CONNECTING AUTONOMOUS RECOMMENDER SYSTEMS THROUGH FEDERATED SEMANTIC GROUNDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Autonomous recommender systems learn from different item offerings and behavioral feedback, and thus acquire complementary knowledge about items and user preferences. However, such knowledge is deeply embedded in system-specific recommendation spaces, making direct cross-system transfer difficult. Federated recommendation enables decentralized knowledge exchange, but existing methods often rely on entity correspondence or compatible model and representation spaces. LLMs offer a shared semantic basis across systems, but their generic semantics do not capture the recommendation meaning learned from local interactions. We propose Federated Semantic Grounding, which grounds shared semantics in local behavior and uses them to transfer recommendation knowledge across systems. The framework applies this principle at both the item and user levels. At the item level, we use shared semantic concepts as anchors to exchange complementary item semantics across different item spaces. At the user level, we re-express locally learned preferences in another system's item context and coordinate their interpretation using mappings learned from local behavior. Experiments across diverse collaboration scenarios show consistent improvements over independent and federated baselines, and further demonstrate that FedSG remains effective when participating systems differ in model architecture or local recommendation setting. These results support semantic grounding as an effective basis for collaboration among autonomous recommender systems. The code is available at: `https://anonymous.4open.science/r/FedSG-683D/`.

1 INTRODUCTION

Recommender systems Zhang et al. (2019); Zangerle & Bauer (2022); Zhang et al. (2025b) commonly operate as autonomous services, with learning and recommendation conducted independently within their respective environments. The knowledge acquired by each system is therefore shaped by its item offering and the behavioral feedback generated through local interactions. Because these environments differ, item concepts and preference patterns well characterized in one system may remain underrepresented in another Zhu et al. (2021a). Cross-system collaboration can thus turn this complementarity into mutual gains, enriching each recommender's understanding of items and user preferences and, in turn, improving local recommendation quality Zhu et al. (2021b).

However, the autonomy that creates such complementarity also makes the resulting knowledge difficult to share. As shown in Figure 1, each recommender serves its own users and items, which lack direct correspondences with those in other systems Zang et al. (2022). Moreover, the representations learned from local interactions are formed within independently constructed model spaces, making them incompatible with direct integration across systems Zhu et al. (2021b). This raises a fundamental question: **how can system-specific recommendation knowledge be transformed into a form that can be effectively utilized across autonomous recommender systems?**

Autonomous recommender systems maintain their interaction data within separate operational domains, which precludes centralized training and makes federated recommendation Sun et al. (2024); Zhang et al. (2025a) the most direct paradigm for cross-system collaboration. Most federated approaches, however, transfer knowledge through model parameters or learned representations Wu et al. (2022); Zhang et al.; Lu et al. (2025), whose effectiveness depends on shared or compatible model spaces, an assumption that does not hold across independently developed recommenders.

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.

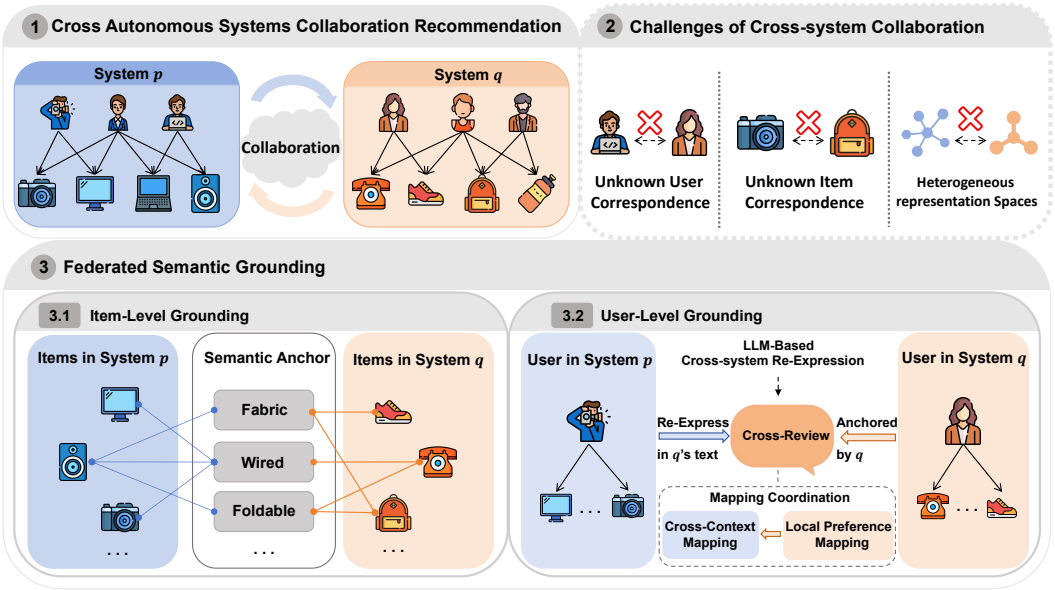


Figure 1: Motivation and overview of **FedSG** for cross-system collaboration among autonomous recommender systems. FedSG bridges non-overlapping users and items and heterogeneous representation spaces through item-level and user-level semantic grounding.

LLMs offer a model-independent semantic basis for expressing recommendation knowledge across systems Guo et al. (2024b); Liu et al. (2025a). However, because their semantics are learned primarily from general language corpora rather than system-specific behavioral feedback, they do not fully capture the recommendation meanings acquired within each system Liu et al. (2024a); Li et al. (2025). Cross-system collaboration therefore requires a knowledge representation that combines semantic commonality with recommendation-specific meaning learned from local behavior.

In this paper, we propose **Fed**erated **S**emantic **G**rounding (**FedSG**), which expresses system-specific recommendation knowledge through shared semantics and grounds these semantics in local behavioral evidence. Rather than exchanging model parameters or learned representations, the framework communicates semantic concepts that can be recognized across systems without aligning their entities or model spaces. Each recommender then uses its own interaction data to determine how these concepts relate to local items and user preferences. In this way, locally acquired recommendation knowledge is transformed into a semantic form that is transferable across systems while preserving its system-specific meaning.

We formulate FedSG around two central dimensions of recommendation knowledge: item semantics and user preferences. At the item level, we use shared semantic concepts as anchors across different item spaces, allowing each recommender to selectively incorporate complementary semantics relevant to its own items. At the user level, we re-express locally learned preferences in the recommendation context of another system, enabling preference knowledge to be communicated without matching users. Each recommender then adapts this cross-system preference knowledge to its own recommendation task using local interaction data. Together, the two mechanisms support complementary knowledge exchange while preserving the autonomy of each recommender. Our main contributions are summarized as follows:

- We formulate autonomous recommender collaboration as semantic knowledge collaboration, decoupling cross-system knowledge exchange from pre-existing entity correspondence and model-space alignment.
- We develop FedSG as the core mechanism for semantic knowledge collaboration, grounding shared semantics in local behavior so that they encode recommendation-specific knowledge transferable across systems. At the item and user levels, FedSG enables complementary item semantics and preference knowledge to be exchanged in a unified framework.
- Extensive experiments across diverse collaboration settings show that FedSG consistently outperforms representative baselines and remains effective under heterogeneous recommender configurations, demonstrating its effectiveness and generality.

