

HyperAim: Hypergraph Contrastive Learning with Adaptive Multi-frequency Filters

Ming Li¹, Ruiting Zhao², Zihao Yan², Lu Bai^{3*}, Lixin Cui⁴, Feilong Cao⁵

¹Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China

²School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China

³School of Artificial Intelligence, Beijing Normal University, Beijing, China

⁴Central University of Finance and Economics, Beijing, China.

⁵School of Mathematical Sciences, Zhejiang Normal University, Jinhua, China

mingli@zjnu.edu.cn, zrt216921@zjnu.edu.cn, yzhaoian2@zjnu.edu.cn, bailu@bnu.edu.cn, cuilixin@cufe.edu.cn, caofeilong88@zjnu.edu.cn

Abstract

Unsupervised hypergraph representation learning has recently gained traction for its ability to model complex high-order interactions without requiring labeled data. However, existing contrastive learning methods typically overlook the frequency diversity inherent in hypergraph signals. To address this issue, we propose **HyperAim**, a contrastive learning framework that integrates adaptive multi-frequency filtering through both decoupled and coupled designs. Specifically, HyperAim employs two decoupled channels with polynomial low-pass and high-pass filters to separately capture distinct frequency components, and a third channel based on framelet decomposition that adaptively fuses multi-frequency signals in a coupled manner. A frequency-aware contrastive learning strategy is introduced to align representations across views using a combination of InfoNCE loss and pseudo-label-guided supervision. Extensive experiments across 12 benchmark datasets, covering both homophilic and heterophilic hypergraphs, demonstrate the consistent superiority of HyperAim over 17 baselines. Ablation studies further confirm the benefits of explicitly modeling and aligning frequency-specific representations.

1 Introduction

Hypergraphs offer a powerful and flexible framework for modeling complex, high-order relationships among entities, as they allow each hyperedge to connect an arbitrary number of nodes. With the increasing need to understand and utilize such structured data, hypergraph representation learning has emerged as a promising approach for learning informative embeddings that preserve the underlying topological and semantic properties of the data.

While significant advances have been made in the development of supervised hypergraph learning algorithms, unsupervised representation learning on hypergraphs remains relatively underexplored. In particular, contrastive learning, which has proven highly effective in unsupervised settings for graphs, has only recently been extended to hypergraphs. Existing methods, such as TriCL (Lee and Shin 2023), HyperGCL (Wei et al. 2022a), and HypeBoy (Kim et al. 2024),

*Corresponding author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

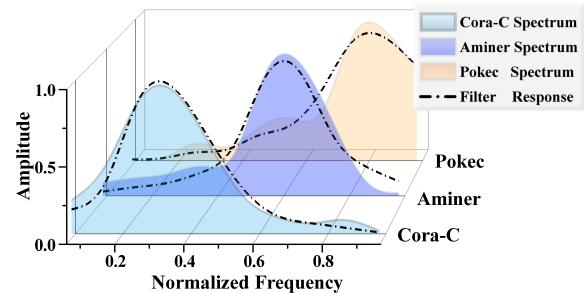


Figure 1: Spectral energy distributions of three representative hypergraph datasets and their corresponding responses to multi-bandpass filters. Matched filters are expected to align with the dominant spectral regions of each dataset.

adopt spatial-based contrastive frameworks that are largely inherited from message-passing paradigms in graph neural networks. These methods primarily emphasize local neighborhood aggregation, which inherently performs low-pass filtering. However, such approaches are limited in their ability to adapt to the diverse spectral properties exhibited by real-world hypergraphs. As illustrated in Figure 1, the spectral energy of different hypergraph datasets, such as Cora-C, Aminer, and Pokec, is concentrated in distinct frequency bands, ranging from low to high frequencies. These observations suggest that a *one-size-fits-all* filtering strategy may be suboptimal, and that adaptive multi-frequency filtering is essential for effective representation learning across diverse hypergraphs.

To address this challenge, we propose **HyperAim**, a novel framework for hypergraph contrastive learning with adaptive multi-frequency filtering. HyperAim integrates both spectrally-inspired and spectral-theoretic perspectives through three complementary branches. The first two branches adopt the design of **decoupled polynomial filters** for frequency-specific message passing. These filters are predefined, fixed-form spectral functions applied independently to extract either low-frequency or high-frequency information. This design enables efficient and disentangled view generation, offering explicit control over different

spectral components. In contrast, the third branch performs joint multi-frequency modeling via a **coupled framelet-based spectral module**, which decomposes node features into multiple frequency bands and adaptively fuses them through learnable spectral weights. This design allows the model to capture cross-frequency interactions and flexibly align with dataset-specific spectral energy distributions. To ensure semantic consistency across these diverse views, we introduce a frequency-aware contrastive learning strategy. Positive pairs are defined by aligning each frequency-specific representation with its fused counterpart, while negative pairs are constructed from mismatched or perturbed views. These are optimized via a binary cross-entropy loss. Additionally, InfoNCE losses are applied within each branch to preserve intra-view consistency. Together, this hybrid objective enables the learning of discriminative, robust, and frequency-consistent node representations. Extensive experiments on 12 benchmark datasets, covering both homophilic and heterophilic hypergraphs, demonstrate that HyperAim consistently outperforms 17 existing methods. Ablation studies further validate the effectiveness of its multi-frequency architecture and contrastive objectives.

In summary, our contributions are three-fold:

- We propose **HyperAim**, a novel hypergraph contrastive learning framework that unifies decoupled polynomial filtering and coupled framelet-based spectral modeling to achieve adaptive multi-frequency representation learning across diverse hypergraph structures.
- We design a self-supervised learning paradigm that constructs frequency-aware contrastive views from both independently filtered and jointly decomposed frequency components. This design enables HyperAim to exploit complementary frequency perspectives, combining explicit disentanglement with multi-scale spectral fusion.
- We conduct extensive experiments on 12 benchmark datasets, showing that HyperAim consistently outperforms 17 state-of-the-art baselines under both homophilic and heterophilic settings. Ablation studies further confirm the contributions of each module and the benefit of integrating multi-frequency information.

