

MOTIF: Motivation-guided Topology Inference for Cold-start Multimodal Recommendation

Yurui Shi^{1†}, Yuchen Miao^{2†[✉]}, Ximing Hu¹, Zijun Wang², and Chang Han²

¹ Taiyuan University of Technology, Taiyuan, China

² Sydney Smart Technology College, Northeastern University, China

[†]Co-first authors; [✉]Corresponding author.

Abstract. Cold-start multimodal recommendation faces three coupled challenges: (i) sparse interactions obscure user intent, (ii) cold items remain topologically isolated, and (iii) similarity-based item graphs may cause semantic drift. To address these issues, we propose MOTIF, a Motivation-guided Topology Inference framework for cold-start multimodal recommendation. MOTIF integrates Semantic Motivation Reasoning, Knowledge-enhanced Graph Reconstruction, Weighted Graph Contrastive Learning, and Semantic-Structural Alignment. It uses offline LLM reasoning to infer motivation semantics, reconstructs transferable item-item topology, and learns robust graph embeddings without injecting generated text into prediction. Experiments on three multimodal benchmarks show consistent gains over graph-based, multimodal, cold-start, and LLM-enhanced baselines, with up to 6.07% relative improvement over the strongest recent baseline.

Keywords: Cold-start Recommendation · Multimodal Recommendation · Graph Contrastive Learning · Large Language Models

1 Introduction

Multimodal recommender systems use item content such as text, images, videos, and attributes to enrich preference modeling. However, cold-start recommendation remains difficult: cold users provide limited behavioral evidence, and cold items have insufficient interaction edges for graph propagation. Effective cold-start multimodal recommendation therefore requires both semantic understanding of sparse evidence and topological reconstruction of incomplete structures.

Existing methods address this problem through graph collaborative filtering [8], graph contrastive learning [25,12,27], multimodal representation learning [23,29,33], cold-start adaptation [20,10,13,31], and LLM-enhanced semantic modeling [7,1,22,17,6]. Yet semantic modeling, graph reconstruction, and representation robustness are often handled separately, which limits their ability to jointly address sparse user intent, cold-item isolation, and noisy multimodal semantics.

Figure 1 illustrates the motivation. Cold-start multimodal methods face limited semantic observability and incomplete topological observability. Sparse interactions may not reveal user motivations, while similarity-based item graphs

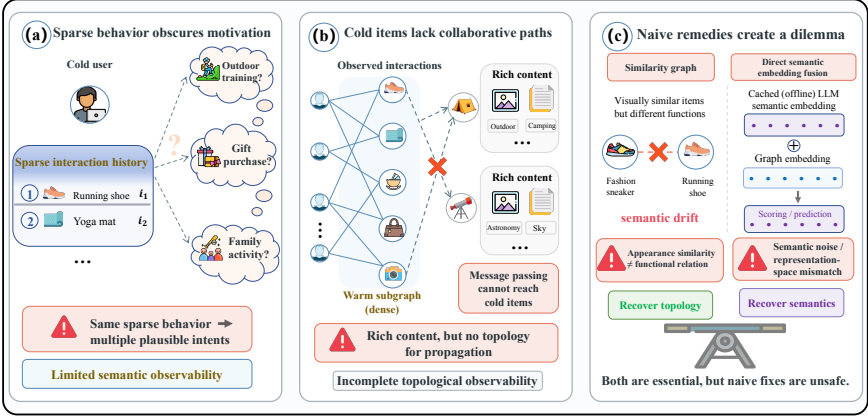


Fig. 1. Motivation of semantic-topological co-reconstruction for cold-start multimodal recommendation.

may connect visually similar but functionally different items, causing semantic drift. Directly feeding LLM-generated text into prediction can also introduce semantic noise. These observations motivate a framework that infers latent motivations, converts them into transferable graph connectivity, and learns robust representations on the reconstructed topology.

To this end, we propose MOTIF, an LLM-enhanced semantic-topological co-reconstruction framework for cold-start multimodal recommendation. MOTIF infers user and item motivations from sparse multimodal contexts, transforms them into knowledge-enhanced item-item topology for cold-item propagation, and learns robust representations through weighted graph contrastive learning and semantic-structural coordination.

The main contributions of this work are summarized as follows:

- We formulate cold-start multimodal recommendation as joint semantic and topological reconstruction over motivations and graph connectivity.
- We use LLM-based motivation reasoning to build cold-start-aware item-item topology for collaborative propagation.
- We integrate graph reconstruction, weighted graph contrastive learning, and semantic-structural alignment into MOTIF for robust sparse-supervision learning.
- We conduct experiments under cold-start multimodal settings to validate MOTIF and its key components.

2 Related Work

2.1 Multimodal Recommendation

Multimodal recommendation incorporates item content such as text, images, video frames, and attributes. MMGCN constructs modality-aware user-item

graphs [23], GRCN refines noisy feedback graphs [24], LATTICE learns modality-aware item graphs [29], and FREEDOM improves robustness by freezing and denoising graph structures [33]. Most methods, however, build item relations from feature similarity or latent correlations [15]. In cold-start scenarios, similar appearance may hide different functions, causing semantic drift. MOTIF instead uses LLM-derived motivation semantics to capture transferable functional relations.

