A Framework of Directed Acyclic Hypergraph Learning

Zhiyuan Dong, Carlos Mundo, Dr. Gonzalo Arce, Dr. Wei Qian, Dr. Daniel Lau

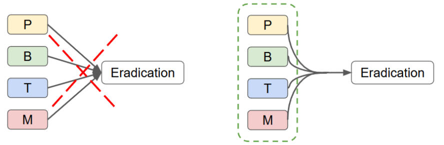
University of Delaware

June 10, 2026

Overview

1. **Motivation** — Higher-order causal interactions in real-world systems
2. **Background** — Structural Causal Models and linear SEM-based DAG learning
3. **Proposed Framework** — Directed Acyclic Hypergraph (DAHG) and multilinear SEM
4. **Acyclicity Constraints** — Naïve tensor collapse vs. t-product–based constraint
5. **Preliminary Results** — Synthetic and Sachs protein network experiments
6. **Future Directions**

Real-World Motivation: *H. pylori* and Combination Therapy



H. pylori colonizes the stomach lining, causing ulcers and gastric cancer risk.

Eradication requires all four drugs jointly:

- PPI + Bismuth + Tetracycline + Metronidazole
- No single agent suffices — outcome depends on **joint action**.

Why DAHG? Classical DAGs model only pairwise effects. The true mechanism is a **multi-cause hyperedge**:

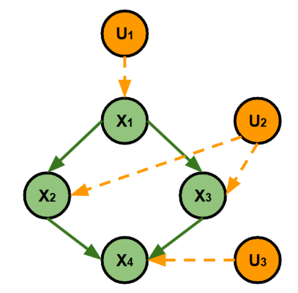
$$\{\text{PPI}, \text{ Bis.}, \text{ Tet.}, \text{ Met.}\} \longrightarrow \text{Eradication}$$

Also: gene regulation, neural circuits, supply-chain co-shocks.

The General Definition for a Causal Structure

Structural Causal Model (SCM):

- $\mathbf{U}$: latent variables determined outside the model
- $\mathbf{X}$: n observed variables determined by others within the model
- $\mathcal{F}$: structural functions encoding direct causal mechanisms

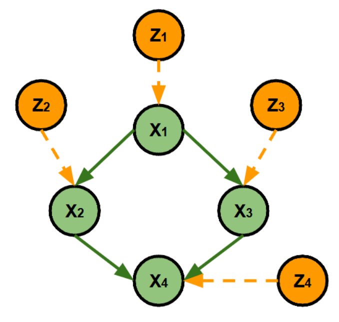


$$X_1 = f_{X_1}(U_1)$$
$$X_2 = f_{X_2}(X_1, U_2)$$
$$X_3 = f_{X_3}(X_1, U_2)$$
$$X_4 = f_{X_4}(X_2, X_3, U_3)$$

Linear SEM for DAG Learning

$A \in \mathbb{R}^{n\times n}$: weighted adjacency matrix; $X \in \mathbb{R}^{n\times d}$: observed data.

$$X = A^\top X + Z \implies X = (I - A^\top)^{-1} Z$$



$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a_{1,2} & 1 & 0 & 0 \\ -a_{1,3} & 0 & 1 & 0 \\ 0 & -a_{2,4} & -a_{3,4} & 1 \end{bmatrix}^{-1} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix}$$

Learning a DAG via Linear Structural Equation Models

Objective

Data $X \in \mathbb{R}^{n \times d}$ satisfies $X \approx A^\top X$.

$$\min_A \frac{1}{2n} \|X - A^\top X\|_F^2 + \lambda \|A\|_1$$

This alone does *not* guarantee a DAG — an **acyclicity constraint** $h(A) = 0$ must be imposed.

Acyclicity Constraints

1. NOTEARS

$$h_{\mathsf{NT}}(A) = \mathrm{tr}\left(e^{A \circ A}\right) - n = 0$$

2. DAGMA

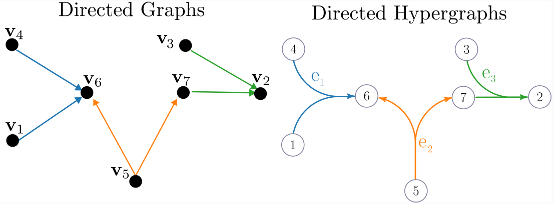
$$h(s, W) = n \log s - \log\det(sI - W \circ W) = 0$$

$s > \rho(W \circ W)$. Under $W \geq 0$ simplifies to $n \log s - \log\det(sI - W) = 0$ (Rey et al.).

