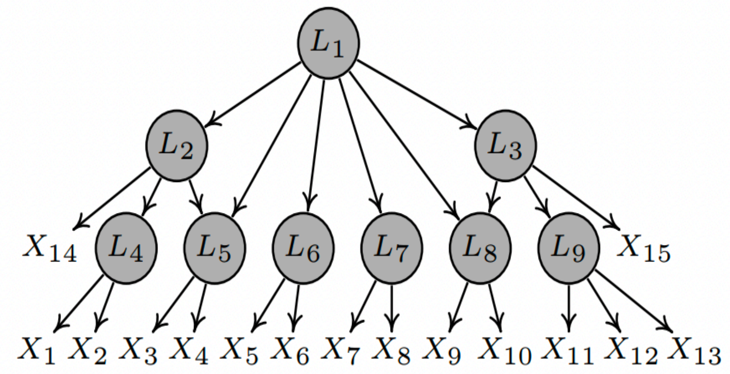


Identification of Linear Latent Hierarchical Structure

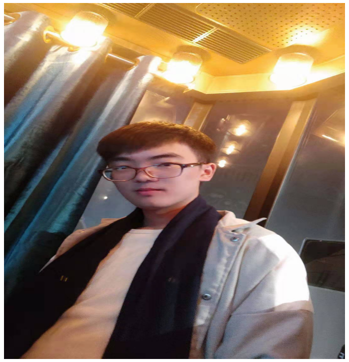
Feng Xie

Department of Applied Statistics, Beijing Technology and Business University
fengxie@btbu.edu.cn

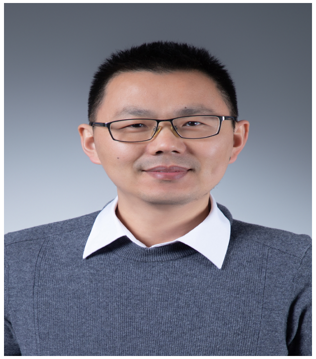
September 17, 2022, PCIC

Acknowledgments

Biwei Huang

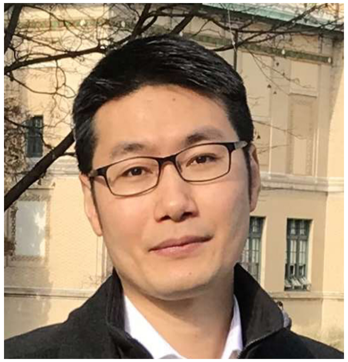


Zhengming Chen



Yangbo He

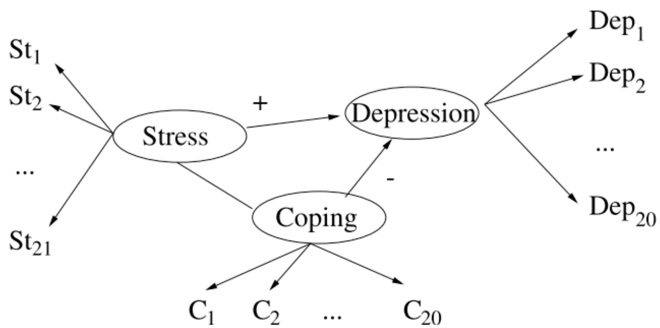
Zhi Geng



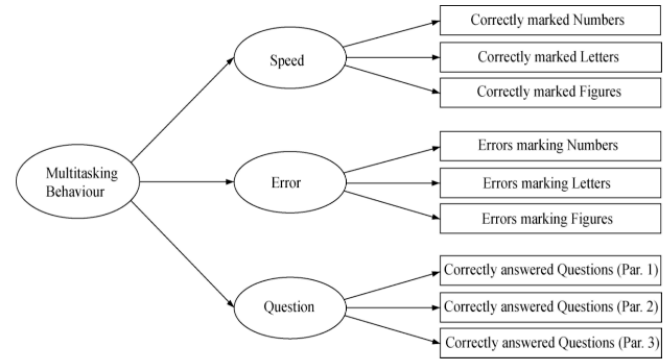
Kun Zhang

Introduction-Latent Causal Structure

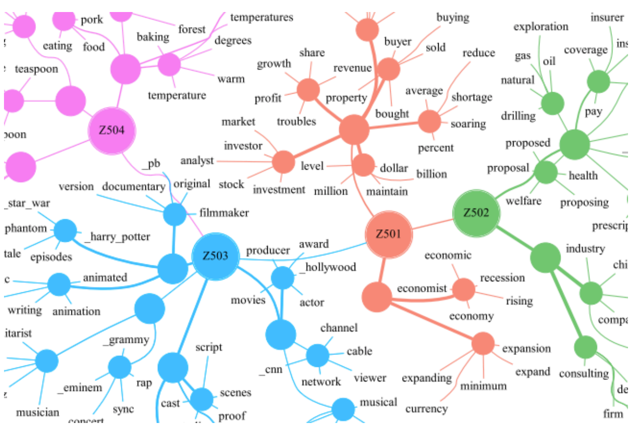
Circle nodes are unobserved



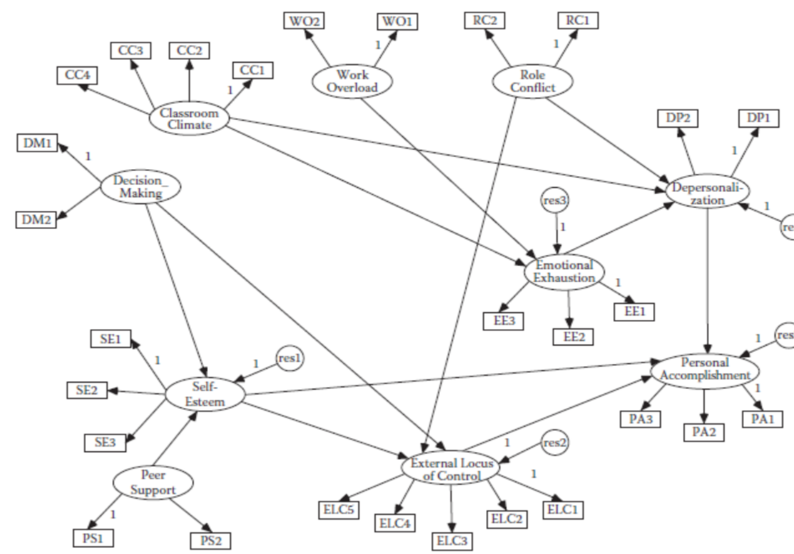
Stress coping model [Silva, JMLR'2006]



SIMKAP model [Himi, Cognition'2019]



Topic model [Chen, AIJ'2017]



Teacher's burnout model [Byrne, 2010]

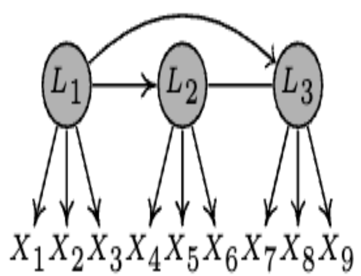
Open Problem: How do we learn the underlying latent structure

Only from observed variables?

