APPENDIX

VIII. PROOF OF THEOREM 1

In this section, we provide proof of Theorem 1 which is given as:

Theorem 1. *There exist configurations on node attributes and parameters on* NOAH *such that the generated hypergraph follows a power-law degree distribution.*

Proof. Let l-th attribute of each core node be drawn from Bernoulli($\mu_{\mathcal{C}}^{(l)}$), and each fringe node from Bernoulli($\mu_{\mathcal{F}}^{(l)}$). Assume the seed probability is uniform: $p_{\text{seed}} = \frac{1}{|\mathcal{C}|}$

Core degree distribution. When $|\mathcal{C}| \to \infty$, the degree of core node v_c is proportional to the attachment probability of v_c to an arbitrary core node. If we adjust $\mu_{c,l}$ so that

$$\frac{\mu_{\mathcal{C}}^{(l)}}{1-\mu_{\mathcal{C}}^{(l)}} = \left(\frac{\mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}}^{(l)}[1,1] + (1-\mu_{\mathcal{C}}^{(l)}) \times \theta_{\mathcal{C}}^{(l)}[0,1]}{\mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}}^{(l)}[1,0] + (1-\mu_{\mathcal{C}}^{(l)}) \times \theta_{\mathcal{C}}^{(l)}[0,0]}\right)^{-\delta},$$

then following Theorem of [35], probability of core nodes having degree d is proportional to $d^{-\delta-\frac{1}{2}}$.

Fringe degree distribution. For core group g generated from seed core v_s, if $\mathbf{x}_s^{(l)} = 1$, the probability of l-th attribute of attached core node v_c having 1 is calculated as:

$$\begin{aligned} &p(\mathbf{x}_c^{(l)} = 1 | v_c \in g, \mathbf{x}_s^{(l)} = 1) \\ &= \frac{\mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}}^{(l)}[1,1]}{(1-\mu_{\mathcal{C}}^{(l)}) \times \theta_{\mathcal{C}}^{(l)}[1,0] + \mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}}^{(l)}[1,1]} \end{aligned}$$

If $\mathbf{x}_s^{(l)} = 0$, the probability of l-th attribute of attached core node v_c having 1 is calculated as:

$$\begin{aligned} &p(\mathbf{x}_c^{(l)} = 1 | v_c \in g, \mathbf{x}_s^{(l)} = 0) \\ &= \frac{\mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}_l}[0,1]}{(1-\mu_{\mathcal{C}}^{(l)}) \times \theta_{\mathcal{C}_l}[0,0] + \mu_{\mathcal{C}}^{(l)} \times \theta_{\mathcal{C}_l}[0,1]} \end{aligned}$$

If we set $\theta_{\mathcal{C}_l}[0,1] = \theta_{\mathcal{C}_l}[1,0]$, and $\theta_{\mathcal{C}_l}[0,1]^2 = \theta_{\mathcal{C}_l}[0,0] \times \theta_{\mathcal{C}_l}[1,1]$, then

$$\begin{aligned} &p(\mathbf{x}_c^{(l)} = 1 | v_c \in g, \mathbf{x}_s^{(l)}) \\ &= p(\mathbf{x}_c^{(l)} = 1 | v_c \in g, \mathbf{x}_s^{(l)}) \end{aligned}$$

Since the attribute distribution of attached core nodes is now independent of the seed core attribute, let's denote it as l-th attribute following Bernoulli($\mu_a^{(l)}$). Moreover, for l-th attribute, the nodes in the core group of the hypergraph follows Bernoulli($\mu_g^{(l)}$), where

$$\mu_g^{(l)} = \frac{\mu_{\mathcal{C}}^{(l)} + \tilde{g} \times \mu_a^{(l)}}{1+\tilde{g}},$$

where $\tilde{g}$ denotes the average core group size. Then, if we adjust $\mu_{\mathcal{F}}^{(l)}$ so that

