LEARNING DISCRETE LATENT VARIABLE STRUCTURES WITH TENSOR RANK CONDITIONS

NeurIPS 2024

Zhengming Chen[1,2] , Ruichu Cai[1,*] , Feng Xie[3] , Jie Qiao[1] ,

Anpeng Wu[4,2], Zijian Li[2], Zhifeng Hao[1,5] , Kun Zhang[2,6,*]

[1]School of Computer Science, Guangdong University of Technology, Guangzhou, China

[2]Machine Learning Department, Mohamed bin Zayed University of Artificial Intelligence
[3]Department of Applied Statistics, Beijing Technology and Business University, Beijing, China

[4]Department of Computer Science and Technology, Zhejiang University
[5]College of Science, Shantou University, Shantou, Guangdong, China

[6]Department of Philosophy, Carnegie Mellon University

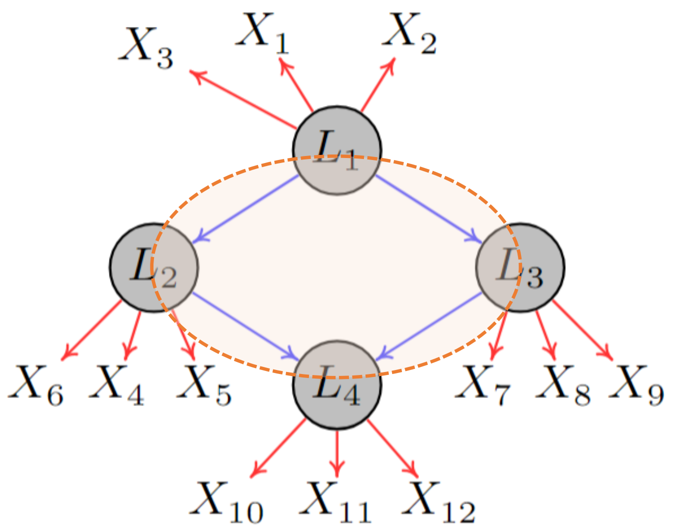
Contents

Problem Definition

	X_1	X_2	$\dots$	X_{19}
1	0	3	$\dots$	2
2	2	5	$\vdots$	3
3	1	4	$\dots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	4	1	$\dots$	4

Observational dataset X
(discrete variable)

Discovery

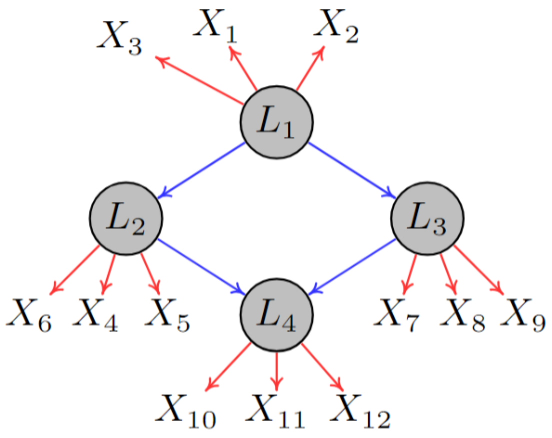


Latent structure

Is it possible to find latent variable L_i and their causal relations only from discrete measured variables X_i?

Discrete Latent Structure Model with Three-Pure Children (3PLSM)

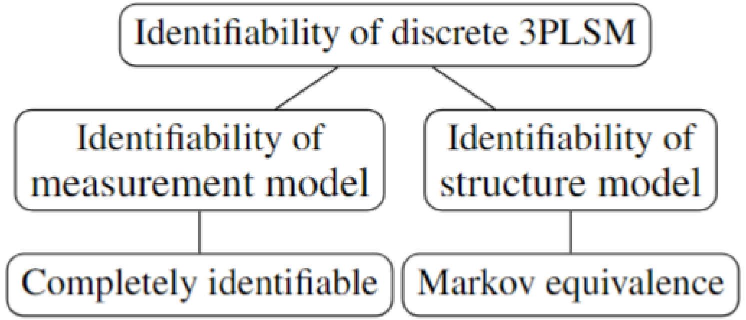
- **Purity Assumption:** there is no direct edges between the observed variables
- **Three-Pure Child Variable Assumption:** each latent variable has at least three pure variables as children
- **Sufficient Observation Assumption:** The cardinality of observed variables support is larger than the cardinality of any latent variables support.



