BEYOND ALIGNMENT: TOWARD STRUCTURAL COMPATIBILITY BETWEEN SEMANTIC AND COLLABORATIVE SPACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Semantic information is widely used to complement collaborative signals in recommender systems, especially to address data sparsity, cold-start, and long-tail recommendation. Existing semantic–collaborative integration methods mainly focus on how to incorporate semantic information through enrichment, alignment, or disentanglement, while paying limited attention to what item–item relations are preserved after fusion. This question matters because the two views organize items differently: collaborative relations are derived from user behavior and shaped by item popularity, leaving tail items with few reliable collaborative neighbors, whereas semantic relations are derived from item content and can provide these items with meaningful neighbors. Consequently, fusion involves a structural trade-off, in which preserving more of one view's structure may weaken that of the other. To study this issue, we introduce *structural compatibility*, which characterizes how well the relational structures of both views are preserved after fusion, and quantify it through three measures: collaborative-neighborhood, semantic-neighborhood, and hierarchical semantic preservation. We further propose **Structure-Preserving Fusion (SPF)**, which injects semantic information into the collaborative space through a controlled residual pathway and preserves semantic structure via hierarchy-aware regularization, thus avoiding global cross-view alignment. Experiments on three datasets show that SPF consistently improves recommendation performance, with larger gains for sparse interactions and tail items. Further analyses reveal that stronger semantic preservation can substantially weaken collaborative-neighborhood preservation without performance gains, indicating that effective fusion requires balancing the preservation of both structures. Overall, our findings establish structural compatibility as a complementary principle for analyzing and designing semantic–collaborative integration.

1 INTRODUCTION

Semantic information has become an important complement to collaborative signals in modern recommender systems, with semantic evidence increasingly derived from knowledge graphs (Wang et al., 2019), multimodal content (Liu et al., 2025a; Wei et al., 2024), large language models (LLMs) (Wu et al., 2024; Ding et al., 2026), and vector-quantized item representations (Hou et al., 2023; Rajput et al., 2023). Such semantic evidence is particularly valuable for mitigating data sparsity and cold-start issues and improving long-tail recommendation (Liu et al., 2025a; Guo et al., 2022). Existing studies integrate this evidence mainly by enriching collaborative representations with auxiliary semantic knowledge (Wang et al., 2019; 2024; Hou et al., 2023), explicitly aligning semantic and collaborative representations (Ren et al., 2024; Ye et al., 2025; Luo et al., 2025; Zheng et al., 2024), or adopting shared–private disentanglement to separate shared from view-specific information (Wang et al., 2026; Yang et al., 2025). Despite their methodological differences, these approaches share an information-centric objective, focusing on what information should be transferred, aligned, shared, or kept view-specific during semantic–collaborative integration (Figure 1(a)).

However, the two views organize the same items in systematically different ways (Figure 1(b)). Collaborative neighborhoods are induced by behavioral co-occurrence (Shenbin & Nikolenko, 2024) and thus inherit the popularity bias of interaction data (Sim & Park, 2026): as Figure 1(c) shows,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

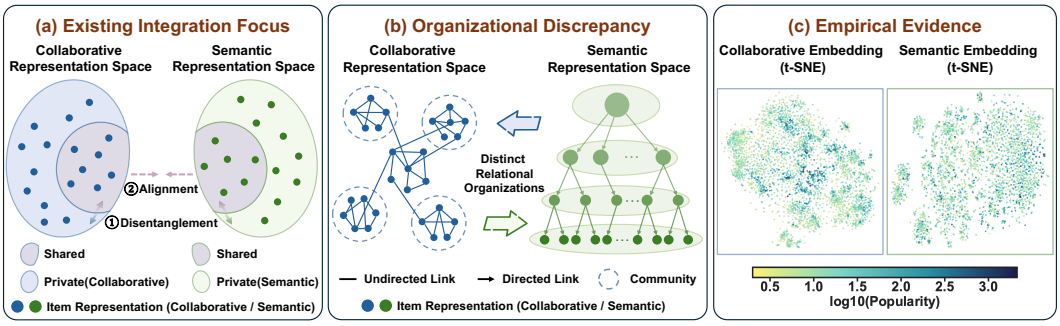


Figure 1: Motivation for structural compatibility. (a) Existing integration methods focus on how semantic information is incorporated. (b) Semantic and collaborative views may organize the same items differently; the semantic tree is illustrative. (c) t-SNE visualizations and mean top-20 neighborhood overlap empirically demonstrate this discrepancy; see Appendix A for details.

collaborative embeddings form a core–periphery structure in which tail items reside at the periphery with few reliable neighbors. Semantic neighborhoods, in contrast, are induced by content and remain largely independent of popularity, supplying precisely the relations that tail items lack. Indeed, on average, only 8.12% of an item's top-20 collaborative neighbors are also among its top-20 semantic neighbors. Fusion therefore implicitly decides which relations survive. Global cross-view alignment assumes the two views share a common relational organization, and thus inevitably reshapes at least one of them: either overriding the semantic relations that tail items depend on, or distorting the behavioral neighborhoods that drive accurate recommendation for popular items. Existing integration strategies mainly govern how semantic information is incorporated, but fail to measure or control which semantic and collaborative relations are retained in the fused space.

To address this gap, we formulate *structural compatibility* as the extent to which the relational organizations of the semantic and collaborative views are retained after fusion. This perspective complements the information-centric objective by making structural preservation and the associated cross-view trade-off explicit targets for modeling and evaluation. We operationalize it via three complementary measures: collaborative-neighborhood preservation, semantic-neighborhood preservation, and hierarchical semantic preservation.

Building on structural compatibility, we propose **Structure-Preserving Fusion (SPF)**, a lightweight framework that combines controlled semantic residual fusion with hierarchy-aware structural regularization. SPF incorporates semantic information through a residual pathway and explicitly constrains the fused space to preserve semantic relational structure, without requiring global cross-view alignment. Experiments on three datasets show that SPF consistently improves recommendation performance, with larger gains under sparse collaborative evidence, particularly for tail recommendation. Across preservation strengths, integration strategies, and collaborative backbones, our analyses further reveal a clear structural trade-off: increasing semantic-structure preservation can substantially reduce collaborative-neighborhood preservation without yielding corresponding recommendation gains. These findings establish structural compatibility as a critical dimension of semantic–collaborative integration, highlighting the importance of balancing semantic and collaborative structure preservation for effective recommendation.

Overall, our contributions can be summarized as follows.

- We introduce a structural perspective on semantic–collaborative integration, termed structural compatibility, and operationalize it through three complementary measures that quantify how collaborative and semantic relational organizations are retained in the fused space.
- We propose SPF, a lightweight integration framework that instantiates structural compatibility through controlled semantic residual fusion and hierarchy-aware structural regularization, without imposing global cross-view alignment.
- We provide systematic empirical evidence for the significance of structural compatibility, showing that effective semantic–collaborative integration requires considering semantic and collaborative structure preservation jointly, particularly when collaborative evidence is sparse.