2 Related Work

Hypergraph representation learning extends conventional graph-based models by capturing higher-order relations through hyperedges. Existing methods under supervised and semi-supervised settings include approaches that approximate hyperedges via clique expansion (Yadati et al. 2019) and those that preserve the native hypergraph topology through hyperedge-based message passing (Feng et al. 2019; Dong, Sawin, and Bengio 2020; Huang and Yang 2021; Chien et al. 2022; Wang et al. 2023a). In the self-supervised setting, early studies often target specific applications such as group discovery (Zhang et al. 2021) and session-based recommendation (Xia, Huang, and Zhang 2022), while more recent works seek general-purpose hypergraph representations. For example, TriCL (Lee and Shin 2023) designs a multi-level contrastive objective over nodes, hyperedges, and incidence relations, and HyperGCL (Wei

et al. 2022b) uses neural view generators to enhance contrastive learning beyond handcrafted perturbations. HypeBoy (Kim et al. 2024) proposes a generative self-supervised framework based on hyperedge completion, offering theoretical insights and improved scalability. However, most of these approaches are built on spatial perturbations or generative schemes, with limited attention to spectral properties of hypergraphs. Although recent graph contrastive learning methods begin to explore spectral perspectives (Chen, Lei, and Wei 2024), such techniques are not yet systematically studied for hypergraph representation learning. This work aims to narrow this gap by incorporating spectral analysis and multi-frequency filtering into the design of hypergraph contrastive learning frameworks.

3 Preliminary and Motivation

Basics for hypergraph. A hypergraph is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, comprising a vertex set \mathcal{V} of size $N = |\mathcal{V}|$, a hyperedge set \mathcal{E} of size $M = |\mathcal{E}|$. Suppose that vertices and hyperedges have feature dimensions d and m , respectively, we have the representation of vertex data as $\mathbf{X} \in \mathbb{R}^{N \times d}$ and hyperedge data as $\mathbf{Y} \in \mathbb{R}^{M \times o}$.

The hypergraph structure, from a vertex perspective, is defined by an incidence matrix $\mathbf{H} \in \{0, 1\}^{N \times M}$ where $\mathbf{H}(v, e) = 1$ if vertex v is contained in hyperedge e , and 0 otherwise, as represented by:

$$\mathbf{H}(v, e) = \begin{cases} 1, & \text{if } v \in e; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The degrees of vertex v and hyperedge e are denoted by diagonal matrices $\mathbf{D}_v \in \mathbb{R}^{N \times N}$ and $\mathbf{D}_e \in \mathbb{R}^{M \times M}$, calculated as $\sum_{e \in \mathcal{E}} \mathbf{H}(v, e)$ and $\sum_{v \in \mathcal{V}} \mathbf{H}(v, e)$, respectively.

The normalized hypergraph Laplacian \mathbf{L} is defined as:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2}.$$

Motivation. To the best of our knowledge, existing self-supervised methods on hypergraphs mainly operate in the spatial domain, relying on incidence-based message passing that effectively captures local interactions but implicitly favors low-pass filtering and overlooks the richer spectral properties encoded in the hypergraph Laplacian. Recent empirical analyses, however, show that hypergraphs exhibit diverse frequency characteristics depending on their topology and semantics: as illustrated in Figure 1, some datasets concentrate energy in low-frequency bands, whereas others possess substantial high-frequency components with non-smooth patterns. This calls for a more flexible representation learning strategy that moves beyond uniform smoothing and can adapt to different spectral regimes. In particular, aligning representations across frequency-aware views, rather than only spatially perturbed ones, can reveal complementary structural cues that standard message passing fails to exploit. Motivated by this, we introduce adaptive multi-frequency analysis into hypergraph contrastive learning: by integrating low-pass, high-pass, and multi-scale spectral representations and designing contrastive objectives across them, we aim to learn robust, semantically consistent, and frequency-aware node embeddings. This motivation underpins the design of **HyperAim**, detailed in the next section.

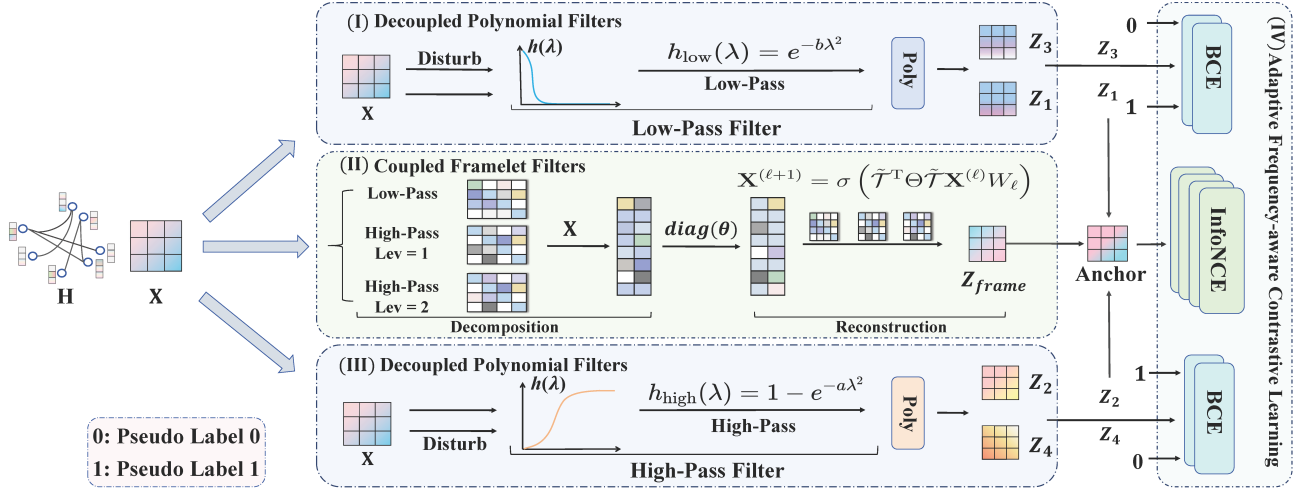


Figure 2: Overall architecture of HyperAim.

