# Attributed Hypergraph Generation with Realistic Interplay Between Structure and Attributes

ICDM 2025
November 12–15, 2025
Washington DC, USA
IEEE International Conference on Data Mining

**Jaewan Chun*** KAIST jjwpalace@kaist.ac.kr
**Seokbum Yoon*** KAIST jing9044@kaist.ac.kr
**Minyoung Choe** KAIST minyoung.choe@kaist.ac.kr
**Geon Lee** KAIST geonlee0325@kaist.ac.kr
**Kijung Shin** KAIST kijungs@kaist.ac.kr

**Best Paper Award**

## Summary

- **Motivation**
  - Real-world group interactions are naturally modeled as a hypergraph.
  - Node attributes may influence hyperedge formation (e.g. homophily).
  - However, most hypergraph generators overlook node attributes.
- **Proposed Algorithm: NoAH**
  - We propose **NoAH**, a node-attribute-based hypergraph generative model that reproduces the *interplay between structure and attributes* observed in real-world hypergraphs.
  - **NoAHFit** tunes its parameters to generate hypergraphs resembling a target hypergraph.
- **Extensive Experiments:**
  - **Effective**: NoAH outperforms existing generators in modeling structure-attribute interplay.
  - **Scalable**: NoAH and NoAHFit are scalable with the numbers of hyperedges and attributes.

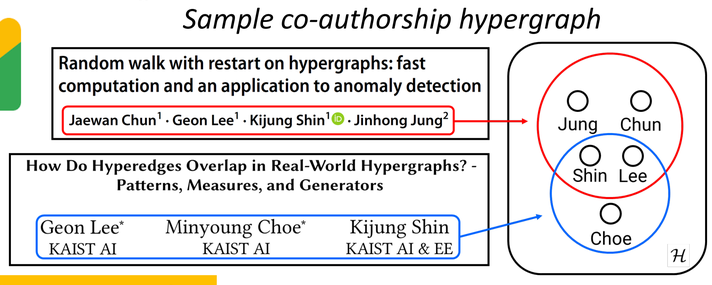
## Introduction

- **Hypergraph**
  - **Hypergraphs** model group interactions among individuals or objects.

*Co-authorship* *Group chat* *Email*

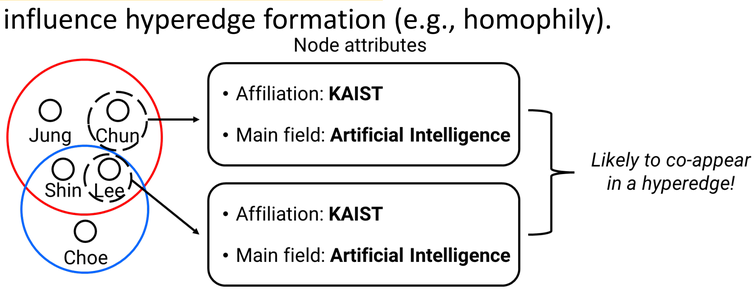
*Examples of real-world group interactions*

*Sample co-authorship hypergraph*




- **Node Attributes and Hyperedge Formation**
  - **Node attributes** can influence hyperedge formation (e.g., homophily).

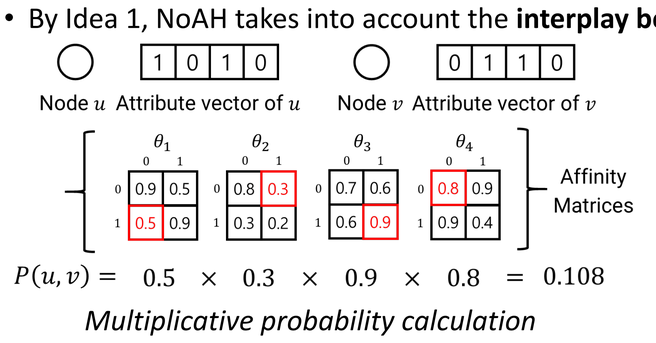
*E.g. co-authorship hypergraph*




- **Hypergraph Generative Models**
  - **Hypergraph generative models** aim to generate realistic hypergraphs with systematic hyperedge formation mechanisms.
    - Enables understanding of underlying mechanism of hyperedge formation.
    - Applications: community detection, hyperedge prediction, and pattern discovery.
  - Most of them **overlook node attributes,** having attribute-independent mechanisms.