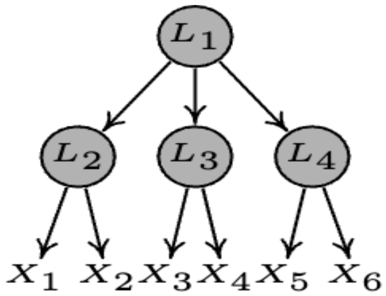
Related works

- **Measurement-based model:** All latent variables have directed measured variables as children in the system [Bollen K A, 1989; Spirtes et al., 2000; Silva et al., JMLR'2006; ; Cui et al., UAI'2018; Shimizu et al., Neurocomputing2009; Kummerfeld and Ramsey, KDD'2016; Cai et al., NeruIPS'2019; Xie et al., NeruIPS'2020 Chen et al., AAAI'2022]…

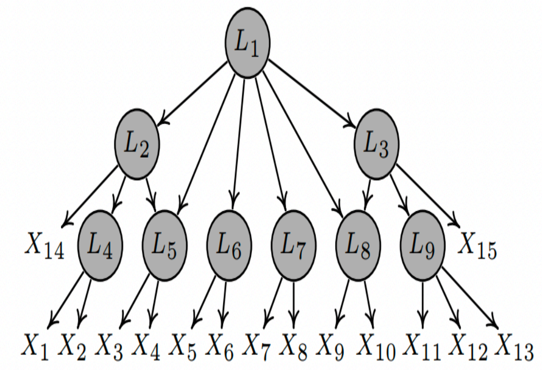
- **Latent tree model:** Each latent variables have at least three neighbors and there is only one path between every pair of variables in the system [Pear, 1988; Choi et al.,JMLR'2011; Zhang, JMLR'2004; Poon et al., ICML'2010; Harmeling & Williams, TPAMI'2010; Mourad et al., JAIR'2013; Zhang & Poon, AAAI'2017; Etesami et al., Neural Computing'2016; Drton et al., Bernoulli'2017]….



(a)
Measurement-based Structure



(b)
Latent Tree Structure



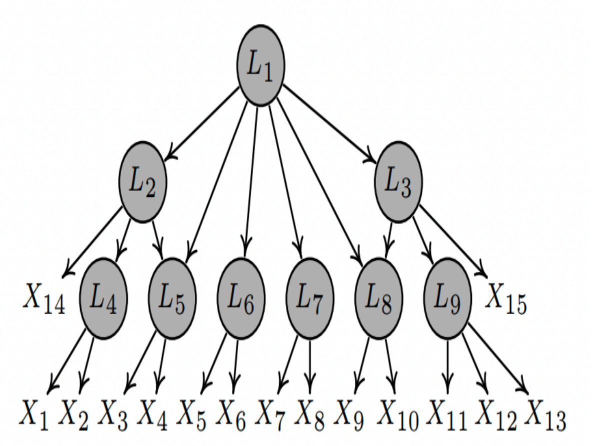
(c)
Latent Hierarchical Structure

Problem Setup

[Linear Latent Hierarchical Structure Model] Let $\mathbf{X}=\{X_1,\ldots, X_m\}$ denote the set of observed variables, and $\mathbf{L}=\{L_1,\ldots, L_n\}$ denote the set of latent variables. All variables $\mathbf{V} = \mathbf{X} \cup \mathbf{L}$ are generated according to a particular type of linear causal model:

$$X_i = \sum_{L_j \in Pa(X_i)} b_{ij} L_j + \varepsilon_{X_i}, \quad (1)$$

$$L_j = \sum_{L_k \in Pa(L_j)} c_{jk} L_k + \varepsilon_{L_j}, \quad (2)$$



An example of latent hierarchical structure

Find the sufficient conditions that render the causal structure of a latent hierarchical model identifiable!

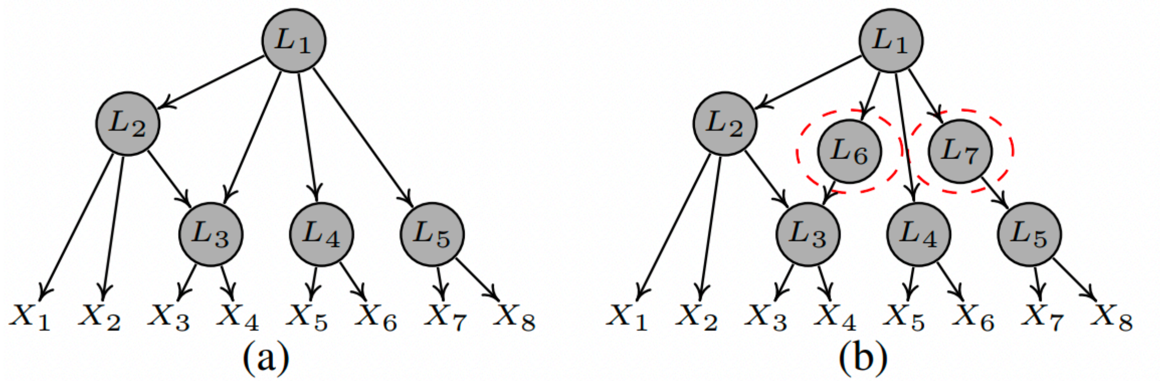
Sufficient Conditions for Model Identification

Condition 1 [Non-Gaussianaity] All noise terms of variables **V** follow non-Gaussian distributions.

Condition 1 is essential to identify causal directions between any two variables.

+

Condition 2 [Minimal Latent Hierarchical Structure] (1) each latent variable has at least three neighbors, and (2) each latent variable has at least two pure children.



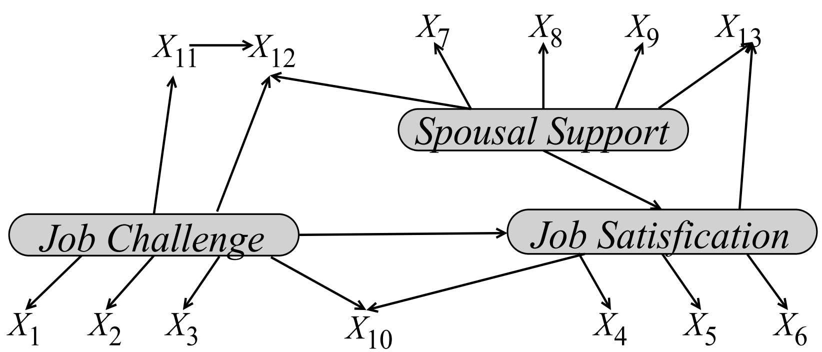
(a) An example of the minimal latent hierarchical structure, (b) A counter-example of the minimal latent hierarchical structure.

Condition 2 ensures that the structure among latent variables does not include any "redundant" latent nodes.

GIN condition-Testing "d-separation" in latent variable model

DEF[Generalized Independent Noise(GIN), condition, Xie et al., 2020]
$(\boldsymbol{Z}, \boldsymbol{Y})$ follows the GIN condition iff there exists non-zeros ω such that $\omega^{\mathrm{T}}\boldsymbol{Y}$ is independent from $\boldsymbol{Z}$, where $\omega^{\mathrm{T}}\mathbb{E}[YZ^{\mathrm{T}}] = 0$.

Graphical criterion: If $(\boldsymbol{Z}, \boldsymbol{Y})$ follows the GIN condition, there is an exogenous subset of the common cause of $\boldsymbol{Y}$ to *d-separate* from $\boldsymbol{Y}$ from $\boldsymbol{Z}$.

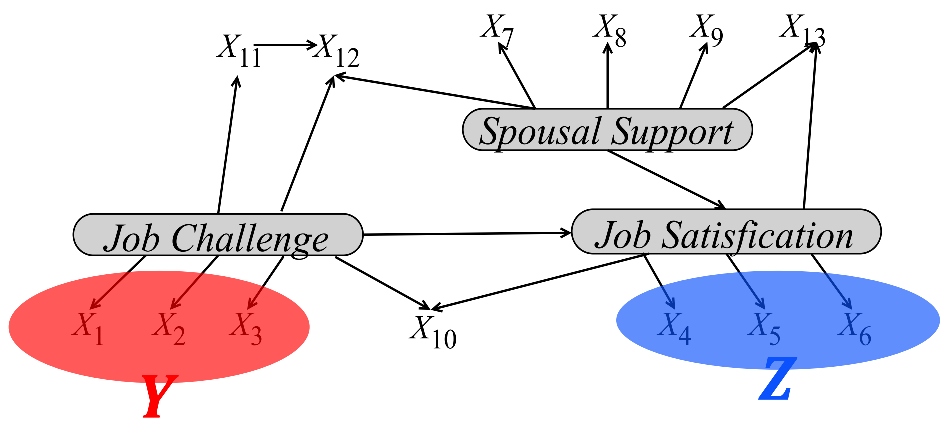


F. Xie, R. Cai, B. Huang, C. Glymour, Z. Hao, and K. Zhang. Generalized Independent Noise Condition for Estimating Latent Variable Causal Graphs. NeurIPS 2020.