$$\frac{\mu_{\mathcal{F}}^{(l)}}{1-\mu_{\mathcal{F}}^{(l)}} = \left(\frac{\mu_g^{(l)} \times \theta_{\mathcal{F}}^{(l)}[1,1] + (1-\mu_g^{(l)}) \times \theta_{\mathcal{F}}^{(l)}[0,1]}{\mu_g^{(l)} \times \theta_{\mathcal{F}}^{(l)}[1,0] + (1-\mu_g^{(l)}) \times \theta_{\mathcal{F}}^{(l)}[0,0]}\right)^{-\delta},$$

following Theorem 7.1 of [35], probability of fringe nodes having degree d is proportional to $d^{-\delta-\frac{1}{2}}$ as $|\mathcal{F}| \to \infty$.

Conclusion. Thus, under the assumptions above, the probability of nodes having degree d is proportional to $d^{-\delta-\frac{1}{2}}$ for both core and fringe nodes, resulting in a power-law distribution for node degree. □

IX. PARAMETER INITIALIZATION OF NOAHFIT

In this section, we provide the parameter initialization method of NOAHFIT.

- **Seed Core Probability $\boldsymbol{p}_{\text{seed}}$:** We initialize the seed core probability to a value proportional to the degree, i.e.,

$$\boldsymbol{p}_{\text{seed}}(v_c) = \frac{|\{e \in \mathcal{E} | v_c \in e\}|}{\sum_{e \in \mathcal{E}} |\mathcal{C}_e|}$$

- **Core and Fringe Affinity Matrices $\Theta_{\mathcal{C}}, \Theta_{\mathcal{F}}$:** For each core and fringe affinity matrices, we initialize to a constant value that produces the mean expected cardinality of (core, fringe) subset same as the target hypergraph, i.e.,

$$\begin{aligned} &\forall l \in \{1,\ldots,k\},\ \forall i \in \{0,1\},\ \forall j \in \{0,1\}, \\ &\theta_{\mathcal{C}}^{(l)}[i,j] = (\overline{c_c} - 1)^{1/k}, \\ &\theta_{\mathcal{F}}^{(l)}[i,j] = \overline{c_f}^{1/k}, \end{aligned}$$

  where $\overline{c_c}, \overline{c_f}$ each denotes the mean cardinality of the core and fringe subset in the target hypergraph.

X. DETAILED MODEL CONFIGURATIONS

- HYPERCL and HYPERLAP: Models utilize degree and hyperedge size distribution.
- HYPERPA: Model utilizes the distribution of hyperedge size and number of new hyperedges for a new node.
- HYPERFF: $p \in \{0.42, 0.45, 0.48\}$ and $q \in \{0.1, 0.2, 0.3\}$
- hyper dK-series: Model utilizes degree and hyperedge size distribution, and additional parameters $(d_v, d_e) \in \{(0,0), (0,1), (1,0)\}$
- THERA: Model utilizes the distribution of hyperedge size. For parameters we tested $C \in \{8, 12, 15\}$, $p \in \{0.5, 0.7, 0.9\}$, $\alpha \in \{2, 6, 10\}$.
- HYCOSBM: $\gamma \in \{0.0, 0.1, \cdots, 0.9\}$
- HYREC and NOAH: For Stack domain, $w_{deg} \in \{10, 100, 1000\}, w_{card} \in [10, 100, 1000]$. For other domains, $w_{deg} \in \{0, 0.1, 0.01\}, w_{card} \in \{0, 1, 2\}$.

XI. PERFORMANCE COMPARISON ON STRUCTURE

In this section, we evaluate NOAH along with 8 baseline generators in Section VI. We evaluate nine structural properties:

- **S1. Degree:** *Degree* of a node v, $d(v)$ is defined as the number of hyperedges containing v, i.e., $d(v) = |\{e \in \mathcal{E} : v \in e\}|$. The degree distribution tends to have a heavy tail in the real world, which can't be reproduced in random models [45].
- **S2. Pair Degree:** *Pair degree* of two nodes u and v is defined as the number of hyperedges containing both u and v, i.e., $|\{e \in \mathcal{E} : u \in e, v \in e\}|$. The distribution of non-zero pair degree tends to have a heavier tail in the real world compared to randomized models [7].