2.2 Graph Collaborative Filtering and Graph Contrastive Learning

Graph collaborative filtering learns representations over user-item graphs. NGCF and LRGCCF use graph neural or residual propagation [21,3], while LightGCN keeps only neighborhood propagation and layer-wise aggregation [8]. Graph contrastive methods such as SGL [25], NCL [12], and SimGCL [27] improve robustness through augmented or perturbed views. However, they usually assume reliable observed neighborhoods, which is weak for sparse users and cold items. WeightedGCL recalibrates final-layer perturbations with squeeze-and-excitation [4]. MOTIF applies this idea on a knowledge-enhanced reconstructed graph for cold-start topology.

2.3 Cold-start Recommendation

Cold-start recommendation mainly relies on side information, representation generation, or rapid adaptation. DropoutNet encourages content-based inference by removing collaborative inputs [20]; MeLU uses few-shot preference estimation [10]; MetaHIN combines heterogeneous information networks with meta-learning [13]; and CVAR generates cold-item representations from side information [31]. These methods rarely reason about latent motivations or reconstruct graph structures for cold-item message passing. MOTIF instead addresses cold-start recommendation from both semantic and topological perspectives.

2.4 LLMs for Recommendation

LLMs provide semantic understanding and reasoning for recommendation [16]. P5 formulates recommendation as text-to-text prediction [7], and TALLRec adapts LLMs through instruction tuning [1]. LLMRec augments profiles and graphs [22], RLMRec aligns LLM-derived semantics with collaborative embeddings [17], and LMMRec extracts fine-grained motivations [6]. The interaction among motivation reasoning, cold-start graph reconstruction, and graph contrastive learning remains underexplored. MOTIF converts LLM-derived motivations into item topology and distills them into graph embeddings without direct semantic injection.

3 Methodology

3.1 Problem Formulation

Let $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$ and $\mathcal{I} = \{i_1, i_2, \dots, i_N\}$ denote the user and item sets. The observed interaction matrix is

$$\mathbf{R} \in \{0, 1\}^{M \times N}, \quad (1)$$

where $R_{ui} = 1$ indicates that user u has interacted with item i . Each item is associated with multimodal information

$$\mathcal{M}_i = \{t_i, v_i, a_i, \dots\}, \quad (2)$$

where t_i , v_i , and a_i denote textual, visual, and attribute-level information.

The goal of top- K recommendation is to learn a scoring function

$$f : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}, \quad (3)$$

which ranks candidate items for each user. In cold-start scenarios, sparse users provide insufficient behavioral evidence, while cold items have limited collaborative connectivity. We formulate this task as a *semantic-topological co-reconstruction* problem: semantic reconstruction infers user and item motivations from sparse multimodal evidence, and topological reconstruction converts such motivations into transferable item-item connectivity. This formulation extends conventional multimodal recommendation by treating multimodal content as motivation-level evidence rather than only auxiliary item features. The reconstructed topology is used as an explicit propagation structure, enabling cold items to receive collaborative signals through semantically and functionally related neighbors.

3.2 Framework Overview

Figure 2 illustrates MOTIF. It contains Semantic Motivation Reasoning, Knowledge-enhanced Graph Reconstruction, Weighted Graph Contrastive Learning, and Semantic-Structural Alignment. LLM-generated semantics guide graph reconstruction and alignment supervision, while final prediction uses only graph embeddings.

3.3 Semantic Motivation Reasoning

Semantic Motivation Reasoning converts sparse multimodal evidence into compact motivation vectors. The LLM is used only offline, and its outputs are encoded before model training.

For item i , the LLM summarizes its functional role, usage scenario, target audience, and possible complementary or substitutable relations from $\mathcal{M}_i$:

$$z_i = g_{\text{LLM}}(\mathcal{M}_i), \quad (4)$$

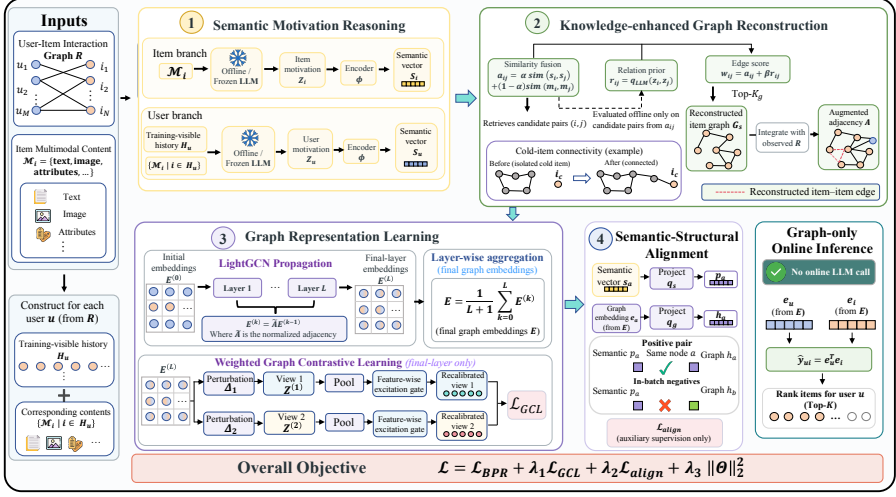


Fig. 2. Overall framework of the proposed semantic-topological co-reconstruction model.