2 PRELIMINARIES

2.1 PROBLEM FORMULATION

Let $\mathcal{G} = (\mathcal{U} \cup \mathcal{I}, \mathcal{E})$ denote the user–item interaction graph, where $\mathcal{U}$ and $\mathcal{I}$ are the user and item sets, and $\mathcal{E} \subseteq \mathcal{U} \times \mathcal{I}$ contains the observed interactions. For each item $i \in \mathcal{I}$, we consider a collaborative representation $\mathbf{h}_i^{cf}$ learned from interaction data and a semantic representation $\mathbf{h}_i^{sem}$ derived from item-side information. The objective of semantic–collaborative integration is to learn a fusion function $\mathbf{z}_i = f_\theta(\mathbf{h}_i^{cf}, \mathbf{h}_i^{sem}) \in \mathbb{R}^d$ for personalized recommendation.

We study this task in the Semantic-ID setting, where each item is assigned a discrete code sequence $\mathbf{c}_i = (c_i^{(1)}, c_i^{(2)}, \ldots, c_i^{(L)})$, with L denoting the length of the Semantic-ID sequence. The first $L-1$ entries are quantization codes, while the final entry is an appended disambiguation code that distinguishes items with otherwise identical quantization sequences. Shared prefixes of the quantization-code sequences define a hierarchy over items, with longer prefixes corresponding to finer semantic groupings and thereby providing an explicit relational prior (Rajput et al., 2023). In our instantiation, the Semantic-ID codes $\mathbf{c}_i$ and the associated continuous semantic representation $\mathbf{h}_i^{sem}$ are both derived from the same quantization of item-content features.

2.2 STRUCTURAL COMPATIBILITY

Given the fusion formulation above, specifying how semantic and collaborative representations are integrated does not characterize whether the relational organization of each view is preserved after fusion. We therefore formalize this preservation property as *structural compatibility*.

Definition 2.1 (Structural Compatibility)**.** Given semantic and collaborative representations of a common item set, structural compatibility characterizes the extent to which the fused item representations preserve the relational organization of each view.

Because the two views may organize the same item set differently, structural compatibility concerns both the preservation of each organization in the fused space and the trade-offs between semantic and collaborative structure preservation.

2.3 OPERATIONALIZING STRUCTURAL COMPATIBILITY

We operationalize structural compatibility using three complementary metrics that quantify collaborative neighborhood preservation, semantic neighborhood preservation, and hierarchical consistency. For the two neighborhood metrics, we use fixed collaborative and semantic references, denoted by $\bar{\mathbf{h}}_i^{cf}$ and $\bar{\mathbf{h}}_i^{sem}$, rather than the model-side representations $\mathbf{h}_i^{cf}$ and $\mathbf{h}_i^{sem}$ that may change during training. The references are obtained from a separately trained collaborative filtering model and quantized item-content representations, respectively. Appendix C details their construction.

We define the top-k neighborhoods of item i in the collaborative reference space, semantic reference space, and fused space as

$$\begin{aligned} \mathcal{N}_k^{cf}(i) &= \mathrm{TopK}_{j \neq i}\, s^{cf}(\bar{\mathbf{h}}_i^{cf}, \bar{\mathbf{h}}_j^{cf}), \\ \mathcal{N}_k^{sem}(i) &= \mathrm{TopK}_{j \neq i}\, s^{sem}(\bar{\mathbf{h}}_i^{sem}, \bar{\mathbf{h}}_j^{sem}), \\ \widehat{\mathcal{N}}_k(i) &= \mathrm{TopK}_{j \neq i}\, s^{z}(\mathbf{z}_i, \mathbf{z}_j), \end{aligned} \tag{1}$$

where s^{cf}, s^{sem} and s^z denote cosine similarities in their respective spaces. All neighborhoods are constructed over the same item set $\mathcal{I}$, and we set $k = 20$ in our experiments.

Collaborative Structure Preservation (CSP). CSP measures the extent to which the collaborative reference neighborhood is preserved in the fused representation:

$$\mathrm{CSP@}k = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{|\mathcal{N}_k^{cf}(i) \cap \widehat{\mathcal{N}}_k(i)|}{k}. \tag{2}$$

Higher CSP indicates stronger preservation of local relations in the collaborative reference space.

Semantic Neighborhood Preservation (SNP). Similarly, SNP measures the extent to which the semantic reference neighborhood is preserved after fusion:

$$\mathrm{SNP@}k = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{|\mathcal{N}_k^{sem}(i) \cap \widehat{\mathcal{N}}_k(i)|}{k}. \tag{3}$$

Higher SNP indicates stronger neighborhood preservation in the continuous semantic space.

Hierarchical Consistency (HC). HC measures the degree of shared Semantic-ID hierarchy among neighboring items in the fused space. Given the Semantic-ID sequence $\mathbf{c}_i$ introduced above, let $\mathrm{prefix}(i, j)$ denote the length of the longest common prefix between $\mathbf{c}_i$ and $\mathbf{c}_j$. We define

$$\mathrm{HC@}k = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{1}{k} \sum_{j \in \widehat{\mathcal{N}}_k(i)} \frac{\mathrm{prefix}(i, j)}{L - 1}. \tag{4}$$

Higher HC indicates that neighboring items in the fused space share longer semantic prefixes and thus better preserve the discrete semantic hierarchy.

Overall, CSP, SNP, and HC characterize complementary aspects of structural compatibility by capturing collaborative neighborhood preservation, semantic neighborhood preservation, and hierarchical consistency, respectively. A joint analysis of these metrics reveals which relational organizations are retained after fusion and the trade-offs in preserving collaborative and semantic organizations.

3 A Structure-Preserving Fusion Framework

3.1 Framework Overview

Motivated by the structural compatibility perspective, we develop a fusion framework termed *Structure-Preserving Fusion (SPF)*. The framework explicitly incorporates structural preservation into semantic–collaborative integration, enabling systematic analysis of its relationship with recommendation performance. To this end, SPF comprises two complementary components. *Controlled Semantic Residual Fusion* incorporates semantic representations into the collaborative space through a residual pathway, while *Hierarchy-Aware Regularization* encourages the fused space to preserve semantic organization over each item's most-specific prefix neighborhood. Both components provide a concrete instantiation of structural compatibility for semantic–collaborative integration.

3.2 Controlled Semantic Residual Fusion

Given the collaborative representation $\mathbf{h}_i^{cf}$ and the semantic representation $\mathbf{h}_i^{sem}$, SPF constructs the fused representation through a residual integration scheme:

$$\mathbf{z}_i = \mathrm{LayerNorm}\left(\mathbf{h}_i^{cf} + \beta(\mathbf{W}_{sem}\mathbf{h}_i^{sem} + \mathbf{b}_{sem})\right), \tag{5}$$

where $\mathbf{W}_{sem}$ and $\mathbf{b}_{sem}$ project semantic representations into the collaborative embedding space, and the coefficient β controls the contribution of semantic information.

SPF implements semantic–collaborative integration as a lightweight additive composition of the two representations. This deliberately constrained formulation limits the influence of architectural complexity, enabling a more direct assessment of the empirical effect of structural regularization.

3.3 Hierarchy-Aware Regularization

Although residual fusion incorporates semantic information, behavior-driven recommendation optimization may still distort semantic organization in the fused space. We therefore introduce a hierarchy-aware regularizer to preserve the hierarchical relations encoded by Semantic IDs.