2 Related Work

Cross-domain recommendation and federated recommendation provide the two most relevant foundations for this work. Cross-domain recommendation studies how knowledge from different item domains can improve recommendation, ranging from behavior-based transfer with shared users and multi-domain sequences, such as C2DSR, Tri-CDR, MGCL, and SyNCRec (Cao et al., 2022; Ma et al., 2024; Xu et al., 2025; Park et al., 2024), to recent semantic approaches that exploit LLM-derived item and preference information, including LLM4CDSR and SemaCDR (Liu et al., 2025b; Zhang et al., 2026). These methods demonstrate the value of complementary cross-domain knowledge, but typically assume shared users, jointly accessible multi-domain behavior, or a unified recommendation pipeline. Federated recommendation instead enables decentralized knowledge exchange while keeping raw interactions local. Existing methods transfer user-side preference representations or prototypes (Wu et al., 2022; Chen et al., 2023; Liu et al., 2023; Tian et al., 2024; Wang et al., 2025a), item-side representations or semantics (Wang et al., 2025c; Guo et al., 2024a; Lu et al., 2025), or model-side knowledge through parameters, representations, or predictions (Wang et al., 2025b; Zhang et al., 2024a;b; Wan et al., 2024; Zheng et al., 2025). Although these approaches support decentralized collaboration, they generally retain assumptions on entity correspondence, compatible representation spaces, or predefined model interfaces; semantic-oriented methods such as PFCR and FFMSR further reduce dependence on shared entities or representations, but primarily focus on item-side semantic transfer (Guo et al., 2024a; Lu et al., 2025). In contrast, FedSG treats shared semantics as a model-independent collaboration layer and grounds them in local behavioral evidence, enabling both item knowledge and user preference knowledge to be coordinated across autonomous recommender systems without requiring entity or model alignment. Further discussion of the related literature is provided in Appendix A.

3 Methodology

3.1 Problem Formulation

Autonomous Recommender System. We characterize an autonomous recommender system p by its local user set $\mathcal{U}^p$, item set $\mathcal{I}^p$, behavioral data $\mathcal{D}^p$, and recommendation model F^p, collectively denoted as $\mathcal{R}^p = (\mathcal{U}^p, \mathcal{I}^p, \mathcal{D}^p, F^p)$. These components define a self-contained recommendation environment in which system p serves its own users and items, observes locally generated behavioral feedback, and independently learns and maintains F^p.

Cross-System Collaboration. Given P autonomous recommender systems, let $\mathcal{R} = \{\mathcal{R}^p\}_{p=1}^{P}$ denote the participating systems. Their collaboration is characterized by two conditions: alignment independence and autonomy preservation.

(C1) Alignment independence. Collaboration assumes no known correspondence between $\mathcal{U}^p$ and $\mathcal{U}^q$, or between $\mathcal{I}^p$ and $\mathcal{I}^q$, for any $p \neq q$; possible overlaps need not be identified. Moreover, each $F^p \in \mathcal{H}^p$ is independently defined, without requiring a common model space or a predefined mapping between $\mathcal{H}^p$ and $\mathcal{H}^q$.

(C2) Autonomy preservation. Each system remains autonomous during collaboration: $\mathcal{D}^p$ stays local, while F^p is independently maintained and optimized within its own environment.

Semantic-Augmented Collaborative Optimization. Under cross-system collaboration, each autonomous recommender retains its own recommendation objective while incorporating semantic knowledge acquired from other systems as complementary information. For system p, let $\mathcal{L}_{\text{rec}}^p$ denote the recommendation loss evaluated on its local behavioral data $\mathcal{D}^p$, and let $\mathcal{K}_{\text{sem}}^p$ denote the collaborative semantic knowledge available to $\mathcal{R}^p$. We formulate the optimization objective as

$$F^{p*} = \arg\min_{F^p} \mathcal{L}_{\text{rec}}^p \left(F^p; \mathcal{D}^p, \mathcal{K}_{\text{sem}}^p\right), \qquad p = 1, \ldots, P. \tag{1}$$

This formulation preserves system-specific optimization: each recommender optimizes F^p locally, while cross-system collaboration contributes complementary knowledge through $\mathcal{K}_{\text{sem}}^p$.

3.2 Federated Semantic Grounding

To enable cross-system knowledge collaboration among autonomous recommender systems, we introduce **Fed**erated **S**emantic **G**rounding (**FedSG**). It uses shared semantics as model-independent knowledge carriers and grounds them in local behavioral evidence, allowing system-specific recommendation knowledge to be encoded in a transferable form. Although the formulation allows P systems, we present the method for a collaborating pair p and q, with $p \neq q$.

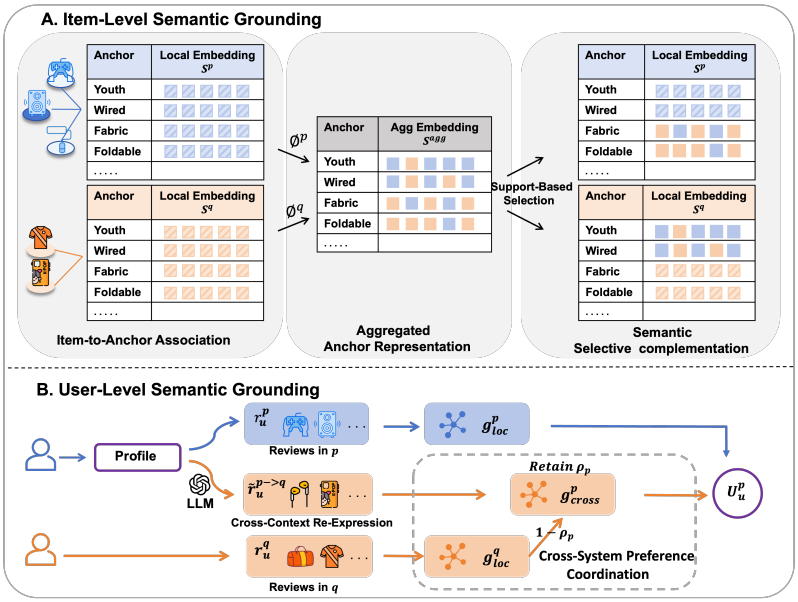


Figure 2: Overview of Federated Semantic Grounding. (a) Item-level grounding performs support-aware coordination over shared semantic anchors and selectively complements underrepresented local semantics. (b) User-level grounding re-expresses local preferences across system contexts and coordinates preference mappings to adapt transferred knowledge for local recommendation.