Discrete Latent Structure Model:
- **Measurement Model:** red edges
- **Structure Model:** blue edges

How to identify the causal structure among latent variables, in a statistically efficient and robust manner?

Identifiability Condition for Discrete 3PLSM

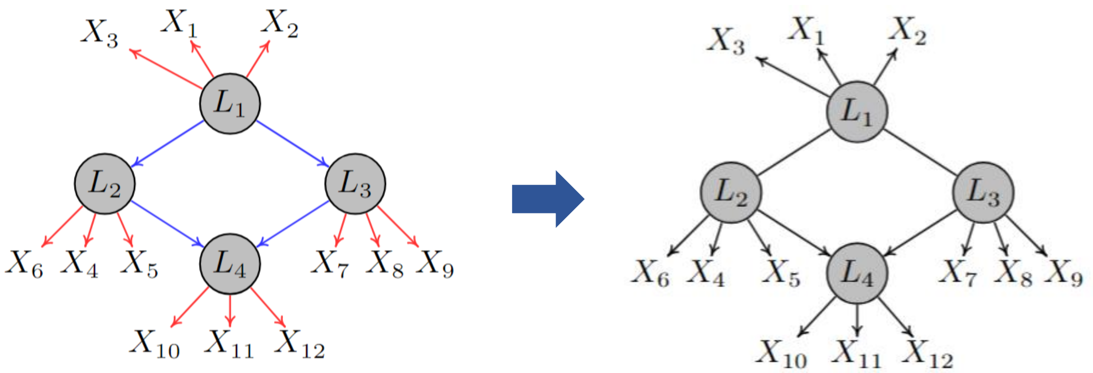


Causal Assumptions:

(1) Markov assumption、Faithfulness assumption.

Full Rank Assumption:

(2) For any conditional probability $\mathbb{P}(X|Pa_X)$, the corresponding contingency table is full rank



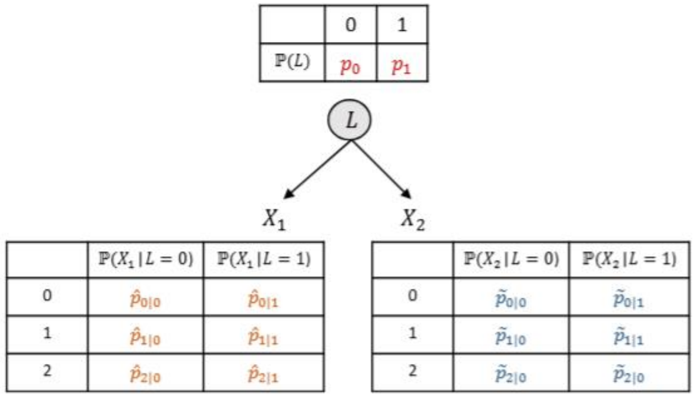
Identifiability results of discrete latent structure model, i.e., the measurement model is full identifiable, and the structure model is identified up to a Markov equivalent class