For each level $\ell \in \{1, \ldots, L-1\}$, let $P^{(\ell)}(i) = \{\, j \in \mathcal{I} : (c_j^{(1)}, \ldots, c_j^{(\ell)}) = (c_i^{(1)}, \ldots, c_i^{(\ell)}) \,\}$ denote the items sharing the first ℓ codes with item i. Because prefix-group sizes vary across the hierarchy, no fixed level suits all items: shallow levels are coarse, whereas deep levels often yield singletons. We therefore select each item's deepest non-singleton level, $\ell_i^* = \max\{\, \ell : |P^{(\ell)}(i)| > 1 \,\}$, and

define its *most-specific prefix neighborhood* as $\mathcal{N}^{pre}(i) = P^{(\ell_i^*)}(i) \setminus \{i\}$. Items with no non-singleton prefix group are excluded, yielding the anchor set $\mathcal{I}^* = \{i \in \mathcal{I} : \ell_i^* \text{ exists}\}$.

To preserve the most specific Semantic-ID relations shared by multiple items, we use $\mathcal{N}^{pre}(i)$ as the positive neighborhood of each anchor $i \in \mathcal{I}^*$ and formulate the hierarchy-aware regularizer as a neighborhood-level InfoNCE objective:

$$\mathcal{L}_{struct} = -\frac{1}{|\mathcal{I}^*|} \sum_{i \in \mathcal{I}^*} \frac{1}{|\mathcal{N}^{pre}(i)|} \sum_{j \in \mathcal{N}^{pre}(i)} \log \frac{\exp\left(s^z(\mathbf{z}_i, \mathbf{z}_j)/\tau\right)}{\sum_{m \in \mathcal{I} \setminus \{i\}} \exp\left(s^z(\mathbf{z}_i, \mathbf{z}_m)/\tau\right)}, \tag{6}$$

where τ is the temperature parameter. The inner average normalizes each anchor's positive-pair loss by its neighborhood size, preventing anchors with more positives from contributing disproportionately. The objective keeps items that share a most-specific prefix mutually close in the fused space relative to all other items, counteracting the drift that behavioral supervision alone would induce.

3.4 Optimization

The recommendation score is computed as $\hat{y}_{ui} = \mathbf{e}_u^\top \mathbf{z}_i$, where $\mathbf{e}_u$ denotes the user representation learned by the collaborative backbone. The overall objective then jointly balances recommendation learning and structural preservation:

$$\mathcal{L} = \mathcal{L}_{rec} + \gamma \mathcal{L}_{struct}, \tag{7}$$

$$\mathcal{L}_{rec} = -\sum_{(u, i^+, i^-) \in \mathcal{D}} \log \sigma(\hat{y}_{ui^+} - \hat{y}_{ui^-}), \tag{8}$$

where $\mathcal{D}$ denotes the set of training triplets (u, i^+, i^-) with positive item i^+ and sampled negative item i^- for user u, $\sigma(\cdot)$ is the sigmoid function, and γ balances recommendation learning and hierarchy-aware regularization. The recommendation objective $\mathcal{L}_{rec}$ retains behavioral supervision for ranking, while $\mathcal{L}_{struct}$ constrains the fused space to respect each item's most specific prefix neighborhood. Combined with the controlled residual fusion, the joint objective makes structural compatibility an explicit consideration in semantic–collaborative integration.

4 Experiments

In this section, we present our experimental setup to answer the following research questions. **RQ1:** How does SPF perform compared with existing methods? **RQ2:** What does each SPF component contribute to recommendation performance? **RQ3:** How are semantic and collaborative structure preservation related to recommendation performance? **RQ4:** How do SPF's relative performance and structural preservation patterns vary across collaborative backbones?

4.1 Experimental Setup

Datasets. We evaluate on MovieLens-100K[1] and subsets of the Books and Games categories of Amazon Reviews 2023 (Hou et al., 2024), denoted ml-100k, Amazon-Books, and Amazon-Games, respectively. To support evaluation in sparse, long-tail settings, we construct the Amazon subsets via user sampling (see Appendix B for preprocessing details). For each seed, we randomly split each user's interactions into training, validation, and test sets in an $8:1:1$ ratio.

Table 1: Statistics of the datasets.

Dataset	#Users	#Items	#Inter.	Density
ml-100k	943	1,682	100,000	6.30%
Amazon-Books	3,763	3,265	35,617	0.29%
Amazon-Games	5,019	4,193	51,113	0.24%

Baselines. We organize the baselines into four groups: collaborative filtering methods without semantic fusion (CF-only), representation alignment, shared–private disentanglement, and Semantic-ID integration. **(i) Collaborative filtering.** LightGCN (He et al., 2020) is the default backbone, while SGL (Wu et al., 2021), SimGCL (Yu et al., 2022), XSimGCL (Yu et al., 2024), and LightGCL (Cai et al., 2023) test generalization across collaborative architectures. **(ii) Representation**

[1] https://github.com/RUCAIBox/RecSysDatasets

alignment. RLMRec-Con (Ren et al., 2024) incorporates semantic information through an auxiliary representation-alignment objective. **(iii) Shared–private disentanglement.** DaRec (Yang et al., 2025) separates shared and view-specific information in semantic and collaborative representations. **(iv) Semantic-ID integration.** Because generative methods such as TIGER (Rajput et al., 2023) are not directly comparable under full-sort evaluation, we construct a full-sort baseline using SPF's Semantic IDs, backbone, and residual-fusion architecture. It learns per-level code embeddings from interactions but does not use SPF's fixed quantized vectors or hierarchy-aware regularizer.

Evaluation Metrics. We evaluate two complementary aspects. **(i) Recommendation metrics.** We report Recall@20 and NDCG@20 under full-sort evaluation. To examine performance across item popularity levels, we additionally report group-wise Recall@20 for *Head*, *Torso*, and *Tail* items. Items are partitioned by cumulative training-interaction mass: Head and Tail contain the highest- and lowest-frequency items accounting for the top and bottom 20% of training interactions, respectively, while Torso contains the remainder. Group-wise Recall@20 uses the same global top-20 list as overall evaluation, without group-specific filtering or reranking. For Head Recall@20, the ground truth contains only Head test items, and users without Head test items are excluded. Torso Recall@20 and Tail Recall@20 are computed in the same way. **(ii) Structural metrics.** We report CSP@20, SNP@20, and HC@20 as defined in Section 2. All three metrics are computed offline after training and are used only for evaluation.

Implementation Details. All methods use 64-dimensional latent representations, two-layer collaborative encoders, identical per-seed data splits, and a shared training and evaluation protocol. Item texts are encoded offline with the pretrained all-MiniLM-L6-v2 model, and a dataset-specific RQ-VAE produces Semantic IDs and quantized vectors that remain fixed during training. SPF uses $\beta = 1.0$, structural-loss temperature $\tau = 0.2$, and regularization coefficient γ=0.01 across datasets and backbones. For each self-supervised collaborative backbone, its auxiliary-objective coefficient is selected on Amazon-Books by five-seed mean validation Recall@20 using CF-only models and then fixed across integration methods. Models are implemented in PyTorch with RecBole 1.0.1 (Zhao et al., 2021). All results are averaged over five seeds; recommendation metrics are reported as means $\pm$ standard deviations and structural metrics as means. Full configurations and reproduction instructions are provided in Appendix C and the anonymous repository[2].