GIN condition-Testing "d-separation" in latent variable model

DEF[Generalized Independent Noise(GIN), condition, Xie et al., 2020]
$(\boldsymbol{Z}, \boldsymbol{Y})$ follows the GIN condition iff there exists non-zeros ω such that $\omega^{\mathrm{T}}\boldsymbol{Y}$ is independent from $\boldsymbol{Z}$, where $\omega^{\mathrm{T}}\mathbb{E}[YZ^{\mathrm{T}}] = 0$.

Graphical criterion: If $(\boldsymbol{Z}, \boldsymbol{Y})$ follows the GIN condition, there is an exogenous subset of the common cause of $\boldsymbol{Y}$ to *d-separate* from $\boldsymbol{Y}$ from $\boldsymbol{Z}$.



$\boldsymbol{Z}$ $\boldsymbol{Y}$
$(\{X_4, X_5, X_6\}, \{X_1, X_2, X_3\})$ follows GIN condition, then the exogenous subset of the commo n cause of $\{X_1, X_2, , X_3\}$, i.e., ***Job Challenge*** *d-separates* $\{X_1, X_2, X_3\}$ from $\{X_4, X_5, X_6\}$。

F. Xie, R. Cai, B. Huang, C. Glymour, Z. Hao, and K. Zhang. Generalized Independent Noise Condition for Estimating Latent Variable Causal Graphs. NeurIPS 2020.

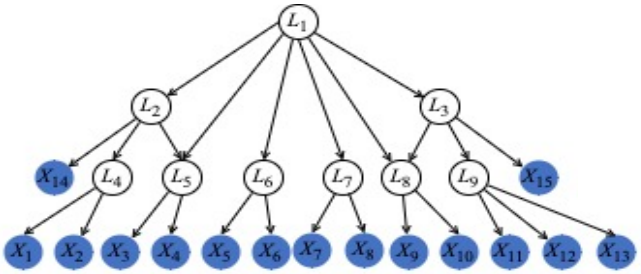
Model Estimation

- **Step 1-Locate all latent variables**
  - P1. Identify **causal clusters** from the active variable set
  - P2. Determine the number of **new latent variables** that need to be introduced for these clusters
  - P3. Update the active variable set

- **Step 2-Infer the causal structure among the identified latent variables**
  - P1. identify the **causal order** among latent variables
  - P2. remove **redundant** edges

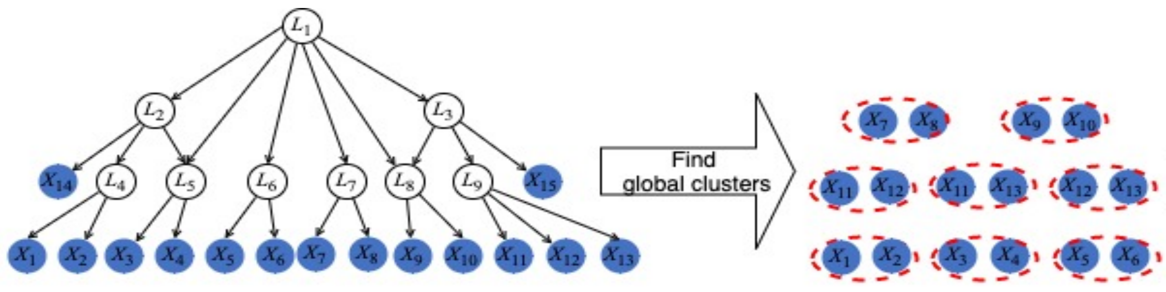
Notice that the number of latent variables and the level of the structure are unknown!

Illustration of Step 1



Ground-truth graph

Illustration of Step 1



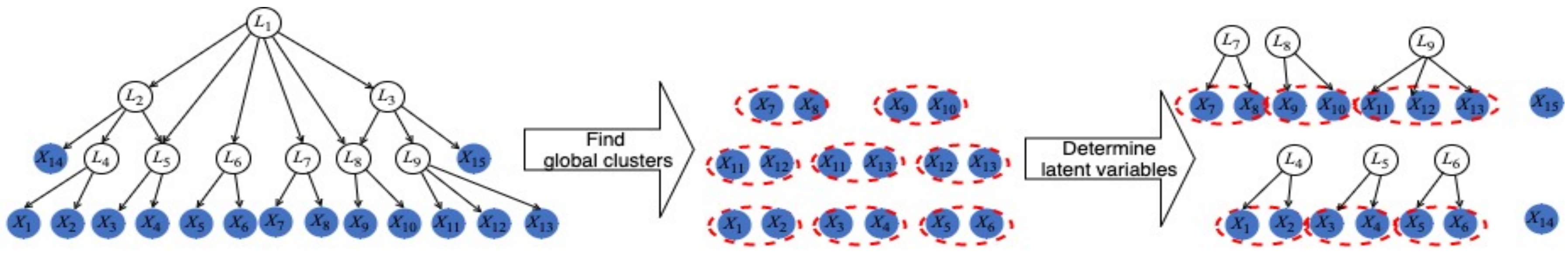
Proposition 1 (Identifying Global Causal Clusters). *Let $\mathcal{A}$ be the active variable set and $\mathbf{Y}$ be a proper subset of $\mathcal{A}$. Then $\mathbf{Y}$ is a global causal cluster if and only if the following two conditions hold: 1) for any subset $\tilde{\mathbf{Y}}$ of $\mathbf{Y}$ with $|\tilde{\mathbf{Y}}| = 2$, $(\mathcal{A} \backslash \mathbf{Y}, \tilde{\mathbf{Y}})$ follows the GIN condition, and 2) no proper subset of $\mathbf{Y}$ satisfies condition 1).*

E.g., $\mathbf{Y} = \{X_1, X_2\}$,we have $(\mathbf{X} \backslash \{X_1, X_2\}, \{X_1, X_2\})$ follows GIN condition.

Implies

$\{X_1, X_2\}$ is a global cluster!

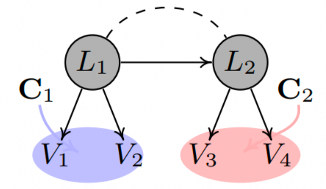
Illustration of Step 1



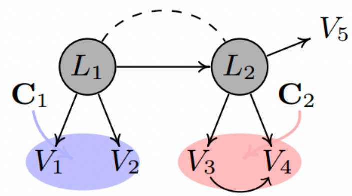
Proposition 2 (Merging Rules). *Let $\mathcal{A}$ be the active variable set and $\mathbf{C}_1$ and $\mathbf{C}_2$ be two global causal clusters. $\mathbf{C}_1$ and $\mathbf{C}_2$ share the common latent parent, if one of the following rules holds.*

- $\mathcal{R}1$. *1) $\mathbf{C}_1$ and $\mathbf{C}_2$ are both pure clusters, and 2) for any subset $\tilde{\mathbf{C}} \subseteq \mathbf{C}_1 \cup \mathbf{C}_2$ with $|\tilde{\mathbf{C}}| = 2$, $(\mathcal{A} \backslash \tilde{\mathbf{C}}, \tilde{\mathbf{C}})$ follows the GIN condition.*
- $\mathcal{R}2$. *1) One of the clusters is a pure cluster and the other is not, e.g., $\mathbf{C}_1$ is pure and $\mathbf{C}_2$ is impure, and 2) for any variable $V_i \in \mathbf{C}_1$ and any variable $V_j \in \mathbf{C}_2$, $(\mathcal{A} \backslash \{\mathbf{C}_2, V_i\}, \{V_i, V_j\})$ follows the GIN condition.*
- $\mathcal{R}3$. *1) $\mathbf{C}_1$ and $\mathbf{C}_2$ both are impure clusters, and 2) for any subset $\tilde{\mathbf{C}} \subseteq \mathbf{C}_1 \cup \mathbf{C}_2$ with $|\tilde{\mathbf{C}}| = 2$, $(\mathcal{A} \backslash \{\mathbf{C}_1 \cup \mathbf{C}_2\}, \tilde{\mathbf{C}})$ follows the GIN condition.*

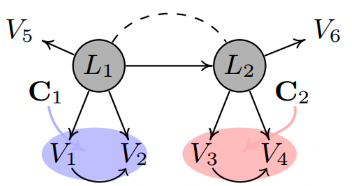
Otherwise, $\mathbf{C}_1$ and $\mathbf{C}_2$ do not share the common latent parent.