where z_i denotes the generated item motivation text. It is encoded as

$$\mathbf{s}_i = \phi(z_i), \quad \mathbf{s}_i \in \mathbb{R}^{d_s}, \quad (5)$$

where $\phi(\cdot)$ is a text encoder.

For user u , let

$$\mathcal{H}_u = \{i \mid R_{ui} = 1\} \quad (6)$$

denote the training-visible interaction history. To avoid information leakage, only training interactions are used. The user motivation text and semantic vector are obtained by

$$z_u = h_{\text{LLM}}(\mathcal{H}_u, \{\mathcal{M}_i \mid i \in \mathcal{H}_u\}), \quad (7)$$

$$\mathbf{s}_u = \phi(z_u), \quad \mathbf{s}_u \in \mathbb{R}^{d_s}. \quad (8)$$

3.4 Knowledge-enhanced Graph Reconstruction

To improve cold-item connectivity, we reconstruct a semantic-aware item graph by combining raw multimodal similarity, LLM-derived motivation semantics, and relation priors. Let $\mathbf{m}_i$ denote the original multimodal representation of item i . For item pair (i, j) , we compute

$$a_{ij} = \alpha \text{sim}(\mathbf{s}_i, \mathbf{s}_j) + (1 - \alpha) \text{sim}(\mathbf{m}_i, \mathbf{m}_j), \quad (9)$$

where $\text{sim}(\cdot, \cdot)$ denotes cosine similarity, and α balances motivation semantics and raw multimodal similarity.

Since feature similarity may connect items with different functions, we introduce an offline relation prior:

$$r_{ij} = q_{\text{LLM}}(z_i, z_j), \quad (10)$$

where $q_{\text{LLM}}(\cdot, \cdot)$ estimates functional relations such as complementarity, substitutability, or shared usage scenarios. For efficiency, r_{ij} is computed only within a candidate set retrieved by a_{ij} .

The final edge score is

$$w_{ij} = a_{ij} + \beta r_{ij}, \quad (11)$$

and each item retains its top- K_g neighbors:

$$\mathcal{N}_i^g = \text{TopK}_{K_g}(\{w_{ij} \mid j \in \mathcal{I}, j \neq i\}). \quad (12)$$

The reconstructed item graph is

$$\mathcal{G}_s = (\mathcal{I}, \mathcal{E}_s), \quad \mathcal{E}_s = \{(i, j) \mid j \in \mathcal{N}_i^g\}, \quad (13)$$

with weighted adjacency matrix $\mathbf{A}_s$.

We integrate the reconstructed item graph with the original interaction graph:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{R} \\ \mathbf{R}^\top & \mathbf{A}_s \end{bmatrix}. \quad (14)$$

Let $\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ be the normalized adjacency matrix. We adopt LightGCN-style propagation:

$$\mathbf{E}^{(k)} = \tilde{\mathbf{A}} \mathbf{E}^{(k-1)}, \quad k = 1, \dots, L, \quad (15)$$

and aggregate layer-wise embeddings:

$$\mathbf{E} = \frac{1}{L+1} \sum_{k=0}^L \mathbf{E}^{(k)}. \quad (16)$$

The final user and item embeddings are denoted as $\mathbf{e}_u$ and $\mathbf{e}_i$.

3.5 Weighted Graph Contrastive Learning

Weighted Graph Contrastive Learning improves robustness against uncertain reconstructed edges.

Instead of perturbing all graph layers, we perturb only the final-layer representations:

$$\mathbf{E}_1^{(L)} = \mathbf{E}^{(L)} + \Delta_1, \quad \mathbf{E}_2^{(L)} = \mathbf{E}^{(L)} + \Delta_2, \quad (17)$$

where Δ_1 and Δ_2 are stochastic perturbations. The two contrastive views are

$$\mathbf{Z}^{(r)} = \frac{1}{L+1} \left(\sum_{k=0}^{L-1} \mathbf{E}^{(k)} + \mathbf{E}_r^{(L)} \right), \quad r \in \{1, 2\}. \quad (18)$$

For node a in view r , let $\mathbf{z}_a^{(r)} \in \mathbb{R}^d$ denote its representation. We compute a compact feature statistic:

$$c_a^{(r)} = \text{Pool}(\mathbf{z}_a^{(r)}) = \frac{1}{d} \sum_{\ell=1}^d z_{a,\ell}^{(r)}. \quad (19)$$