4.2 Results

Overall Recommendation Performance (RQ1). Table 2 reports the recommendation performance of SPF and all baselines across the three datasets. SPF achieves the highest Recall@20 and NDCG@20 on all three datasets, improving Recall@20 over the strongest baseline by approximately 11% on Books and 9% on Games. Among the baselines incorporating semantic information, DaRec and RLMRec-Con underperform LightGCN across all three datasets, while Semantic-ID improves on certain datasets. This suggests that recommendation gains depend not simply on adding semantic information, but on how semantic and collaborative signals are integrated. SPF performs particularly well on the two sparse Amazon datasets, achieving the highest Tail Recall@20, whereas its gains on the denser ml-100k dataset are mainly reflected in overall ranking. This pattern demonstrates the effectiveness of its semantic–collaborative integration across varying sparsity levels, with greater benefits for tail recommendation when collaborative evidence is limited.

Ablation Study and Regularization Effects (RQ2). Having established the overall performance of SPF, we next examine the contributions of its two components. Table 3 isolates their effects through the progression from LightGCN to Controlled Semantic Residual Fusion (RF) and full SPF. On the recommendation metrics, RF improves ranking performance on two datasets, while adding structural regularization produces further gains across all three, with larger improvements under sparse collaborative evidence. The tail results follow a similar pattern, indicating that explicit structural preservation is particularly useful for items with limited interaction support. On the structural metrics, adding $\mathcal{L}_{\text{struct}}$ increases SNP@20 and HC@20 relative to the RF variant but lowers CSP@20. Thus, the regularizer does not simply provide additional supervision; it systematically shifts the fused representation toward stronger semantic neighborhood and hierarchical preservation, accompanied by reduced collaborative-neighborhood preservation. For the full SPF configuration,

[2] https://anonymous.4open.science/r/azxyq/

Table 2: Recommendation performance with a shared LightGCN backbone. For ranking metrics, bold/underline mark the best/second-best results within each dataset block; $*$ denotes $p < 0.05$ in a two-sided paired t-test over five seeds against the strongest baseline in the same column.

Dataset	Method	Family	R@20	N@20	Tail R@20
ml-100k *(dense; density 6.30%)*	LightGCN	CF	0.340±.042	0.289±.038	0.018±.013
	DaRec	disent.	0.334±.007	0.282±.007	0.022±.009
	RLMRec-Con	align	0.101±.005	0.074±.004	**0.057**±.011
	Semantic-ID	Sem-ID	0.360±.003	0.296±.005	0.046±.006
	SPF	ours	**0.367***±.005	**0.305***±.005	0.038±.002
Amazon-Books *(sparse; density 0.29%)*	LightGCN	CF	0.145±.004	0.070±.001	0.065±.007
	DaRec	disent.	0.106±.016	0.046±.008	0.016±.007
	RLMRec-Con	align	0.094±.035	0.043±.018	0.061±.011
	Semantic-ID	Sem-ID	0.198±.004	0.095±.002	0.106±.010
	SPF	ours	**0.219***±.008	**0.107***±.004	**0.126***±.012
Amazon-Games *(sparse; density 0.24%)*	LightGCN	CF	0.116±.016	0.050±.007	0.019±.002
	DaRec	disent.	0.075±.003	0.032±.002	0.004±.002
	RLMRec-Con	align	0.045±.002	0.019±.001	0.033±.010
	Semantic-ID	Sem-ID	0.112±.005	0.049±.003	0.030±.004
	SPF	ours	**0.127**±.003	**0.055**±.002	**0.044**±.003

Table 3: SPF component ablation with a LightGCN backbone. ✓ denotes enabled components.

Dataset	LightGCN	RF	$\mathcal{L}_{struct}$	R@20	N@20	Tail R@20	CSP@20	SNP@20	HC@20
ml-100k	✓	–	–	0.340±.042	0.289±.038	0.018±.013	–	–	–
	✓	✓	–	0.364±.004	0.300±.006	**0.038**±.009	**0.397**	0.074	0.041
	✓	✓	✓	**0.367**±.005	**0.305***±.005	**0.038***±.002	0.311	**0.197**	**0.109**
Amazon-Books	✓	–	–	0.145±.004	0.070±.001	0.065±.007	–	–	–
	✓	✓	–	0.206±.007	0.100±.003	0.107±.014	**0.296**	0.082	0.058
	✓	✓	✓	**0.219***±.008	**0.107***±.004	**0.126***±.012	0.255	**0.183**	**0.128**
Amazon-Games	✓	–	–	0.116±.016	0.050±.007	0.019±.002	–	–	–
	✓	✓	–	0.115±.005	0.051±.003	0.027±.004	**0.311**	0.064	0.065
	✓	✓	✓	**0.127**±.003	**0.055***±.002	**0.044***±.003	0.254	**0.172**	**0.154**

this structural shift is accompanied by improved recommendation performance, indicating complementary roles for residual fusion and structural regularization.

The ablation above examines the effect of introducing structural regularization; Figure 2 further investigates how its strength affects recommendation performance and structural preservation by varying γ. A moderate preservation strength yields the best overall recommendation performance, whereas stronger regularization continues to increase SNP@20 and HC@20 while reducing CSP@20 and eventually degrading Recall@20. This systematic shift shows that γ provides direct control over the trade-off between semantic and collaborative structure preservation. Notably, on Amazon-Games, the sparsest of the three evaluation datasets, Tail Recall@20 continues to improve even after overall Recall@20 begins to decline. This divergence indicates that stronger semantic preservation affects different popularity groups differently, benefiting low-frequency items without necessarily improving the global ranking. Overall, the relationship between preservation strength and recommendation performance is non-monotonic and differs between overall and tail ranking.

Structural Preservation and Compatibility Analysis (RQ3). RQ2 characterizes the trade-off between semantic and collaborative structure preservation within SPF. We next extend this analysis across integration strategies by comparing SPF with RLMRec-Con, which exhibits a contrasting preservation profile with substantially higher SNP@20 but markedly lower CSP@20. Table 4 compares structural preservation and recommendation performance across RLMRec-Con alignment strengths, with SPF at its default γ=0.01 included for comparison. Introducing alignment sharply increases SNP@20 and HC@20 while substantially reducing CSP@20 compared with the alignment-free setting, but further increasing the alignment strength provides no consistent improvement in

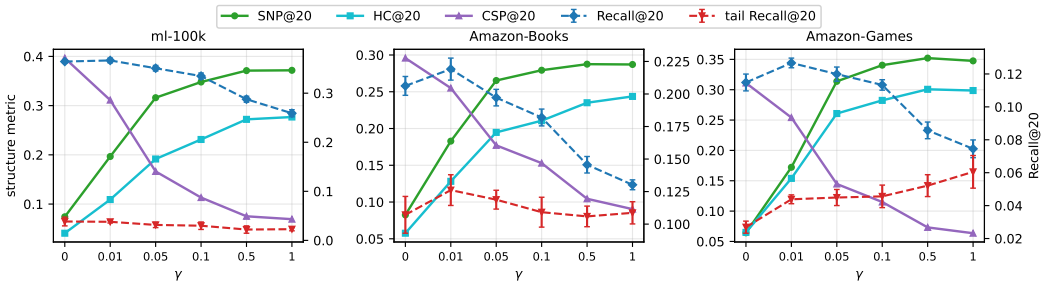


Figure 2: Effect of preservation strength γ on structural metrics (left axis) and Recall@20 and Tail Recall@20 (right axis), with the three datasets ordered by density. Lines show five-seed means; error bars for the recommendation metrics indicate $\pm$ one standard deviation.