Framework Overview. For each system p, FedSG represents the knowledge involved in cross-system collaboration as behavior-grounded semantic knowledge:

$$\mathcal{K}_{\text{sem}}^p = \left(\mathcal{K}_{\text{item}}^p, \mathcal{K}_{\text{pref}}^p\right). \tag{2}$$

Here, $\mathcal{K}_{\text{item}}^p$ and $\mathcal{K}_{\text{pref}}^p$ capture two complementary dimensions of recommendation knowledge. The former supports semantic knowledge exchange across independently defined item spaces, while the latter enables preference knowledge transfer across different recommendation contexts. These semantic components are derived from local recommendation environments and coordinated across systems, whereas the behavioral data $\mathcal{D}^p$ and recommendation model F^p remain local. The resulting complementary knowledge is then incorporated into each system's own recommendation process. The two grounding pathways and their cross-system coordination are illustrated in Figure 2.

3.3 Item-Level Semantic Grounding

Within FedSG, item-level semantic grounding enables collaboration across independently defined item spaces by expressing local item knowledge through shared semantic concepts as cross-system anchors. These anchors allow item knowledge to be related and coordinated across systems without matching item identities, providing complementary semantics for each local recommender.

Shared Semantic Anchoring. Since item identities are independently defined across autonomous recommenders, we construct a shared hierarchical semantic schema as cross-system anchors. An LLM induces K high-level concepts and their fine-grained descriptors from the participating item-domain descriptions, with the prompt template provided in Appendix B.1:

$$\mathcal{C} = \{C_k\}_{k=1}^{K}, \qquad \mathcal{A}_k = \{a_{k,j}\}_{j=1}^{N_k}, \tag{3}$$

where C_k denotes the k-th semantic concept and $\mathcal{A}_k$ its descriptor set. For example, *Functional Purpose* may include *communication*, *entertainment*, and *productivity*. The resulting schema provides shared semantic dimensions for associating items across systems without identity correspondence.

Item-to-Anchor Association. For each item $i \in \mathcal{I}^p$, the same LLM maps its textual information (e.g., title, category, and brand) to descriptors in the shared schema; the detailed prompt is provided in Appendix B.2. We define $b_{i,k,j}^p \in \{0,1\}$, where $b_{i,k,j}^p = 1$ indicates that item i is associated with descriptor $a_{k,j}$ under concept C_k, and 0 otherwise.

Semantic Anchor Representation. System p parameterizes each descriptor with a learnable embedding $\mathbf{s}^p_{k,j} \in \mathbb{R}^{d_s}$, forming the local semantic table $\mathbf{S}^p = \{\mathbf{s}^p_{k,j}\}_{k,j}$. For item i, associated descriptor embeddings are aggregated within each concept and concatenated across concepts:

$$\begin{aligned} \mathbf{e}^p_{i,k} &= \frac{\sum_{j=1}^{N_k} b^p_{i,k,j}\mathbf{s}^p_{k,j}}{\sum_{j=1}^{N_k} b^p_{i,k,j}} \in \mathbb{R}^{d_s}, \\ \mathbf{e}^p_i &= \mathbf{e}^p_{i,1} \| \cdots \| \mathbf{e}^p_{i,K} \in \mathbb{R}^{Kd_s}. \end{aligned} \tag{4}$$

Selective Semantic Complementation. Descriptor coverage can differ substantially across systems, making uniform coordination potentially harmful to well-supported local semantics. We measure the local support of descriptor $a_{k,j}$ in system p as $\phi^p_{k,j} = \sum_{i \in \mathcal{I}^p} b^p_{i,k,j}$, where lower support indicates greater need for complementation. The support statistic both weights cross-system aggregation and determines whether complementation is applied. The server computes

$$\mathbf{s}^{\text{agg}}_{k,j} = \frac{\phi^p_{k,j}\mathbf{s}^p_{k,j} + \phi^q_{k,j}\mathbf{s}^q_{k,j}}{\phi^p_{k,j} + \phi^q_{k,j} + \epsilon}, \tag{5}$$

where ϵ is a small constant for stability, assigning greater influence to the better-supported system. System p then selectively uses the aggregated representation for under-supported descriptors:

$$\bar{\mathbf{s}}^p_{k,j} = \begin{cases} \mathbf{s}^{\text{agg}}_{k,j}, & \phi^p_{k,j} < \tau_p, \\ \mathbf{s}^p_{k,j}, & \phi^p_{k,j} \geq \tau_p, \end{cases} \tag{6}$$

where τ_p controls the complementation criterion. The resulting table $\bar{\mathbf{S}}^p = \{\bar{\mathbf{s}}^p_{k,j}\}_{k,j}$ yields $\bar{\mathbf{e}}^p_i$ through the same concept-wise aggregation and concatenation, preserving well-supported local semantics while complementing underrepresented ones.

3.4 User-Level Semantic Grounding

User preference knowledge is closely tied to the item offering and recommendation context of each autonomous recommender, making locally learned preferences difficult to transfer directly across systems. To address this, we first derive semantic preference profiles from local behavioral evidence and re-express them under another system's recommendation context. We then coordinate preference mappings across systems so that the re-expressed preferences can incorporate knowledge learned in the corresponding context. The complete prompt for constructing the preference profile and its cross-context re-expression is provided in Appendix B.3.

Local Preference Profiling. For each user $u \in \mathcal{U}^p$, we summarize the local interaction history and available semantic feedback into a semantic preference profile R^p_u capturing the user's principal preference characteristics. A pretrained text encoder E_{text} then maps it to

$$\mathbf{r}^p_u = E_{\text{text}}(R^p_u), \qquad \mathbf{r}^p_u \in \mathbb{R}^{d_r}. \tag{7}$$

The resulting $\mathbf{r}^p_u$ provides a semantic representation of user preference in system p.

Cross-Context Preference Re-expression. Because the local preference profile R^p_u is conditioned by system p's item context, we re-express it with respect to the item offering of system q. Let $\mathcal{M}^q$ denote the textual semantic information describing system q's items. Using an LLM and the same text encoder, we obtain

$$\begin{aligned} \widetilde{R}^{p\to q}_u &= G_{\text{LLM}}\left(R^p_u, \mathcal{M}^q\right), \\ \widetilde{\mathbf{r}}^{p\to q}_u &= E_{\text{text}}\left(\widetilde{R}^{p\to q}_u\right) \in \mathbb{R}^{d_r}. \end{aligned} \tag{8}$$

The resulting representation expresses system p's preference knowledge in system q's item context without requiring user correspondence.

