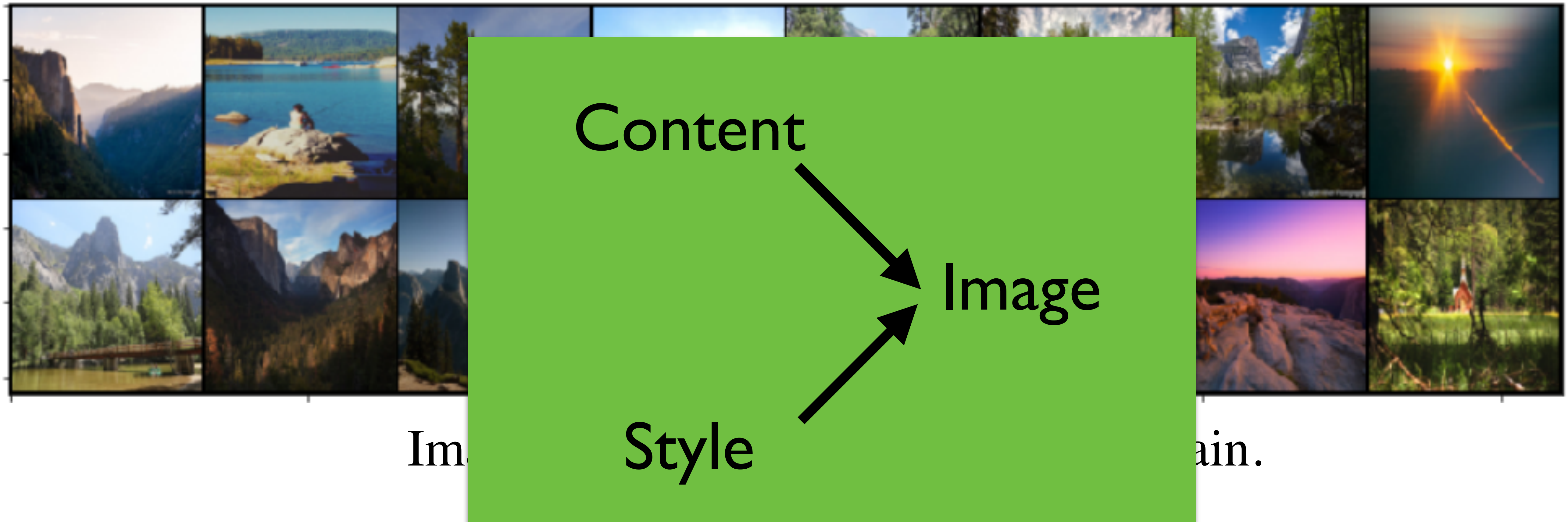
- Underlying components $\mathbf{Z}_S$ may change across domains
- Changing components $\mathbf{Z}_S$ are identifiable; invariant part $\mathbf{Z}_C$ are identifiable up to its subspace
- Using invariant part $\mathbf{Z}_C$ and transformed changing part $\tilde{\mathbf{Z}}_S$ for prediction

Models	→ Art	→ Clipart	→ Product	→ Realworld	Avg
Source Only (He et al., 2016)	64.58±0.68	52.32±0.63	77.63±0.23	80.70±0.81	68.81
DANN (Ganin et al., 2016)	64.26±0.59	58.01±1.55	76.44±0.47	78.80±0.49	69.38
DANN+BSP (Chen et al., 2019)	66.10±0.27	61.03±0.39	78.13±0.31	79.92±0.13	71.29
DAN (Long et al., 2015)	68.28±0.45	57.92±0.65	78.45±0.05	81.93±0.35	71.64
MCD (Saito et al., 2018)	67.84±0.38	59.91±0.55	79.21±0.61	80.93±0.18	71.97
M3SDA (Peng et al., 2019)	66.22±0.52	58.55±0.62	79.45±0.52	81.35±0.19	71.39
DCTN (Xu et al., 2018)	66.92±0.60	61.82±0.46	79.20±0.58	77.78±0.59	71.43
MIAN (Park & Lee, 2021)	69.39±0.50	63.05±0.61	79.62±0.16	80.44±0.24	73.12
MIAN-γ (Park & Lee, 2021)	69.88±0.35	**64.20±0.68**	80.87±0.37	81.49±0.24	74.11
iMSDA (Ours)	**75.77±0.21**	60.83±0.73	**84.13±0.09**	**84.83±0.12**	**76.39**