Table 4: Structural preservation and recommendation performance under varying RLMRec-Con alignment weights w with a LightGCN backbone. SPF at default γ=0.01 is included for comparison.

Dataset	Method	w	R@20	Head R@20	Tail R@20	CSP@20	SNP@20	HC@20
Amazon-Books	RLMRec-Con	0.0	0.144±.008	**0.419**±.029	0.062±.005	**0.283**	0.063	0.045
		0.05	0.162±.046	0.320±.132	0.065±.008	0.128	0.478	0.202
		0.1	0.094±.035	0.131±.099	0.061±.011	0.080	0.537	0.220
		0.5	0.077±.003	0.079±.010	0.063±.012	0.074	**0.549**	**0.221**
		1.0	0.076±.004	0.078±.011	0.063±.012	0.073	**0.549**	**0.221**
	SPF (γ=0.01)	–	**0.219***±.008	0.410±.015	**0.126***±.012	0.255	0.183	0.128
Amazon-Games	RLMRec-Con	0.0	0.115±.016	**0.345**±.027	0.017±.002	**0.401**	0.050	0.052
		0.05	0.086±.034	0.202±.133	0.025±.010	0.083	**0.561**	**0.260**
		0.1	0.045±.002	0.065±.026	0.033±.010	0.053	0.535	0.246
		0.5	0.045±.002	0.071±.023	0.027±.012	0.055	0.517	0.246
		1.0	0.044±.002	0.070±.023	0.027±.012	0.055	0.517	0.246
	SPF (γ=0.01)	–	**0.127**±.003	0.294±.008	**0.044**±.003	0.254	0.172	0.154

either overall or tail recommendation. In contrast, SPF achieves higher overall and Tail Recall@20 than RLMRec-Con under all tested alignment weights, while retaining substantially more collaborative structure than the alignment-enabled settings, despite exhibiting lower semantic-structure preservation. We argue that this advantage stems from SPF incorporating semantic relations while preserving more of the interaction-derived organization, thereby avoiding the substantial loss of collaborative locality induced by stronger alignment. This indicates that recommendation performance depends on the joint preservation of semantic and collaborative structures, not semantic preservation strength alone. A complementary visualization is provided in Appendix D.

Generalization across Collaborative Backbones (RQ4). To examine whether the recommendation and structural patterns observed under LightGCN persist across different collaborative backbones, we extend the comparison to five collaborative encoders on Amazon-Books. Table 5 summarizes the recommendation and structural results of the four integration methods across these backbones. Against Semantic-ID, SPF achieves stronger semantic-neighborhood and hierarchical preservation while maintaining comparable collaborative preservation. RLMRec-Con instead attains much higher semantic preservation but retains substantially less collaborative locality, together with lower recommendation performance. DaRec exhibits relatively weak preservation of both semantic and collaborative organization and correspondingly weaker recommendation performance. Importantly, these relative structural and recommendation patterns recur across all five backbones, with SPF achieving the highest Tail Recall@20 with each collaborative backbone. These results indicate that both the recommendation benefits and structural preservation behavior of SPF extend beyond a particular collaborative encoder.

Table 5: Backbone robustness on Amazon-Books. Each block shares one collaborative backbone.

Backbone	Integration	Group R@20			Overall		Structure		
		Tail	Torso	Head	R@20	N@20	SNP@20	CSP@20	HC@20
LightGCN	+RLMRec-Con	0.061±.011	0.097±.030	0.131±.099	0.094±.035	0.043±.018	**0.537**	0.080	**0.220**
	+DaRec	0.016±.007	0.060±.016	0.347±.029	0.106±.016	0.046±.008	0.041	0.073	0.038
	+Semantic-ID	0.106±.010	0.176±.004	0.373±.019	0.198±.004	0.095±.002	0.112	**0.270**	0.085
	+SPF	**0.126***±.012	**0.190***±.007	**0.410***±.015	**0.219***±.008	**0.107***±.004	0.183	0.255	0.128
SGL	+RLMRec-Con	0.069±.012	0.090±.010	0.093±.012	0.085±.010	0.039±.004	**0.538**	0.091	**0.219**
	+DaRec	0.022±.006	0.067±.007	**0.323**±.013	0.107±.007	0.046±.003	0.047	0.103	0.043
	+Semantic-ID	0.134±.012	0.185±.004	0.255±.013	0.186±.004	0.092±.003	0.103	0.446	0.076
	+SPF	**0.156***±.013	**0.198***±.006	0.284±.015	**0.206***±.008	**0.102***±.003	0.148	**0.460**	0.102
SimGCL	+RLMRec-Con	0.064±.010	0.087±.004	0.095±.013	0.083±.005	0.038±.004	**0.508**	0.097	**0.215**
	+DaRec	0.020±.006	0.072±.010	0.349±.023	0.115±.008	0.049±.004	0.045	0.111	0.040
	+Semantic-ID	0.128±.012	0.181±.004	0.318±.014	0.195±.005	0.096±.002	0.101	**0.434**	0.074
	+SPF	**0.146***±.012	**0.193***±.006	**0.364**±.021	**0.217***±.007	**0.107***±.004	0.158	0.433	0.109
XSimGCL	+RLMRec-Con	0.068±.009	0.103±.027	0.125±.061	0.098±.026	0.045±.015	**0.404**	0.116	**0.191**
	+DaRec	0.030±.008	0.082±.010	**0.322**±.025	0.116±.005	0.051±.003	0.047	0.115	0.041
	+Semantic-ID	0.148±.010	0.178±.008	0.245±.010	0.184±.004	0.092±.003	0.098	0.464	0.071
	+SPF	**0.161***±.014	**0.190**±.005	0.294±.023	**0.205***±.008	**0.103***±.004	0.124	**0.490**	0.085
LightGCL	+RLMRec-Con	0.057±.012	0.145±.004	0.298±.011	0.152±.004	0.071±.002	**0.532**	0.120	**0.215**
	+DaRec	0.022±.007	0.075±.006	0.345±.014	0.116±.004	0.051±.004	0.052	0.126	0.045
	+Semantic-ID	0.110±.005	0.165±.005	0.353±.007	0.189±.003	0.091±.002	0.129	**0.349**	0.100
	+SPF	**0.127***±.010	**0.176**±.007	**0.368***±.007	**0.203***±.007	**0.100***±.004	0.192	0.335	0.133

5 Related Work

Existing studies on semantic-enhanced recommendation broadly fall into three categories. The first category enriches collaborative recommendation with external semantic signals, such as knowledge graphs, multimodal content, or vector-quantized item semantics, to alleviate sparsity and improve representation quality (Wang et al., 2019; Guo et al., 2022; Hou et al., 2023; Liu et al., 2025a;b). The second category emphasizes explicit cross-view alignment, where semantic and collaborative representations are mapped into a shared latent space through contrastive or matching objectives (Ren et al., 2024; Zheng et al., 2024; Luo et al., 2025; Ye et al., 2025). The third category questions whether global alignment is sufficient and argues that shared and private information should be separated rather than collapsed into a single view (Wang et al., 2026; Yang et al., 2025). Across these lines of work, a central question is how semantic information should be incorporated into collaborative recommendation, whether through enrichment, alignment, or shared–private disentanglement.