Tensor Rank Condition for Discrete Causal Models

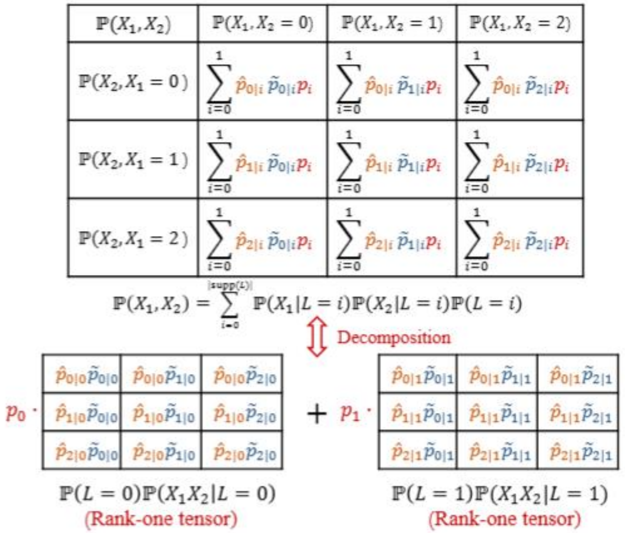
Graphical Criteria

Theorem 3.3 (Graphical implication of tensor rank condition). *In the discrete causal model, suppose Assumptions 2.2 ~ Assumption 2.4 hold. Consider an observed variable set* $\mathbf{X}_p = \{X_1, \cdots, X_n\}$ *(*$\mathbf{X}_p \subseteq \mathbf{X}$ *and* $n \geq 2$*) and the corresponding* n*-way probability tensor* $\mathcal{T}_{(\mathbf{X}_p)}$ *that is the tabular representation of the joint probability mass function* $\mathbb{P}(X_1, \cdots, X_n)$*. Then,* $\text{Rank}(\mathcal{T}_{(\mathbf{X}_p)}) = r$ $(r > 1)$ *if and only if (i) there exist a conditional set* $\mathbf{S} \subset \mathbf{V}$ *with* $|\text{supp}(\mathbf{S})| = r$ *that d-separates any pair of variables in* $\{X_1, \cdots, X_n\}$*, and (ii) does not exist conditional set* $\tilde{\mathbf{S}}$ *that satisfies* $|\text{supp}(\tilde{\mathbf{S}})| < r$*.*

Example



(a) A latent structure with conditional probability tables.



(b) The decomposition of the joint distribution.

Algorithm for Estimating Discrete 3PLSM

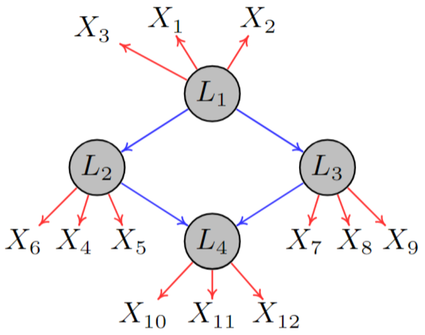
- **Step I:** Identify Causal Cluster
  - Find **causal clusters** from the observed variable set by Tensor Rank Condition

Proposition 4.3 (Identification of causal cluster). *In the discrete 3PLSM mode, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let $r = |\mathrm{supp}(L_i)|$ denote the cardinality of the latent support. Given three disjoint observed variables $X_i, X_j, X_k \in \mathbf{X}$,*

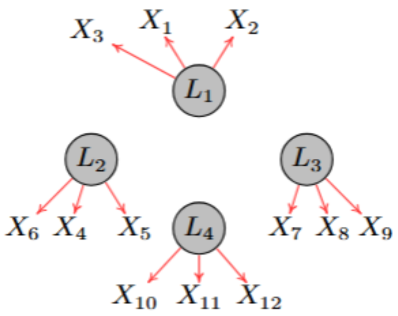
- *$\mathcal{R}ule1$: if the rank of tensor $\mathcal{T}_{(X_i,X_j,X_k)}$ is not equal to r, i.e., $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k)}) \neq r$, then X_i, X_j and X_k belong to the different latent parents.*
- *$\mathcal{R}ule2$: for any X_s, $X_s \in \mathbf{X} \setminus \{X_i, X_j, X_k\}$, if the rank of tensor $\mathcal{T}_{(X_i,X_j,X_k,X_s)}$ is r, i.e., $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k,X_s)}) = r$, then $\{X_i, X_j, X_k\}$ share the same latent parent.*