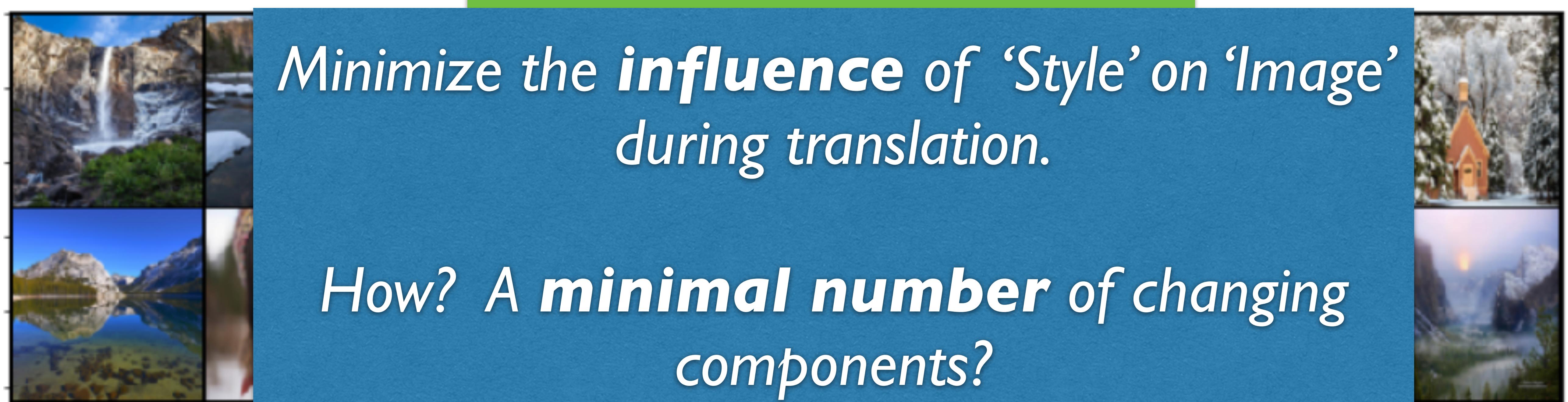
Table 2. Classification results on Office-Home. Backbone: Resnet-50. Baseline results are taken from (Park & Lee, 2021).

- *Kong, Xie, Yao, Zheng, Chen, Stojanov, Akinwande, Zhang, Partial disentanglement for domain adaptation, ICML 2022*

Unsupervised Image-to-Image Translation



Im ain.



Minimize the ***influence*** *of 'Style' on 'Image' during translation.*

How? A ***minimal number*** *of changing components?*

Images from the winter season domain.

Published as a conference paper at ICLR 2023

MULTI-DOMAIN IMAGE GENERATION AND TRANSLATION WITH IDENTIFIABILITY GUARANTEES

Shaoan Xie[1], **Lingjing Kong**[1], **Mingming Gong**[3,2], and **Kun Zhang**[1,2]

[1] Carnegie Mellon University
[2]Mohamed bin Zayed University of Artificial Intelligence
[3]The University of Melbourne
shaoan@cmu.edu, lingjingkong@cmu.edu,
mingming.gong@unimelb.edu.au, kunz1@cmu.edu

ABSTRACT

Multi-domain image generation and unpaired image-to-to-image translation are two important and related computer vision problems. The common technique for the two tasks is the learning of a joint distribution from multiple marginal distributions. However, it is well known that there can be infinitely many joint distributions that can derive the same marginals. Hence, it is necessary to formulate suitable constraints to address this highly ill-posed problem. Inspired by the recent advances in nonlinear Independent Component Analysis (ICA) theory, we propose a new method to learn the joint distribution from the marginals by enforcing a specific type of minimal change across domains. We report one of the first results connecting multi-domain generative models to identifiability and shows

Sample Images Generated by Generative Adversarial Networks (GANs)

Images generated by a GAN created by NVIDIA.

