Attributed Hypergraph Generation with Realistic Interplay Between Structure and Attributes (Extended Abstract)*

Jaewan Chun[1] , **Seokbum Yoon**[1] , **Minyoung Choe**[1] , **Geon Lee**[1] and **Kijung Shin**[1]
[1]KAIST
{jjwpalace, jing9044, minyoung.choe, geonlee0325, kijungs}@kaist.ac.kr

Abstract

In many real-world scenarios, interactions happen in a group-wise manner with multiple entities, and therefore, hypergraphs are a suitable tool to accurately represent such interactions. Various generative models have been proposed to explore fundamental mechanisms underlying hyperedge formation. However, most existing models do not account for node attributes, which can play a significant role in hyperedge formation. As a result, they fail to capture the interactions between structure and node attributes. To address this issue, we propose NOAH, a stochastic hypergraph generative model for attributed hypergraphs. NOAH utilizes the core–fringe node hierarchy to model hyperedge formation as a series of node attachments and determines attachment probabilities based on node attributes. We further introduce NOAHFIT, a parameter learning procedure for fitting NOAH to a given real-world hypergraph. Through experiments, we show that NOAH with NOAHFIT more accurately reproduces the structure–attribute interplay than baseline hypergraph generative models. Our code, supplementary materials, and datasets are provided at https://github.com/jaewan01/NoAH.

1 Introduction

Many real-world interactions occur in groups, such as co-authorship among researchers and group discussions on online Q&A sites. Hypergraphs, which consist of hyperedges, naturally represent group interactions involving an arbitrary number of individuals or entities. Hypergraph modeling has shown effectiveness in applications such as clustering [Kumar *et al.*, 2020; Hayashi *et al.*, 2020], classification [Yu *et al.*, 2012], and anomaly detection [Chun *et al.*, 2024].

Real-world hypergraphs exhibit systematic structural patterns, such as high-degree nodes [Do *et al.*, 2020], densely overlapping hyperedges [Lee *et al.*, 2021], and high transitivity [Kim *et al.*, 2023]. Building upon these findings, a number of hypergraph generative models have been proposed to reproduce realistic hypergraph structure. These models help explain real-world hypergraphs, and they are also employed in various data-mining applications, including community detection [Ruggeri *et al.*, 2023; Badalyan *et al.*, 2024] and hyperedge prediction [Contisciani *et al.*, 2022].

Despite their success, most hypergraph generative models overlook the interplay between hypergraph structure and node attributes. Node attributes can offer valuable information such as affiliation and field of study in co-authorship hypergraphs, where nodes represent authors and hyperedges represent co-authored publications. Moreover, as exemplified by homophily [McPherson *et al.*, 2001], such node attributes can influence the formation of hyperedges.

Thus, we propose NOAH (**No**de **A**ttribute based **H**ypergraph generator), a novel hypergraph generative model based on node attributes. To avoid considering exponentially many hyperedge candidates, NOAH forms each hyperedge through sequential attachments of nodes to its seed node(s). Attachment probabilities are computed by multiplying affinity scores derived from the values of each node attribute. In addition, NOAH incorporates a core–fringe node hierarchy to enhance realism. We also introduce NOAHFIT, which learns the parameters of NOAH to capture the structure-attribute interplay in a given hypergraph.

Through experiments on nine real-world hypergraphs using six structure-attribute measures, we show that NOAH with NOAHFIT outperforms eight existing hypergraph generative models in the overall assessment across the measures.

2 Related Work and Preliminaries

We review related work and introduce the notation used throughout the paper.

Related Work. (Hyper)graph generative models aim to reproduce structural patterns observed in real-world (hyper)graphs [Chakrabarti and Faloutsos, 2006; Lee *et al.*, 2025], often to uncover mechanisms behind their formation. Several models exploit node hierarchies [Kim *et al.*, 2023; Papachristou and Kleinberg, 2022]; our model follows this line through a simple yet effective core–fringe structure.