Cross-System Preference Coordination. Although $\widetilde{\mathbf{r}}^{p\to q}_u$ expresses system p's preference in the item context of system q, its recommendation interpretation should incorporate behavioral knowledge learned in q. Accordingly, system p uses g^p_{cross} to process the re-expressed preference, while g^q_{loc} captures the corresponding preference interpretation learned from local users in system q. The server coordinates these mappings as

$$g^p_{\text{cross}} \leftarrow \rho_p\, g^p_{\text{cross}} + (1-\rho_p)\, g^q_{\text{loc}}, \qquad p \neq q, \tag{9}$$

where $\rho_p \in [0,1]$ balances retained cross-context knowledge and the local interpretation contributed by system q. The update preserves the former while incorporating recommendation knowledge learned in the corresponding item context.

3.5 Local Recommendation and Federated Optimization

We incorporate item- and user-level semantic knowledge from cross-system collaboration into each autonomous recommender. Collaboratively grounded item semantics augment local item representations, while local and cross-context preference knowledge enrich user representations, enabling each system to exploit complementary knowledge for local recommendation.

Item Semantic Integration. For item $i \in \mathcal{I}^p$, we combine its ID-based embedding with locally grounded and collaboratively complemented semantics:

$$\mathbf{v}_i^p = \mathbf{v}_{i,\mathrm{id}}^p + \psi_{\mathrm{loc}}^p\left(\mathbf{e}_i^p\right) + \psi_{\mathrm{col}}^p\left(\bar{\mathbf{e}}_i^p\right), \tag{10}$$

where ψ_{loc}^p and ψ_{col}^p are separate projections for local and collaborative semantic knowledge.

User Preference Integration. For user $u \in \mathcal{U}^p$, we integrate the ID-based embedding $\mathbf{u}_{u,\mathrm{id}}^p$ with local and cross-context preference representations through their corresponding mappings:

$$\mathbf{u}_u^p = \alpha_{u,1}^p\, \mathbf{u}_{u,\mathrm{id}}^p + \alpha_{u,2}^p\, g_{\mathrm{loc}}^p\left(\mathbf{r}_u^p\right) + \alpha_{u,3}^p\, g_{\mathrm{cross}}^p\left(\widetilde{\mathbf{r}}_u^{p\to q}\right), \tag{11}$$

where $\alpha_{u,1}^p$, $\alpha_{u,2}^p$, and $\alpha_{u,3}^p$ control the contributions of identity-specific, local-context, and cross-context preference information, respectively.

Local Recommendation Learning. Each system optimizes its local recommender F^p on $\mathcal{D}^p$ under its own objective $\mathcal{L}_{\mathrm{rec}}^p$. Since the semantic components participate in the integrated user and item representations, they are jointly optimized by the same local behavioral supervision, while the recommender architecture and task-specific loss remain system-specific.

Federated Optimization. Training proceeds in communication rounds between the autonomous recommenders and a coordinating server; the complete procedure is provided in Algorithm 1 in Appendix C. **(1) Local learning.** Each system optimizes its local recommender and semantic components on $\mathcal{D}^p$ under $\mathcal{L}_{\mathrm{rec}}^p$, grounding the collaborative components in local behavioral feedback. **(2) Semantic coordination.** Each system uploads the required semantic components, after which the server performs selective item-level complementation (Section 3.3) and preference-mapping coordination (Section 3.4). **(3) Knowledge redistribution.** The coordinated semantic components are returned to the corresponding systems for the next round of local learning. Throughout training, behavioral data and complete recommendation models remain local, while only the exchanged semantic knowledge is coordinated by the server.

4 Experiment

4.1 Experimental Setup

Datasets. We construct three collaboration scenarios from Amazon Review Hou et al. (2026), each involving two autonomous recommender systems with independently defined local item spaces. Each system contains two product domains: Scenario 1 uses Video/Electronics–Phones/Sports, Scenario 2 uses Home/Kindle–Books/Tools, and Scenario 3 uses Movies/CDs–Office/Arts. These combinations provide diverse recommendation environments and cross-system domain configurations. Detailed dataset statistics are provided in Appendix D.1.

Baselines. We instantiate FedSG with C2DSR and SyNCRec to evaluate its effectiveness across different recommendation backbones. We compare against four groups of methods: independent recommendation (C2DSR Cao et al. (2022)/SyNCRec Park et al. (2024) and their semantic-enhanced variants), conventional federated learning (FedAvg McMahan et al. (2017), FedAtt Chai et al. (2020), and FedMoE Yi et al. (2026)), semantic-augmented federated learning (Sem-FedAvg, Sem-FedAtt, and Sem-FedMoE), and federated semantic recommendation (FFMSR Lu et al. (2025)). Detailed descriptions of all baselines are provided in Appendix D.2.

Evaluation. For cross-domain backbones, users shared across local domains are retained within each system, while no user correspondence is assumed across systems. We adopt leave-one-out evaluation and report MRR, HR@10, and NDCG@10 He et al. (2017). Complete preprocessing, optimization, implementation, and hyperparameter settings are provided in Appendix D.3.

4.2 Overall Performance

Tables 1–2 report the main results across three collaboration scenarios. For conciseness, the SyNCRec results for Scenarios 2 and 3 are deferred to Appendix E. We make four observations. **First,**

Table 1: Main recommendation performance on Scenario 1 (Video/Electronics–Phones/Sports). "(a)"–"(d)" denote independent recommendation, conventional federated learning, semantic-augmented federated learning, and federated semantic recommendation, respectively. The best results are highlighted in bold, and the second-best results are underlined.