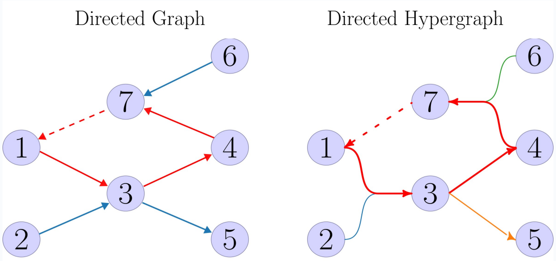
[1] Zheng et al., NeurIPS 2018 [2] Bello et al., NeurIPS 2022 [3] Rey et al., ICASSP 2025

Extending the Framework to Directed Acyclic Hypergraphs

1. High Order SEM



2. Acyclicity Constraint for DAHGs



Multilinear Structural Equation Model

For node v_i:

$$x_i = \sum_{(j_2,\ldots,j_M)\in\mathcal{T}_i} \mathcal{W}_{i\leftarrow j_2,\ldots,j_M} \prod_{m=2}^{M} x_{j_m} + z_i$$

- $\mathcal{T}_i$: incoming tail-tuples to node i
- $\mathcal{W}_{i\leftarrow j_2,\ldots,j_M}$: joint causal strength of $\{v_{j_2},\ldots,v_{j_M}\}$ on v_i
- z_i: independent exogenous noise

Extend the linear SEM to **higher-order interactions** over hyperedges.

Adjacency Tensor $\mathcal{A} \in \mathbb{R}^{N^M}$, M-th Order

M = max hyperarc size; $\mathcal{A}$ has M indices, first = head.

For B-hyperarc $e_i = (T(e_i), \{v_k\})$, fill all $(M-1)$-permutations of $T(e_i)$:

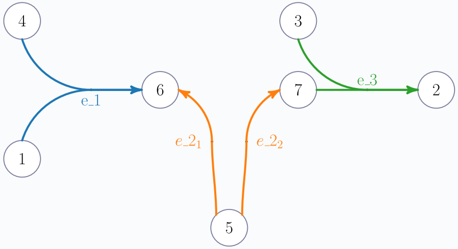
$$a_{k,\,p_2,\dots,p_M} = \frac{|T(e_i)|}{\alpha_i}$$

α_i: # ordered $(M-1)$-permutations of $T(e_i)$

SEM weight entry:

$$\mathcal{W}_{k,p_2,\dots,p_M} = w_{k\leftarrow T(e_i)} \cdot a_{k,p_2,\dots,p_M}$$

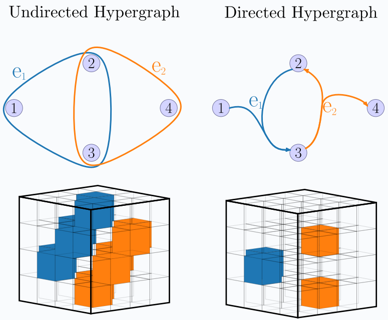
Gallo et al. (2022) — uniform DHG only; **our work**: non-uniform, arbitrary M.



$M = 3$: $\mathcal{A} \in \mathbb{R}^{N\times N\times N}$, indices (i_1, i_2, i_3), i_1=head, standard entry $w = a = 1$

$$e_1 : \{v_1, v_4\} \to \{v_6\}:\ a_{6,1,4} = a_{6,4,1} = 1$$
$$e_{2_1} : \{v_5\} \to \{v_6\}:\ a_{6,5,5} = 1$$
$$e_{2_2} : \{v_5\} \to \{v_7\}:\ a_{7,5,5} = 1$$
$$e_3 : \{v_3, v_7\} \to \{v_2\}:\ a_{2,3,7} = a_{2,7,3} = 1$$

Adjacency Tensor and t-Product Formulation



Hyperedge weights stored in an order-M adjacency tensor:

$$\mathcal{W} \in \mathbb{R}^{N \times N \times \cdots \times N} \quad (M \text{ modes})$$

$$\mathcal{W}_{k,p_2,\ldots,p_M} = w_{k \leftarrow T(e)} \cdot a_{k,p_2,\ldots,p_M}$$