$\mathcal{R}_1$

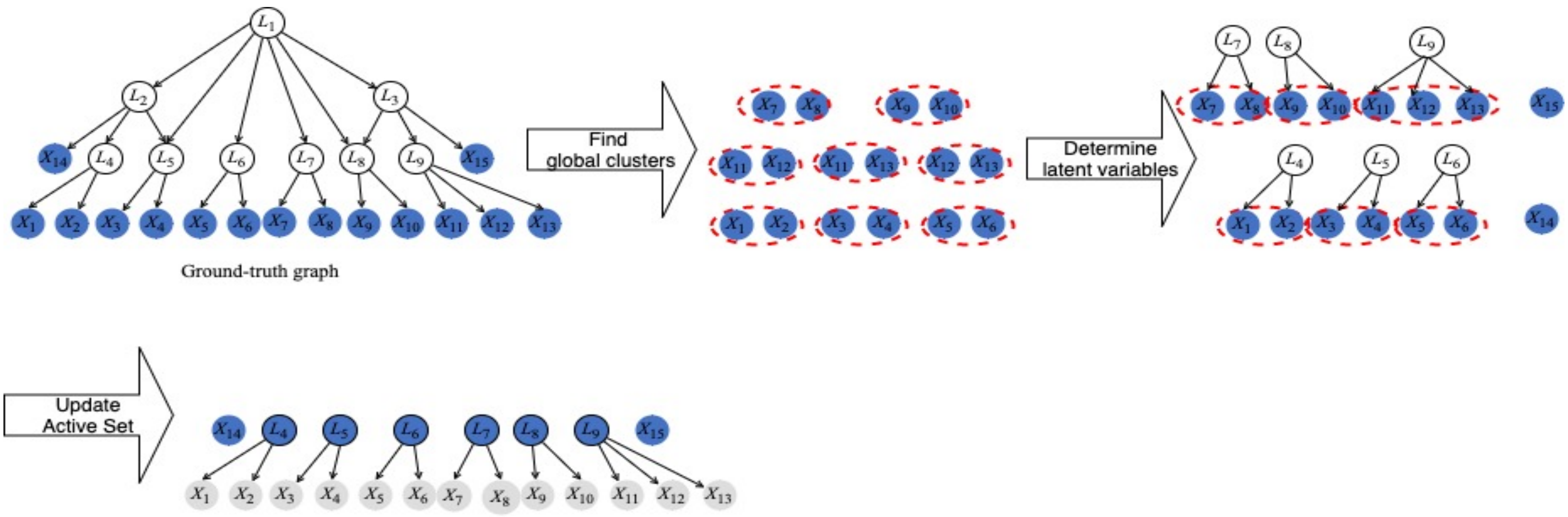


$\mathcal{R}_2$



$\mathcal{R}_3$

Illustration of Step 1



Proposition 3 (Active Variable Set Update)**.** *Let $\mathcal{A}$ be the current active variable set and $\mathcal{L}$ be the latent variable sets discovered in the current iteration. Then the new active variable set $\mathcal{A}' = \mathcal{A} \cup \mathcal{L} \setminus \mathbf{Ch}(\mathcal{L})$. Moreover, the GIN conditions over variables in $\mathcal{A}'$ are equivalent to those that replace $V \in \mathcal{A}'$ by any variable in its corresponding cluster identified in the latest iteration.*

E.g., $(\mathbf{X} \setminus \{X_1, X_2\}, \{L_4, X_{14}\})$ follows GIN condition.

iff

$(\mathbf{X} \setminus \{X_1, X_2\}, \{X_1, X_{14}\})$ follows GIN condition.