GANs

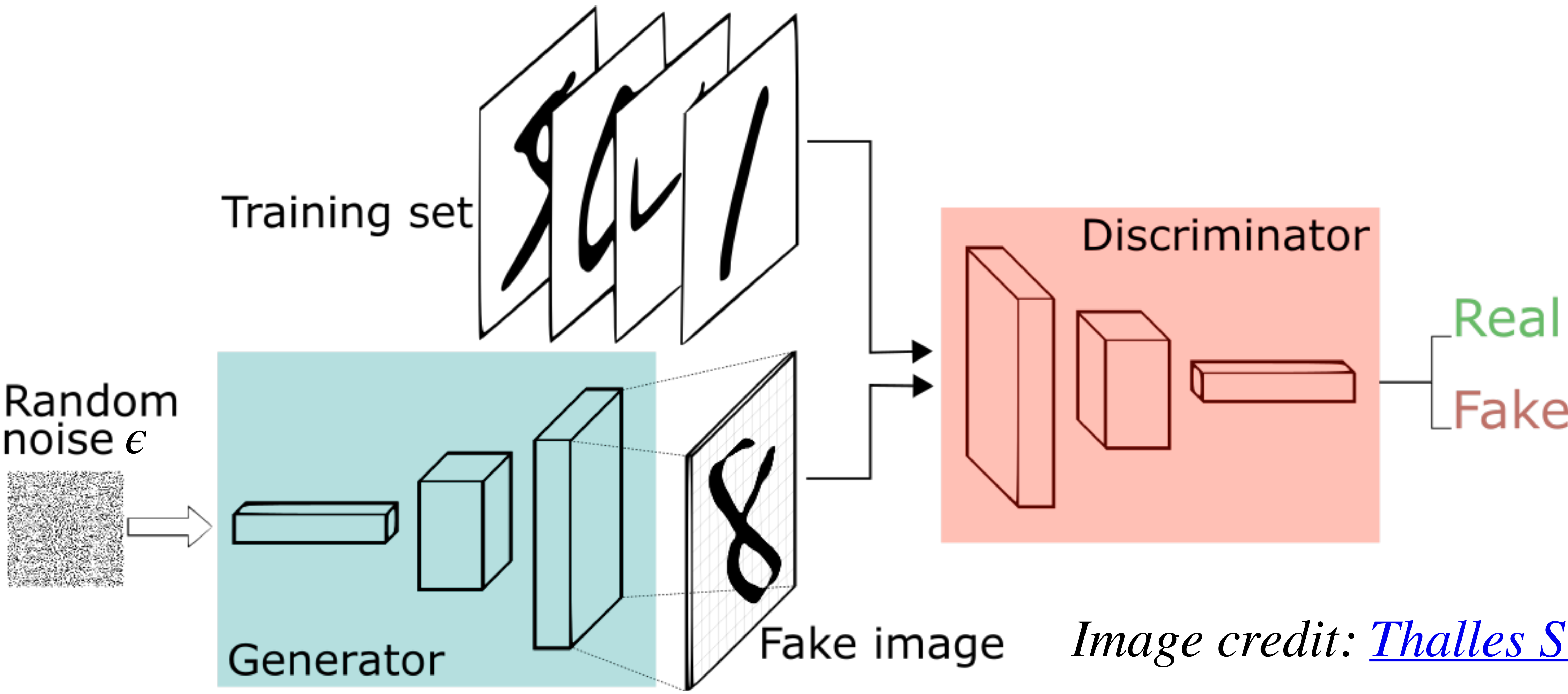
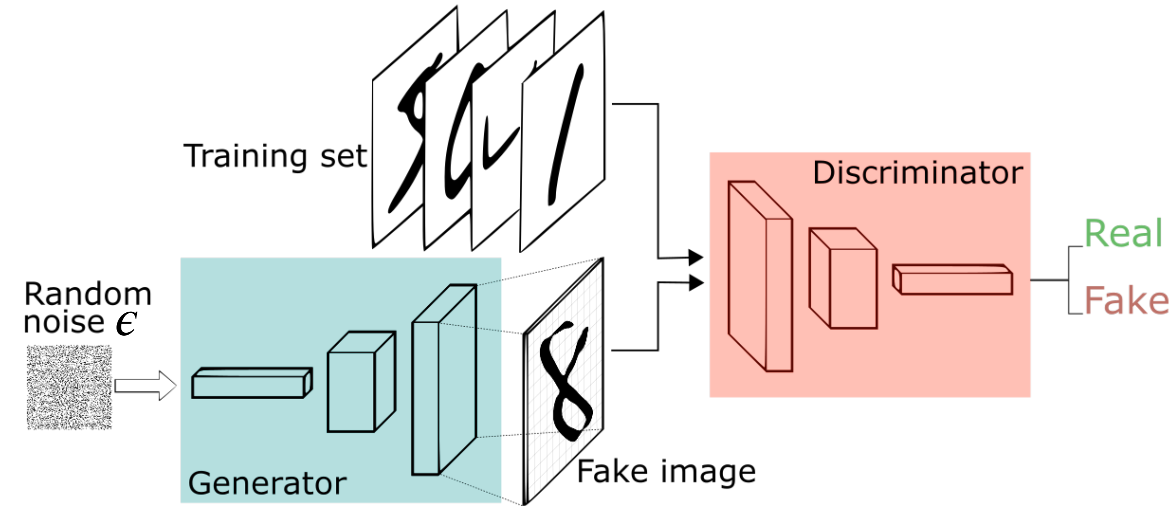