Category	Model	Video			Electronics			Phones			Sports		
		MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
(a)	C2DSR	0.1810	0.3581	0.2060	0.1324	0.2496	0.1419	0.1426	0.2779	0.1569	0.1433	0.2406	0.1488
	Sem-C2DSR	0.2307	0.4612	0.2682	0.1682	0.3233	0.1863	0.1863	0.3486	0.2078	0.1618	0.3069	0.1777
(b)	FedAvg	0.2104	0.4067	0.2405	0.1460	0.2804	0.1590	0.1577	0.2998	0.1744	0.1538	0.2517	0.1589
	FedAtt	0.2169	0.4002	0.2430	0.1443	0.2955	0.1615	0.1682	0.2944	0.1807	0.1537	0.2455	0.1571
	FedMoE	0.2022	0.3961	0.2300	0.1396	0.2668	0.1511	0.1616	0.2899	0.1746	0.1488	0.2585	0.1562
(c)	Sem-FedAvg	0.2365	0.4637	0.2739	0.1581	0.3226	0.1776	0.1817	0.3551	0.2053	0.1737	0.3207	0.1904
	Sem-FedAtt	0.2165	0.4344	0.2506	0.1680	0.3261	0.1864	0.1844	0.3614	0.2091	0.1684	0.3195	0.1861
	Sem-FedMoE	0.2205	0.4466	0.2571	0.1601	0.3199	0.1788	0.1868	0.3596	0.2101	0.1711	0.3250	0.1897
(d)	FFMSR	0.2324	0.4238	0.2610	0.1572	0.2928	0.1702	0.1802	0.3252	0.1969	0.1684	0.3033	0.1830
(Ours)	**FedSG**	**0.2622**	**0.5179**	**0.3068**	**0.1769**	**0.3608**	**0.2020**	**0.2016**	**0.3856**	**0.2283**	**0.1824**	**0.3520**	**0.2046**
(a)	SyNCRec	0.1504	0.3390	0.1741	0.1336	0.2758	0.1473	0.1247	0.2455	0.1342	0.0980	0.2094	0.1054
	Sem-SyNCRec	0.2441	0.5509	0.2998	0.1587	0.3656	0.1873	0.1719	0.3804	0.1923	0.1376	0.3288	0.1620
(b)	FedAvg	0.1494	0.3317	0.1721	0.1407	0.2730	0.1521	0.1291	0.2292	0.1343	0.0763	0.1541	0.0761
	FedAtt	0.1513	0.3301	0.1731	0.1114	0.2397	0.1256	0.1280	0.2545	0.1389	0.0898	0.1802	0.0923
	FedMoE	0.1440	0.3236	0.1655	0.1083	0.2444	0.1244	0.1241	0.2418	0.1332	0.0810	0.1765	0.0840
(c)	Sem-FedAvg	0.2636	0.5583	0.3177	0.1816	0.3916	0.2123	0.1830	0.4139	0.2193	0.1549	0.3462	0.1805
	Sem-FedAtt	0.2629	0.5607	0.3177	0.1760	0.3830	0.2054	0.1838	0.4167	0.2204	0.1521	0.3250	0.1727
	Sem-FedMoE	0.2523	0.5721	0.3126	0.1778	0.3962	0.2103	0.1934	0.4130	0.2269	0.1626	0.3611	0.1908
(d)	FFMSR	0.2497	0.5575	0.3058	0.1573	0.3629	0.1855	0.1875	0.3976	0.2180	0.1349	0.3089	0.1580
(Ours)	**FedSG**	**0.2760**	**0.5754**	**0.3314**	**0.1842**	**0.4012**	**0.2158**	**0.1968**	**0.4213**	**0.2280**	**0.1670**	**0.3754**	**0.1973**

Table 2: Main recommendation performance on Scenarios 2 and 3 with C2DSR as the backbone.

Scenario 2: Home/Kindle–Books/Tools

Category	Model	Home			Kindle			Books			Tools		
		MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
(a)	C2DSR	0.1690	0.2746	0.1784	0.1488	0.2279	0.1516	0.2345	0.4348	0.2643	0.1504	0.2755	0.1628
	Sem-C2DSR	0.1924	0.3176	0.2078	0.1747	0.3070	0.1886	0.2647	0.5165	0.3089	0.1590	0.3049	0.1765
(b)	FedAvg	0.1468	0.2439	0.1541	0.0842	0.1674	0.0866	0.2415	0.4351	0.2713	0.1500	0.2787	0.1633
	FedAtt	0.1627	0.2766	0.1754	0.1308	0.2465	0.1410	0.2452	0.4404	0.2751	0.1541	0.2814	0.1671
	FedMoE	0.1892	0.2930	0.1988	0.1282	0.2279	0.1341	0.2298	0.4428	0.2626	0.1506	0.2850	0.1654
(c)	Sem-FedAvg	0.1811	0.3012	0.1934	0.1573	0.2698	0.1680	0.2787	0.5362	0.3244	0.1626	0.3093	0.1798
	Sem-FedAtt	0.1939	0.3217	0.2103	0.1755	0.2558	0.1750	0.2761	0.5160	0.3171	0.1619	0.3035	0.1777
	Sem-FedMoE	0.1795	0.2992	0.1918	0.1561	0.2605	0.1635	0.2681	0.4934	0.3059	0.1550	0.2840	0.1680
(d)	FFMSR	0.1924	0.3135	0.2064	0.1724	0.2698	0.1785	0.2579	0.4779	0.2939	0.1555	0.2934	0.1710
	FedSG	**0.2002**	**0.3361**	**0.2177**	**0.1891**	**0.3302**	**0.2067**	**0.3187**	**0.5780**	**0.3660**	**0.1688**	**0.3334**	**0.1898**

Scenario 3: Movies/CDs–Office/Arts

Category	Model	Movies			CDs			Office			Arts		
		MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
(a)	C2DSR	0.1631	0.2786	0.1736	0.1549	0.2718	0.1655	0.1635	0.2775	0.1746	0.1427	0.2629	0.1536
	Sem-C2DSR	0.1971	0.3553	0.2171	0.1727	0.3214	0.1900	0.1789	0.2802	0.1856	0.1756	0.3300	0.1939
(b)	FedAvg	0.1788	0.3081	0.1929	0.1761	0.3311	0.1963	0.1549	0.2294	0.1567	0.1303	0.2167	0.1342
	FedAtt	0.1800	0.3143	0.1953	0.1686	0.3147	0.1862	0.1575	0.2349	0.1590	0.1272	0.2057	0.1288
	FedMoE	0.1718	0.3212	0.1899	0.1537	0.2895	0.1681	0.1750	0.2788	0.1831	0.1439	0.2398	0.1485
(c)	Sem-FedAvg	0.1944	0.3503	0.2141	0.1792	0.3340	0.1980	0.1816	0.2857	0.1889	0.1791	0.3081	0.1926
	Sem-FedAtt	0.1868	0.3335	0.2040	0.1649	0.3097	0.1807	0.1784	0.2981	0.1895	0.1703	0.3146	0.1876
	Sem-FedMoE	0.1812	0.3270	0.1980	0.1683	0.3190	0.1852	0.1850	0.2775	0.1898	0.1697	0.3091	0.1845
(d)	FFMSR	0.1850	0.3187	0.2001	0.1716	0.3271	0.1909	0.1747	0.2679	0.1788	0.1666	0.2794	0.1744
	FedSG	**0.2410**	**0.4369**	**0.2714**	**0.2447**	**0.4603**	**0.2796**	**0.1916**	**0.3269**	**0.2059**	**0.1972**	**0.3498**	**0.2164**

FedSG is consistently effective across recommendation environments and backbone architectures. Its advantage holds across all three domain compositions and both C2DSR and SyNCRec, indicating that the collaboration mechanism is not tied to a particular local recommender or cross-

Table 3: Domain-level performance under heterogeneous local recommender configurations on Scenario 1. Panel (a) considers two cross-domain recommenders with different backbones. Panel (b) considers a cross-domain recommender collaborating separately with the two single-domain recommenders. The single-domain results are reported individually, while the FedSG result for the cross-domain recommender is averaged over the two corresponding collaboration runs.