Illustration of Step 1

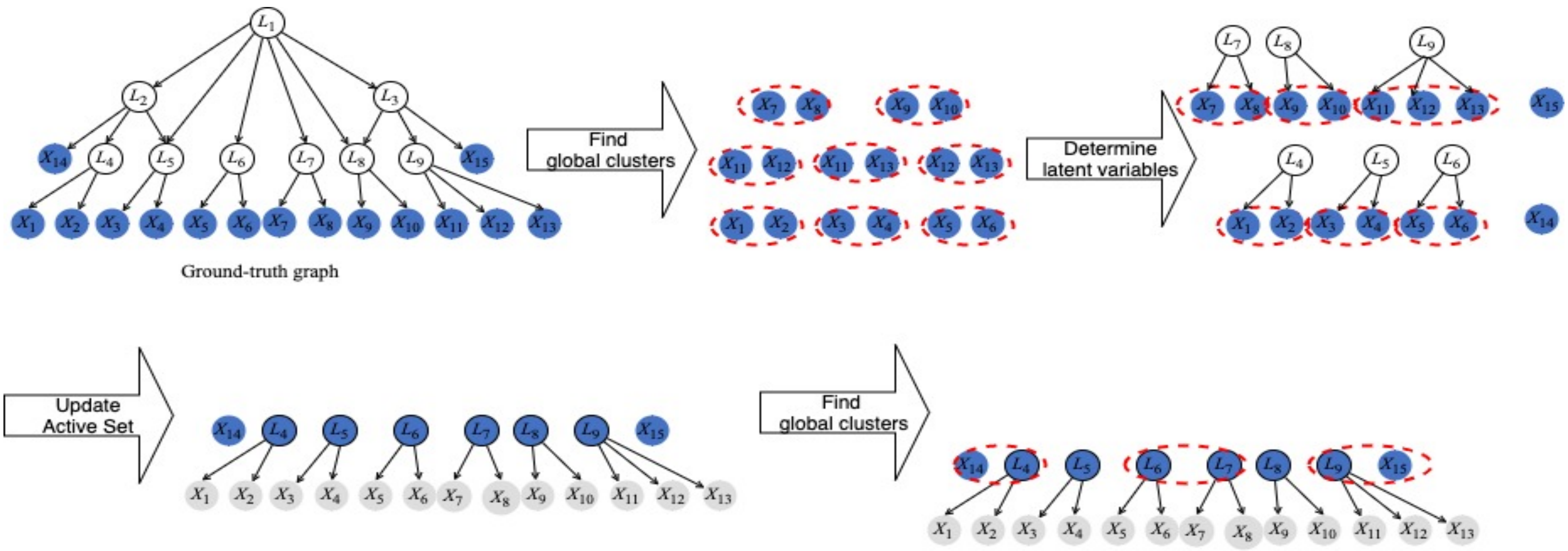


Illustration of Step 1

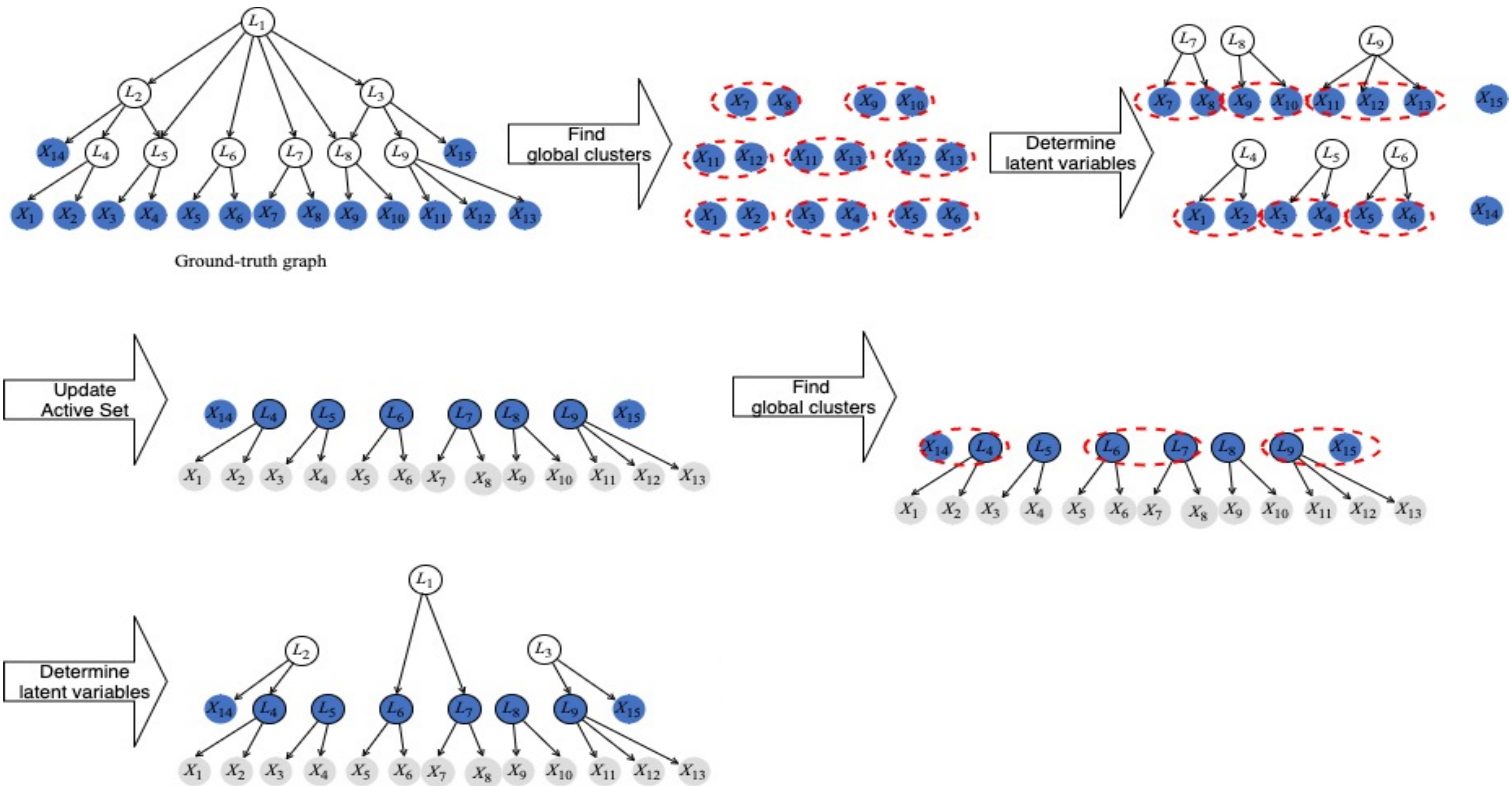


Illustration of Step 1

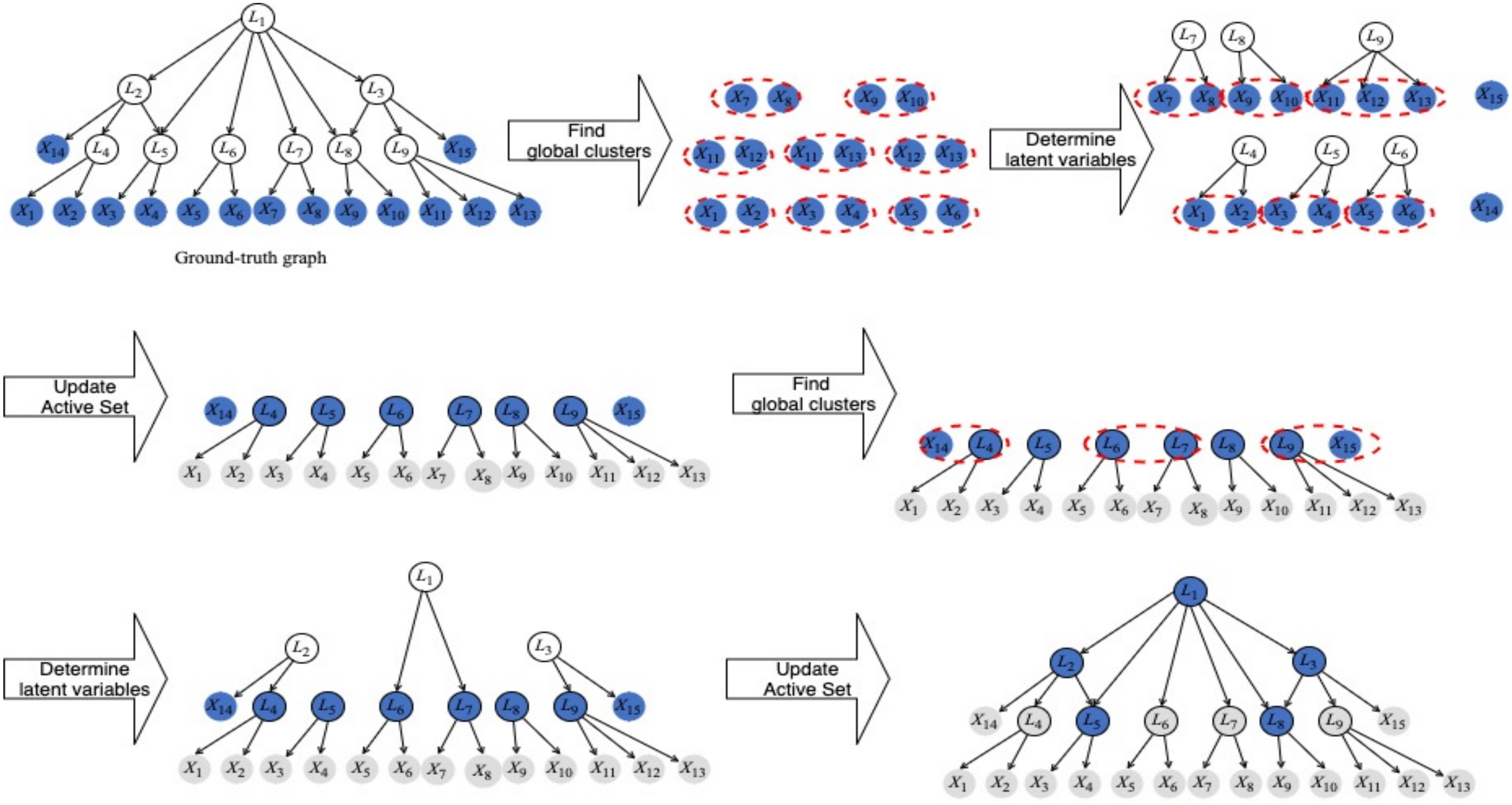