Recently, various deep learning based (hyper)graph generative models, including variational autoencoders [Simonovsky and Komodakis, 2018], generative adversarial networks [Bojchevski *et al.*, 2018; De Cao and Kipf, 2018], and

*This is an abridged version of a paper that won the Best Paper Award at ICDM 2025 [Chun *et al.*, 2025]

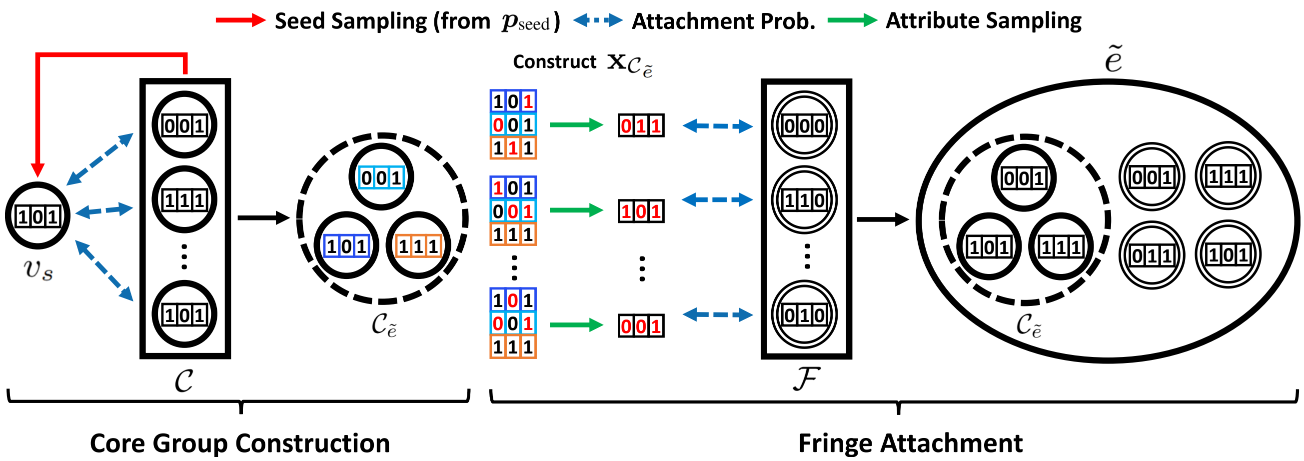

Figure 1: NoAH generates each hyperedge in two steps: (1) Core group construction: sample a seed core node v_s according to $\boldsymbol{p}_{\text{seed}}$ and attach additional core nodes to v_s to form a core group $\mathcal{C}_{\tilde{e}}$, and (2) Fringe attachment: attach fringe nodes to core group $\mathcal{C}_{\tilde{e}}$ based on mixed attribute vector $\boldsymbol{x}_{\mathcal{C}_{\tilde{e}}}$ sampled attribute-wise from $\mathbf{X}_{\mathcal{C}_{\tilde{e}}}$. The attached core and fringe nodes, together with the seed node, form a hyperedge.

diffusion models [Liu *et al.*, 2023], have also been proposed. However, these models typically require a large collection of graph instances for training.

For graphs, several generative models incorporate node attributes, which can play a crucial role in graph generation, as demonstrated by homophily [McPherson *et al.*, 2001]. Examples include the exponential random graph model [Robins *et al.*, 2007], the stochastic block model [Airoldi *et al.*, 2008], and the latent space model [Wang *et al.*, 2023]. Among them, the Multiplicative Attribute Graph (MAG) model [Kim and Leskovec, 2012] is especially relevant, as our attribute-affinity formulation is inspired by its multiplicative edge probability.

In contrast, very few hypergraph generative models incorporate node attributes. One exception is a stochastic block model variant [Badalyan *et al.*, 2024], where attributes are conditionally independent of the hypergraph structure given latent communities and thus do not directly drive hyperedge formation. Unlike these models, NoAH forms hyperedges explicitly based on attribute relationships among nodes, extending the attribute-driven perspective of MAG from pairwise edges to group-wise interactions.