4 Proposed Method: HyperAim

4.1 Framework Overview

The overall architecture of the proposed **HyperAim** framework is illustrated in Figure 2. HyperAim is a self-supervised hypergraph contrastive learning framework that captures diverse spectral semantics via adaptive multi-frequency filtering, involving three key components for generating and aligning frequency-aware representations:

- **Decoupled Polynomial Filters for Frequency-Specific Message Passing.** The first and third branches apply predefined polynomial low-pass and high-pass filters in two independent channels (i.e., Modules **I** and **III** in Figure 2). These filters generate disentangled frequency-specific views by applying targeted perturbations, and are supervised using binary pseudo-labels via a contrastive binary cross-entropy (BCE) loss.
- **Coupled Framelet Filters for Multi-frequency Representation Learning.** The middle branch (i.e., Module **II** in Figure 2) performs joint spectral analysis using framelet transforms, decomposing signals into low-pass and multiple levels of high-pass components. These are then adaptively fused via learnable frequency coefficients to form a unified representation, which serves as the anchor for multi-frequency contrast.
- **Adaptive Frequency-aware Contrastive Learning Across Channels.** To align representations across the three frequency-aware views, HyperAim introduces a hybrid loss combining InfoNCE and BCE objectives (i.e., Module **IV** in Figure 2). This design promotes both intra-branch consistency and cross-branch alignment, enabling robust and semantically coherent node representations across heterogeneous spectral regimes.

4.2 Decoupled Polynomial Filters for Frequency-Specific Message Passing

To efficiently generate frequency-specific views for contrastive learning, HyperAim introduces two *independent*

message-passing branches, each operating on a predefined spectral band, one low-pass and one high-pass, based on the normalized hypergraph Laplacian. These two branches are *structurally decoupled* and correspond to the top and bottom modules (**I** and **III**) in Figure 2. Although the filters originate from spectral theory, they are implemented using polynomial approximations, which avoid explicit eigendecomposition and make them compatible with scalable message-passing architectures.

Given node features $\mathbf{X} \in \mathbb{R}^{n \times d}$, the filtering operation is defined as: $\mathbf{Y} = h(\mathbf{L})\mathbf{X}$, where $h(\lambda)$ is a frequency-selective function applied to the spectrum of the normalized Laplacian \mathbf{L} . To isolate distinct spectral ranges, we employ exponential filters:

$$h_{\text{low}}(\lambda) = e^{-b\lambda^2}, h_{\text{high}}(\lambda) = 1 - e^{-a\lambda^2}, \quad a, b > 0, \quad (2)$$

where a and b determine the selectivity of each filter. The low-pass filter preserves smooth and globally coherent signals, while the high-pass filter captures fine-grained variations and structural discontinuities.

To implement these filters efficiently, we approximate them using Chebyshev polynomials up to order K , where $T_k(\cdot)$ denotes the k -th order Chebyshev polynomial and $\hat{\mathbf{L}}$ is the rescaled Laplacian. The filtered outputs are:

$$\mathbf{Z}_{\text{L}} = f_{\theta} \left(\sum_{k=0}^K w_k^{\text{low}} T_k(\hat{\mathbf{L}})\mathbf{X} \right), \quad (3)$$

$$\mathbf{Z}_{\text{H}} = f_{\theta} \left(\sum_{k=0}^K w_k^{\text{high}} T_k(\hat{\mathbf{L}})\mathbf{X} \right), \quad (4)$$

and $w_k^{\text{low}}, w_k^{\text{high}}$ are filter coefficients derived from the respective functions, and $f_{\theta}(\cdot)$ is a shared MLP used to project the filtered features into a common embedding space.

It is important to note that these low-pass and high-pass filters are applied in two distinct branches without interaction, which means that they are predefined, frequency-isolated channels. This decoupled structure provides clear

and controllable contrastive views, allowing the model to isolate low- and high-frequency information explicitly and supervise them independently. Such a design facilitates efficient training and targeted representation learning. To enrich this design with a more unified and flexible spectral modeling capacity, we further introduce a framelet-based spectral module in the next subsection. Unlike the polynomial filters, this module jointly decomposes signals into multi-level frequency components and enables adaptive fusion across frequency bands, offering a complementary perspective to the independently constructed contrastive views above.

4.3 Coupled Framelet Filters for Multi-frequency Representation Learning

While the previous branches rely on predefined, decoupled polynomial filters to isolate low- and high-frequency information, they do not capture the interactions or joint effects across frequency bands. To address this, we adopt a multiresolution spectral analysis perspective and employ a framelet-based hypergraph neural network module (Li et al. 2025a) with coupled and adaptive multi-frequency filters. Unlike polynomial filters that operate independently, framelet transforms enable the simultaneous extraction of low-pass and multiple levels of high-pass components (Li et al. 2024), which are subsequently fused in a learnable manner. This design allows the model to represent hypergraph signals in a frequency-coordinated and data-adaptive fashion, providing a complementary channel to the fixed filtering structures described in **Section 4.2**.

Formally, a framelet system comprises two essential components: a *filter bank* $\eta = \{a; b^{(1)}, \dots, b^{(K)}\}$ and a corresponding set of real-valued scaling functions $\Psi = \{\gamma; \delta^{(1)}, \dots, \delta^{(K)}\}$, where a denotes the low-pass filter and $b^{(k)}$ represents the k -th high-pass filter for $k = 1, \dots, K$. These filters satisfy the classical refinement conditions in the Fourier domain for all $\xi \in \mathbb{R}$, as follows:

$$\hat{\gamma}(2\xi) = \hat{a}(\xi)\hat{\gamma}(\xi), \quad \hat{\delta}^{(k)}(2\xi) = \hat{b}^{(k)}(\xi)\hat{\gamma}(\xi).$$

Given that the hypergraph Laplacian $\mathbf{L} \in \mathbb{R}^{N \times N}$ is a real symmetric and positive semi-definite matrix, it admits an eigendecomposition of the form $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U} = [u_1, \dots, u_N] \in \mathbb{R}^{N \times N}$ is an orthogonal matrix whose columns are the eigenvectors of \mathbf{L} , and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N) \in \mathbb{R}^{N \times N}$ is a diagonal matrix of corresponding eigenvalues with $\lambda_\ell \geq 0$. At dilation level $j = 1, \dots, J$ and for a reference node $p \in V$, the framelets are defined as:

$$\varphi_{j,p}(v) = \sum_{\ell=1}^N u_\ell(v) \hat{\gamma}\left(\frac{\lambda_\ell}{2^j}\right) u_\ell(p), \quad (5)$$

$$\psi_{j,p,k}(v) = \sum_{\ell=1}^N u_\ell(v) \hat{\delta}^{(k)}\left(\frac{\lambda_\ell}{2^j}\right) u_\ell(p), \quad (6)$$

where $\varphi_{j,p}$ and $\psi_{j,p,k}$ are the low-pass and k -th high-pass framelets centered at node p , respectively.