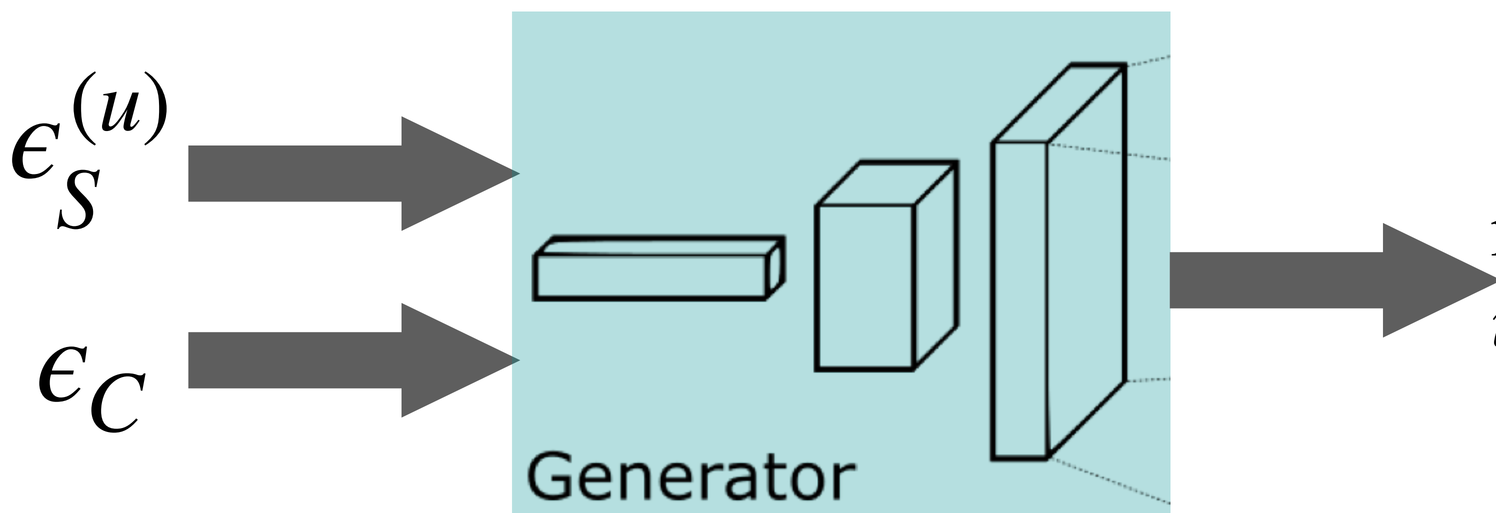
Image credit: [Thalles Silva]