Preliminaries. An *attributed hypergraph* $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ consists of a set of nodes $\mathcal{V} = \{v_1, \cdots, v_{|\mathcal{V}|}\}$, a set of hyperedges $\mathcal{E} = \{e_1, \cdots, e_{|\mathcal{E}|}\}$, and a node attribute matrix $\mathbf{X} \in \{0,1\}^{|\mathcal{V}|\times k}$. Each hyperedge $e \in \mathcal{E}$ is a non-empty subset of nodes, and $\mathbf{x}_i^{(l)}$ denotes the l-th attribute value of node v_i. In this work, we assume binary node attributes, which simplifies both model design and implementation while remaining valid for many real-world datasets. Moreover, categorical and continuous attributes can also be converted into binary ones via one-hot encoding and thresholding, respectively, as in our experiments.

3 Proposed Method

We introduce NoAH, a node attribute-based hypergraph generator, and NoAHFit, its parameter fitting algorithm. The generation process is illustrated in Figure 1.

3.1 Proposed Generation Method: NoAH

NoAH generates a hypergraph by stochastically sampling nodes to attach to each hyperedge. The node set $\mathcal{V}$ is partitioned into two disjoint subsets: the core node set $\mathcal{C} \subseteq \mathcal{V}$ and the fringe node set $\mathcal{F} \subseteq \mathcal{V}$, s.t. $\mathcal{C} \cup \mathcal{F} = \mathcal{V}$. NoAH generates each hyperedge in a two-step process. *First*, a subset of core nodes form a core group. *Second*, fringe nodes attach to the core group according to attribute affinity with the group.

Step 1. Core Group Construction. NoAH initiates hyperedge construction by forming a subset of core nodes $\mathcal{C}_{\tilde{e}} \subseteq \mathcal{C}$, which plays a role as the *structural nucleus* of the hyperedge. Specifically, it first samples a *seed core node* $v_s \in \mathcal{C}$ from the seed probability distribution $\boldsymbol{p}_{\text{seed}} \in [0,1]^{|\mathcal{C}|}$. Each remaining core node $v_c \in \mathcal{C} \setminus \{v_s\}$ is then added based on its attribute affinity with the seed node v_s, computed as:

$$P_{\mathcal{C}}(v_c|v_s, \Theta_{\mathcal{C}}) = \prod_{l=1}^{k} \boldsymbol{\theta}_{\mathcal{C}}^{(l)}[\mathbf{x}_s^{(l)}, \mathbf{x}_c^{(l)}], \qquad (1)$$

where $\Theta_{\mathcal{C}} = \{\boldsymbol{\theta}_{\mathcal{C}}^{(1)}, \ldots, \boldsymbol{\theta}_{\mathcal{C}}^{(k)}\}$ denotes core attribute affinity matrices, and each $\boldsymbol{\theta}_{\mathcal{C}}^{(l)} \in \mathbb{R}^{2\times 2}$ captures the affinity between binary values of the l-th attribute.

Step 2. Fringe Attachment. Given the core group $\mathcal{C}_{\tilde{e}}$, NoAH constructs a binary attribute vector $\mathbf{x}_{\mathcal{C}_{\tilde{e}}} \in \{0,1\}^k$. Each l-th attribute $\mathbf{x}_{\mathcal{C}_{\tilde{e}}}^{(l)}$ is independently sampled as:

$$\mathbf{x}_{\mathcal{C}_{\tilde{e}}}^{(l)} \sim \text{Bernoulli}\left(\frac{1}{|\mathcal{C}_{\tilde{e}}|}\sum_{v_i \in \mathcal{C}_{\tilde{e}}} \mathbf{x}_i^{(l)}\right). \qquad (2)$$