(a) Different Cross-domain Recommenders

Scenario	Method	Cross-domain (C2DSR)				Cross-domain (SyNCRec)			
		Domain	MRR	H@10	N@10	Domain	MRR	H@10	N@10
Scenario 1	Independent	Video	0.2307	0.4612	0.2682	Phones	0.1719	0.3804	0.1923
	FedSG		**0.2632**	**0.4988**	**0.3030**		**0.1784**	**0.3958**	**0.2050**
	Independent	Electronics	0.1682	0.3233	0.1863	Sports	0.1376	0.3288	0.1620
	FedSG		**0.1745**	**0.3679**	**0.2011**		**0.1446**	**0.3418**	**0.1712**

(b) Cross-domain and Single-domain Recommenders

Scenario	Method	Cross-domain (C2DSR)				Single-domain (SASRec)			
		Domain	MRR	H@10	N@10	Domain	MRR	H@10	N@10
Scenario 1	Independent	Video	0.2307	0.4612	0.2682	Phones	0.1823	0.3634	0.2084
	FedSG		**0.2548**	**0.4874**	**0.2932**		**0.1911**	**0.3837**	**0.2192**
	Independent	Electronics	0.1682	0.3233	0.1863	Sports	0.1890	0.3531	0.2106
	FedSG		**0.1723**	**0.3498**	**0.1952**		**0.1925**	**0.3590**	**0.2159**

system domain configuration. This is consistent with FedSG, which communicates knowledge through shared semantics rather than backbone-specific model structures. **Second, local semantic augmentation consistently strengthens independent recommendation.** The gains of Sem-C2DSR and Sem-SyNCRec show that semantic descriptions of items and user preferences provide complementary information beyond interaction patterns, improving local recommendation even before cross-system collaboration. **Third, conventional model-level federation yields less consistent gains, whereas semantic augmentation makes federation substantially more effective.** Even with the same backbone, systems trained on different item domains and behavioral distributions learn distinct interaction patterns, making direct parameter aggregation less reliable. The stronger performance of Sem-FedAvg, Sem-FedAtt, and Sem-FedMoE suggests that semantics provide a more transferable basis for cross-system knowledge sharing. **Finally, effective semantic collaboration requires more than simply introducing semantic information.** Semantic-augmented federated baselines already incorporate the same local semantic components, while FFMSR also coordinates semantic representations. The further gains of FedSG indicate that grounding shared semantics in local behavioral evidence enables more effective transfer of recommendation-specific knowledge across autonomous systems.

4.3 Cross-System Collaboration with Heterogeneous Local Recommenders

In practice, autonomous recommender systems may differ in both model architecture and local recommendation formulation, requiring collaboration without standardizing their local models. We therefore consider two heterogeneous settings: different cross-domain recommenders, instantiated with C2DSR and SyNCRec, and cross-domain/single-domain collaboration, instantiated with C2DSR and SASRec. In both settings, we compare FedSG with independent training of the corresponding local recommenders. Table 3 reports the domain-level results on Scenario 1, while the corresponding results on Scenarios 2 and 3 are provided in Appendix F.

The results lead to three observations. **First, FedSG remains effective across different recommendation architectures.** In the cross-domain/cross-domain setting, it consistently outperforms independent training across all scenarios and metrics, showing that collaboration does not require a shared backbone. **Second, this effectiveness extends to different local recommendation settings.** FedSG also improves performance when cross-domain and single-domain recommenders collaborate, indicating that participating systems need not share the same recommendation formulation. **Third, the gains of FedSG do not systematically diminish with greater heterogeneity.** Although the cross-domain/single-domain setting differs in both model architecture and local recommendation formulation, its improvements do not consistently shrink relative to the different-backbone setting, suggesting that semantic collaboration remains effective despite heterogeneity in both dimensions.

Table 4: Ablation results of FedSG with C2DSR as the backbone. The best results are highlighted in bold, and the second-best results are underlined.

Variant	Scenario 1			Scenario 2			Scenario 3		
	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
w/o Item Coordination	<u>0.1948</u>	<u>0.3778</u>	<u>0.2207</u>	0.1993	0.3622	0.2218	0.1850	0.3324	0.2024
w/o Preference Coordination	0.1905	0.3703	0.2152	<u>0.2113</u>	<u>0.3838</u>	<u>0.2361</u>	<u>0.2078</u>	<u>0.3804</u>	<u>0.2319</u>
Avg. Item Coordination	0.1840	0.3569	0.2071	0.2013	0.3524	0.2205	0.1878	0.3381	0.2061
FedSG	**0.2058**	**0.4041**	**0.2354**	**0.2192**	**0.3944**	**0.2450**	**0.2186**	**0.3935**	**0.2433**

4.4 Ablation Studies

We further examine FedSG through ablation studies with C2DSR. We disable cross-system coordination on either the item or preference side, yielding w/o Item Coordination and w/o Preference Coordination, and replace support-aware selective item complementation with weighted averaging, yielding Avg. Item Coordination. Results are averaged over the four local domains in each scenario.

As shown in Table 4 the results lead to three observations. **First, item- and preference-level coordination provide complementary benefits.** Removing either pathway consistently degrades performance, indicating that item semantics and preference knowledge contribute distinct information to cross-system collaboration. **Second, their relative contributions vary across scenarios.** Preference coordination is more important in some settings, whereas item coordination contributes more in others, showing that their relative importance depends on the recommendation environment. **Third, item-side performance depends on how semantic knowledge is coordinated.** Replacing selective complementation with simple averaging consistently reduces performance, supporting the preservation of well-supported local semantics while selectively incorporating complementary knowledge.

4.5 Hyperparameter Analysis

We analyze four parameters governing semantic construction and coordination in FedSG: the semantic complementation criterion τ_p, preference coordination coefficient ρ_p, number of shared semantic concepts K, and number of cross-context preference examples. The first two control the extent of item- and preference-level coordination, while the latter two determine semantic breadth and contextual detail. We vary one parameter at a time and select values by validation performance. Detailed sensitivity curves and cross-scenario analyses are provided in Appendix G.