- **Step II:** Identify Causal Structure among Latent Variables
  - Identify the **d-separation relations** among latent variables by Tensor Rank Condition

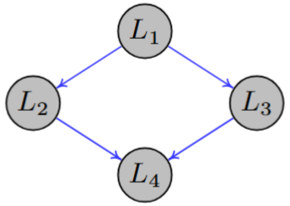
Theorem 4.7 (d-separation among latent variable). *In the discrete 3PLSM, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let r denote the cardinality of the latent support. Then, $L_i \perp\!\!\!\perp L_j | \mathbf{L}_p$ if and only if $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,\mathbf{X}_{p1},\mathbf{X}_{p2})}) = r^{|\mathbf{L}_p|}$, where X_i and X_j are the pure children of L_i and L_j, $\mathbf{X}_{p1}$ and $\mathbf{X}_{p2}$ are two disjoint child sets of $\mathbf{L}_p$ that satisfy $\forall L_i \in \mathbf{L}_p, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p1} \neq \emptyset, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p2} \neq \emptyset$.*



Ground truth



Measurement model



Structure model

Algorithm for Estimating Discrete 3PLSM

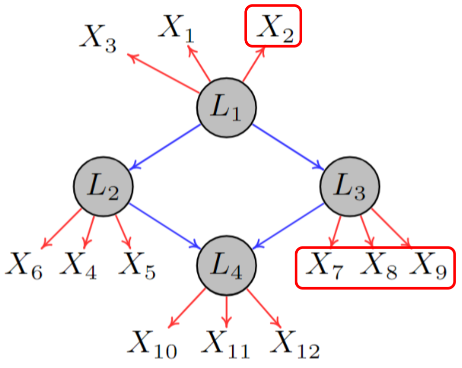
- **Step I:** Identify Causal Cluster
  - Find **causal clusters** from the observed variable set by Tensor Rank Condition

Proposition 4.3 (Identification of causal cluster). *In the discrete 3PLSM mode, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let $r = |\mathrm{supp}(L_i)|$ denote the cardinality of the latent support. Given three disjoint observed variables $X_i, X_j, X_k \in \mathbf{X}$,*

- *Rule1: if the rank of tensor $\mathcal{T}_{(X_i,X_j,X_k)}$ is not equal to r, i.e., $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k)}) \neq r$, then X_i, X_j and X_k belong to the different latent parents.*
- *Rule2: for any X_s, $X_s \in \mathbf{X} \setminus \{X_i, X_j, X_k\}$, if the rank of tensor $\mathcal{T}_{(X_i,X_j,X_k,X_s)}$ is r, i.e., $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k,X_s)}) = r$, then $\{X_i, X_j, X_k\}$ share the same latent parent.*

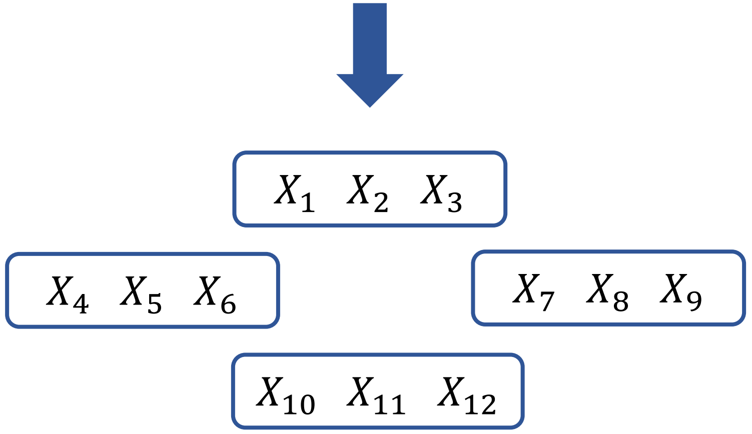
- **Step II:** Identify Causal Structure among Latent Variables
  - Identify the d-separation relations among latent variables by Tensor Rank Condition