## Proposed Hypergraph Generator: NoAH

- **Overview of NoAH**
  - We propose **NoAH**: **No**de **A**ttribute based **H**ypergraph Generator.
  - NoAH aims to capture the **structure-attribute interplay** observed in real-world hypergraphs.
  - NoAH is built upon three key ideas below.
- **Idea 1: Node Attribute based Probability Design**
  - NoAH models hyperedge formation probability based on node attributes.
  - Specifically, edge formation probability is computed as the product of **affinity matrices**, where each entry represents the contribution of an attribute value pair.
  - By Idea 1, NoAH takes into account the **interplay between structure and attributes**.




*Multiplicative probability calculation*

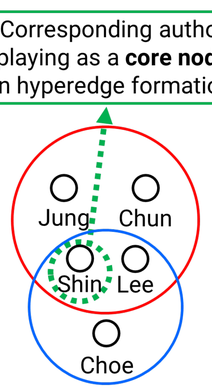
Connection probability (**Eq. (1)**):

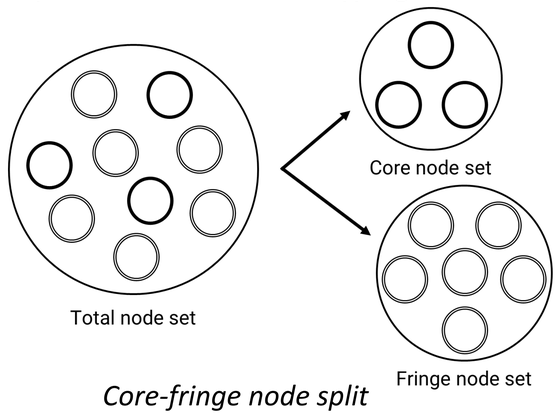
$$P(u,v) = \prod_{l=1}^{k} \Theta_l \left[\mathbf{x}_u^{(l)}, \mathbf{x}_v^{(l)}\right]$$

- $k$: the dimension of node attributes
- $\mathbf{x}_u$: attribute vector of node $u$

- **Idea 2: Core-fringe Node Hierarchy**
  - NoAH splits nodes into core (i.e. central) and fringe (i.e. peripheral) nodes.
  - By Idea 2, NoAH captures **hierarchical structure** prevalent in real-world hypergraphs.

*Core-fringe node hierarchy example in co-authorship hypergraph*

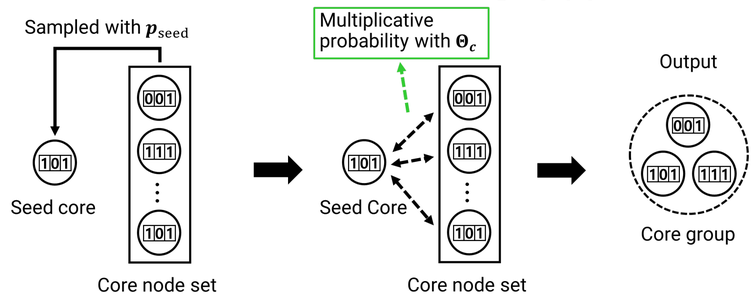


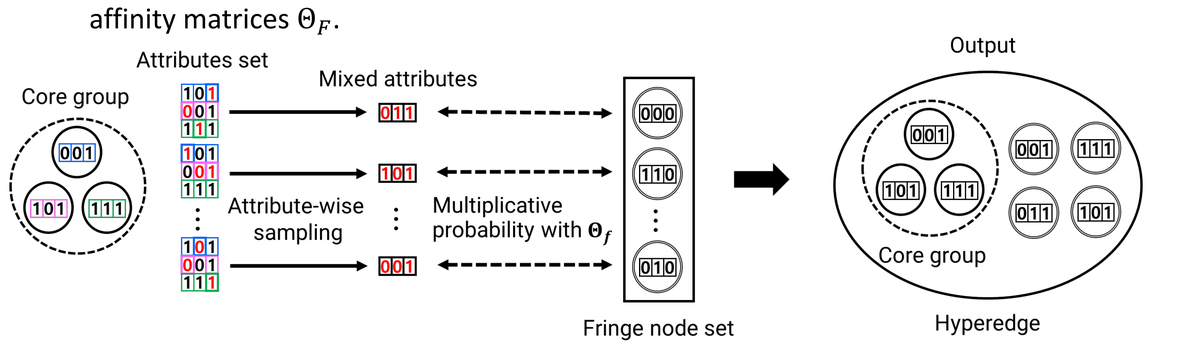

*Core-fringe node split*

- **Idea 3: Hyperedge Formation via Series of Attachments**
  - NoAH models formation of each hyperedge as a series of attachments: (1) core group construction, and (2) fringe attachment
  - By Idea 3, NoAH reduces search space for hyperedge construction, achieving tractability.