This samples each attribute according to its average presence among core-group members. Each fringe node $v_f \in \mathcal{F}$ is then attached with probability:

$$P_{\mathcal{F}}(v_f|\mathcal{C}_{\tilde{e}}, \Theta_{\mathcal{F}}) = \prod_{l=1}^{k} \boldsymbol{\theta}_{\mathcal{F}}^{(l)}[\mathbf{x}_{\mathcal{C}_{\tilde{e}}}^{(l)}, \mathbf{x}_f^{(l)}], \qquad (3)$$

where $\Theta_{\mathcal{F}} = \{\boldsymbol{\theta}_{\mathcal{F}}^{(1)}, \ldots, \boldsymbol{\theta}_{\mathcal{F}}^{(k)}\}$ is the set of fringe affinity matrices, and each $\boldsymbol{\theta}_{\mathcal{F}}^{(l)} \in \mathbb{R}^{2\times 2}$ captures the affinity between binary values of the l-th attribute.

	T2	T3	T4	HE	HOHE	NHS	A.R.
HyperCL	6,816	10,702	10,672	19.81	122.02	19.94	6.7
HyperPA	6,871	10,739	10,559	25.90	103.76	22.46	7.7
HyperFF	6,472	10,483	10,677	17.60	66.97	15.23	4.0
HyperLAP	6,757	10,737	10,311	18.33	55.40	19.43	5.0
hyper dK	6,968	10,234	10,154	13.38	102.83	16.09	3.7
THera	6,450	10,498	10,476	18.04	53.58	18.59	3.8
HyCoSBM	6,285	10,613	10,278	19.89	135.30	38.60	6.0
HyRec	6,996	10,840	10,271	20.01	51.37	20.04	6.2
NoAH	5,734	10,157	10,151	24.31	44.26	15.39	2.3
NoAH-CF	9,036	14,550	13,515	48.26	106.32	121.04	9.7

(a) Citeseer (NoAH ranks first overall)

	T2	T3	T4	HE	HOHE	NHS	A.R.
HyperCL	20.4	51.4	106.4	1.180	1.369	1.730	7.2
HyperPA	20.6	52.6	107.8	1.175	1.421	1.654	8.0
HyperFF	20.2	61.6	106.5	1.017	1.290	1.785	7.8
HyperLAP	20.3	51.6	102.0	1.187	0.977	1.714	6.0
hyper dK	20.0	51.4	97.8	0.931	1.429	1.716	5.8
THera	19.9	51.2	99.8	1.166	0.781	1.611	3.5
HyCoSBM	3.9	52.1	97.9	1.819	1.414	1.797	6.7
HyRec	19.8	54.0	92.8	0.740	0.960	1.637	4.3
NoAH	12.2	37.8	89.6	0.628	1.273	1.374	2.0
NoAH-CF	19.2	50.3	103.9	0.682	1.176	1.635	3.7

(b) High School (NoAH ranks first overall)

	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	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

(c) 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	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

Average Rank over Nine Datasets (NoAH ranks first overall)

Table 1: NoAH outperforms baselines in reproducing structure-attribute interplay on three representative datasets and achieves the best average rank over nine datasets. Top three results are highlighted in blue (first), green (second), and yellow (third). A.R.: average rank.

3.2 Proposed Fitting Method: NoAHFit

NoAHFit fits NoAH to a given attributed hypergraph by learning $\boldsymbol{p}_{\text{seed}}$, $\Theta_{\mathcal{C}}$, and $\Theta_{\mathcal{F}}$.

Core and Fringe Partition. Given a hypergraph $\mathcal{H}$, NoAHFit identifies core nodes $\mathcal{C}$ using UMHS [Amburg *et al.*, 2021] and defines fringe nodes as its complement, i.e., $\mathcal{F} = \mathcal{V} \setminus \mathcal{C}$. Each hyperedge e is decomposed into core and fringe subsets, $\mathcal{C}_e = e \cap \mathcal{C}$ and $\mathcal{F}_e = e \cap \mathcal{F}$.