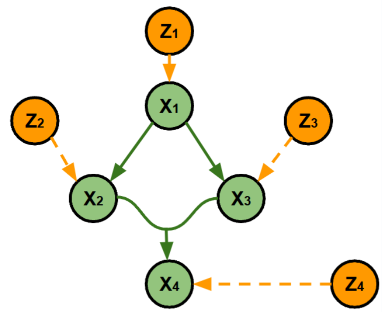
Cross-Interaction Feature Tensor:

$$\mathcal{X} = \sum_d (x^{(d)} \circ \cdots \circ x^{(d)}) \in \mathbb{R}^{N \times D \times N^{M-2}}$$

SEM via Hypergraph Representation:

$$\mathbf{X} = (\mathcal{W} * \mathcal{X})^{(1)} = \sum_{k=1}^{N^{M-2}} \mathcal{W}^{[k]} \mathcal{X}^{[k]}$$

Second-Order Interaction Case ($M = 3$)



Hyperedges: $X_1 \to X_2$, $X_1 \to X_3$, $\{X_2, X_3\} \to X_4$

$$X_i = \sum_{j,k} \mathcal{W}_{ijk} X_j X_k + Z_i$$

Each hyperedge has a *single head node* (B-hypergraph).

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} Z_1 \\ w_{1\to 2} X_1^2 + Z_2 \\ w_{1\to 3} X_1^2 + Z_3 \\ w_{\{2,3\}\to 4} X_2 X_3 + Z_4 \end{bmatrix}$$

Naïve Tensor Collapse for Acyclicity

Collapse to a frontal slice sum:

$$\bar{W} \equiv \sum_{k=1}^{n} \mathcal{W}_{::k} \in \mathbb{R}^{n \times n}, \qquad h(\bar{W}) = \mathrm{tr}\big(\exp(\bar{W} \circ \bar{W})\big) - n = 0$$

Remark (tail symmetry)

If $\mathcal{W}_{ijk} = \mathcal{W}_{ikj}$, collapsing along any tail axis is redundant: one frontal-slice sum suffices before applying the acyclicity constraint.

Limitations of Naïve Tensor Collapsing

Issue 1: Loss of Higher-Order Structure

$$\bar{W} = \sum_{k=1}^{n} \mathcal{W}_{::k}$$

mixes all hyperedge modes into one matrix, discarding slice-specific interactions.

Issue 2: Sign Cancellation
If $\mathcal{W}^{[k_1]} > 0$ and $\mathcal{W}^{[k_2]} < 0$,

$$\sum_{k} \mathcal{W}^{[k]} \approx 0$$

even though each slice encodes a valid higher-order hyperpath.

Intuition of the t-Product for Adjacency Tensors

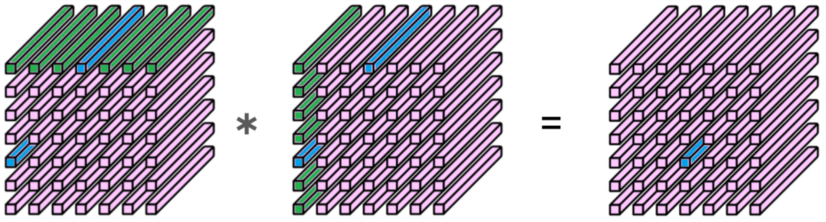
Matrix case: W^i counts i-step paths in a DAG.
Hypergraph case: The natural analogue of A^2 via t-product:

$$\mathcal{W} * \mathcal{W} \quad \Longleftrightarrow \quad \text{all 2-step directed hyperpaths}$$

For each pair (i, j):

$$(\mathcal{W} * \mathcal{W})_{ij:} = \sum_{k=1}^{n} \mathcal{W}_{ik:}\, \mathcal{W}_{kj:}$$

accumulates all 2-step hyperpaths $i \leftarrow k \leftarrow j$ across third-mode slices.