The excitation network produces a dimension-wise gate:

$$\mathbf{g}_a^{(r)} = \sigma \left(\mathbf{W}_2 \delta \left(\mathbf{W}_1 c_a^{(r)} \right) \right), \quad \mathbf{g}_a^{(r)} \in \mathbb{R}^d. \quad (20)$$

The recalibrated representation is

$$\hat{\mathbf{z}}_a^{(r)} = \mathbf{z}_a^{(r)} \odot \mathbf{g}_a^{(r)}. \quad (21)$$

For node a , the graph contrastive loss is

$$\mathcal{L}_{\text{GCL}} = - \sum_{a \in \mathcal{B}} \log \frac{\exp(\text{sim}(\hat{\mathbf{z}}_a^{(1)}, \hat{\mathbf{z}}_a^{(2)})/\tau)}{\sum_{b \in \mathcal{B}} \exp(\text{sim}(\hat{\mathbf{z}}_a^{(1)}, \hat{\mathbf{z}}_b^{(2)})/\tau)}, \quad (22)$$

where $\mathcal{B}$ is a mini-batch of nodes and τ is the contrastive temperature.

3.6 Semantic-Structural Alignment

LLM-derived semantic vectors and graph embeddings come from different representation spaces. To avoid direct fusion noise, we use an auxiliary alignment objective. For node a , let $\mathbf{s}_a$ and $\mathbf{e}_a$ denote its semantic vector and graph embedding. They are projected into a shared space:

$$\mathbf{p}_a = q_s(\mathbf{s}_a), \quad \mathbf{h}_a = q_g(\mathbf{e}_a), \quad (23)$$

where $q_s(\cdot)$ and $q_g(\cdot)$ are learnable projection functions. The alignment loss is

$$\mathcal{L}_{\text{align}} = - \sum_{a \in \mathcal{B}} \log \frac{\exp(\text{sim}(\mathbf{p}_a, \mathbf{h}_a)/\tau_s)}{\sum_{b \in \mathcal{B}} \exp(\text{sim}(\mathbf{p}_a, \mathbf{h}_b)/\tau_s)}, \quad (24)$$

where τ_s is the alignment temperature.

3.7 Prediction and Optimization

The final recommendation score is computed only from graph embeddings:

$$\hat{y}_{ui} = \mathbf{e}_u^\top \mathbf{e}_i. \quad (25)$$

Given a training triplet (u, i^+, i^-) , where i^+ is an observed item and i^- is a sampled negative item, we adopt the Bayesian Personalized Ranking loss [19]:

$$\mathcal{L}_{\text{BPR}} = - \sum_{(u, i^+, i^-) \in \mathcal{D}} \log \sigma(\hat{y}_{ui^+} - \hat{y}_{ui^-}). \quad (26)$$

The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{BPR}} + \lambda_1 \mathcal{L}_{\text{GCL}} + \lambda_2 \mathcal{L}_{\text{align}} + \lambda_3 \|\Theta\|_2^2. \quad (27)$$

Thus, prediction remains graph-based, while LLM-derived semantics supervise reconstruction and alignment.

4 Experiments

We evaluate MOTIF in terms of overall accuracy, cold-start robustness, component effectiveness, graph quality, hyperparameter sensitivity, and efficiency.

4.1 Experimental Setup

Datasets. We conduct experiments on Amazon-Baby, Amazon-Sports, and MicroLens-50K. Amazon-Baby and Amazon-Sports contain user-item interactions with textual and visual item information, while MicroLens-50K contains short-video interactions with textual and visual-frame information. Interactions are split into training, validation, and test sets. User-side and item-side cold-start settings are constructed from the training graph. User-level LLM reasoning uses only training-visible interactions, and item-level reasoning uses only item-side multimodal content to prevent validation and test leakage.

Table 1. Dataset statistics after preprocessing.

Dataset	#Users	#Items	#Interactions	Sparsity
Amazon-Baby	19 445	7 050	160 792	99.88%
Amazon-Sports	35 598	18 357	296 337	99.95%
MicroLens-50K	50 000	19 220	359 708	99.96%

Cold-start Protocol. For user-side cold-start evaluation, users with 1–2, 3–5, and more than 5 visible training interactions are regarded as extreme-cold, cold, and warm users, respectively. For item-side evaluation, cold items are created by masking their training interactions while keeping multimodal content available. The masked interactions are used only for validation and testing.

Baselines. We compare with representative baselines from five groups. Graph collaborative filtering baselines include MF-BPR [19], NGCF [21], LightGCN [8], UltraGCN [14], LayerGCN [32], and FKAN-GCF [26]. Graph contrastive learning baselines include SGL [25], NCL [12], SimGCL [27], LightGCL [2], DCCF [18], RecDCL [28], BIGCF [30], WeightedGCL [4], and LightCSCF [9]. Multimodal recommendation baselines include MMGCN [23], GRCN [24], LATTICE [29], FREEDOM [33], and CRANE [5]. Cold-start baselines include DropoutNet [20] and CVAR [31]. LLM-enhanced baselines include LLMRec [22], RLMLRec [17], LMMRec [6], and RecGOAT [11].