Derivation of the Likelihood of Each Hyperedge. For each hyperedge e, NoAHFit decomposes its likelihood into core group construction and fringe attachment:

$$P(e) = P_{\text{core}}(e) \cdot P_{\text{fringe}}(e). \tag{4}$$

Here, $P_{\text{core}}(e)$ accounts for selecting a seed core and attaching the remaining cores in $\mathcal{C}_e$, while $P_{\text{fringe}}(e)$ accounts for attaching fringe nodes in $\mathcal{F}_e$.

- **Likelihood of Sampling Core Nodes:** Given a hyperedge e, the likelihood $P_{\text{core}}(e)$ of its core group $\mathcal{C}_e$ is:

  $$P_{\text{core}}(e) = \sum\nolimits_{v_s \in \mathcal{C}_e} \boldsymbol{p}_{\text{seed}}(v_s) \cdot P_{\mathcal{C}}(\mathcal{C}_e \setminus \{v_s\} | v_s),$$

  where $P_{\mathcal{C}}(\mathcal{C}_e \setminus \{v_s\} | v_s)$ is the likelihood of sampling the remaining core nodes $\mathcal{C}_e \setminus \{v_s\} \subset \mathcal{C}$ given the seed node v_s.

- **Likelihood of Sampling Fringe Nodes:** Given the core group $\mathcal{C}_e$, the likelihood of its fringe subset $\mathcal{F}_e$ is:

  $$P_{\text{fringe}}(e) = \prod_{v_f \in \mathcal{F}_e} P_{\mathcal{F}}(v_f | \mathcal{C}_e) \cdot \prod_{v_f \in \mathcal{F} \setminus \mathcal{F}_e} (1 - P_{\mathcal{F}}(v_f | \mathcal{C}_e)),$$

  where $P_{\mathcal{F}}(v_f | \mathcal{C}_e)$ is the probability of attaching fringe node v_f to the core group $\mathcal{C}_e$, obtained by marginalizing over the stochastic binary attribute vector $\mathbf{x}_{\mathcal{C}_e}$ of $\mathcal{C}_e$.

Update of the Parameters of NoAH. NoAHFit learns the parameters by minimizing

$$\mathcal{L} = \mathcal{L}_{edge} + w_{deg}\mathcal{L}_{deg} + w_{card}\mathcal{L}_{card}. \tag{5}$$

$\mathcal{L}_{edge}$ is the negative log-likelihood over hyperedges. $\mathcal{L}_{deg}$ and $\mathcal{L}_{card}$ are the MSE losses matching expected degree and cardinality distributions with those of the input hypergraph.

4 Experiments

We present experimental results demonstrating the effectiveness of NoAH and NoAHFit.

4.1 Experimental Settings

Datasets. We use nine real-world hypergraphs from four domains: Citeseer & Cora [Yadati *et al.*, 2019] (Academic Paper), High School [Chodrow *et al.*, 2021] & Workspace [Génois and Barrat, 2018] (Contact), Amazon Music [Ni *et al.*, 2019], Yelp Restaurant & Yelp Bar [Amburg *et al.*, 2020] (Review), Devops & Patents (Online Q&A)[1]. Note that our experiments are done with varying numbers of attributes, scaling from 5 (Workspace) to 3,703 (Citeseer).

Baselines. We consider eight baseline generative models: HyperCL, HyperLAP [Lee *et al.*, 2021], HyperPA [Do *et*