Theorem 4.7 (d-separation among latent variable). *In the discrete 3PLSM, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let r denote the cardinality of the latent support. Then, $L_i \perp\!\!\!\perp L_j | \mathbf{L}_p$ if and only if $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,\mathbf{X}_{p1},\mathbf{X}_{p2})}) = r^{|\mathbf{L}_p|}$, where X_i and X_j are the pure children of L_i and L_j, $\mathbf{X}_{p1}$ and $\mathbf{X}_{p2}$ are two disjoint child sets of $\mathbf{L}_p$ that satisfy $\forall L_i \in \mathbf{L}_p, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p1} \neq \emptyset, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p2} \neq \emptyset$.*



Ground truth

For example, $\mathrm{Rank}(\mathbb{P}(X_7, X_8, X_9, X_2)) = |\mathrm{supp}(L_3)|$,
since L_3 d-separates all variables in $\{X_7, X_8, X_9, X_2\}$



Algorithm for Estimating Discrete 3PLSM

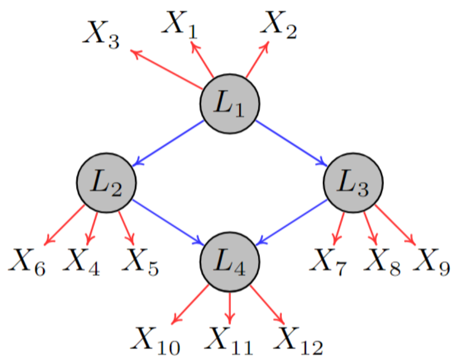
- **Step I:** Identify Causal Cluster
  - Find **causal clusters** from the observed variable set by Tensor Rank Condition

Proposition 4.3 (Identification of causal cluster). *In the discrete 3PLSM mode, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let* $r = |\mathrm{supp}(L_i)|$ *denote the cardinality of the latent support. Given three disjoint observed variables* $X_i, X_j, X_k \in \mathbf{X}$,

- *Rule1: if the rank of tensor* $\mathcal{T}_{(X_i,X_j,X_k)}$ *is not equal to* r, *i.e.,* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k)}) \neq r$, *then* X_i, X_j *and* X_k *belong to the different latent parents.*
- *Rule2: for any* X_s, $X_s \in \mathbf{X} \setminus \{X_i, X_j, X_k\}$, *if the rank of tensor* $\mathcal{T}_{(X_i,X_j,X_k,X_s)}$ *is* r, *i.e.,* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k,X_s)}) = r$, *then* $\{X_i, X_j, X_k\}$ *share the same latent parent.*

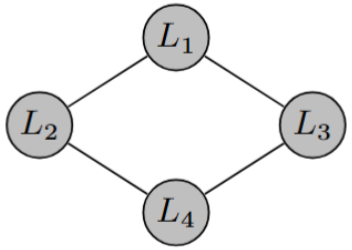
- **Step II:** Identify Causal Structure among Latent Variables
  - Identify the d-separation relations among latent variables by Tensor Rank Condition

Theorem 4.7 (d-separation among latent variable). *In the discrete 3PLSM, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let* r *denote the cardinality of the latent support. Then,* $L_i \perp\!\!\!\perp L_j | \mathbf{L}_p$ *if and only if* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,\mathbf{X}_{p1},\mathbf{X}_{p2})}) = r^{|\mathbf{L}_p|}$, *where* X_i *and* X_j *are the pure children of* L_i *and* L_j, $\mathbf{X}_{p1}$ *and* $\mathbf{X}_{p2}$ *are two disjoint child sets of* $\mathbf{L}_p$ *that satisfy* $\forall L_i \in \mathbf{L}_p, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p1} \neq \emptyset, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p2} \neq \emptyset$.