Our work instead examines what happens to the relational organization of each view after integration. Semantic and collaborative representations can organize the same items according to different relational principles, yet these organizations may be reshaped when the two representations are integrated. We therefore study which semantic and collaborative relations are retained after fusion and how their preservation relates to recommendation performance. In this sense, SPF turns structural preservation into an explicit modeling and evaluation target for semantic–collaborative integration.

6 Conclusion

This study shows that relational-organization preservation is an important consideration in semantic–collaborative integration. We formalize this property as structural compatibility, introduce complementary measures for jointly assessing preservation across the two views, and develop Structure-Preserving Fusion (SPF) as a structure-preserving integration framework. Experiments indicate that SPF is particularly effective when collaborative evidence is sparse, especially for tail items. Sensitivity and cross-backbone analyses further demonstrate that balancing semantic and collaborative structure preservation is more beneficial for recommendation than simply strengthening semantic preservation. Taken together, these findings establish structural compatibility as a useful principle for understanding and designing semantic–collaborative integration. While instantiated here with Semantic-ID hierarchies, extending structural compatibility to other semantic organizations and more adaptive preservation mechanisms remains an important direction for future work.

AI USE STATEMENT

In this work, generative AI tools were used solely for translation and language editing to improve the clarity and readability of the manuscript. All AI-assisted text was reviewed and verified by the authors, with any identified inaccuracies or inconsistencies manually corrected. The authors take full responsibility for the final content of the manuscript.

REPRODUCIBILITY STATEMENT

We support reproducibility at both the manuscript and artifact levels. Section 2 defines structural compatibility and its three evaluation measures, Sections 3.2 and 3.3 specify the two components of SPF, and Section 4.1 describes the datasets, splits, baselines, evaluation protocol, implementation framework, and five-seed reporting procedure. The appendices provide further details on the t-SNE and neighborhood-overlap analysis in Figure 1(c) (Appendix A), dataset construction and metadata preprocessing (Appendix B), semantic-asset generation, seed-matched collaborative references, optimization and model configurations, and sources of run-to-run variability (Appendix C), as well as parameter sensitivity (Appendix D). An anonymized code repository with reproduction instructions is available at `https://anonymous.4open.science/r/azxyq/`. It includes implementations of SPF and the baselines, experiment configurations, environment specifications, and scripts for dataset construction, semantic-asset generation and index alignment, model training, recommendation and structural evaluation, and result aggregation. Raw third-party datasets are not redistributed; instead, the repository links to their official sources and provides deterministic preprocessing procedures with fixed random seeds, enabling reconstruction of the experimental data and reported results under the original terms of use.

REFERENCES

Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. Lightgcl: Simple yet effective graph contrastive learning for recommendation. In *The Eleventh International Conference on Learning Representations, ICLR 2023*. OpenReview.net, 2023.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2318–2335, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.137. URL `https://aclanthology.org/2024.findings-acl.137/`.

Zhihao Ding, Jinming Li, Shuai Mu, and Jieming Shi. Token-efficient long-term interest sketching and internalized reasoning for llm-based recommendation. In *International Conference on Learning Representations*, volume 2026, pp. 124150–124170, 2026.

Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A survey on knowledge graph-based recommender systems. *IEEE Trans. Knowl. Data Eng.*, 34(8): 3549–3568, 2022. doi: 10.1109/TKDE.2020.3028705.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 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, SIGIR 2020*, pp. 639–648. ACM, 2020. doi: 10.1145/3397271.3401063.

Yupeng Hou, Zhankui He, Julian J. McAuley, and Wayne Xin Zhao. Learning vector-quantized item representation for transferable sequential recommenders. In *Proceedings of the ACM Web Conference 2023, WWW 2023*, pp. 1162–1171. ACM, 2023.

Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*, 2024.

Qidong Liu, Jiaxi Hu, Yutian Xiao, Xiangyu Zhao, Jingtong Gao, Wanyu Wang, Qing Li, and Jiliang Tang. Multimodal recommender systems: A survey. *ACM Comput. Surv.*, 57(2):26:1–26:17, 2025a. doi: 10.1145/3695461.

Shuo Liu, An Zhang, Guoqing Hu, Hong Qian, and Tat-seng Chua. Preference diffusion for recommendation. In *International Conference on Learning Representations*, volume 2025, pp. 79844–79881, 2025b.

Xinchen Luo, Jiangxia Cao, Tianyu Sun, Jinkai Yu, Rui Huang, Wei Yuan, Hezheng Lin, Yichen Zheng, Shiyao Wang, Qigen Hu, Changqing Qiu, Jiaqi Zhang, Xu Zhang, Zhiheng Yan, Jingming Zhang, Simin Zhang, Mingxing Wen, Zhaojie Liu, and Guorui Zhou. QARM: quantitative alignment multi-modal recommendation at kuaishou. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM 2025*, pp. 5915–5922. ACM, 2025. doi: 10.1145/3746252.3761502.

Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Hulikal Keshavan, Trung Vu, Lukasz Heldt, Lichan Hong, Yi Tay, Vinh Q. Tran, Jonah Samost, Maciej Kula, Ed H. Chi, and Mahesh Sathiamoorthy. Recommender systems with generative retrieval. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023*, 2023.

Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. Representation learning with large language models for recommendation. In *Proceedings of the ACM Web Conference 2024, WWW 2024*, pp. 3464–3475. ACM, 2024. doi: 10.1145/3589334.3645458.

Ilya Shenbin and Sergey Nikolenko. Implicitslim and how it improves embedding-based collaborative filtering. In *International Conference on Learning Representations*, volume 2024, pp. 2946–2967, 2024.

Wooseok Sim and Hogun Park. Low-pass personalized subgraph federated recommendation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 11488–11518, 2026.

Maolin Wang, Dongze Wu, Jianing Zhou, Hongyu Chen, Beining Bao, Yu Jiang, Chenbin Zhang, Chang Wang, Jian Liu, and Lei Sha. Rethinking semantic–collaborative integration: Why alignment is not enough. *arXiv preprint arXiv:2604.22195*, 2026.

Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019*, pp. 950–958. ACM, 2019.

Yuhan Wang, Qing Xie, Mengzi Tang, Lin Li, Jingling Yuan, and Yongjian Liu. A dual perspective framework of knowledge-correlation for cross-domain recommendation. *ACM Trans. Knowl. Discov. Data*, 18(6), April 2024. ISSN 1556-4681. doi: 10.1145/3652520.

Tianxin Wei, Bowen Jin, Ruirui Li, Hansi Zeng, Zhengyang Wang, Jianhui Sun, Qingyu Yin, Hanqing Lu, Suhang Wang, Jingrui He, et al. Towards unified multi-modal personalization: Large vision-language models for generative recommendation and beyond. In *International Conference on Learning Representations*, volume 2024, pp. 9912–9930, 2024.

Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 726–735. ACM, 2021. doi: 10.1145/3404835.3462862.

Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, and Enhong Chen. A survey on large language models for recommendation. *World Wide Web (WWW)*, 27(5):60, 2024. doi: 10.1007/S11280-024-01291-2.