[1]https://archive.org/download/stackexchange

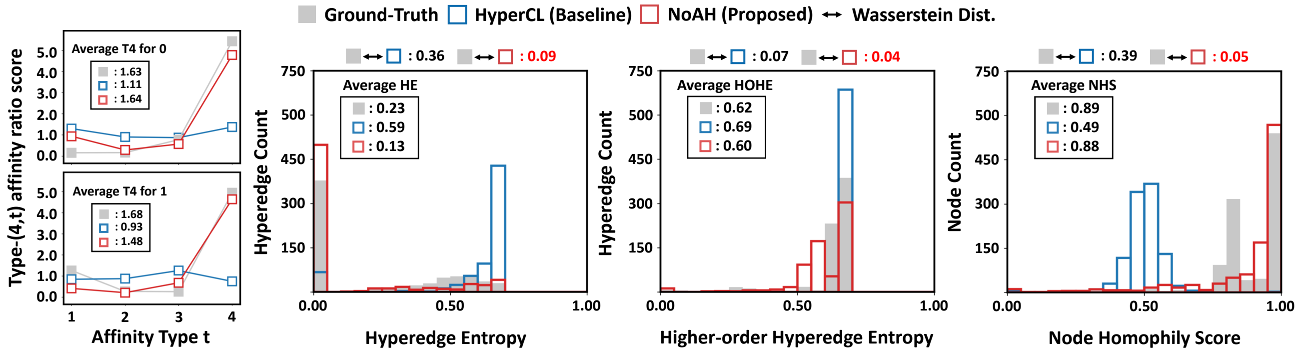

Figure 2: Structure–attribute interplay with respect to the first node attribute in the Amazon Music dataset. Our proposed generative model, NoAH, fitted by NoAHFit, effectively captures the interplay between structure and node attributes, outperforming HyperCL.

al., 2020], HyperFF [Kook *et al.*, 2020], hyper dK [Nakajima *et al.*, 2021], THera [Kim *et al.*, 2023], HyCoSBM [Badalyan *et al.*, 2024], HyRec [Choe *et al.*, 2025]. Among the baselines, HyCoSBM explicitly utilizes node attributes. For HyperPA, HyperFF, THera, and HyRec, where node identities are not preserved during the generation process, we assign node attributes randomly. For identity-preserving baselines, we assign attributes according to the corresponding original nodes.

Evaluation. We compare models in terms of their ability to reproduce the structure-attribute interplay using six metrics. Type-s affinity ratio scores (**T2**, **T3**, **T4**) capture fine-grained attribute compositions within size-s hyperedges, while hyperedge entropy (**HE**), higher-order hyperedge entropy (**HOHE**), and node homophily score (**NHS**) capture attribute homogeneity at the hyperedge and node levels. For all metrics, lower values indicate smaller discrepancies from the ground-truth hypergraph. Detailed descriptions of the metrics are provided in the full paper [Chun *et al.*, 2025].

4.2 Performance Comparison

In Table 1, we report results on three datasets (Citeseer, High School, Amazon Music) from different domains, along with the average rank over all nine datasets. Averaged over nine datasets, NoAH ranks first in four metrics (type-3 affinity score, type-4 affinity score, hyperedge entropy, and node homophily score), second in type-2 affinity score, and third in higher-order hyperedge entropy. NoAH achieves the best average rank on 5 out of 9 datasets. These results demonstrate the overall superiority of NoAH in capturing the structure–attribute interplay of real-world hypergraphs.

4.3 Case Study

We conduct a case study on the Amazon Music dataset, where the first attribute indicates whether a reviewer has reviewed music in the New York Blues genre. As shown in Figure 2, the ground-truth hypergraph exhibits strong homophily: hyperedge entropy is skewed toward 0, node homophily is skewed toward 1, and the type-$(4,4)$ affinity ratio score is high for both attribute values. HyperCL fails to reproduce these patterns, whereas NoAH closely matches them. The learned affinity matrices in Figure 3 further show that NoAHFit captures this homophily by assigning higher affinities to same-value attribute pairs.

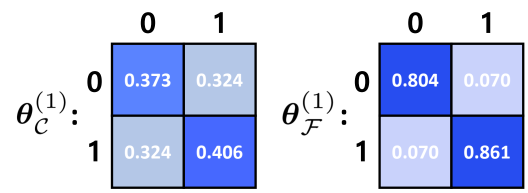

Figure 3: $\theta_{\mathcal{C}}^{(1)}$ and $\theta_{\mathcal{F}}^{(1)}$ estimated by NoAHFit on the Amazon Music dataset. NoAHFit captures homophily by assigning higher affinities to same attribute value pairs ($0 \leftrightarrow 0$ and $1 \leftrightarrow 1$) than to different attribute value pairs ($0 \leftrightarrow 1$).