- **Details of Hyperedge Formation by NoAH**
  - **1. Core group construction**
    - 1-1. Sample a seed core from core nodes based on seed core probability $\boldsymbol{p}_{\text{seed}}$.
    - 1-2. Attach other core nodes to the seed core using **Eq. (1)** with core affinity matrices $\Theta_C$.




  - **2. Fringe attachment**
    - 2-1. Generate representative attribute of the core group by attribute-wise sampling.
    - 2-2. Attach fringe nodes to the core group using **Eq. (1)** with mixed attribute and fringe affinity matrices $\Theta_F$.




## NoAHFit: Parameter Fitting Algorithm for NoAH

- **Overview of NoAHFit**
  - The formation probability of each hyperedge in NoAH can be parameterized through:
    - $\boldsymbol{p}_{\text{seed}}$: seed core probability
    - $\Theta_C, \Theta_F$: core and fringe affinity matrices
  - **NoAHFit** fits these parameters to generate hypergraphs resembling a given hypergraph.
  - NoAHFit aims to maximize the likelihood of the given hypergraph (i.e., the product of the likelihood of each hyperedge) and employs gradient descent.

## Experimental Results

- **Reproduction of Structure-attribute Interplay**
  - We evaluate the **structure-attribute interplay** reproduced by hypergraph generators.
  - 10 hypergraph generators, including:
    - **NoAH** (fitted by NoAHFit) & **NoAH-CF** (a variant without core-fringe node hierarchy)
  - 9 datasets from 4 different domains: academic paper, contact, review, and online Q&A.
  - 3 complementary metrics:
    - Type-$s$ affinity ratio scores (**T2, T3, T4**): fine-grained patterns in hyperedge attribute dist.
    - Hyperedge entropy (**HE, HOHE**): coarse-grained patterns in hyperedge attribute dist.
    - Node homophily score (**NHS**): node-level patterns of attribute dist.

Rank 1 Rank 2 Rank 3 *A.R.: average ranking

| | T2 | T3 | T4 | HE | HOHE | NHS | A.R. |
|---|---|---|---|---|---|---|---|
| HyperCL | 27.3 | 53.0 | 63.6 | 1.016 | 0.382 | 1.053 | 6.8 |
| HyperPA | 27.3 | 55.4 | 71.2 | 1.154 | 0.527 | 1.096 | 8.8 |
| HyperFF | 24.1 | 54.3 | 60.5 | 0.449 | 0.299 | 1.055 | 5.2 |
| HyperLAP | 27.3 | 52.4 | 68.3 | 1.026 | 0.361 | 1.042 | 6.3 |
| hyper dK-series | 31.4 | 52.4 | 61.6 | 1.249 | 0.450 | 1.026 | 7.3 |
| THera | 26.0 | 50.6 | 67.0 | 0.976 | 0.394 | 1.003 | 5.0 |
| HyCoSBM | 11.8 | 57.9 | 72.1 | 0.306 | 0.371 | 0.900 | 4.7 |
| HyRec | 25.3 | 50.6 | 61.8 | 1.138 | 0.402 | 0.982 | 5.3 |
| NoAH | 21.0 | 47.8 | 55.1 | 0.275 | 0.394 | 0.229 | 1.8 |
| NoAH-CF | 21.8 | 49.7 | 58.0 | 0.363 | 1.188 | 0.402 | 3.7 |