To ensure tightness and perfect reconstruction, the filters

must satisfy the following energy partition condition:

$$\left| \hat{a}\left(\frac{\lambda_\ell}{2^j}\right) \right|^2 + \sum_{k=1}^K \left| \hat{b}^{(k)}\left(\frac{\lambda_\ell}{2^j}\right) \right|^2 = 1. \quad (7)$$

A common example is the Haar-type filter bank (Dong 2017), defined by: $\hat{a}(x) = \cos(x/2)$, $\hat{b}^{(1)}(x) = \sin(x/2)$. Given a signal $f \in \mathbb{R}^N$ defined on the hypergraph, the corresponding framelet coefficients are computed via: $\mathbf{v}_0 = \{\langle \varphi_{j,p}, f \rangle\}_{p \in V}$, $\mathbf{w}_j^k = \{\langle \psi_{j,p,k}, f \rangle\}_{p \in V}$, which can be expressed in matrix form as:

$$\mathbf{v}_0 = \mathbf{U} \hat{a}\left(\frac{\mathbf{\Lambda}}{2}\right) \mathbf{U}^\top f, \quad \mathbf{w}_j^k = \mathbf{U} \hat{b}^{(k)}\left(\frac{\mathbf{\Lambda}}{2^{j+1}}\right) \mathbf{U}^\top f.$$

With the recursive operators defined, the complete framelet transform system can be written as:

$$\begin{aligned} \mathcal{T}_{k,1} &= \mathbf{U} \hat{b}^{(k)}\left(\frac{\mathbf{\Lambda}}{2^M}\right) \mathbf{U}^\top, \quad j = 1 \\ \mathcal{T}_{k,j} &= \mathbf{U} \hat{b}^{(k)}\left(\frac{\mathbf{\Lambda}}{2^{M+j-1}}\right) \hat{a}\left(\frac{\mathbf{\Lambda}}{2^{M+j-2}}\right) \cdots \hat{a}\left(\frac{\mathbf{\Lambda}}{2^M}\right) \mathbf{U}^\top, \\ & \quad j = 2, \dots, J \end{aligned}$$

where M is a constant chosen such that the largest eigenvalue $\lambda_{\max} \leq 2^M \pi$, ensuring that all scaled spectral responses lie within a valid frequency range.

With the recursive operators defined, the complete framelet transform system can be written as:

$$\mathcal{T} = [\mathcal{T}_{0,J}, \mathcal{T}_{1,1}, \dots, \mathcal{T}_{1,J}, \dots, \mathcal{T}_{K,J}] \quad (8)$$

where each $\mathcal{T}_{k,j}$ corresponds to the transform operator for the k -th high-pass filter at scale j , and $\mathcal{T}_{0,J}$ represents the low-pass component at the coarsest level.

To ensure computational efficiency in practical implementations, particularly on large-scale graphs where computing the full eigendecomposition of the Laplacian is infeasible, we adopt Chebyshev polynomial approximation (Hammond, Vandergheynst, and Gribonval 2011) to approximate the spectral filters efficiently. Let \mathcal{C}_r denote the r -th order Chebyshev polynomial. The approximated framelet transform matrix is constructed as:

$$\tilde{\mathcal{T}} = [\tilde{\mathcal{T}}_{0,J}, \tilde{\mathcal{T}}_{1,1}, \dots, \tilde{\mathcal{T}}_{1,J}, \dots, \tilde{\mathcal{T}}_{K,J}] \quad (9)$$

where the approximate filter operators are given by:

$$\tilde{\mathcal{T}}_{k,1} = \mathcal{C}_r(2^{-M}\mathbf{L}),$$

$$\tilde{\mathcal{T}}_{k,j} = \mathcal{C}_r(2^{-M-j+1}\mathbf{L})\mathcal{C}_0(2^{-M-j+2}\mathbf{L}) \cdots \mathcal{C}_0(2^{-M}\mathbf{L}),$$

for $j = 2, \dots, J$.

Based on the constructed framelet system $\tilde{\mathcal{T}}$, a spectral-based hypergraph neural network with multi-frequency filters can be established. The forward propagation at the ℓ -th layer is defined as:

$$\mathbf{X}^{(\ell+1)} = \sigma\left(\tilde{\mathcal{T}}^\top \Theta \tilde{\mathcal{T}} \mathbf{X}^{(\ell)} W_\ell\right)$$

where $\mathbf{X}^{(0)} := \mathbf{X}$ denotes the initial input feature matrix, W_ℓ is a trainable weight matrix, and $\sigma(\cdot)$ represents a non-linear activation function (e.g., ReLU). The diagonal matrix $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_N)$ contains learnable spectral filter coefficients that modulate the contributions of low- and high-frequency components.

4.4 Adaptive Frequency-aware Contrastive Learning Across Channels

To fully exploit the complementary nature of the three frequency-aware views generated by **HyperAim**, we design an adaptive contrastive learning strategy that aligns node representations across channels and perturbations. This strategy combines two complementary components: an InfoNCE-based instance discrimination loss and a pseudo-label-guided binary classification loss. Together, they encourage both intra-view consistency and inter-view semantic alignment across frequency bands.