4.4 Ablation Study

To assess the contribution of the core–fringe node hierarchy in NoAH, we compare its performance with a variant, NoAH-CF, which omits this hierarchical structure. In NoAH-CF, each hyperedge is generated by sampling a seed node from the entire node set and attaching additional nodes directly, without distinguishing between core and fringe roles. As reported in Table 1, NoAH consistently outperforms NoAH-CF across all datasets, showing the importance of the core–fringe hierarchy in realistic hyperedge formation.

5 Conclusions

In this work, we proposed NoAH, a stochastic generative model for attributed hypergraphs that reproduces realistic interplay between structure and node attributes. By leveraging a two-level node hierarchy, core and fringe nodes, NoAH formulates hyperedge generation as the sequential attachment of nodes (first cores, then fringes), where attachment probabilities are governed by node attributes. We also introduced NoAHFit, a parameter estimation algorithm that fits NoAH to a given hypergraph by estimating affinity matrices and seed core probabilities. Through extensive experiments on nine real-world hypergraphs across four diverse domains, we demonstrate that NoAH with NoAHFit more accurately reproduces the structure–attribute interplay than eight existing hypergraph generative models across six metrics.

Acknowledgments

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00871 / RS-2022-II220871, Development of AI Autonomy and Knowledge Enhancement for AI Agent Collaboration, 50%) (No. RS-2024-00457882, AI Research Hub Project, 40%) (RS-2019-II190075, Artificial Intelligence Graduate School Program (KAIST), 10%).

Contribution Statement

Jaewan Chun and Seokbum Yoon contributed equally to this work.

References

[Airoldi *et al.*, 2008] Edo M Airoldi, David Blei, Stephen Fienberg, and Eric Xing. Mixed membership stochastic blockmodels. In *NeurIPS*, 2008.

[Amburg *et al.*, 2020] Ilya Amburg, Nate Veldt, and Austin R Benson. Fair clustering for diverse and experienced groups. *arXiv:2006.05645*, 2020.

[Amburg *et al.*, 2021] Ilya Amburg, Jon Kleinberg, and Austin R Benson. Planted hitting set recovery in hypergraphs. *Journal of Physics: Complexity*, 2(3):035004, 2021.

[Badalyan *et al.*, 2024] Anna Badalyan, Nicolò Ruggeri, and Caterina De Bacco. Structure and inference in hypergraphs with node attributes. *Nature Communications*, 15(1):7073, 2024.

[Bojchevski *et al.*, 2018] Aleksandar Bojchevski, Oleksandr Shchur, Daniel Zügner, and Stephan Günnemann. Netgan: Generating graphs via random walks. In *ICML*, 2018.

[Chakrabarti and Faloutsos, 2006] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Computing Surveys*, 38(1):2–es, 2006.

[Chodrow *et al.*, 2021] Philip S Chodrow, Nate Veldt, and Austin R Benson. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, 7(28):eabh1303, 2021.

[Choe *et al.*, 2025] Minyoung Choe, Jihoon Ko, Taehyung Kwon, Kijung Shin, and Christos Faloutsos. Kronecker generative models for power-law patterns in real-world hypergraphs. In *WWW*, 2025.

[Chun *et al.*, 2024] Jaewan Chun, Geon Lee, Kijung Shin, and Jinhong Jung. Random walk with restart on hypergraphs: fast computation and an application to anomaly detection. *Data Mining and Knowledge Discovery*, 38(3):1222–1257, 2024.

[Chun *et al.*, 2025] Jaewan Chun, Seokbum Yoon, Minyoung Choe, Geon Lee, and Kijung Shin. Attributed hypergraph generation with realistic interplay between structure and attributes. In *ICDM*, 2025.