- **S3. Size:** *Size* of a hyperedge e, $s(e)$ is defined as the number of nodes contained in e, i.e., $s(e) = |\{v \in e\}|$. The size distribution tends to have a heavy tail in the real world, which can't be reproduced in random models [45].
- **S4. Intersection Size (Int. Size):** *Intersection size* of two hyperedges e_1 and e_2 is defined as the number of nodes contained in both e_1 and e_2, i.e., $|\{v \in \mathcal{V} : v \in e_1, v \in e_2\}|$. The distribution of non-zero intersection size tends to be heavy-tailed in real-world hypergraphs [45].
- **S5. Singular Values (SV):** *Singular values* are derived from the incidence matrix of a hypergraph. Specifically, for each $i \in 1, \ldots, R$, we calculate the ratio $\frac{s_i^2}{\sum_{k=1}^{R} s_k^2}$, where s_i denotes the i-th largest singular value and R denotes the rank of the incidence matrix. These values reflect the variance captured by the associated singular vectors [48], and in many real-world hypergraphs, they are often highly skewed [45].
- **S6. Connected Component Size (CC):** We consider the proportion of nodes contained within each i-th largest connected component in the clique expansion of a hypergraph. The clique expansion of a hypergraph refers to the undirected graph formed by substituting each hyperedge $e \in \mathcal{E}$ with a clique with the nodes in e. In a real-world hypergraph, clique expansion has a giant connected component comprising a majority of nodes [6].
- **S7. Global Clustering Coefficient (GCC):** C_v, the *Local clustering coefficient* of a node v, is defined as follows:

$$C_v := 2 \times \frac{\text{the number of triangles involving } v}{\text{the number of connected triplets involving } v}, \quad (10)$$

  We estimate the *global clustering coefficient* by averaging local clustering coefficients in the clique expansion (defined in S6) of a hypergraph using [49]. This statistic tends to be larger in real-world hypergraphs than in uniform random hypergraphs [6].
- **S8. Density:** We consider the *density* which is defined as ratio of the number of hyperedges to the number of nodes, i.e., $\frac{|\mathcal{V}|}{|\mathcal{E}|}$ [50]. Hypergraphs within the same domain tend to exhibit similar levels of density significance [7].
- **S9. Overlapness:** We consider the *overlapness* which is defined as $\sum_{e \in \mathcal{E}} \frac{|e|}{|\mathcal{E}|}$ [7]. Hypergraphs within the same domain tend to exhibit similar levels of overlapness significance [7].
- **S10. Effective Diameter:** We consider the *effective diameter* which is defined as the smallest $d \in \mathbb{Z}$ such that the paths of length at most d in the clique expansion (defined in S6) connect 90% of reachable pairs of nodes [51]. This statistic tends to be small in real-world hypergraphs [45].

We used the Kolmogorov-Smirnov (D-statistic) for degree, size, pair degree, and intersection size, root mean square error (RMSE) for singular values, clustering coefficients, density, and overlapness, and relative difference for effective diameter. The results are summarized in Table IV. Note that although baseline models such as HYPERCL, HYPERPA, hyper dK-series, and THERA explicitly incorporate node degree and hyperedge size distributions in their generative processes, NOAH still demonstrates competitive performance in these experiments.

TABLE IV

STRUCTURAL METRIC EVALUATION ACROSS 9 DATASETS. A.R. DENOTES AVERAGE RANK.