The results reveal four patterns. ***First***, τ_p has a clear non-monotonic effect: overly restrictive complementation limits cross-system enrichment, whereas overly broad complementation may involve already well-supported local semantics, supporting selective item-level coordination. ***Second***, the preferred ρ_p varies across domains, showing that the balance between retained cross-context knowledge and newly introduced local interpretation is environment-dependent. ***Third***, increasing K does not monotonically improve performance; broader semantic coverage can help, but excessive concepts may introduce redundant or weakly relevant information. ***Finally***, cross-context preference examples provide additional contextual evidence beyond summary-level profiles, but their marginal benefit is domain-dependent and may saturate or decline. Overall, effective semantic grounding requires balancing the extent of cross-system coordination with the richness of semantic knowledge, with preferred configurations varying across recommendation environments.

5 Conclusion

This work explores how autonomous recommender systems can collaborate when their users, items, behavioral data, and models are independently maintained. We propose Federated Semantic Grounding (FedSG), which uses behavior-grounded semantics as a shared layer for expressing and exchanging recommendation knowledge across systems without relying on entity or model alignment. By coordinating complementary item and preference knowledge, FedSG allows each recommender to benefit from other systems while preserving its own local modeling process. Experiments across multiple collaboration scenarios and recommender architectures consistently demonstrate advantages over independent learning and conventional model-level federation, and further show that the proposed collaboration remains effective when participating systems differ in both architecture and local recommendation setting. These results establish semantic grounding as a promising foundation for connecting independently developed recommender systems through knowledge-level rather than model-level collaboration.

AI USE STATEMENT

We used generative AI tools to assist with drafting and revising manuscript text, improving language clarity and concision, and refining the presentation of figures and tables. All research ideas, experimental design, implementation, analysis, and conclusions were developed and verified by the authors.

REFERENCES

Jiangxia Cao, Xin Cong, Jiawei Sheng, Tingwen Liu, and Bin Wang. Contrastive cross-domain sequential recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, 2022. doi: 10.1145/3511808.3557262.

Zheng Chai, Yujing Chen, Liang Zhao, Yue Cheng, and Huzefa Rangwala. Fedat: A communication-efficient federated learning method with asynchronous tiers under non-iid data. *ArXivorg*, 2020.

Gaode Chen, Xinghua Zhang, Yijun Su, Yantong Lai, Ji Xiang, Junbo Zhang, and Yu Zheng. Win-win: A privacy-preserving federated framework for dual-target cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4149–4156, 2023.

Lei Guo, Ziang Lu, Junliang Yu, Quoc Viet Hung Nguyen, and Hongzhi Yin. Prompt-enhanced federated content representation learning for cross-domain recommendation. In *Proceedings of the ACM Web Conference 2024*, 2024a. doi: 10.1145/3589334.3645337.

Lei Guo, Ziang Lu, Junliang Yu, Quoc Viet Hung Nguyen, and Hongzhi Yin. Prompt-enhanced federated content representation learning for cross-domain recommendation. In *Proceedings of the ACM Web Conference 2024*, pp. 3139–3149, 2024b.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pp. 173–182, 2017.

Yupeng Hou, Jiacheng Li, Xiangjun Fu, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation: Benchmarking llms as semantic encoders. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2026.

Jun Hu, Wenwen Xia, Xiaolu Zhang, Chilin Fu, Weichang Wu, Zhaoxin Huan, Ang Li, Zuoli Tang, and Jun Zhou. Enhancing sequential recommendation via llm-based semantic embedding learning. In *Companion Proceedings of the ACM Web Conference 2024*, 2024. doi: 10.1145/3589335.3648307.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining*, pp. 197–206, 2018. doi: 10.1109/ICDM.2018.00035.

Guanghan Li, Xun Zhang, Yufei Zhang, Yifan Yin, Guojun Yin, and Wei Lin. Semantic convergence: Harmonizing recommender systems via two-stage alignment and behavioral semantic tokenization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12040–12048, 2025.

Qidong Liu, Xian Wu, Yejing Wang, Zijian Zhang, Feng Tian, Yefeng Zheng, and Xiangyu Zhao. Llm-esr: Large language models enhancement for long-tailed sequential recommendation. *Advances in Neural Information Processing Systems*, 37:26701–26727, 2024a.

Qidong Liu, Xian Wu, Yejing Wang, Zijian Zhang, Feng Tian, Yefeng Zheng, and Xiangyu Zhao. LLM-ESR: Large language models enhancement for long-tailed sequential recommendation. *arXiv preprint arXiv:2405.20646*, 2024b.

Qidong Liu, Xiangyu Zhao, Yejing Wang, Zijian Zhang, Howard Zhong, Chong Chen, Xiang Li, Wei Huang, and Feng Tian. Bridge the domains: Large language models enhanced cross-domain sequential recommendation. In *Proceedings of the 48th international ACM SIGIR conference on research and development in information retrieval*, pp. 1582–1592, 2025a.

Qidong Liu, Xiangyu Zhao, Yejing Wang, Zijian Zhang, Howard Zhong, Chong Chen, Xiang Li, Wei Huang, and Feng Tian. Bridge the domains: Large language models enhanced cross-domain sequential recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025b. doi: 10.1145/3726302.3729911.

Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Jianwei Yin, Yanchao Tan, and Longfei Zheng. Federated probabilistic preference distribution modelling with compactness co-clustering for privacy-preserving multi-domain recommendation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 2206–2214, 2023.

Ziang Lu, Lei Guo, Xu Yu, Zhiyong Cheng, Xiaohui Han, and Lei Zhu. Federated semantic learning for privacy-preserving cross-domain recommendation. *arXiv preprint arXiv:2503.23026*, 2025.

Haokai Ma, Ruobing Xie, Lei Meng, Xin Chen, Xu Zhang, Leyu Lin, and Jie Zhou. Triple sequence learning for cross-domain recommendation. *ACM Transactions on Information Systems*, 42(4), 2024. doi: 10.1145/3638351.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. Pmlr, 2017.

Chung Park, Taesan Kim, Hyungjun Yoon, Junui Hong, Yelim Yu, Mincheol Cho, Minsung Choi, and Jaegul Choo. Pacer and runner: Cooperative learning framework between single- and cross-domain sequential recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2071–2080, 2024. doi: 10.1145/3626772.3657710.

Zehua Sun, Yonghui Xu, Yong Liu, Wei He, Lanju Kong, Fangzhao Wu, Yali Jiang, and Lizhen Cui. A survey on federated recommendation systems. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1):6–20, 2024.

Changxin Tian, Yuexiang Xie, Xu Chen, Yaliang Li, and Xin Zhao. Privacy-preserving cross-domain recommendation with federated graph learning. *ACM Transactions on Information Systems*, 2024. doi: 10.1145/3653448.