Minimax game which G wants to minimize V while D wants to maximize it:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

GAN-Based Implementations





model the data distribution in u-th domain

- Match the data distribution across domains, while the dimensionality of $\epsilon_S^{(u)}$ is as small as possible (minimal changes across domains **controlled by ƛ;** no penalty when ƛ=0)

- Correspondence relations among domains are identifiable

Multi-domain Image Generation & Translation with Identifiability Guarantees

- Idea: Matching the distributions across domains with a minimal number of changing components
- Correspondence info (joint distribution) identifiable under mild assumptions
- Example: Generating female & males images with the same "content"

Ours (λ=0.1)

StyleGAN2-ADA

TGAN

- *Xie, Kong, Gong, Zhang, "Multi-domain image generation and translation with identifiability guarantees", ICLR 2023*

More results…

Figure 10: CelebA-HQ. Without the sparsity regularization, i.e., $\lambda = 0$, we observe some unnecessary changes between the image tuples in each row. For example, e.g.,the added sun-glasses and skin color change in the first row. TGAN changes the background (first row of third panel). CoGAN changes the skin color (second row, second panel).

More results…

Figure 11: AFHQ. StyleGAN2-ADA changes animal poses in many examples, e.g., second and third row of first panel. Our base ($\lambda = 0$) also changes the poses, e.g., first and third row of second panel. CoGAN and TGAN are slightly better in preserving poses but we can observe that some generated images are unrealistic. For example, the wolf (first row, third panel of TGAN) and the dog (third row, third panel of CoGAN).

Outline

- Why?
- How?
  - IID case
    - Linear-Gaussian case
    - Linear, non-Gaussian case
    - Nonlinear case
  - From multiple distributions
  - **With temporal information**

Temporally Disentangled Representation Learning

Weiran Yao
CMU
weiran@cmu.edu

Guangyi Chen
CMU & MBZUAI
guangyichen1994@gmail.com

Kun Zhang
CMU & MBZUAI
kunz1@cmu.edu

Abstract

Recently in the field of unsupervised representation learning, strong identifiability results for disentanglement of causally-related latent variables have been established by exploiting certain side information, such as class labels, in addition to independence. However, most existing work is constrained by functional form assumptions such as independent sources or further with linear transitions, and distribution assumptions such as stationary, exponential family distribution. It is unknown whether the underlying latent variables and their causal relations are identifiable if they have arbitrary, nonparametric causal influences in between. In this work, we establish the identifiability theories of nonparametric latent causal processes from their nonlinear mixtures under fixed temporal causal influences and analyze how distribution changes can further benefit the disentanglement. We propose **TDRL**, a principled framework to recover time-delayed latent causal vari-

Learning Latent Causal Dynamics

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

Learn the underlying causal dynamics from their mixtures?
"Time-delayed" influence renders latent processes & their relations identifiable

Time-series Inputs $\{x_t\}_{t=0}^{T}$

Unsupervised Representation Learning

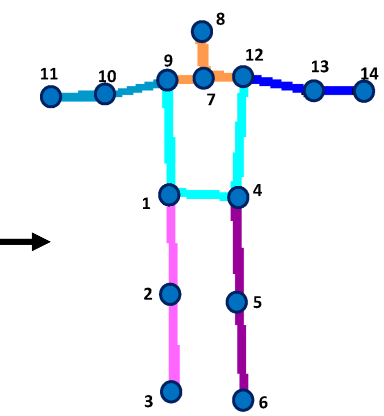
$$\mathbf{x}_t = \mathbf{g}(\mathbf{z}_t)$$

Latent processes

Latent temporal causal processes z_{it} can be recovered if they follow

- completely **nonparametric** model; or furthermore,
- **non-stationary** noise; or
- **non-stationary** causal influence, or
- **Parametric** constraints