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.195	0.183	0.000	0.220	0.038	0.286	0.120	0.289	0.289	0.555	4.8
HYPERPA	0.127	0.105	0.033	0.011	0.144	0.297	0.323	0.780	0.811	0.429	5.9
HYPERFF	0.366	0.056	0.396	0.209	0.104	0.301	0.164	2.247	1.153	0.031	6.4
HYPERLAP	0.244	0.030	0.000	0.024	0.054	0.260	0.196	0.227	0.227	0.021	3.2
hyper dK-series	0.001	0.185	0.148	0.222	0.067	0.252	0.118	0.039	0.003	0.533	4.5
THERA	0.456	0.079	0.013	0.006	0.057	0.129	0.307	0.350	0.389	0.370	4.7
HYCOSBM	0.484	0.176	0.114	0.177	0.085	0.298	0.050	0.110	0.843	0.704	6.1
HYREC	0.238	0.032	0.162	0.127	0.048	0.337	0.202	0.018	0.136	0.286	4.3
NOAH	0.111	0.172	0.513	0.171	0.118	0.075	0.304	0.517	0.097	0.204	5.0

(a) Citeseer (NOAH ranks **sixth** overall)

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.189	0.115	0.000	0.412	0.034	0.286	0.228	0.276	0.276	0.464	4.7
HYPERPA	0.137	0.258	0.020	0.172	0.273	0.298	0.414	1.674	1.661	0.368	6.3
HYPERFF	0.441	0.199	0.251	0.351	0.041	0.298	0.258	4.410	2.328	0.117	6.6
HYPERLAP	0.177	0.034	0.000	0.334	0.024	0.279	0.162	0.236	0.236	0.361	2.6
hyper dK-series	0.171	0.119	0.001	0.417	0.057	0.288	0.225	0.176	0.174	0.439	4.7
THERA	0.601	0.053	0.015	0.236	0.104	0.257	0.091	1.227	1.162	0.376	4.7
HYCOSBM	0.349	0.112	0.039	0.387	0.091	0.297	0.209	0.291	0.617	0.599	5.9
HYREC	0.124	0.121	0.171	0.019	0.037	0.622	0.113	0.410	0.208	0.515	4.7
NOAH	0.096	0.110	0.272	0.409	0.046	0.191	0.310	0.447	0.140	0.398	4.6

(b) Cora

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.031	0.272	0.000	0.034	0.029	0.000	0.409	0.000	0.000	0.284	2.7
HYPERPA	0.189	0.165	0.009	0.018	0.099	0.000	0.472	0.063	0.066	0.254	3.5
HYPERFF	0.899	0.100	0.115	0.061	0.260	0.000	0.126	0.907	0.896	0.935	5.8
HYPERLAP	0.034	0.067	0.000	0.020	0.016	0.000	0.264	0.000	0.000	0.312	1.9
hyper dK-series	0.327	0.251	0.245	0.030	0.122	0.000	0.335	0.093	0.003	0.295	4.6
THERA	0.266	0.373	0.008	0.093	0.056	0.000	0.066	0.000	0.006	0.090	3.4
HYCOSBM	1.000	1.000	0.458	0.239	0.194	0.000	0.986	0.924	6.376	0.658	7.9
HYREC	0.237	0.267	0.248	0.109	0.118	0.143	0.265	0.197	0.143	0.361	6.2
NOAH	0.495	0.349	0.530	0.035	0.120	0.000	0.731	0.000	0.304	0.270	5.4