Evaluation Metrics. We adopt Recall@ K and NDCG@ K with $K \in \{10, 20\}$. For each user, all items not observed in the training set are treated as candidate items. Each experiment is repeated with five random seeds, and the final results are reported as average values. Statistical significance is evaluated by paired t -test.

Implementation Details. All methods are implemented in PyTorch. The embedding dimension is set to 64. The number of graph propagation layers is tuned in $\{1, 2, 3, 4\}$, the learning rate in $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}\}$, and the batch size in $\{1024, 2048, 4096\}$. For our method, LLM reasoning, semantic encoding, and graph reconstruction are performed offline, and no LLM call is used during online inference.

4.2 Overall Performance

Tables 2 and 3 report overall performance against graph-based, multimodal, cold-start, and LLM-enhanced baselines.

Table 2. Overall performance compared with graph-based baselines. Results are averaged over five random seeds; standard deviations are omitted for readability.

Method	Amz-Baby		Amz-Sports		MicroLens	
	R@20	N@20	R@20	N@20	R@20	N@20
<i>Graph Collaborative Filtering</i>						
MF-BPR	0.0487	0.0210	0.0533	0.0241	0.0460	0.0207
NGCF	0.0591	0.0258	0.0650	0.0287	0.0552	0.0249
LightGCN	0.0698	0.0301	0.0782	0.0344	0.0641	0.0290
UltraGCN	0.0734	0.0322	0.0815	0.0363	0.0673	0.0307
LayerGCN	0.0716	0.0314	0.0828	0.0359	0.0688	0.0314
FKAN-GCF	0.0759	0.0331	0.0846	0.0372	0.0700	0.0319
<i>Graph Contrastive Learning</i>						
SGL	0.0741	0.0325	0.0837	0.0369	0.0709	0.0323
NCL	0.0768	0.0338	0.0861	0.0381	0.0724	0.0330
SimGCL	0.0796	0.0349	0.0892	0.0397	0.0743	0.0338
LightGCL	0.0785	0.0344	0.0884	0.0390	0.0751	0.0342
DCCF	0.0812	0.0357	0.0905	0.0403	0.0762	0.0347
RecDCL	0.0827	0.0361	0.0921	0.0408	0.0769	0.0351
BIGCF	0.0834	0.0366	0.0917	0.0412	0.0775	0.0354
WeightedGCL	0.0858	0.0377	0.0948	0.0421	0.0796	0.0363
LightCSCF	0.0871	0.0382	0.0960	0.0427	0.0805	0.0368
MOTIF	0.1083	0.0472	0.1146	0.0508	0.0938	0.0429
Improv.	5.45%	6.07%	5.23%	5.83%	4.92%	5.15%

MOTIF consistently outperforms all baseline groups, validating the benefit of reconstructing cold-start topology and aligning motivation semantics with collaborative graph embeddings.

4.3 Cold-start Performance

To evaluate cold-start robustness, Table 4 reports performance on extreme-cold users, cold users, and cold items. We compare with representative baselines from graph collaborative filtering, graph contrastive learning, multimodal recommendation, and LLM-enhanced recommendation.

MOTIF achieves the best performance across all cold-start settings, indicating that motivation reasoning benefits sparse users and reconstructed topology benefits cold items.

Table 3. Overall performance compared with multimodal, cold-start, LLM-enhanced, and recent baselines. Results are averaged over five random seeds; standard deviations are omitted for readability.

Method	Amz-Baby		Amz-Sports		MicroLens	
	R@20	N@20	R@20	N@20	R@20	N@20
<i>Multimodal Recommendation</i>						
MMGCN	0.0641	0.0277	0.0712	0.0314	0.0598	0.0271
GRCN	0.0739	0.0320	0.0808	0.0357	0.0664	0.0303
LATTICE	0.0859	0.0375	0.0950	0.0424	0.0731	0.0334
FREEDOM	0.0976	0.0420	0.1034	0.0455	0.0784	0.0357
CRANE	0.0993	0.0431	0.1055	0.0468	0.0810	0.0371
<i>Cold-start Recommendation</i>						
DropoutNet	0.0612	0.0264	0.0681	0.0298	0.0570	0.0258
CVAR	0.0708	0.0309	0.0774	0.0341	0.0649	0.0294
<i>LLM-enhanced Recommendation</i>						
LLMRec	0.0934	0.0408	0.1012	0.0447	0.0828	0.0376
RLMRec	0.0961	0.0419	0.1039	0.0459	0.0850	0.0388
LLMRec	0.1008	0.0438	0.1070	0.0473	0.0879	0.0401
RecGOAT	0.1027	0.0445	0.1089	0.0480	0.0894	0.0408
MOTIF	0.1083	0.0472	0.1146	0.0508	0.0938	0.0429
Improv.	5.45%	6.07%	5.23%	5.83%	4.92%	5.15%

Table 4. Cold-start performance comparison in terms of Recall@20 and NDCG@20. Results are averaged over five random seeds; standard deviations are below 0.0016 and omitted for readability.