Causal Skeleton Recovery



Recovered latent processes

$$\underbrace{\mathbf{x}_t = \mathbf{g}(\mathbf{z}_t)}_{\text{Nonlinear mixing}}, \quad \underbrace{z_{it} = f_i\left(\{z_{j,t-\tau} | z_{j,t-\tau} \in \mathbf{Pa}(z_{it})\}, \epsilon_{it}\right)}_{\text{Stationary nonparametric transition}} \; with \; \underbrace{\epsilon_{it} \sim p_{\epsilon_i}}_{\text{Stationary noise}}.$$

- *Yao, Chen, Zhang, "Causal Disentanglement for Time Series," NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, "Learning Temporally causal latent processes from general temporal data," ICLR 2022*

Comparisons

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

Learn the underlying causal dynamics from their mixtures?
"Time-delayed" influence renders latent processes & their relations identifiable

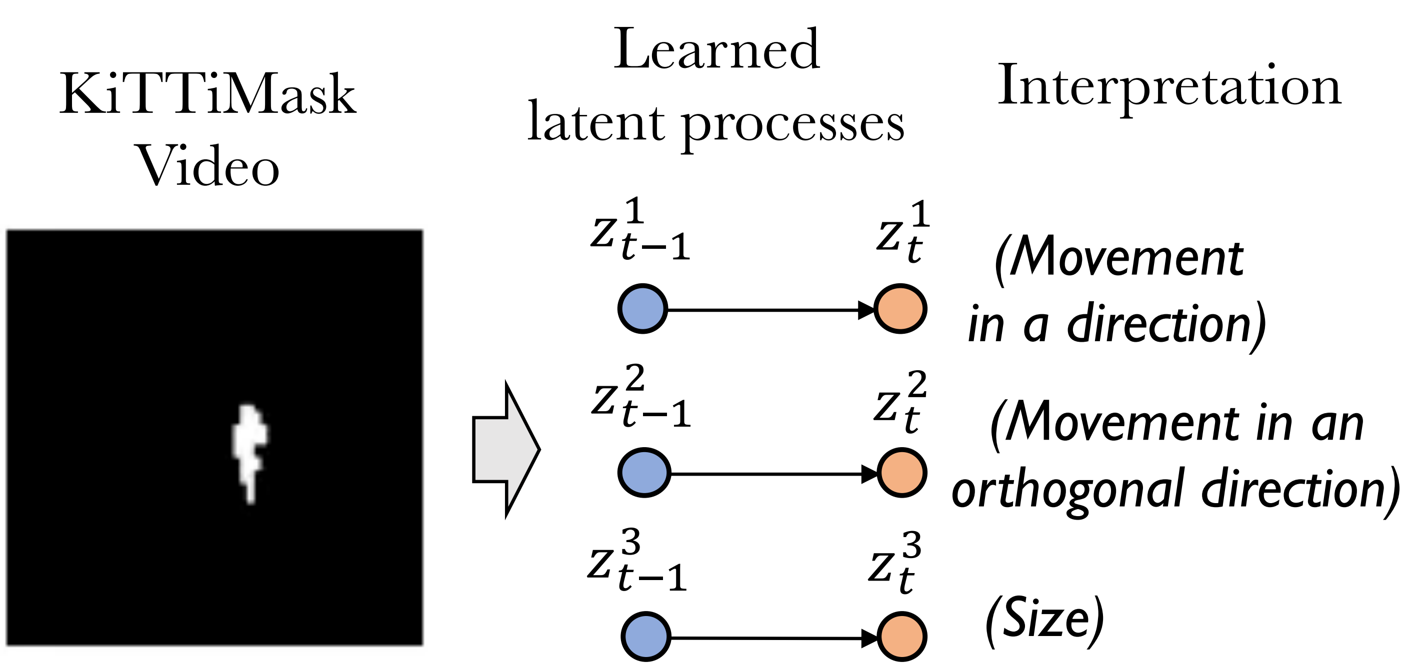
Table 1: Attributes of nonlinear ICA theories for time-series. A check denotes that a method has an attribute or can be applied to a setting, whereas a cross denotes the opposite. † indicates our approach.