(c) High School

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.087	0.012	0.000	0.002	0.016	0.000	0.412	0.000	0.000	0.142	2.3
HYPERPA	0.086	0.087	0.007	0.006	0.048	0.000	0.411	0.028	0.024	0.040	3.9
HYPERFF	0.826	0.017	0.056	0.032	0.167	0.000	0.206	0.811	0.805	1.624	5.8
HYPERLAP	0.065	0.021	0.000	0.004	0.019	0.000	0.331	0.000	0.000	0.139	2.3
hyper dK-series	0.043	0.078	0.335	0.015	0.026	0.000	0.035	0.114	0.016	0.190	3.8
THERA	0.185	0.296	0.011	0.038	0.027	0.000	0.082	0.000	0.006	0.199	4.0
HYCOSBM	0.978	0.999	0.590	0.282	0.200	0.000	1.347	0.266	2.696	0.609	7.7
HYREC	0.512	0.429	0.286	0.188	0.284	0.342	0.241	0.516	0.446	0.655	7.4
NOAH	0.272	0.019	0.505	0.000	0.050	0.000	0.535	0.000	0.199	0.080	4.2

(d) Workspace

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.097	0.051	0.000	0.102	0.027	0.002	0.027	0.063	0.063	0.088	2.2
HYPERPA	0.225	0.313	0.034	0.162	0.210	0.002	0.304	1.874	1.712	0.040	5.5
HYPERFF	0.139	0.040	0.740	0.442	0.048	0.002	0.150	6.974	0.210	1.218	5.5
HYPERLAP	0.126	0.033	0.000	0.015	0.031	0.002	0.108	0.067	0.067	0.227	2.6
hyper dK-series	0.497	0.261	0.420	0.357	0.254	0.002	0.618	0.000	0.008	0.093	5.5
THERA	0.456	0.123	0.025	0.181	0.224	0.002	0.299	0.611	0.517	0.087	5.3
HYCOSBM	0.431	0.305	0.501	0.380	0.089	0.005	0.018	0.119	0.684	0.343	6.6
HYREC	0.315	0.039	0.083	0.033	0.079	0.002	0.194	0.001	0.042	0.098	3.4
NOAH	0.233	0.156	0.280	0.242	0.119	0.018	0.228	0.063	0.458	0.603	6.1

(e) Amazon Music

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.093	0.103	0.000	0.035	0.053	0.000	0.025	0.024	0.024	0.124	2.2
HYPERPA	0.549	0.368	0.023	0.187	0.253	0.000	0.294	0.326	0.320	0.308	7.0
HYPERFF	0.352	0.069	0.479	0.126	0.085	0.000	0.068	1.158	0.271	1.479	6.1
HYPERLAP	0.090	0.042	0.000	0.031	0.053	0.000	0.036	0.025	0.025	0.033	1.9
hyper dK-series	0.198	0.192	0.010	0.101	0.085	0.000	0.332	0.000	0.014	0.134	4.1
THERA	0.177	0.149	0.020	0.169	0.148	0.000	0.200	0.000	0.042	0.311	4.7
HYCOSBM	0.347	0.193	0.265	0.083	0.063	0.004	0.100	0.287	0.448	0.083	6.1
HYREC	0.356	0.186	0.231	0.236	0.090	0.000	0.013	0.034	0.357	0.111	5.4
NOAH	0.252	0.027	0.285	0.082	0.053	0.066	0.032	0.058	0.267	0.494	4.9

(f) Yelp Restaurant

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.140	0.201	0.000	0.062	0.028	0.000	0.189	0.019	0.019	0.280	3.4
HYPERPA	0.653	0.343	0.029	0.196	0.293	0.000	0.280	0.935	0.917	0.072	6.7
HYPERFF	0.569	0.078	0.604	0.219	0.160	0.000	0.115	1.516	0.418	1.768	6.6
HYPERLAP	0.125	0.109	0.000	0.031	0.026	0.000	0.024	0.016	0.016	0.046	1.7
hyper dK-series	0.177	0.259	0.008	0.101	0.057	0.000	0.369	0.000	0.012	0.282	4.0
THERA	0.216	0.076	0.018	0.150	0.085	0.000	0.077	0.038	0.058	0.241	3.9
HYCOSBM	0.774	0.334	0.311	0.189	0.101	0.005	0.044	0.336	0.648	0.075	6.6
HYREC	0.332	0.061	0.259	0.236	0.116	0.000	0.024	0.265	0.078	0.042	4.3
NOAH	0.381	0.159	0.393	0.066	0.176	0.106	0.006	0.067	0.294	0.222	5.6