[Contisciani *et al.*, 2022] Martina Contisciani, Federico Battiston, and Caterina De Bacco. Inference of hyperedges and overlapping communities in hypergraphs. *Nature Communications*, 13(1):7229, 2022.

[De Cao and Kipf, 2018] Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.

[Do *et al.*, 2020] Manh Tuan Do, Seeun Yoon, Bryan Hooi, and Kijung Shin. Structural patterns and generative models of real-world hypergraphs. In *KDD*, 2020.

[Génois and Barrat, 2018] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[Hayashi *et al.*, 2020] Koby Hayashi, Sinan G. Aksoy, Cheong Hee Park, and Haesun Park. Hypergraph random walks, laplacians, and clustering. In *CIKM*, 2020.

[Kim and Leskovec, 2012] Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. *Internet Mathematics*, 8(1-2):113–160, 2012.

[Kim *et al.*, 2023] Sunwoo Kim, Fanchen Bu, Minyoung Choe, Jaemin Yoo, and Kijung Shin. How transitive are real-world group interactions? - measurement and reproduction. In *KDD*, 2023.

[Kook *et al.*, 2020] Yunbum Kook, Jihoon Ko, and Kijung Shin. Evolution of real-world hypergraphs: Patterns and models without oracles. In *ICDM*, 2020.

[Kumar *et al.*, 2020] Tarun Kumar, Sankaran Vaidyanathan, Harini Ananthapadmanabhan, Srinivasan Parthasarathy, and Balaraman Ravindran. Hypergraph clustering by iteratively reweighted modularity maximization. *Applied Network Science*, 5(1):52, 2020.

[Lee *et al.*, 2021] Geon Lee, Minyoung Choe, and Kijung Shin. How do hyperedges overlap in real-world hypergraphs? - patterns, measures, and generators. In *WWW*, 2021.

[Lee *et al.*, 2025] Geon Lee, Fanchen Bu, Tina Eliassi-Rad, and Kijung Shin. A survey on hypergraph mining: Patterns, tools, and generators. *ACM Computing Surveys*, 57(8):1–36, 2025.

[Liu *et al.*, 2023] Chengyi Liu, Wenqi Fan, Yunqing Liu, Jiatong Li, Hang Li, Hui Liu, Jiliang Tang, and Qing Li. Generative diffusion models on graphs: Methods and applications. In *IJCAI*, 2023.

[McPherson *et al.*, 2001] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[Nakajima *et al.*, 2021] Kazuki Nakajima, Kazuyuki Shudo, and Naoki Masuda. Randomizing hypergraphs preserving degree correlation and local clustering. *IEEE Transactions on Network Science and Engineering*, 9(3):1139–1153, 2021.

[Ni *et al.*, 2019] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *EMNLP*, 2019.

[Papachristou and Kleinberg, 2022] Marios Papachristou and Jon Kleinberg. Core-periphery models for hypergraphs. In *KDD*, 2022.

[Robins *et al.*, 2007] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, 2007.

[Ruggeri *et al.*, 2023] Nicolò Ruggeri, Martina Contisciani, Federico Battiston, and Caterina De Bacco. Community detection in large hypergraphs. *Science Advances*, 9(28):eadg9159, 2023.

[Simonovsky and Komodakis, 2018] Martin Simonovsky and Nikos Komodakis. Graphvae: Towards generation of small graphs using variational autoencoders. In *ICANN*, 2018.

[Wang *et al.*, 2023] Selena Wang, Subhadeep Paul, and Paul De Boeck. Joint latent space model for social networks with multivariate attributes. *Psychometrika*, 88(4):1197–1227, 2023.

[Yadati *et al.*, 2019] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In *NeurIPS*, 2019.

[Yu *et al.*, 2012] Jun Yu, Dacheng Tao, and Meng Wang. Adaptive hypergraph learning and its application in image classification. *IEEE Transactions on Image Processing*, 21(7):3262–3272, 2012.