Method	Amz-Baby		Amz-Sports		MicroLens	
	R@20	N@20	R@20	N@20	R@20	N@20
<i>Extreme-cold users</i>						
LightGCN	0.0291	0.0118	0.0327	0.0138	0.0254	0.0106
WeightedGCL	0.0348	0.0146	0.0389	0.0167	0.0301	0.0129
FREEDOM	0.0412	0.0173	0.0461	0.0198	0.0338	0.0145
LLMRec	0.0467	0.0201	0.0518	0.0227	0.0402	0.0176
LMMRec	0.0521	0.0224	0.0573	0.0251	0.0441	0.0193
MOTIF	0.0628	0.0271	0.0689	0.0304	0.0529	0.0233
Improv.	20.54%	20.98%	20.24%	21.12%	19.95%	20.73%
<i>Cold users</i>						
LightGCN	0.0438	0.0184	0.0489	0.0211	0.0372	0.0162
WeightedGCL	0.0507	0.0216	0.0561	0.0244	0.0428	0.0188
FREEDOM	0.0579	0.0247	0.0638	0.0279	0.0471	0.0207
LLMRec	0.0612	0.0263	0.0681	0.0298	0.0524	0.0231
LMMRec	0.0668	0.0287	0.0729	0.0321	0.0563	0.0249
MOTIF	0.0741	0.0319	0.0808	0.0357	0.0627	0.0277
Improv.	10.93%	11.15%	10.84%	11.21%	11.37%	11.25%
<i>Cold items</i>						
LightGCN	0.0223	0.0091	0.0251	0.0104	0.0198	0.0082
WeightedGCL	0.0276	0.0114	0.0308	0.0131	0.0237	0.0099
FREEDOM	0.0389	0.0164	0.0432	0.0186	0.0314	0.0134
LLMRec	0.0431	0.0183	0.0479	0.0208	0.0367	0.0159
LMMRec	0.0488	0.0209	0.0537	0.0234	0.0408	0.0178
MOTIF	0.0613	0.0264	0.0671	0.0297	0.0512	0.0226
Improv.	25.61%	26.32%	24.95%	26.92%	25.49%	26.97%

4.4 Ablation Study

We conduct ablation studies to verify the contribution of each core component. The variants include removing LLM-based reasoning (**w/o LLM-R**), removing

knowledge-enhanced graph reconstruction (**w/o KGR**), replacing the reconstructed graph with a similarity-based graph (**Similarity-Graph**), removing weighted graph contrastive learning (**w/o WGCL**), removing feature-wise recalibration (**w/o SE**), removing semantic-structural alignment (**w/o Align**), and directly fusing LLM semantic embeddings with graph embeddings for prediction (**Direct-Fusion**).

Table 5. Ablation study on key components. Results are averaged over five random seeds; per-cell standard deviations are below 0.0013 and omitted for readability.

Variant	Amz-Baby		Amz-Sports		MicroLens	
	R@20	N@20	R@20	N@20	R@20	N@20
w/o LLM-R	0.1016	0.0443	0.1080	0.0479	0.0850	0.0389
w/o KGR	0.1006	0.0438	0.1028	0.0456	0.0875	0.0400
Similarity-Graph	0.1030	0.0449	0.1099	0.0487	0.0902	0.0413
w/o WGCL	0.1036	0.0452	0.1101	0.0488	0.0881	0.0403
w/o SE	0.1068	0.0465	0.1133	0.0502	0.0920	0.0421
w/o Align	0.1046	0.0456	0.1113	0.0493	0.0913	0.0417
Direct-Fusion	0.1021	0.0445	0.1093	0.0485	0.0907	0.0415
MOTIF	0.1083	0.0472	0.1146	0.0508	0.0938	0.0429

The ablation results confirm the contribution of each component. Removing LLM reasoning or graph reconstruction degrades performance, and the gap between **Similarity-Graph** and MOTIF shows the benefit of motivation-aware reconstruction. The decrease of **Direct-Fusion** supports using LLM semantics as auxiliary supervision rather than prediction input.

4.5 Reconstructed Graph Analysis

We compare the original interaction graph, a multimodal similarity graph, and the proposed reconstructed graph to examine the quality of the reconstructed topology.

Table 6. Diagnostic analysis of reconstructed item graphs.

Graph Type	Iso. Cold	Cold Deg.	NU@10	Build Time
Original Graph	0.412	0.7	0.083	–
Similarity Graph	0.061	14.3	0.214	86.4s
Reconstructed Graph	0.017	19.8	0.317	173.5s

Table 6 shows that the reconstructed graph reduces isolated cold items, increases their average degree, and achieves higher NU@10, indicating more useful cold-item topology than similarity-based graphs.

4.6 Hyperparameter Sensitivity

We analyze five important hyperparameters: the graph contrastive loss weight λ_1 , the semantic-structural alignment weight λ_2 , the semantic-modality balance coefficient α , the relation-prior weight β , and the number of reconstructed neighbors K_g . Each hyperparameter is varied while the others are fixed to their best validation values.