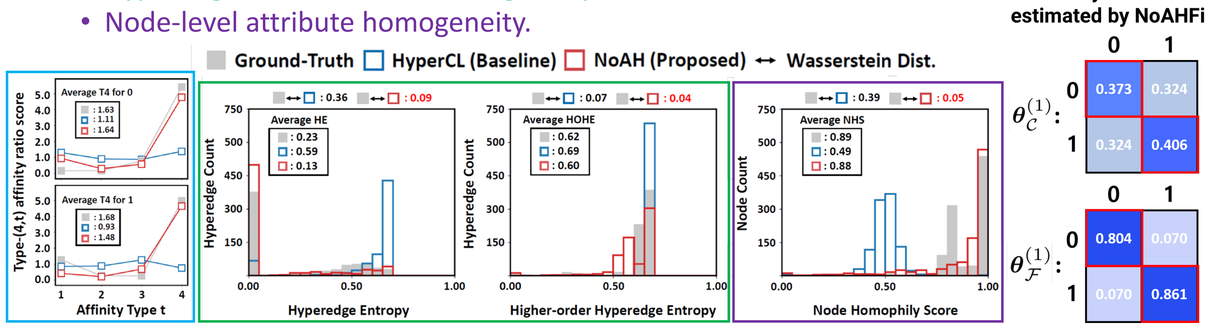
(e) Amazon Music (NoAH ranks **first** overall)

| | T2 | T3 | T4 | HE | HOHE | NHS | A.R. |
|---|---|---|---|---|---|---|---|
| HyperCL | 6.9 | 5.7 | 5.0 | 6.6 | 5.3 | 6.6 | 6.2 |
| HyperPA | 7.8 | 8.6 | 7.4 | 7.7 | 8.0 | 7.2 | 9.7 |
| HyperFF | 5.2 | 6.3 | 6.2 | 5.1 | 5.4 | 6.1 | 5.7 |
| HyperLAP | 5.3 | 5.3 | 5.3 | 6.2 | 2.8 | 5.6 | 4.7 |
| hyper dK-series | 6.1 | 3.9 | 4.7 | 4.3 | 5.4 | 4.6 | 3.8 |
| THera | 4.9 | 4.0 | 5.8 | 5.2 | 4.0 | 5.0 | 3.8 |
| HyCoSBM | 1.2 | 4.8 | 6.4 | 4.1 | 6.0 | 6.7 | 5.3 |
| HyRec | 7.7 | 7.1 | 3.9 | 5.6 | 5.1 | 4.0 | 5.0 |
| NoAH | 2.9 | 2.1 | 3.9 | 3.4 | 4.4 | 1.4 | 1.5 |
| NoAH-CF | 7.0 | 7.2 | 6.3 | 6.8 | 8.4 | 7.9 | 9.0 |

(j) Average Rank over Nine Datasets (NoAH ranks **first** overall)

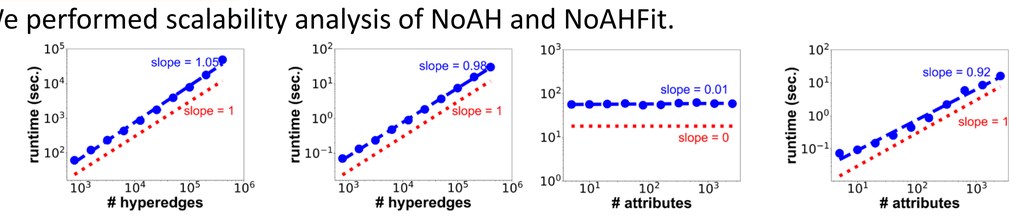
***NoAH reproduces most realistic structure-attribute interplay***

- **Case Study**
  - The interplay between the structure and an attribute is closely examined in the **(1) original hypergraph** (Amazon Music), hypergraphs generated by **(2) HyperCL**, and **(3) NoAH**.
  - Each results below indicate NoAH successfully reproduces:
    - Detailed attribute distributions in size-4 hyperedges.
    - Hyperedge-level attribute homogeneity.
    - Node-level attribute homogeneity.




***NoAH models the characteristic of each attribute (here, homophily) through affinity matrices***

- **Scalability Analysis**
  - We performed scalability analysis of NoAH and NoAHFit.




(a) NoAHFit (fitting) w.r.t. number of hyperedge (b) NoAH (generation) w.r.t. number of hyperedge (c) NoAHFit (fitting) w.r.t. number of attributes (d) NoAH (generation) w.r.t. number of attributes

***Both NoAH and NoAHFit are scalable w.r.t. the number of hyperedges and attributes***

DM540