Contrastive InfoNCE Loss. We first encode node embeddings from the three frequency-specific channels introduced in the previous subsections: the *low-frequency encoder* from the decoupled low-pass polynomial filter (see **Section 4.2**), the *high-frequency encoder* from the high-pass polynomial filter (also in **Section 4.2**), and the *multi-frequency encoder* from the coupled framelet-based spectral module (see **Section 4.3**). Formally, we obtain the following embeddings:

$$\mathbf{z}^{(H)} = \text{Enc}_{\mathcal{H}}(\mathbf{X}), \quad \mathbf{z}^{(L)} = \text{Enc}_{\mathcal{L}}(\mathbf{X}), \quad \mathbf{z}^{(U)} = \text{Enc}_{\mathcal{U}}(\mathbf{X}),$$

where $\mathbf{z}^{(H)}$, $\mathbf{z}^{(L)}$, and $\mathbf{z}^{(U)}$ correspond to the high-pass, low-pass, and framelet-based multi-frequency embeddings, respectively. To obtain a unified spatial representation, the low- and high-frequency views are linearly combined with learnable weights:

$$\mathbf{z}_1 = \frac{\sigma(\alpha\mathbf{z}^{(H)} + \beta\mathbf{z}^{(L)})}{\|\sigma(\alpha\mathbf{z}^{(H)} + \beta\mathbf{z}^{(L)})\|_2}, \quad (10)$$

where α and β are trainable scalars, and $\sigma(\cdot)$ denotes a nonlinear activation function (e.g., ReLU). This spatial fusion captures frequency-coordinated patterns from decoupled branches.

The fused spatial embedding \mathbf{z}_1 is concatenated with the multi-frequency spectral embedding $\mathbf{z}^{(U)}$, and projected into a contrastive space: $\mathbf{z}_2 = \text{Proj}([\mathbf{z}_1; \mathbf{z}^{(U)}])$.

Given an augmented input $\tilde{\mathbf{X}}$, we obtain its projected embedding $\tilde{\mathbf{z}}_2$. The InfoNCE loss is then defined as:

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{n} \sum_{i=1}^n \log \frac{\exp(s_{\text{pos}}^{(i)}/\tau)}{\exp(s_{\text{pos}}^{(i)}/\tau) + \sum_{j \in \mathcal{N}_i} \exp(s_{\text{neg}}^{(i,j)}/\tau)},$$

where $s_{\text{pos}}^{(i)} = \langle \mathbf{z}_2^{(i)}, \tilde{\mathbf{z}}_2^{(i)} \rangle$ denotes the cosine similarity between a node’s representation and its augmented counterpart, and $s_{\text{neg}}^{(i,j)} = \langle \mathbf{z}_2^{(i)}, \tilde{\mathbf{z}}_2^{(j)} \rangle$ measures similarity with negative examples. τ is a temperature parameter, and \mathcal{N}_i denotes the set of negatives for node i .

Binary Cross-Entropy Loss with Pseudo-Labels. To further promote alignment between frequency views, we introduce a supervised signal using pseudo-labels that indicate whether a pair of representations originates from matched or mismatched frequency channels. Positive pairs are constructed by pairing each frequency-specific embedding (e.g., $\mathbf{z}^{(H)}$, $\mathbf{z}^{(L)}$) with the fused multi-frequency anchor \mathbf{z}_2 , while negative pairs are generated through channel mismatch or perturbation.

Each candidate pair is passed through an MLP classifier to predict its alignment label. Let z_i denote the predicted logit for the i -th pair, and $y_i \in \{0, 1\}$ its pseudo-label. The binary cross-entropy loss is then given by:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^n [y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i))],$$

where $\sigma(\cdot)$ denotes the sigmoid activation. This loss explicitly enforces semantic alignment across frequency-aware views.

Total Training Objective. The final training loss is a weighted combination of the two objectives:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \lambda_c \cdot \mathcal{L}_{\text{InfoNCE}}, \quad (11)$$

where λ_c balances the instance-level contrast and frequency-level supervision. Empirical studies on the sensitivity of this hyperparameter are presented in **Section 5.6**. This hybrid learning paradigm enables **HyperAim** to capture both intra-channel invariance and cross-frequency consistency, leading to more generalizable hypergraph representations.

5 Experiments

To systematically evaluate the effectiveness of HyperAim, we design a series of experiments centered around the following five research questions:

- **RQ1:** How does HyperAim perform under the fine-tuning evaluation protocol compared to state-of-the-art hypergraph learning methods?
- **RQ2:** Can HyperAim learn transferable representations that generalize well in the linear evaluation setting?
- **RQ3:** Is HyperAim robust and effective on heterophilic hypergraphs, where connected nodes may have dissimilar features or class labels?
- **RQ4:** What is the contribution of each component in HyperAim’s architecture, including the decoupled filters, framelet-based spectral module, and contrastive objectives?
- **RQ5:** How sensitive is HyperAim to the choice of its key hyperparameters, such as contrastive loss weight and filter configuration?

5.1 Experimental Setups

Datasets. We evaluate HyperAim on 12 publicly available hypergraph datasets spanning a range of domains, including co-citation networks, co-authorship graphs, computer vision tasks, and political and social networks. Specifically, Cora, Citeseer, Pubmed, and Cora-CA are citation-based datasets from (Yadati et al. 2019); IMDB is from (Wang et al. 2019); AMiner is introduced in (Zhang et al. 2019); MN-40 originates from the 3D vision dataset in (Wu et al. 2015); and House is a political network dataset from (Chien et al. 2022). These benchmarks reflect varying graph sizes, sparsity levels, and homophily ratios. Additionally, we include **four heterophilic hypergraph** datasets: Actor, Amazon, Twitch, and Pokec, introduced in (Li et al. 2025b).

Baselines. We compare HyperAim against 17 baseline models, categorized into three groups. The first group includes