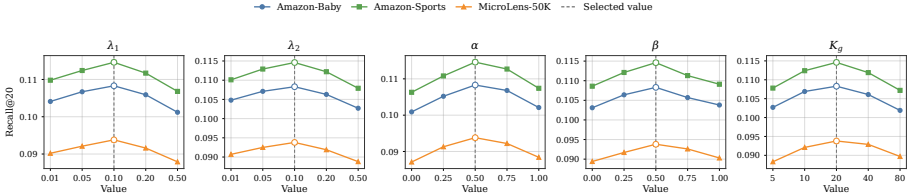


Fig. 3. Hyperparameter sensitivity analysis on Recall@20. Candidate indices correspond to increasing values of each hyperparameter, and the dashed vertical lines indicate the selected values.

Figure 3 shows that moderate hyperparameter values generally perform best. Very small λ_1 or λ_2 weakens contrastive regularization or semantic alignment, whereas overly large values may dominate recommendation supervision. Intermediate α , β , and K_g better balance semantic connectivity and graph noise.

4.7 Efficiency Analysis

We report offline preprocessing cost and online training/inference cost. Offline cost includes LLM reasoning, semantic encoding, and graph reconstruction, while online cost includes trainable parameters, training time, GPU memory, and inference latency. This separation is important because the expensive semantic reasoning stage is executed before training and does not participate in online recommendation.

Table 7. Efficiency comparison. Offline LLM reasoning and graph reconstruction are performed before training and are not used during online inference.

Method	Params	GPU Mem.	Train	Inference	Offline LLM	Recon.
LightGCN	1.4M	2.1GB	11s	0.8s	—	—
WeightedGCL	1.6M	2.4GB	14s	0.9s	—	—
FREEDOM	2.3M	3.7GB	23s	1.1s	—	2.4min
LLMRec	2.1M	3.2GB	19s	1.0s	6.8h	3.1min
LMMRec	2.7M	3.9GB	27s	1.2s	9.4h	4.7min
RecGOAT	3.1M	4.3GB	31s	1.3s	11.2h	5.9min
MOTIF	1.9M	2.8GB	17s	0.9s	8.1h	6.3min

Table 7 shows that MOTIF introduces additional offline cost due to motivation reasoning and topology reconstruction, but these costs are amortized after

preprocessing. Once semantic vectors and the reconstructed graph are cached, the model uses only graph embeddings for prediction, avoids online LLM calls, and keeps inference latency comparable to lightweight graph recommenders.

5 Conclusion

This paper proposed MOTIF, an LLM-enhanced semantic-topological co-reconstruction framework for cold-start multimodal recommendation. MOTIF uses offline LLM reasoning to infer motivation semantics, reconstruct transferable item-item topology, and learn robust graph embeddings through weighted graph contrastive learning and semantic-structural alignment. Experiments on three datasets show consistent gains over graph-based, multimodal, cold-start, and LLM-enhanced baselines; ablations further validate motivation-aware reconstruction and auxiliary alignment. The results show that translating LLM-derived knowledge into graph structure and representation supervision outperforms direct semantic fusion. Prediction uses only graph embeddings, avoiding online LLM calls. Future work will explore adaptive graph priors and more efficient semantic reasoning.

Acknowledgments. This work was supported by the National College Student Innovation and Entrepreneurship Training Program under Project No. 202619145026.

References

1. Bao, K., Zhang, J., Zhang, Y., Wang, W., Feng, F., He, X.: TALLRec: An effective and efficient tuning framework to align large language model with recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems. pp. 1007–1014 (2023)
2. Cai, X., Huang, C., Xia, L., Ren, X.: LightGCL: Simple yet effective graph contrastive learning for recommendation. In: International Conference on Learning Representations (2023)
3. Chen, L., Wu, L., Hong, R., Zhang, K., Wang, M.: Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 27–34 (2020)
4. Chen, Z., Xu, J., Wei, Y., Peng, Z.: WeightedGCL: Squeeze and excitation: A weighted graph contrastive learning for collaborative filtering. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2769–2773 (2025)
5. Dai, J., Fang, Q., Hu, J., Cai, D., Yang, Y., Zhao, C.: CRANE: Cross-modal attention network with dual graph learning in multimodal recommendation. ACM Transactions on Recommender Systems (2026)
6. Di, Y., Zhang, Z., Wang, Y., Liu, J., Yan, J., Wei, J., Chen, X., Liu, Y.: LMM-Rec: LLM-driven motivation-aware multimodal recommendation. arXiv preprint arXiv:2602.05474 (2026)