(g) Yelp Bar

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.328	0.024	0.129	0.011	0.041	0.225	0.068	0.369	0.495	0.127	3.9
HYPERPA	0.350	0.103	0.127	0.126	0.102	0.194	0.383	0.272	0.389	0.335	4.8
HYPERFF	0.537	0.273	0.273	0.061	0.165	0.293	0.682	1.165	1.726	0.366	7.7
HYPERLAP	0.324	0.023	0.129	0.011	0.042	0.221	0.066	0.368	0.495	0.126	3.1
hyper dK-series	0.073	0.024	0.205	0.007	0.048	0.152	0.764	0.103	0.088	0.201	3.6
THERA	0.433	0.021	0.134	0.001	0.232	0.215	0.124	0.000	0.084	0.481	4.1
HYCOSBM	0.815	0.019	0.277	0.035	0.119	0.293	0.243	0.098	2.662	0.615	6.4
HYREC	0.398	0.329	0.216	0.137	0.112	0.254	0.682	1.685	2.545	0.097	6.8
NOAH	0.339	0.048	0.175	0.015	0.139	0.176	0.111	0.353	0.194	0.187	4.4

(h) Devops

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	0.354	0.012	0.010	0.010	0.133	0.064	0.077	0.446	0.477	0.070	3.1
HYPERPA	0.383	0.116	0.016	0.141	0.181	0.041	0.495	0.398	0.396	0.586	5.5
HYPERFF	0.630	0.099	0.208	0.012	0.237	0.114	0.010	4.588	3.914	0.412	6.3
HYPERLAP	0.345	0.013	0.010	0.008	0.133	0.063	0.096	0.435	0.465	0.084	2.9
hyper dK-series	0.009	0.009	0.167	0.014	0.133	0.050	0.378	0.087	0.005	0.095	3.1
THERA	0.590	0.140	0.016	0.105	0.348	0.413	0.071	0.000	0.003	0.783	5.9
HYCOSBM	0.767	0.017	0.293	0.065	0.165	0.113	0.005	0.218	1.434	0.505	5.8
HYREC	0.463	0.304	0.213	0.059	0.263	0.280	0.314	4.086	4.262	0.272	7.4
NOAH	0.437	0.034	0.126	0.014	0.135	0.112	0.308	0.369	0.403	0.254	4.6

(i) Patents

	Degree	Pair Degree	Size	Int. Size	SV	CC	GCC	Density	Overlapness	Diameter	A.R.
HYPERCL	2.9	4.6	2.1	3.9	1.8	6.2	4.3	4.4	3.8	4.6	2.8
HYPERPA	4.7	7.1	3.8	5.0	6.9	6.1	7.8	6.6	6.3	4.0	6.3
HYPERFF	7.1	4.2	7.6	6.3	6.4	7.7	4.6	8.8	7.4	6.8	7.3
HYPERLAP	2.6	2.2	2.1	2.1	2.0	5.8	3.4	4.1	3.4	2.8	1.5
hyper dK-series	3.3	5.6	5.2	5.4	4.9	5.3	6.3	2.7	1.9	5.3	4.1
THERA	6.2	4.8	3.7	4.9	6.3	6.0	4.2	4.0	4.1	5.3	4.4
HYCOSBM	8.2	6.4	7.6	6.8	6.0	7.6	4.4	5.3	8.1	6.9	7.6
HYREC	5.4	5.4	6.6	6.2	5.4	8.0	4.3	5.7	5.7	4.7	5.8
NOAH	4.6	4.7	7.4	4.3	5.9	5.8	5.6	5.0	4.4	4.7	4.6

(j) Average Rank over Nine Datasets (NOAH ranks **fifth** overall)