Method	Citeseer	Cora	Pubmed	Cora-CA	AMiner	IMDB	MN-40	House	A.R.
R.G.	18.1 (0.9)	17.4 (1.0)	36.0 (0.7)	17.8 (0.7)	10.2 (0.2)	33.9 (0.7)	3.7 (0.2)	26.7 (0.3)	19.0
MLP	32.5 (7.0)	27.9 (7.0)	62.1 (3.7)	34.8 (5.1)	22.3 (1.7)	39.1 (2.4)	89.4 (1.5)	72.2 (3.9)	15.0
HGNN	41.9 (7.8)	50.0 (7.2)	72.9 (5.0)	50.2 (5.7)	30.3 (2.5)	42.2 (2.9)	88.0 (1.4)	52.7 (3.8)	12.0
HyperGCN	31.4 (9.5)	33.1 (10.2)	63.5 (14.4)	37.1 (9.1)	26.4 (3.6)	37.9 (4.5)	55.1 (7.8)	49.8 (3.5)	16.9
HNNH	43.1 (8.7)	50.0 (7.9)	72.1 (5.4)	48.3 (6.2)	30.0 (2.4)	42.3 (3.4)	86.1 (1.6)	49.7 (2.2)	13.5
UniGCN	44.2 (8.1)	49.1 (8.4)	74.4 (3.9)	51.3 (6.3)	32.7 (1.8)	41.6 (3.5)	89.1 (1.0)	51.1 (2.4)	11.1
UniGIN	40.4 (9.1)	47.8 (7.7)	69.8 (5.6)	48.3 (6.1)	30.2 (1.4)	41.4 (2.7)	88.2 (1.8)	51.1 (3.0)	14.4
UniGCNII	44.2 (9.0)	48.5 (7.4)	74.1 (3.9)	54.8 (7.5)	32.5 (1.7)	42.5 (3.9)	90.8 (1.1)	50.8 (4.3)	10.1
AllSet	43.5 (8.0)	47.6 (4.2)	72.4 (4.5)	57.5 (5.7)	29.3 (1.2)	42.3 (2.4)	92.1 (0.6)	54.1 (3.4)	11.0
ED-HNN	40.3 (8.0)	47.6 (7.7)	72.7 (4.7)	54.8 (5.4)	30.0 (2.1)	41.4 (3.0)	90.7 (0.9)	71.3 (3.7)	11.6
PhenomNN	49.8 (9.6)	56.4 (9.6)	76.1 (3.5)	60.8 (6.2)	33.8 (2.0)	44.1 (3.7)	95.9 (0.8)	70.4 (5.6)	4.31
GraphMAE2	41.1 (10.0)	49.3 (8.3)	72.9 (4.2)	55.4 (8.4)	32.8 (1.9)	43.3 (2.7)	90.1 (0.7)	52.8 (3.5)	9.63
MaskGAE	49.6 (10.1)	57.1 (8.8)	72.8 (4.3)	57.8 (5.9)	33.7 (1.6)	44.5 (2.5)	90.0 (0.9)	51.8 (3.3)	7.44
TriCL	51.7 (9.8)	60.2 (7.9)	76.2 (3.6)	64.3 (5.5)	33.1 (2.2)	46.9 (2.9)	90.3 (1.0)	69.7 (4.9)	4.63
HyperGCL	47.0 (9.2)	60.3 (7.4)	76.8 (3.7)	62.0 (5.1)	33.2 (1.6)	43.9 (3.6)	91.2 (0.8)	69.2 (4.9)	4.94
H-GD	45.4 (9.9)	50.6 (8.2)	74.5 (3.5)	58.8 (6.2)	32.6 (2.2)	43.0 (3.3)	90.0 (1.0)	69.7 (5.1)	7.75
HyperGRL	42.3 (9.3)	49.1 (8.8)	73.0 (4.3)	55.8 (8.0)	33.0 (1.8)	43.1 (2.7)	90.1 (0.8)	52.5 (3.3)	9.88
HypeBoy	56.7 (9.8)	62.3 (7.7)	77.0 (3.4)	66.3 (4.6)	34.1 (2.2)	47.6 (2.5)	90.4 (0.9)	70.4 (4.8)	2.94
HyperAim	59.9 (3.1)	65.2 (5.6)	78.1 (2.0)	66.4 (2.6)	43.6 (1.1)	57.1 (1.0)	96.4 (0.6)	74.4 (2.8)	1.00

Table 1: Performance comparison under the **fine-tuning** protocol. **Bold** denotes the best result; underlined denotes the second-best result. A.R. refers to the average rank across all datasets. Values in parentheses indicate the standard deviation.

ten supervised hypergraph learning models, such as GCN, UniGCNII, and recent approaches like ED-HNN (Wang et al. 2023a) and PhenomNN (Wang et al. 2023b). The second group comprises two graph self-supervised learning (SSL) methods originally developed for simple graphs: GraphMAE2 (Hou et al. 2023) and MaskGAE (Li et al. 2023). The third group includes five self-supervised methods designed or adapted for hypergraphs, including HyperGRL (Du et al. 2022), HyperGCL (Wei et al. 2022a), TriCL (Lee and Shin 2023), HypeBoy (Kim et al. 2024), and H-GD (Zheng et al. 2022). For fair comparison, all SSL models adopt GCN (Kipf and Welling 2017) or UniGCNII (Huang and Yang 2021) as backbone encoders, depending on whether they are graph-based or hypergraph-based.

5.2 Evaluation with Fine-tuning Protocol (RQ1)

To assess the quality of representations learned in a self-supervised manner, we follow the standard fine-tuning protocol widely adopted in prior work (Wei et al. 2022a; Lee and Shin 2023). Each model is first pretrained without label supervision and then fine-tuned using a small fraction (1%) of labeled nodes, with 1% for validation and the remaining 98% for testing. We repeat the process across 20 different random splits, each with 5 random initializations, and report the mean and standard deviation.

As shown in Table 1, **HyperAim** consistently achieves state-of-the-art performance across all eight benchmark datasets. Notably, it outperforms the best-performing supervised hypergraph model (e.g., UniGCNII) by large margins, achieving +16.7% on Cora (65.2 vs. 48.5) and +11.1% on AMiner (43.6 vs. 32.5). Compared with recent self-supervised approaches, HyperAim surpasses HypeBoy and TriCL in average accuracy, achieving the best average rank (A.R.). It also attains leading results on all datasets, particu-

larly excelling on structurally diverse or challenging datasets like IMDB (57.1), MN-40 (96.4), and House (74.4). These findings confirm the effectiveness of HyperAim’s adaptive multi-frequency filtering and dual-channel contrastive objectives in producing representations that are highly transferable and robust to dataset heterogeneity.

5.3 Evaluation under Linear Protocol (RQ2)

To further assess the generalization ability of the learned embeddings, we adopt a standard linear evaluation protocol. In this setting, the encoder is frozen after self-supervised pre-training, and a simple linear classifier is trained on top of the fixed node representations.