Xihong Yang, Heming Jing, Zixing Zhang, Jindong Wang, Huakang Niu, Shuaiqiang Wang, Yu Lu, Junfeng Wang, Dawei Yin, Xinwang Liu, En Zhu, Defu Lian, and Erxue Min. DaRec: a disentangled alignment framework for large language model and recommender system. In *41st IEEE International Conference on Data Engineering, ICDE 2025*, pp. 904–917. IEEE, 2025. doi: 10.1109/ICDE65448.2025.00073.

Wencai Ye, Mingjie Sun, Shaoyun Shi, Peng Wang, Wenjin Wu, and Peng Jiang. DAS: dual-aligned semantic ids empowered industrial recommender system. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM 2025*, pp. 6217–6224. ACM, 2025.

Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. Are graph augmentations necessary?: Simple graph contrastive learning for recommendation. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1294–1303. ACM, 2022. doi: 10.1145/3477495.3531937.

Junliang Yu, Xin Xia, Tong Chen, Lizhen Cui, Nguyen Quoc Viet Hung, and Hongzhi Yin. Xsimgcl: Towards extremely simple graph contrastive learning for recommendation. *IEEE Trans. Knowl. Data Eng.*, 36(2):913–926, 2024. doi: 10.1109/TKDE.2023.3288135. URL https://doi.org/10.1109/TKDE.2023.3288135.

Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, Yingqian Min, Zhichao Feng, Xinyan Fan, Xu Chen, Pengfei Wang, Wendi Ji, Yaliang Li, Xiaoling Wang, and Ji-Rong Wen. RecBole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pp. 4653–4664. ACM, 2021. doi: 10.1145/3459637.3482016.

Bowen Zheng, Yupeng Hou, Hongyu Lu, Yu Chen, Wayne Xin Zhao, Ming Chen, and Ji-Rong Wen. Adapting large language models by integrating collaborative semantics for recommendation. In *40th IEEE International Conference on Data Engineering, ICDE 2024*, pp. 1435–1448. IEEE, 2024. doi: 10.1109/ICDE60146.2024.00118.

A t-SNE Visualization and Neighborhood Overlap

This section expands upon the t-SNE visualization and mean top-20 neighborhood-overlap result reported in Figure 1(c) and provides the corresponding reproduction details. Specifically, we describe the item representations, popularity-stratified sampling procedure, dimensionality-reduction settings, and neighborhood-overlap computation.

Dataset and representations. The visualization is conducted on the Amazon Reviews'23 Games 5-core dataset, which contains 94,762 users, 25,612 items, and 814,586 interactions (Hou et al., 2024). Item popularity is computed exclusively from the training interactions. For the semantic view, the title, category, description, and feature fields of each item are encoded into 1,024-dimensional representations using BGE-M3 (Chen et al., 2024). For the collaborative view, 64-dimensional item representations are learned from the user–item interaction graph using LightGCN (He et al., 2020). The two embedding matrices are aligned using the same item identifiers.

Sampling and visualization. To retain both popular and long-tail items, we construct a popularity-stratified sample of 4,000 items. Specifically, 800 items are sampled from the top 20% most popular items, while the remaining 3,200 items are sampled from the other items. Sampling is performed without replacement using random seed 42, and the same item set is used for both views. The resulting semantic and collaborative embeddings are independently L2-normalized and projected into two dimensions using t-SNE, with perplexity 40, PCA initialization, and random seed 42.

As shown in Figure 1(c), both panels visualize the same set of items, with one point per item in each panel. Points are colored by $\log_{10}(p_i + 1)$, where p_i denotes the number of training interactions associated with item i. The visualization is used only for a qualitative comparison of the neighborhood organizations induced by the semantic and collaborative views and is not intended to provide a quantitative measure of structural similarity.

Neighborhood overlap. To complement the qualitative visualization, we compute the mean top-20 neighborhood overlap in the original high-dimensional embedding spaces before applying t-SNE. For each of the 4,000 items, collaborative and semantic neighbors are retrieved by cosine similarity from the same 4,000-item candidate set, excluding the item itself. We define the mean overlap as $\text{Overlap@20} = \frac{1}{4000} \sum_{i=1}^{4000} \frac{|\mathcal{N}_{20}^{cf}(i) \cap \mathcal{N}_{20}^{sem}(i)|}{20}$, which yields 8.12%.

B Dataset Construction

This section expands upon the dataset description in Section 4.1 and specifies the sampling, filtering, and item-metadata procedures used to construct the datasets summarized in Table 1.

Dataset sampling and filtering. We retain the complete MovieLens-100K dataset without subsampling or frequency filtering, comprising $100{,}000$ interactions from 943 users and $1{,}682$ items. For Amazon-Books and Amazon-Games, drawn from the corresponding Amazon Reviews 2023 categories (Hou et al., 2024), we uniformly sample users without replacement using a fixed seed and retain their interactions and induced items. We then apply iterative 5-core filtering, repeatedly removing users and items with fewer than five remaining interactions until convergence, and remap the surviving identifiers to contiguous tokens. This yields $3{,}763$ users, $3{,}265$ items, and $35{,}617$ interactions for Amazon-Books, and $5{,}019$ users, $4{,}193$ items, and $51{,}113$ interactions for Amazon-Games. No additional frequency filtering is applied during data loading. The Amazon subsets are constructed once and held fixed across recommendation methods, collaborative backbones, and training seeds, and are subsequently partitioned using the protocol in Section 4.1.

Item metadata. For Amazon items, the semantic input is constructed from the nonempty title, category, description, and feature fields. For MovieLens-100K items, we use the title, release year, and genre fields. Field values are concatenated in a fixed order, and the same item texts and identifier mappings are used by SPF and all semantic baselines.

C Implementation Details

This section provides implementation details for the experimental setup, including the training protocol, the construction of semantic assets and structural references, and model configurations.

Training protocol. Unless otherwise specified, all models use 64-dimensional embeddings and two collaborative propagation layers. Optimization uses Adam with a learning rate of 10^{-3}, no optimizer weight decay, a batch size of 2048, and one uniformly sampled negative item per positive interaction. Embedding regularization with coefficient 10^{-4} is applied together with the BPR loss. For backbones with an auxiliary self-supervised objective, the corresponding objective is retained alongside the BPR loss. Training is performed for at most 200 epochs, with early stopping after 10 consecutive epochs without improvement in validation Recall@20. Each experimental configuration is trained and evaluated with five random seeds, $\{2024, 2025, 2026, 2027, 2028\}$.

Semantic asset construction. Item text is encoded by the all-MiniLM-L6-v2 sentence transformer into 384-dimensional, L_2-normalized features. We train a TIGER-style RQ-VAE (Rajput et al., 2023) with four residual codebooks of 32 entries each, latent dimension 64, encoder hidden dimensions $(512, 256, 128)$, commitment weight 0.25, learning rate 10^{-3}, batch size 256, and 1000 epochs. Training uses k-means initialization and dead-code revival, with a revival threshold of 0.02 relative to uniform code usage and residual scaling disabled. Four quantization codes are assigned to each item, and a disambiguation code is appended to distinguish items with identical four-code sequences, resulting in Semantic IDs of length $L = 5$. The continuous semantic reference $\bar{\mathbf{h}}_i^{sem}$ is constructed as the L_2-normalized sum of the four selected quantization codewords; the disambiguation code is not included in this representation. The resulting Semantic IDs and quantized representations are mapped to recommender item indices using item identifiers. A single semantic asset set is generated per dataset and reused throughout all downstream recommendation experiments, so the reported variation across random seeds excludes variability from retraining the RQ-VAE.