Using the Fourier Property of the t-Product for DAHG Acyclicity

The t-product is diagonalized by the FFT along the third mode:

$$\widehat{\mathcal{W}}^{(i)} = \text{fft}\left(\mathcal{W}\right)_{::i}, \qquad i = 1, \ldots, n_s$$

Each Fourier slice behaves as an **independent adjacency matrix** at frequency i. **Tensor DAGMA constraint** (sum over frequency slices):

$$\sum_{i=1}^{n_s} h\left(s, \widehat{\mathcal{W}}^{(i)}\right) = \sum_{i=1}^{n_s} \left(n \log s - \log\det\left(sI - \hat{\mathbf{W}}^{(i)} \circ \bar{\hat{\mathbf{W}}}^{(i)}\right)\right)$$

Equivalent tensor-level expression:

$$= n\, n_s \log s - \log\det\left(s\mathcal{I} - \mathcal{W}^{\diamond}\right) \equiv h(s, \mathcal{W})$$

$\mathcal{W}^{\diamond}$: self-product of $\mathcal{W}$

Structure Learning Algorithm for DAHG

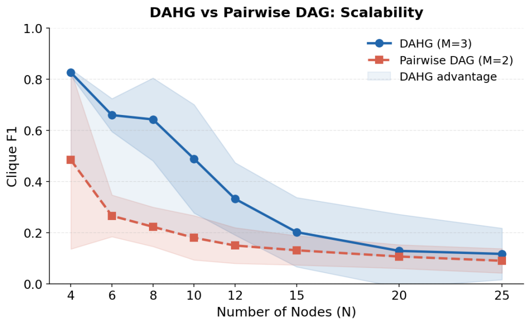
Goal: Estimate $\mathcal{W}$ such that $X \approx (\mathcal{W} * \mathcal{X})^{(1)}$ with DAHG acyclicity.
Objective:

$$\min_{\mathcal{W}} \left\| X - (\mathcal{W} * \mathcal{X})^{(1)} \right\|_F^2 + \lambda \|\mathcal{W}\|_1 \quad \text{s.t.} \quad h(s, \mathcal{W}) = 0$$

Solved via augmented Lagrangian with gradient-based optimizer.
$\|\mathcal{W}\|_1 = \sum_{ijk} \left| w_{ijk} \right|$: entrywise ℓ_1 norm, which promotes sparsity in $\mathcal{W}$.

Synthetic Experiment: DAHG vs. Pairwise DAG Scalability

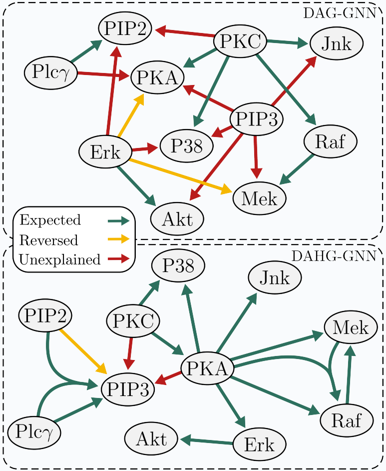
Setup. Random DAHGs via topological ordering; weights in $[0.5, 2.0]$, i.i.d. Gaussian noise. Metric: Clique F1.



Observations.

- DAHG ($M=3$) consistently outperforms pairwise baseline ($M=2$).
- Gap is largest at small N, where higher-order signal is identifiable.
- Both degrade as N grows — combinatorial difficulty of hyperedge recovery.

Real-Data Experiment: Sachs Protein Signaling Network



Dataset. Sachs et al. (Science, 2005). Ground truth: 17 refined edges (Kleinegesse et al., 2022).

Results.

- DAHG: **13/17** edges vs. DAG-GNN: 8/17
- Extra hyperedges recovered:
  - $\{\text{PKA, Mek}\} \to \text{Raf}$
  - $\{\text{PIP2, Plc}\gamma\} \to \text{PIP3}$
- Both are biologically validated.

Takeaway. DAHG recovers joint multi-protein mechanisms invisible to pairwise DAG models.

Future Work

- **Noise Modeling & Latent Confounders.** Extend to heterogeneous noise distributions and hidden confounders within the hypergraph causal framework.
- **Scalability.** Candidate hyperedges grow combinatorially with N and M; efficient pruning and greedy expansion strategies are needed.
- **Nonlinear Hypergraph SEMs.** Extend beyond multilinear interactions to kernel or neural function classes over hyperedges.

Thank You!

Questions welcome.

Zhiyuan Dong ▪ Dongz@udel.edu