As shown in Table 2, **HyperAim** achieves the best performance on five out of eight datasets (Cora, AMiner, IMDB, MN-40, and House), and secures the top position in terms of average rank (1.50). It also delivers the second-best accuracy on Cora-CA and Citeseer, and remains competitive on Citeseer, with only a 2.2% gap from the best-performing method. These results clearly indicate that HyperAim can extract general-purpose, linearly separable representations that transfer effectively across diverse downstream tasks. The strong performance under both fine-tuning and linear evaluation protocols underscores the benefit of combining decoupled polynomial filters, coupled framelet-based encoders, and frequency-aware contrastive learning.

5.4 Study on Heterophilic Hypergraphs (RQ3)

To further assess generalization under challenging structural conditions, we evaluate **HyperAim** on four heterophilic hypergraph datasets: Actor, Amazon-Ratings, Twitch-Gamers, and Pokec, using a 10%/10%/80% train/validation/test split, averaged over 20 random runs.

Method	Citeseer	Cora	Pubmed	Cora-CA	AMiner	IMDB	MN-40	House	A.R.
Naive X	27.8 (7.0)	32.4 (4.6)	62.8 (2.8)	31.9 (5.5)	21.4 (1.2)	38.1 (1.9)	<u>91.9</u> (1.1)	<u>71.3</u> (5.4)	6.38
GraphMAE2	29.2 (6.5)	37.5 (7.0)	55.5 (9.5)	38.2 (9.1)	27.3 (2.7)	36.6 (3.5)	89.1 (1.8)	51.7 (3.5)	7.50
MaskGAE	47.2 (11.1)	56.8 (9.3)	62.6 (5.5)	56.0 (4.8)	33.2 (2.0)	44.1 (3.9)	90.5 (0.9)	50.0 (2.8)	5.38
TriCL	53.3 (10.0)	62.1 (8.8)	<u>74.5</u> (4.1)	63.6 (5.2)	<u>35.0</u> (3.6)	48.0 (3.2)	80.0 (5.1)	69.1 (5.5)	3.50
HyperGCL	42.6 (8.6)	61.8 (8.3)	67.6 (8.0)	58.1 (6.3)	<u>33.3</u> (2.2)	47.5 (2.8)	84.1 (2.8)	67.1 (5.4)	4.88
H-GD	35.6 (7.8)	37.6 (6.8)	58.0 (8.2)	48.6 (7.4)	33.8 (5.0)	35.2 (2.9)	76.6 (4.4)	68.3 (5.7)	6.38
HyperGRL	35.3 (8.2)	35.4 (8.8)	50.2 (8.7)	39.4 (8.1)	28.0 (2.8)	34.8 (3.0)	89.4 (1.5)	52.0 (3.7)	7.25
HypeBoy	59.6 (9.9)	63.5 (9.4)	75.0 (3.4)	66.0 (4.6)	34.3 (3.2)	<u>48.8</u> (1.8)	89.2 (2.2)	69.4 (5.4)	2.25
HyperAim	<u>57.4</u> (3.7)	67.7 (4.6)	71.0 (1.9)	<u>64.4</u> (3.9)	43.1 (1.0)	56.4 (0.8)	94.2 (1.9)	72.3 (4.0)	1.50

Table 2: Performance comparison under the **linear** protocol. **Bold** denotes the best result; underlined denotes the second-best result. A.R. refers to the average rank across all datasets.

As shown in Table 3, **HyperAim** consistently outperforms HypeBoy, the strongest baseline, with substantial gains on Actor (+19.06%), Pokec (+6.39%), and Amazon (+2.19%). These results validate the effectiveness of HyperAim’s multi-frequency design in capturing informative signals even when local neighborhoods are less predictive due to feature-label dissimilarity.

Method	Actor	Amazon	Twitch	Pokec
TriCL	64.55 (0.53)	26.43 (0.40)	<u>50.61</u> (0.49)	55.43 (0.87)
GraphMAE2	62.31 (0.47)	26.24 (0.34)	50.03 (0.23)	50.82 (0.52)
HyperGRL	63.28 (0.36)	26.73 (0.27)	50.15 (0.30)	51.61 (0.44)
HypeBoy	61.98 (0.63)	27.04 (0.76)	50.21 (0.64)	50.56 (1.30)
HyperAim	81.04 (0.51)	29.23 (0.31)	51.46 (0.44)	56.95 (0.70)

Table 3: Evaluation on heterophilic hypergraph datasets.

5.5 Ablation Study (RQ4)

VARIANT	Cora	Cora-CA	Pokec	Twitch
w/o HP	61.1 (6.12)	64.2 (3.31)	56.84 (0.62)	51.43 (0.46)
w/o LP	59.8 (5.33)	64.4 (2.83)	56.85 (0.64)	51.37 (0.51)
w/o FRA	62.8 (6.63)	66.2 (3.41)	56.84 (0.81)	51.42 (0.52)
FULL	65.2 (5.64)	66.4 (2.62)	56.95 (0.74)	51.46 (0.41)

Table 4: Ablation study results.

To evaluate the effectiveness of each module in **HyperAim**, we conduct ablation experiments by removing the high-pass spatial filter (w/o HP), the low-pass spatial filter (w/o LP), and the framelet-based encoder (w/o FRA) individually. As shown in Table 4, the full model (FULL) achieves the best performance on all four datasets, confirming the complementary roles of all components. On Cora, removing HP or LP leads to notable accuracy drops (-4.1% and -5.4%), indicating the importance of both spatial filters for multi-frequency representation. Excluding the framelet encoder also degrades performance by 2.4%, validating its contribution. While the margins are smaller on Cora-CA, Pokec, and Twitch, the full model still consistently outper-

forms its ablated variants, highlighting the synergy among the decoupled and coupled frequency channels.

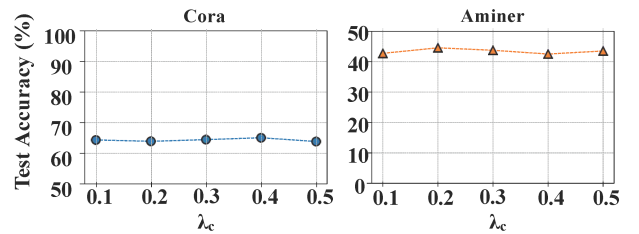


Figure 3: Sensitivity of test accuracy to contrast loss weight λ_c on the Cora and Aminer dataset.