Ground truth

For example, $\mathrm{Rank}(\mathbb{P}(X_4, X_7, X_1, X_2)) = |\mathrm{supp}(L_1)|$, since L_1 d-separates L_2 from L_3



Algorithm for Estimating Discrete 3PLSM

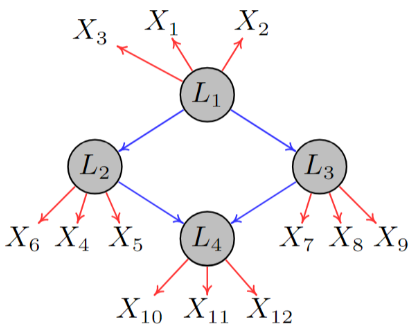
- **Step I:** Identify Causal Cluster
  - Find **causal clusters** from the observed variable set by <u>Tensor Rank Condition</u>

Proposition 4.3 (Identification of causal cluster). *In the discrete 3PLSM mode, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let* $r = |\mathrm{supp}(L_i)|$ *denote the cardinality of the latent support. Given three disjoint observed variables* $X_i, X_j, X_k \in \mathbf{X}$,

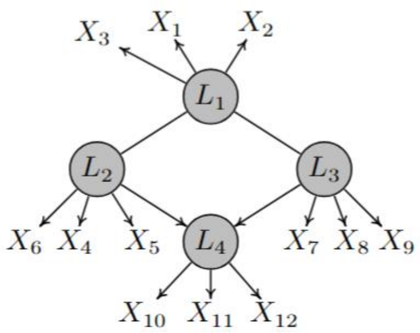
- *Rule1: if the rank of tensor* $\mathcal{T}_{(X_i,X_j,X_k)}$ *is not equal to r, i.e.,* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k)}) \neq r$, *then* X_i, X_j *and* X_k *belong to the different latent parents.*
- *Rule2: for any* X_s, $X_s \in \mathbf{X} \setminus \{X_i, X_j, X_k\}$, *if the rank of tensor* $\mathcal{T}_{(X_i,X_j,X_k,X_s)}$ *is r, i.e.,* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,X_k,X_s)}) = r$, *then* $\{X_i, X_j, X_k\}$ *share the same latent parent.*

- **Step II:** Identify Causal Structure among Latent Variables
  - Identify the d-separation relations among latent variables by <u>Tensor Rank Condition</u>

Theorem 4.7 (d-separation among latent variable). *In the discrete 3PLSM, suppose Assumption 2.2 ~ Assumption 2.4 hold. Let r denote the cardinality of the latent support. Then,* $L_i \perp\!\!\!\perp L_j | \mathbf{L}_p$ *if and only if* $\mathrm{Rank}(\mathcal{T}_{(X_i,X_j,\mathbf{X}_{p1},\mathbf{X}_{p2})}) = r^{|\mathbf{L}_p|}$, *where* X_i *and* X_j *are the pure children of* L_i *and* L_j, $\mathbf{X}_{p1}$ *and* $\mathbf{X}_{p2}$ *are two disjoint child sets of* $\mathbf{L}_p$ *that satisfy* $\forall L_i \in \mathbf{L}_p, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p1} \neq \emptyset, \mathrm{Ch}_{L_i} \cap \mathbf{X}_{p2} \neq \emptyset$.



Ground truth



Output of our algorithm

The latent structure is identified up to a Markov equivalent class!

Identifiability Results

Algorithm 1 Finding the causal cluster

Input: Data from a set of measured variables $\mathbf{X}_{\mathcal{G}}$, and the dimension of latent support r