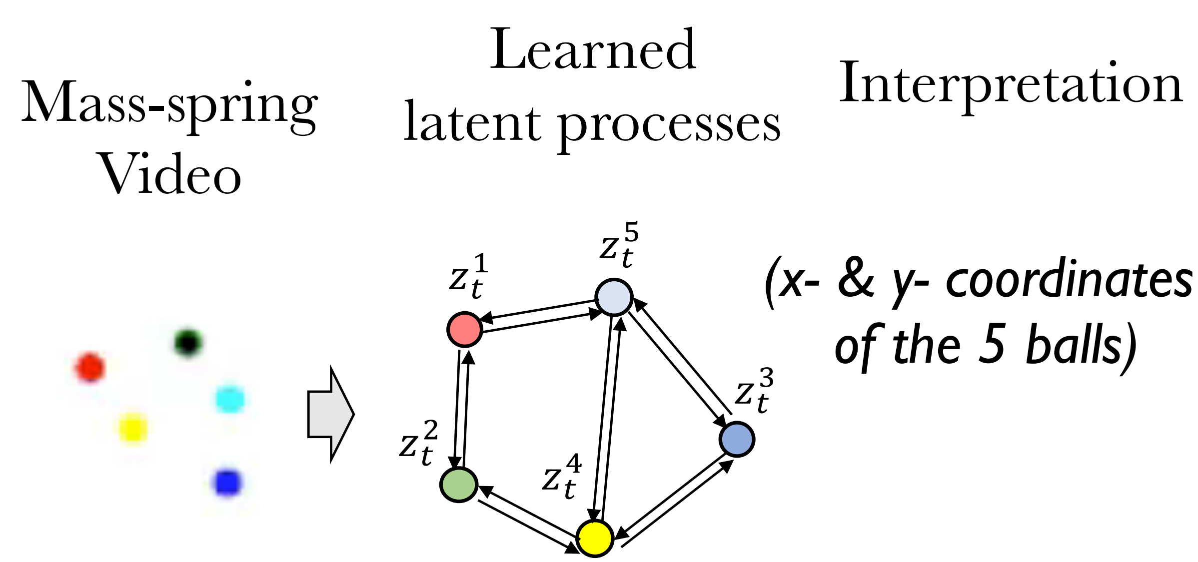
Theory	Time-varying Relation	Causally-related Process	Partitioned Subspace	Nonparametric Transition	Applicable to Stationary Environment
PCL	✗	✗	✗	✓	✓
GCL	✓	✗	✗	✓	✓
HM-NLICA	✗	✗	✗	✗	✗
SlowVAE	✗	✗	✗	✗	✓
SNICA	✓	✓	✗	✗	✗
i-VAE	✓	✗	✗	✗	✗
LEAP	✗	✓	✗	✓	✗
TDRL †	✓	✓	✓	✓	✓

- *Yao, Chen, Zhang, "Causal Disentanglement for Time Series," NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, "Learning Temporally causal latent processes from general temporal data," ICLR 2022*

Results on Video Data

- For easy interpretation, consider two simple video data sets
  - KiTTiMask: a video dataset of binary pedestrian masks
  - Mass-spring system: a video dataset with ball movement and invisible springs





- *Yao, Chen, Zhang, "Learning Latent Causal Dynamics," NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, "Learning Temporally causal latent processes from general temporal data," ICLR 2022*

Causal Representation Learning: A Summary

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	No	No	(Different types of) equivalence class
		Yes	
	Yes	No	Unique identifiability (under structural conditions)
		Yes	
Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!
I., but non-I.D.	No	No	More informative than MEC (CD-NOD)
	Yes		May have unique identifiability
	No	Yes	Changing subspace identifiable
	Yes		Variables in changing relations identifiable

Summary

- Essential to learn hidden causal variables in many cases!
- Possible to achieve even in the IID case
- Benefit from distribution changes and temporal information
- Future work
    - Efficient procedure?
    - Necessary and sufficient identifiability conditions?
    - Changing relations among hidden variables?