5.6 Hyperparameter Sensitivity Analysis (RQ5)

We analyze the sensitivity of HyperAim to the contrastive loss weight λ_c . Figure 3 presents results on the CORA and AMINER datasets. Across the tested range $[0.1, 0.5]$, the model maintains stable performance with minor fluctuations, indicating that HyperAim is robust to the choice of λ_c . This robustness suggests that the contrastive objective contributes consistently to representation learning without requiring fine-grained tuning of its relative weight.

6 Conclusion

In this work, we proposed HyperAim, a self-supervised hypergraph contrastive learning framework that models frequency diversity through the joint use of decoupled and coupled multi-frequency filtering. The decoupled channels apply predefined polynomial low-pass and high-pass filters to capture frequency-specific signals, while the coupled channel uses framelet transforms to adaptively encode and fuse multi-frequency components. A promising direction for future work is hypergraph contrastive learning under heterophily, where connected nodes often belong to different classes and exhibit feature dissimilarity. Although we report preliminary results on heterophilic hypergraphs, the current framework is not explicitly tailored to this setting; heterophily-aware view generation and frequency-adaptive objectives may further enhance representation learning in such scenarios.

Acknowledgements

This work was supported in part by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (No. 2024C03262), and the National Natural Science Foundation of China (No. 62172370, No. 62536006, No. U21A20473, No. 62576371).

References

- Chen, J.; Lei, R.; and Wei, Z. 2024. PolyGCL: Graph contrastive Learning via learnable spectral polynomial filters. In *ICLR*.
- Chien, E.; Pan, C.; Peng, J.; and Milenkovic, O. 2022. You are allset: A multiset function framework for hypergraph neural networks. In *ICLR*.
- Dong, B. 2017. Sparse representation on graphs by tight wavelet frames and applications. *Applied and Computational Harmonic Analysis*, 42(3): 452–479.
- Dong, Y.; Sawin, W.; and Bengio, Y. 2020. Hnhn: Hypergraph networks with hyperedge neurons. In *ICML Workshop on Graph Representation Learning and Beyond (GRL+)*.
- Du, B.; Yuan, C.; Barton, R.; Neiman, T.; and Tong, H. 2022. Self-supervised hypergraph representation learning. In *2022 IEEE International Conference on Big Data (Big Data)*, 505–514. IEEE.
- Feng, Y.; You, H.; Zhang, Z.; Ji, R.; and Gao, Y. 2019. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 3558–3565.
- Hammond, D. K.; Vandergheynst, P.; and Gribonval, R. 2011. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2): 129–150.
- Hou, Z.; He, Y.; Cen, Y.; Liu, X.; Dong, Y.; Kharlamov, E.; and Tang, J. 2023. GraphMAE2: A Decoding-Enhanced Masked Self-Supervised Graph Learner. In *WWW*.
- Huang, J.; and Yang, J. 2021. Unignn: a unified framework for graph and hypergraph neural networks. In *IJCAI*.
- Kim, S.; Kang, S.; Bu, F.; Lee, S. Y.; Yoo, J.; and Shin, K. 2024. HypeBoy: Generative Self-Supervised Representation Learning on Hypergraphs. In *ICLR*.
- Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.
- Lee, D.; and Shin, K. 2023. I’m me, we’re us, and i’m us: Tri-directional contrastive learning on hypergraphs. In *AAAI*, 8456–8464.
- Li, J.; Wu, R.; Sun, W.; Chen, L.; Tian, S.; Zhu, L.; Meng, C.; Zheng, Z.; and Wang, W. 2023. What’s Behind the Mask: Understanding Masked Graph Modeling for Graph Autoencoders. In *KDD*.
- Li, J.; Zheng, R.; Feng, H.; Li, M.; and Zhuang, X. 2024. Permutation equivariant graph framelets for heterophilous graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9): 11634–11648.
- Li, M.; Fang, Y.; Wang, Y.; Feng, H.; Gu, Y.; Bai, L.; and Lio, P. 2025a. Deep hypergraph neural networks with tight framelets. In *AAAI*, 18385–18392.
- Li, M.; Gu, Y.; Wang, Y.; Fang, Y.; Bai, L.; Zhuang, X.; and Lio, P. 2025b. When hypergraph meets heterophily: New benchmark datasets and baseline. In *AAAI*, 18377–18384.
- Wang, P.; Yang, S.; Liu, Y.; Wang, Z.; and Li, P. 2023a. Equivariant hypergraph diffusion neural operators. In *ICLR*.
- Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference, 2022–2032*.
- Wang, Y.; Gan, Q.; Qiu, X.; Huang, X.; and Wipf, D. 2023b. From hypergraph energy functions to hypergraph neural networks. In *ICML*.
- Wei, T.; You, Y.; Chen, T.; Shen, Y.; He, J.; and Wang, Z. 2022a. Augmentations in hypergraph contrastive learning: Fabricated and generative. *Advances in neural information processing systems*, 35: 1909–1922.
- Wei, T.; You, Y.; Chen, T.; Shen, Y.; He, J.; and Wang, Z. 2022b. Augmentations in Hypergraph Contrastive Learning: Fabricated and Generative. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 1909–1922. Curran Associates, Inc.
- Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; and Xiao, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition, 1912–1920*.
- Xia, L.; Huang, C.; and Zhang, C. 2022. Self-supervised hypergraph transformer for recommender systems. In *KDD*, 2100–2109.
- Yadati, N.; Nimishakavi, M.; Yadav, P.; Nitin, V.; Louis, A.; and Talukdar, P. 2019. HyperGCN: A new method for training graph convolutional networks on hypergraphs. *Advances in neural information processing systems*, 32.
- Zhang, C.; Song, D.; Huang, C.; Swami, A.; and Chawla, N. V. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 793–803.
- Zhang, J.; Gao, M.; Yu, J.; Guo, L.; Li, J.; and Yin, H. 2021. Double-scale self-supervised hypergraph learning for group recommendation. In *CIKM*, 2557–2567.
- Zheng, Y.; Pan, S.; Lee, V.; Zheng, Y.; and Yu, P. S. 2022. Rethinking and scaling up graph contrastive learning: An extremely efficient approach with group discrimination. In *NeurIPS*.