Reference representations for structural evaluation. For each trained semantic–collaborative integration model, a CF-only model is trained separately using the same data split, collaborative backbone, training protocol, and random seed. The item representations produced by this model are used as the collaborative reference $\bar{\mathbf{h}}_i^{cf}$ for CSP@20. Because the collaborative structure learned by a CF-only model varies across random seeds, using a single fixed reference for all seeds would confound CSP@20 with variation in the reference model itself. We therefore match references by seed: an integration model trained with seed s is evaluated against the CF-only model trained with the same seed s. CSP@20 is computed separately for each seed and then aggregated across the five seeds. The semantic reference $\bar{\mathbf{h}}_i^{sem}$ used by SNP@20 is the fixed RQ-VAE quantized representation described above, while the Semantic-ID sequences used by HC@20 are generated from the same RQ-VAE outputs. These semantic references are shared across integration methods, random seeds, and collaborative backbones within each dataset. The collaborative reference $\bar{\mathbf{h}}_i^{cf}$ is used only for structural evaluation and is distinct from the model-side collaborative representation $\mathbf{h}_i^{cf}$ optimized during integration training. For SPF, the same fixed quantized vector is used as both the semantic reference $\bar{\mathbf{h}}_i^{sem}$ and the model-side semantic representation $\mathbf{h}_i^{sem}$.

SPF and ablation configurations. SPF applies a trainable 64×64 affine projection to the semantic representation $\mathbf{h}_i^{sem}$, followed by residual fusion with coefficient $\beta = 1.0$ and layer normalization. For structural regularization, we set the temperature to $\tau = 0.2$. The implementation permits up to $K = 300$ positive neighbors per anchor; when fewer neighbors are available, the loss is averaged only over the valid neighbors. This limit is non-binding in our experiments, as the largest neighborhoods contain 102, 176, and 274 items on ml-100k, Amazon-Books, and Amazon-Games, respectively. For the component ablation, RF retains the complete SPF fusion module but omits $\mathcal{L}_{\text{struct}}$, making structural regularization the only difference between RF and SPF. To determine the structural regularization strength, we conduct a grid search over $\gamma \in \{0, 0.01, 0.05, 0.1, 0.5, 1\}$ on all three datasets. The setting γ=0.01 achieves the highest mean validation Recall@20 over five seeds on both Amazon subsets while remaining near-optimal on ml-100k, with a difference of only approximately 0.001 from the dataset-specific optimum. Accordingly, we use γ=0.01 as the default setting across datasets and collaborative backbones unless otherwise specified.

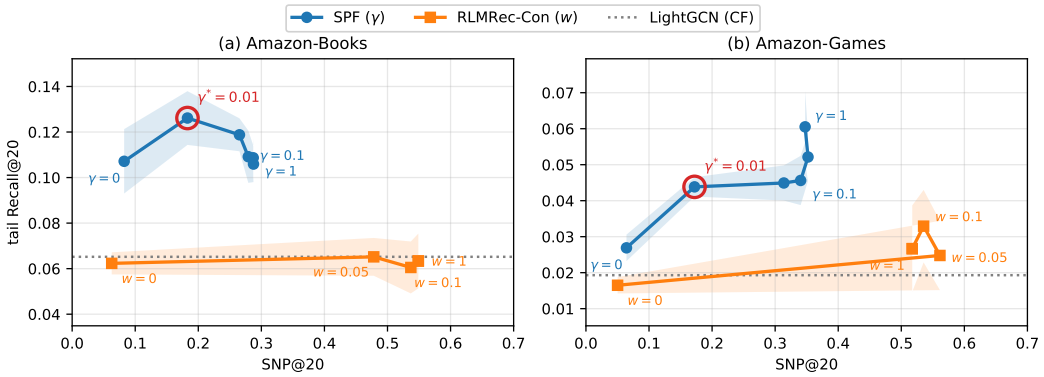


Figure 3: Semantic-neighborhood preservation versus Tail Recall@20 on Amazon-Books and Amazon-Games with a LightGCN backbone. SPF varies γ, whereas RLMRec-Con varies w. Markers show five-seed means; shaded bands indicate $\pm$ one standard deviation in Tail Recall@20; red circles mark the default γ=0.01; and dotted lines denote the LightGCN reference.

Self-supervised backbone configurations. For SGL, SimGCL, XSimGCL, and LightGCL, the coefficient of the auxiliary self-supervised objective is selected on Amazon-Books from $\{0.005, 0.01, 0.05, 0.1, 0.5\}$. Each candidate value is evaluated over five CF-only training seeds, and the value with the highest mean validation Recall@20 is selected and then fixed for all subsequent experiments using the corresponding backbone. The resulting configurations are as follows: SGL with edge-dropout ratio 0.1, temperature 0.5, and contrastive weight 0.1; SimGCL with perturbation magnitude 0.1, temperature 0.2, and contrastive weight 0.01; XSimGCL with perturbation magnitude 0.2, temperature 0.2, contrastive layer 1, and contrastive weight 0.05; and LightGCL with SVD rank 5, temperature 0.8, and contrastive weight 0.005.

Semantic–collaborative integration baseline configurations. RLMRec-Con uses collaborative representations for item ranking and incorporates semantic information through an item-side unidirectional InfoNCE distillation objective, with distillation weight $w = 0.1$ and temperature 0.2. DaRec uses a shared–private trade-off coefficient of $\lambda = 0.1$, 8 preference centers in the local structure alignment module, and auxiliary loss coefficients of 1.0. Repeated GPU executions of DaRec under a fixed random seed are not bitwise identical, but the resulting variation remains within its observed seed-to-seed variability and does not change the ordering of the compared methods. The Semantic-ID baseline uses the same affine projection, residual fusion, and layer normalization as SPF, but employs trainable semantic representations optimized jointly with the recommendation objective and excludes structural regularization.

D Parameter Sensitivity Analysis

Figure 3 provides a complementary visualization for RQ3 by examining the relationship between semantic-neighborhood preservation and Tail Recall@20 under varying control strengths on Amazon-Books and Amazon-Games. For RLMRec-Con, the alignment weight w controls the strength of semantic–collaborative alignment, with $w = 0$ corresponding to the alignment-free setting. Moving from $w = 0$ to nonzero alignment produces a large increase in SNP@20, whereas the corresponding Tail Recall@20 does not exhibit a consistent improvement as the alignment strength increases. In comparison, varying the structural regularization strength γ in SPF results in a different preservation–performance pattern: higher Tail Recall@20 is attained at substantially lower SNP@20 levels than those produced by the alignment-enabled RLMRec-Con settings.

The figure focuses on SNP@20 and Tail Recall@20 to make this relationship directly visible. The complete numerical comparison, including CSP@20, HC@20, and overall Recall@20, is reported in Table 4 in the main text. These results provide additional evidence that stronger semantic-neighborhood preservation alone is insufficient to improve recommendation performance.