Output: Causal cluster $\mathcal{C}$

Initialize the causal cluster set $\mathcal{C} := \varnothing$, and $\mathcal{G}' = \varnothing$;
// Identify Causal Skeleton
Begin the recursive procedure
repeat
  for each X_i, X_j and $X_k \in \mathbf{X}$ **do**
    if $\text{Rank}(\mathcal{T}_{\{X_i, X_j, X_k\}}) \neq r$ **then**
      Continue; *// Rule1 of Prop. 4.3*
    end if
    if $\text{Rank}(\mathcal{T}_{\{X_i, X_j, X_k, X_s\}}) = r$, for all $X_s \in \mathbf{X} \setminus \{X_i, X_j, X_k\}$ **then**
      $\mathbf{C} = \mathbf{C} \cup \{\{X_i, X_j, X_k\}\}$;
    end if
  end for
until no causal cluster is found.
// Merging cluster and introducing latent variables
Merge all the overlapping sets in $\mathbf{C}$ by Prop. 4.5.
for each $C_i \in \mathbf{C}$ **do**
  Introduce a latent variable L_i for C_i;
  $\mathcal{G} = \mathcal{G} \cup \{L_i \to X_j | X_j \in C_i\}$.
end for
return Graph $\mathcal{G}$ and causal cluster $\mathcal{C}$.

- **Theorem** (*Identification of Measurement Model*). In the discrete 3PLSM model, suppose Assumption 2.2 ~ Assumption 2.4 hold. The measurement model is fully identifiable by Algorithm 1.

Algorithm 2 PC-TENSOR-RANK

Input: Data set $\mathbf{X} = \{X_1, \ldots, X_m\}$ and causal cluster $\mathcal{C}$

Output: A partial DAG $\mathcal{G}$.

Initialize the maximal conditions set dimension k;
Let L_i denote as C_i, $C_i \in \mathcal{C}$;
Form the complete undirected graph $\mathcal{G}$ on the latent variable set $\mathbf{L}$;
for $\forall L_i, L_j \in \mathbf{L}$ and adjacent in $\mathcal{G}$ **do**
  //Test the CI relations among latent variables by Theorem 4.7
  if $\exists \mathbf{L}_p \subseteq \mathbf{L} \setminus \{L_i, L_j\}$ and $(|\mathbf{L}_p| < k)$ such that $L_i \perp\!\!\!\perp L_j | \mathbf{L}_p$ hold **then**
    delete edge $L_i - L_j$ from G;
  end if
end for
Search V structures and apply meek rules Meek (1995).
return a partial DAG $\mathcal{G}$ of latent variables.

- **Theorem** (*Identification of Structure Model*). In the discrete 3PLSM, suppose Assumption 2.2 ~ Assumption 2.4 hold. Given the measurement model, the causal structure over the latent variable is identified up to a Markov equivalent class by the PC-TENSOR-RANK algorithm.

Experimental Results

Table 2: Results on learning pure measurement models, where the data is generated by the discrete 3PLSM. Lower value means higher accuracy.

Algorithm		Latent omission				Latent commission				Mismeasurements			
		Our	BayPy	LTM	BPC	**Our**	BayPy	LTM	BPC	**Our**	BayPy	LTM	BPC
$SM_1 + MM_1$	5k	0.15(3)	0.10(2)	0.15(3)	0.96(10)	0.00(0)	0.10(2)	0.00(0)	0.00(0)	0.05(1)	0.00(0)	0.00(0)	0.00(0)
	10k	0.05(1)	0.05(1)	0.10(2)	0.90(10)	0.00(0)	0.05(1)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)
	50k	0.00(0)	0.00(0)	0.00(0)	0.90(10)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)
$SM_2 + MM_1$	5k	0.23(5)	0.19(6)	0.26(6)	0.90(10)	0.00(0)	0.19(6)	0.03(1)	0.00(0)	0.05(2)	0.19(6)	0.23(6)	0.00(0)
	10k	0.13(4)	0.13(4)	0.13(4)	0.86(10)	0.00(0)	0.03(4)	0.00(0)	0.00(0)	0.00(0)	0.13(4)	0.13(4)	0.00(0)
	50k	0.06(2)	0.10(3)	0.10(3)	0.86(10)	0.00(0)	0.13(4)	0.00(0)	0.00(0)	0.00(0)	0.13(4)	0.10(3)	0.00(0)
$SM_2 + MM_2$	5k	0.12(2)	0.19(6)	0.21(5)	0.90(10)	0.00(0)	0.19(6)	0.00(0)	0.00(0)	0.03(1)	0.16(6)	0.21(5)	0.00(0)
	10k	0.03(1)	0.13(4)	0.10(3)	0.86(10)	0.00(0)	0.13(4)	0.00(0)	0.00(0)	0.00(0)	0.11(4)	0.10(3)	0.00(0)
	50k	0.00(0)	0.07(2)	0.07(2)	0.83(10)	0.00(0)	0.07(2)	0.00(0)	0.00(0)	0.00(0)	0.07(2)	0.06(2)	0.00(0)
$SM_3 + MM_1$	5k	0.25(6)	0.30(6)	0.55(10)	0.86(10)	0.00(0)	0.30(6)	0.00(0)	0.00(0)	0.12(5)	0.20(6)	0.55(10)	0.00(0)
	10k	0.17(5)	0.25(5)	0.50(10)	0.83(10)	0.00(0)	0.25(5)	0.00(0)	0.00(0)	0.05(3)	0.16(5)	0.50(10)	0.00(0)
	50k	0.08(3)	0.20(4)	0.50(10)	0.83(10)	0.00(0)	0.20(4)	0.00(0)	0.00(0)	0.03(2)	0.13(4)	0.50(10)	0.00(0)