Illustration of Step 1

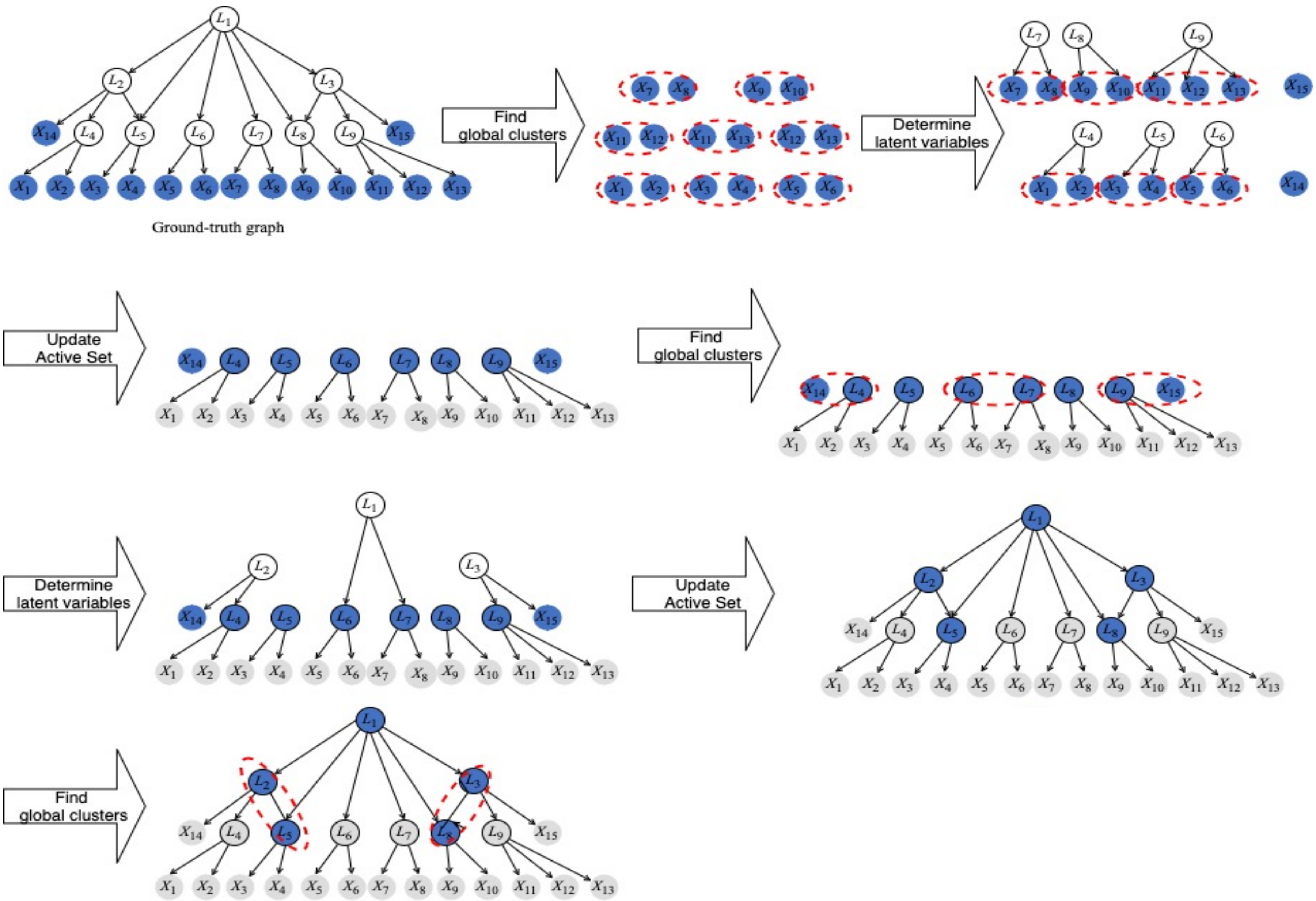


Illustration of Step 1

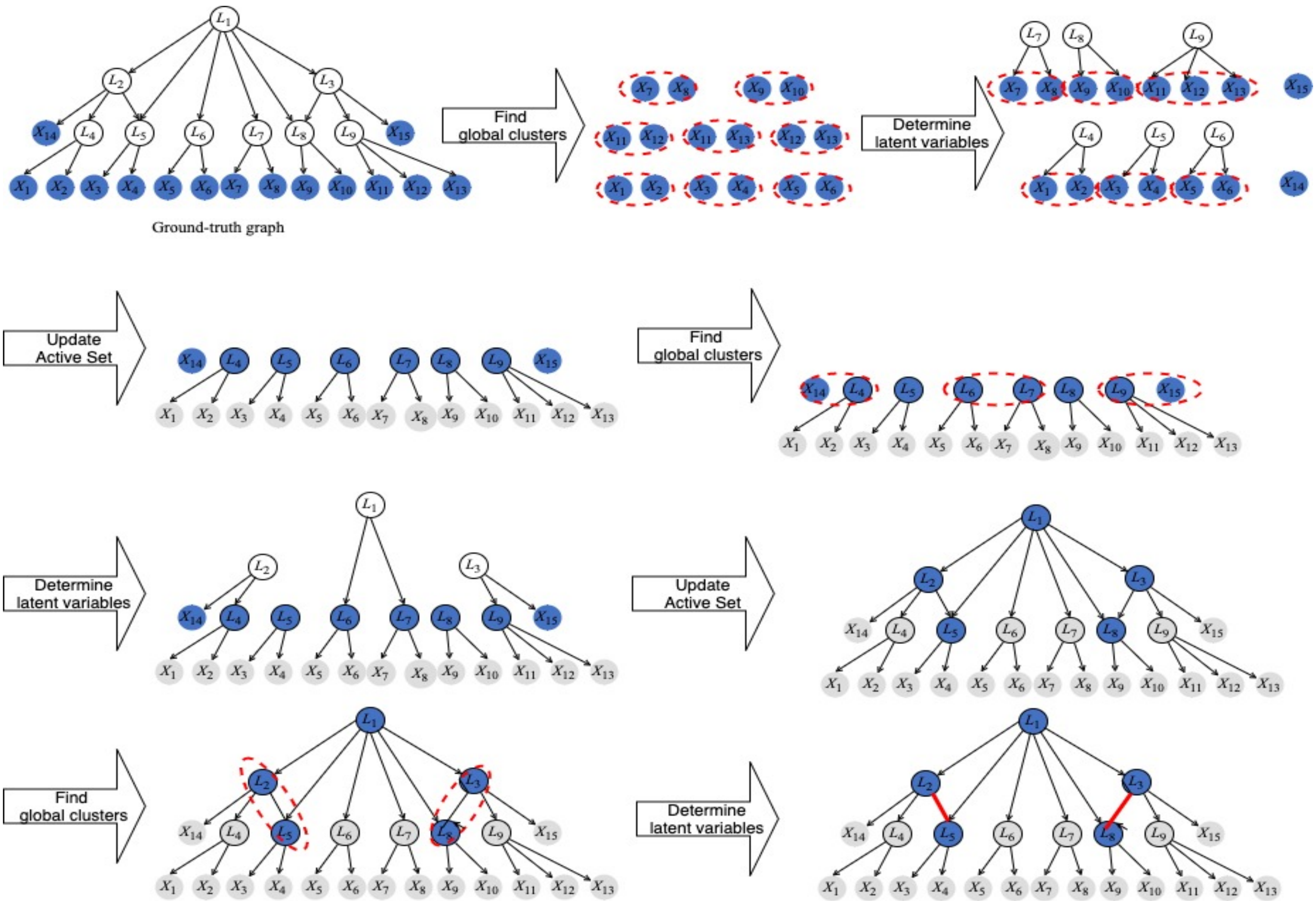
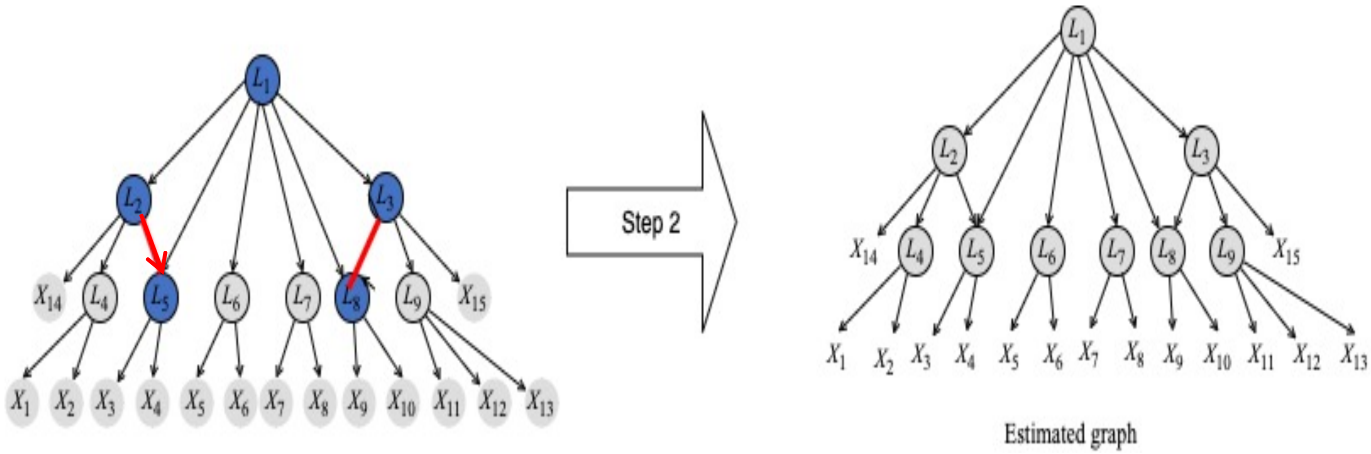
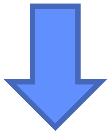