7. Geng, S., Liu, S., Fu, Z., Ge, Y., Zhang, Y.: P5: Recommendation as language processing: A unified pretrain, personalized prompt and predict paradigm. In: Proceedings of the 16th ACM Conference on Recommender Systems. pp. 299–315 (2022)
8. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 639–648 (2020)
9. Kai, F., Zhang, Y., Wang, K., Sang, L., Zhang, Y.: LightCSCF: Revisiting contrastive learning in collaborative filtering via parallel graph filters. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 14991–14999 (2026)
10. Lee, H., Im, J., Jang, S., Cho, H., Chung, S.: MeLU: Meta-learned user preference estimator for cold-start recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1073–1082 (2019)
11. Li, Y., Ju, H., Song, Z., Yang, W., Lu, C., Jiang, P., Gai, K.: RecGOAT: Graph optimal adaptive transport for LLM-enhanced multimodal recommendation with dual semantic alignment. arXiv preprint arXiv:2602.00682 (2026)
12. Lin, Z., Tian, C., Hou, Y., Zhao, W.X.: NCL: Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In: Proceedings of the ACM Web Conference 2022. pp. 2320–2329 (2022)
13. Lu, Y., Fang, Y., Shi, C.: MetaHIN: Meta-learning on heterogeneous information networks for cold-start recommendation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1563–1573 (2020)
14. Mao, K., Zhu, J., Xiao, X., Lu, B., Wang, Z., He, X.: UltraGCN: Ultra simplification of graph convolutional networks for recommendation. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management. pp. 1253–1262 (2021)
15. Miao, Y.: Universal refinement without interaction: Order-optimal 1-bit mean estimation. arXiv preprint arXiv:2607.24358 (2026)
16. Miao, Y., Cui, M., Zhu, Y., Wang, Y., Xu, S.: R³-REC: Reasoning-driven recommendation via retrieval-augmented LLMs over multi-granular interest signals. In: ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 5951–5955 (2026)
17. Ren, X., Wei, W., Xia, L., Su, L., Cheng, S., Wang, J., Yin, D., Huang, C.: RLM-Rec: Representation learning with large language models for recommendation. In: Proceedings of the ACM Web Conference 2024. pp. 3464–3475 (2024)
18. Ren, X., Xia, L., Zhao, J., Yin, D., Huang, C.: DCCF: Disentangled contrastive collaborative filtering. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1137–1146 (2023)
19. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. pp. 452–461 (2009)
20. Volkovs, M., Yu, G., Poutanen, T.: DropoutNet: Addressing cold start in recommender systems. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
21. Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: NGCF: Neural graph collaborative filtering. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 165–174 (2019)

22. Wei, W., Ren, X., Tang, J., Wang, Q., Su, L., Cheng, S., Wang, J., Yin, D., Huang, C.: LLMRec: Large language models with graph augmentation for recommendation. In: *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. pp. 806–815 (2024)
23. Wei, Y., He, X., Wang, X., Hong, R., Nie, L., Chua, T.S.: MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In: *Proceedings of the 27th ACM International Conference on Multimedia*. pp. 1437–1445 (2019)
24. Wei, Y., Wang, X., Nie, L., He, X., Chua, T.S.: GRCN: Graph-refined convolutional network for multimedia recommendation with implicit feedback. In: *Proceedings of the 28th ACM International Conference on Multimedia*. pp. 3541–3549 (2020)
25. Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., Xie, X.: SGL: Self-supervised graph learning for recommendation. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 726–735 (2021)
26. Xu, J., Chen, Z., Li, J., Yang, S., Wang, W., Hu, X., Ngai, E.C.H.: FourierKAN-GCF: Enhancing graph collaborative filtering with FourierKAN feature transformation. In: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. pp. 5376–5380 (2025)
27. Yu, J., Yin, H., Xia, X., Chen, T., Cui, L., Nguyen, Q.V.H.: SimGCL: Are graph augmentations necessary? simple graph contrastive learning for recommendation. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1294–1303 (2022)
28. Zhang, D., Geng, Y., Gong, W., Qi, Z., Chen, Z., Tang, X., Shan, Y., Dong, Y., Tang, J.: RecDCL: Dual contrastive learning for recommendation. In: *Proceedings of the ACM Web Conference 2024*. pp. 3655–3666 (2024)
29. Zhang, J., Zhu, Y., Liu, Q., Wu, S., Wang, S., Wang, L.: LATTICE: Mining latent structures for multimedia recommendation. In: *Proceedings of the 29th ACM International Conference on Multimedia*. pp. 3872–3880 (2021)
30. Zhang, Y., Sang, L., Zhang, Y.: BIGCF: Exploring the individuality and collectivity of intents behind interactions for graph collaborative filtering. In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1253–1262 (2024)
31. Zhao, X., Ren, Y., Du, Y., Zhang, S., Wang, N.: CVAR: Improving item cold-start recommendation via model-agnostic conditional variational autoencoder. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 2595–2600 (2022)
32. Zhou, X., Lin, D., Liu, Y., Miao, C.: LayerGCN: Layer-refined graph convolutional networks for recommendation. In: *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering*. pp. 1247–1259 (2023)
33. Zhou, X., Shen, Z.: FREEDOM: A tale of two graphs: Freezing and denoising graph structures for multimodal recommendation. In: *Proceedings of the 31st ACM International Conference on Multimedia*. pp. 935–943 (2023)