Table 3: Results on learning the structure model. The symbol '-' indicates that the current method does not output this information. Lower value means higher accuracy.

Algorithm		Edge omission				Edge commission				Orientation omission			
		Our	BayPy	LTM	BPC	**Our**	BayPy	LTM	BPC	**Our**	BayPy	LTM	BPC
$Collider + MM_1$	5k	0.00(0)	1.00(10)	0.26(8)	1.00(10)	0.10(1)	0.00(0)	0.00(0)	0.00(0)	0.10(1)	1.00(10)	–	1.00(0)
	10k	0.00(0)	1.00(10)	0.23(6)	1.00(10)	0.00(0)	0.02(1)	0.0(0)	0.00(0)	0.00(0)	1.00(10)	–	1.00(0)
	50k	0.00(0)	1.00(10)	0.10(3)	1.00(10)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	1.00(10)	–	1.00(0)
$SM_2 + MM_1$	5k	0.15(3)	1.00(10)	0.16(6)	1.00(10)	0.10(1)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
	10k	0.05(1)	1.00(10)	0.13(4)	1.00(10)	0.01(1)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
	50k	0.00(0)	1.00(10)	0.10(3)	1.00(10)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
$Star + MM_1$	5k	0.10(3)	1.00(10)	0.25(5)	1.00(10)	0.20(5)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
	10k	0.06(2)	1.00(10)	0.15(3)	1.00(10)	0.08(3)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
	50k	0.03(1)	1.00(10)	0.15(3)	1.00(10)	0.05(2)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	–	0.00(0)
$SM_3 + MM_1$	5k	0.22(7)	1.00(10)	0.50(10)	1.00(10)	0.40(6)	0.00(0)	0.02(1)	0.00(0)	0.20(2)	1.00(10)	–	1.00(10)
	10k	0.15(5)	1.00(10)	0.50(10)	1.00(10)	0.10(2)	0.00(0)	0.00(0)	0.00(0)	0.10(1)	1.00(10)	–	1.00(10)
	50k	0.05(2)	1.00(10)	0.50(10)	1.00(10)	0.05(1)	0.00(0)	0.00(0)	0.00(0)	0.00(0)	1.00(10)	–	1.00(10)

- Setup: different measurement model (MM) and different structure model (SM)

- Can we recover the ground-truth structure, including the measurement model and the structure model?

Conclusions and Future work

- **Establish a connection between the tensor rank condition and the graphical patterns**
- **Provide the simple but efficient algorithm for learning discrete latent structure model**
- **Future work: hierarchical structure, impure structure condition…**

THANK YOU FOR YOUR ATTENTION!