Illustration of Step 2

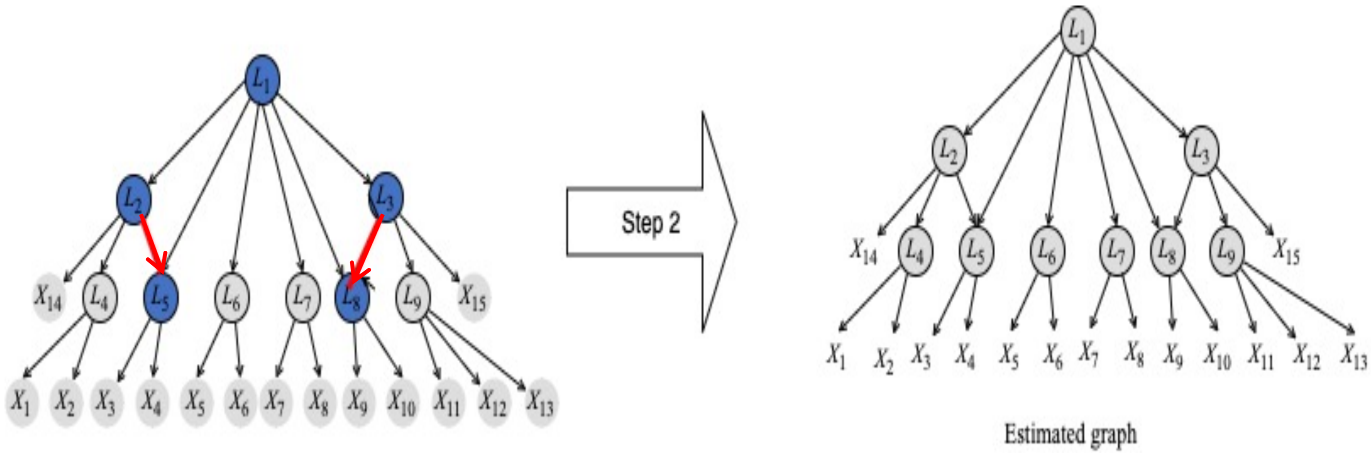


$(\{L_2, L_7\}, \{L_2, L_5, L_6\})$ follows GIN condition
while $(\{L_5, L_7\}, \{L_2, L_5, L_6\})$ violates GIN condition.



$L_2 \succ L_5$

Illustration of Step 2



$(\{L_3, L_7\}, \{L_3, L_8, L_6\})$ follows GIN condition
while $(\{L_8, L_7\}, \{L_3, L_8, L_6\})$ violates GIN condition.

Imply

$L_3 \succ L_8$

Identification Result

Theorem 1 (Identifiability of Latent Hierarchical Structure)**.**
Suppose that the input data $\mathbf{X}$ *follows LiNGLaH with the minimal latent hierarchical structure. Then the underlying causal graph* $\mathcal{G}$ *is fully identifiable with LaHME, including latent variables and their causal relationships.*

The latent hierarchical structure is identifiable under assumptions of non-Gaussianity and minimal latent hierarchical structure.

Simulation Results

- 4 cases, with different latent structures, including measurement-based (Case 1~2) and tree-based (Case 3) structures
- Can we recover the ground-truth structure, including causal direction?
  - Structure Recovery Error Rate: measure falsely recovered the graph
  - Error in Hidden Variable: measure omitted latent variables
  - Correct Ordering Rate: measure the correction of the causal directions

Table 1. Performance of LaHME, GIN, FOFC, BPC, CLRG and CLNJ on learning latent hierarchical structure.

		Structure Recovery Error Rate ↓						Error in Hidden Variables ↓						Correct-Ordering Rate ↑					
Algorithm		LaHME	GIN	FOFC	BPC	CLRG	CLNJ	LaHME	GIN	FOFC	BPC	CLRG	CLNJ	LaHME	GIN	FOFC	BPC	CLRG	CLNJ
Case 1	1k	0.1	0.2	1.0	1.0	1.0	1.0	0.1	0.1	0.5	0.6	2.0	2.0	0.96	0.92	-	-	-	-
	5k	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.1	2.0	2.0	1.0	1.0	-	-	-	-
	10k	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	2.0	2.0	1.0	1.0	-	-	-	-
Case 2	1k	0.2	1.0	1.0	1.0	1.0	1.0	0.2	3.2	3.8	3.9	4.0	4.0	0.9	0.08	-	-	-	-
	5k	0.1	1.0	1.0	1.0	1.0	1.0	0.1	3.0	3.6	3.8	4.0	4.0	0.96	0.1	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	3.0	3.5	3.8	4.0	4.0	1.0	0.1	-	-	-	-
Case 3	1k	0.1	1.0	1.0	1.0	1.0	1.0	0.2	1.3	3.0	3.1	3.0	3.0	0.92	0.0	-	-	-	-
	5k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.2	3.0	3.2	3.0	3.0	1.0	0.0	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	3.2	3.4	3.0	3.0	1.0	0.0	-	-	-	-
Case 4	1k	0.3	1.0	1.0	1.0	1.0	1.0	0.4	3.4	7.0	7.2	8.0	8.0	0.9	0.0	-	-	-	-
	5k	0.2	1.0	1.0	1.0	1.0	1.0	0.2	3.2	6.6	6.9	8.0	8.0	0.94	0.0	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	3.1	5.8	6.7	8.0	8.0	1.0	0.0	-	-	-	-

Simulation Results

- 4 cases, with different latent structures, including measurement-based (Case 1~2) and tree-based (Case 3) structures
- Can we recover the ground-truth structure, including causal direction?
  - Structure Recovery Error Rate: measure falsely recovered the graph
  - Error in Hidden Variable: measure omitted latent variables
  - Correct Ordering Rate: measure the correction of the causal directions

Table 1. Performance of LaHME, GIN, FOFC, BPC, CLRG and CLNJ on learning latent hierarchical structure.

		Structure Recovery Error Rate ↓						Error in Hidden Variables ↓						Correct-Ordering Rate ↑					
Algorithm		LaHME	GIN	FOFC	BPC	CLRG	CLNJ	LaHME	GIN	FOFC	BPC	CLRG	CLNJ	LaHME	GIN	FOFC	BPC	CLRG	CLNJ
Case 1	1k	0.1	0.2	1.0	1.0	1.0	1.0	0.1	0.1	0.5	0.6	2.0	2.0	0.96	0.92	-	-	-	-
	5k	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.1	2.0	2.0	1.0	1.0	-	-	-	-
	10k	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	2.0	2.0	1.0	1.0	-	-	-	-
Case 2	1k	0.2	1.0	1.0	1.0	1.0	1.0	0.2	3.2	3.8	3.9	4.0	4.0	0.9	0.08	-	-	-	-
	5k	0.1	1.0	1.0	1.0	1.0	1.0	0.1	3.0	3.6	3.8	4.0	4.0	0.96	0.1	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	3.0	3.5	3.8	4.0	4.0	1.0	0.1	-	-	-	-
Case 3	1k	0.1	1.0	1.0	1.0	1.0	1.0	0.2	1.3	3.0	3.1	3.0	3.0	0.92	0.0	-	-	-	-
	5k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.2	3.0	3.2	3.0	3.0	1.0	0.0	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	3.2	3.4	3.0	3.0	1.0	0.0	-	-	-	-
Case 4	1k	0.3	1.0	1.0	1.0	1.0	1.0	0.4	3.4	7.0	7.2	8.0	8.0	0.9	0.0	-	-	-	-
	5k	0.2	1.0	1.0	1.0	1.0	1.0	0.2	3.2	6.6	6.9	8.0	8.0	0.94	0.0	-	-	-	-
	10k	0.0	1.0	1.0	1.0	1.0	1.0	0.0	3.1	5.8	6.7	8.0	8.0	1.0	0.0	-	-	-	-

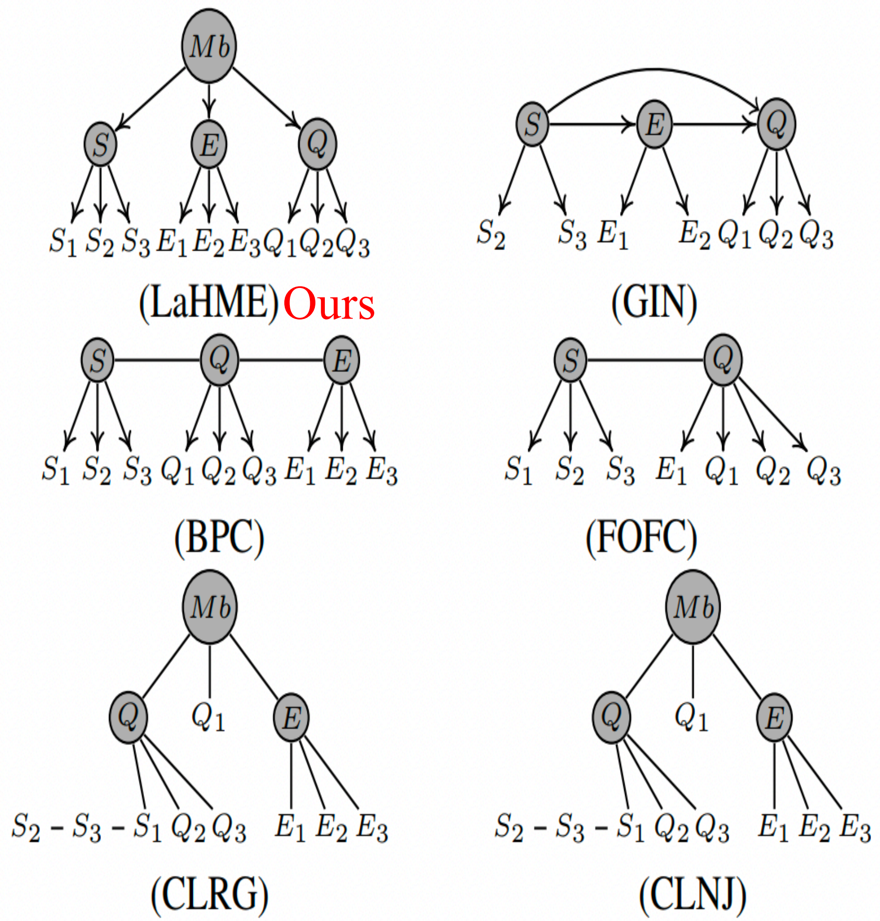
Ours Ours Ours

Application to multitasking behavior Data

- The data set consists of 202 samples

Latent Factors	Children (Indicators)
Speed (S)	Correctly marked Numbers (S1), Correctly marked Latters (S2), and Correctly marked Figures (S3)
Error (E)	Errors marking Numbers (E1), Errors marking Latters (E2), and Errors marking Figures (E3)
Question (Q)	Correctly answered Questions Par.1 (Q1), Correctly answered Questions Par.2 (Q2), and Correctly answered Questions Par.3 (Q3)
Multitasking behavior (Mb)	Speed, Error, and Question

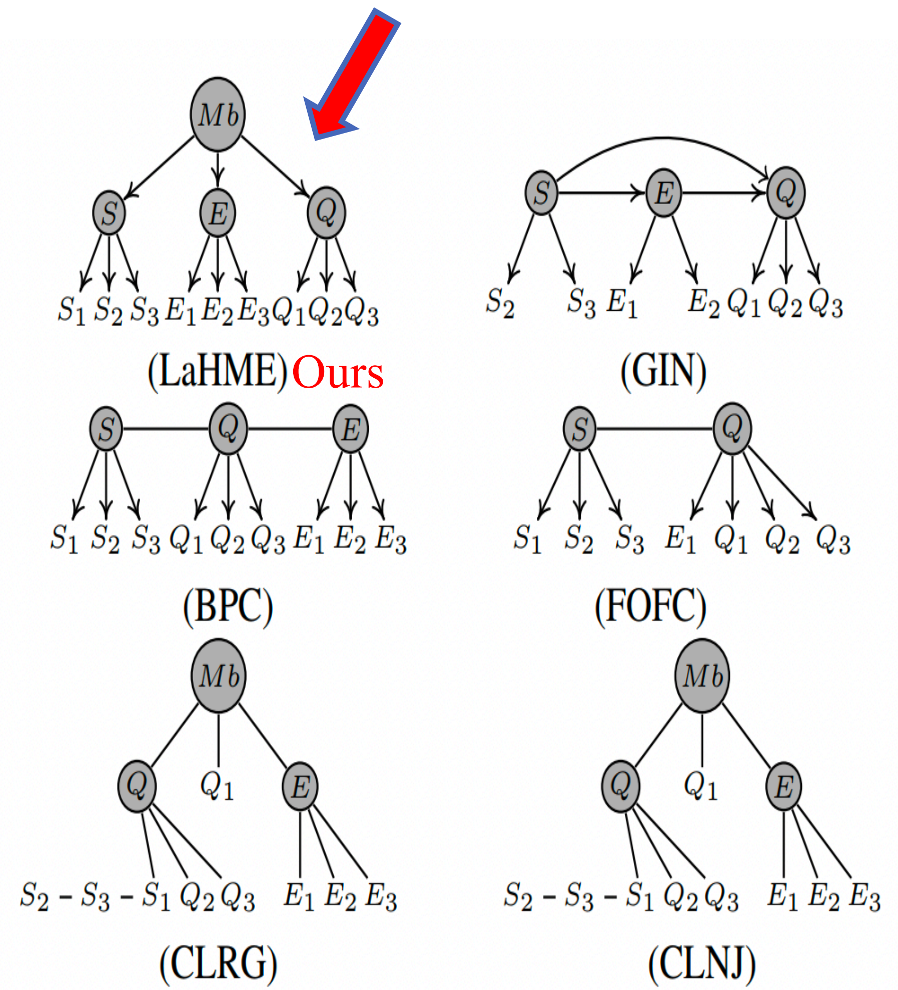
- Consistent with the hypothesized model given in Himi et al., 2019



Application to multitasking behavior Data

- The data set consists of 202 samples

Latent Factors	Children (Indicators)
Speed (S)	Correctly marked Numbers (S1), Correctly marked Latters (S2), and Correctly marked Figures (S3)
Error (E)	Errors marking Numbers (E1), Errors marking Latters (E2), and Errors marking Figures (E3)
Question (Q)	Correctly answered Questions Par.1 (Q1), Correctly answered Questions Par.2 (Q2), and Correctly answered Questions Par.3 (Q3)
Multitasking behavior (Mb)	Speed, Error, and Question



- Consistent with the hypothesized model given in Himi et al., 2019

Conclusion

- Essential to learn linear latent hierarchical structure
- Provide sufficient conditions for structural identifiability
- Future work: n-factor model, nonlinear hierarchical structure…

Thank you for your attention!