Sheng Wan, Dashan Gao, Hanlin Gu, and Daning Hu. FedPDD: A privacy-preserving double distillation framework for cross-silo federated recommendation. *arXiv preprint arXiv:2305.06272*, 2024.

Li Wang, Shoujin Wang, Quangui Zhang, Qiang Wu, and Min Xu. Federated user preference modeling for privacy-preserving cross-domain recommendation. *IEEE Transactions on Multimedia*, 27:5324–5336, 2025a. doi: 10.1109/TMM.2025.3543106.

Qingren Wang, Yuchuan Zhao, Yi Zhang, Yiwen Zhang, Shuiguang Deng, and Yun Yang. Federated contrastive learning for cross-domain recommendation. *IEEE Transactions on Services Computing*, 18(2), 2025b. doi: 10.1109/TSC.2025.3528325.

Zhihao Wang, He Bai, Wenke Huang, Duantengchuan Li, Jian Wang, and Bing Li. Federated recommendation with explicitly encoding item bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12792–12800, 2025c.

Meihan Wu, Li Li, Chang Tao, Eric Rigall, Xiaodong Wang, and Chengzhong Xu. FedCDR: Federated cross-domain recommendation for privacy-preserving rating prediction. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, 2022. doi: 10.1145/3511808.3557320.

Zitao Xu, Shu Chen, Weike Pan, and Zhong Ming. A multi-view graph contrastive learning framework for cross-domain sequential recommendation. *ACM Transactions on Recommender Systems*, 3(4), 2025. doi: 10.1145/3695884.

Liping Yi, Han Yu, Gang Wang, Xiaoguang Liu, and Qinghua Hu. pfedmoe: Data-level personalization with mixture of experts in model-heterogeneous personalized federated learning. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. A survey on cross-domain recommendation: taxonomies, methods, and future directions. *ACM Transactions on Information Systems*, 41(2):1–39, 2022.

Eva Zangerle and Christine Bauer. Evaluating recommender systems: survey and framework. *ACM computing surveys*, 55(8):1–38, 2022.

Chunxu Zhang, Guodong Long, Zijian Zhang, Zhiwei Li, Honglei Zhang, Qiang Yang, and Bo Yang. Personalized recommendation models in federated settings: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2025a.

Chunxu Zhang, Shanqiang Huang, Zijian Zhang, Jiahong Liu, Linsong Yu, Ruiqi Wan, Bo Yang, and Irwin King. SemaCDR: Llm-powered transferable semantics for cross-domain sequential recommendation. In *Proceedings of the ACM Web Conference 2026*, pp. 6795–6806, 2026.

Hongyu Zhang, Dongyi Zheng, Xu Yang, Jiyuan Feng, and Qing Liao. *FedDCSR: Federated Cross-domain Sequential Recommendation via Disentangled Representation Learning*, pp. 535–543. doi: 10.1137/1.9781611978032.62.

Hongyu Zhang, Dongyi Zheng, Xu Yang, Jiyuan Feng, and Qing Liao. FedDCSR: Federated cross-domain sequential recommendation via disentangled representation learning. In *Proceedings of the 2024 SIAM International Conference on Data Mining*, pp. 535–543, 2024a.

Hongyu Zhang, Dongyi Zheng, Lin Zhong, Xu Yang, Jiyuan Feng, Yunqing Feng, and Qing Liao. FedHCDR: Federated cross-domain recommendation with hypergraph signal decoupling. In *Machine Learning and Knowledge Discovery in Databases: Research Track*, pp. 350–366, 2024b. doi: 10.1007/978-3-031-70341-6_21.

Kaike Zhang, Qi Cao, Fei Sun, Yunfan Wu, Shuchang Tao, Huawei Shen, and Xueqi Cheng. Robust recommender system: a survey and future directions. *ACM Computing Surveys*, 58(1):1–38, 2025b.

Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*, 52(1):1–38, 2019.

Dongyi Zheng, Hongyu Zhang, Jianyang Zhai, Zhong Lin, Lingzhi Wang, Jiyuan Feng, Xiangke Liao, Yonghong Tian, Nong Xiao, and Qing Liao. FedCSR: A federated framework for multi-platform cross-domain sequential recommendation with dual contrastive learning. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 8699–8713, 2025.

Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. Cross-domain recommendation: Challenges, progress, and prospects. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 4721–4728. International Joint Conferences on Artificial Intelligence Organization, 8 2021a. doi: 10.24963/ijcai.2021/639. URL https://doi.org/10.24963/ijcai.2021/639. Survey Track.

Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):1171–1184, 2021b.

APPENDIX

A RELATED WORK

A.1 SEQUENTIAL AND CROSS-DOMAIN RECOMMENDATION

Sequential recommendation models evolving user preferences from chronologically ordered interactions and provides the basic recommendation setting considered in this work. SASRec introduces self-attention to identify historical interactions relevant to next-item prediction (Kang & McAuley, 2018). Recent studies further incorporate textual semantics into sequential recommendation. SAID uses LLM-derived textual information to learn semantically aligned item ID embeddings, while LLM-ESR combines LLM semantics with collaborative signals for item modeling and enhances long-tail user representations through retrieval-augmented self-distillation (Hu et al., 2024; Liu et al., 2024b). These methods enrich recommendation with external semantic knowledge, but operate within a single recommendation environment rather than transferring knowledge across domains.

Cross-domain sequential recommendation (CDSR) extends sequential modeling by exploiting behavioral information from multiple domains. C2DSR jointly captures intra-sequence sequential dependencies and inter-sequence item relationships, and introduces a contrastive cross-domain infomax objective to relate single-domain and cross-domain user representations (Cao et al., 2022). Tri-CDR separately models source-domain, target-domain, and mixed interaction sequences and coordinates them through cross-domain attention and contrastive learning (Ma et al., 2024). MGCL models complementary intra- and cross-domain sequential dependencies through multiple graph views and contrastive learning (Xu et al., 2025), while SyNCRec jointly exploits single-domain and cross-domain recommenders to mitigate negative transfer between domains (Park et al., 2024). Despite their different modeling strategies, these methods generally formulate knowledge transfer within a unified cross-domain recommendation pipeline in which multi-domain behavioral information can be jointly modeled.

More recent CDSR methods introduce LLM-derived semantics to establish transferable connections beyond behavioral co-occurrence. LLM4CDSR learns unified semantic item representations from LLMs and constructs hierarchical preference profiles from cross-domain interaction histories (Liu et al., 2025b). SemaCDR distinguishes domain-specific and domain-agnostic semantics and integrates them with cross-domain behavioral information to learn transferable item and preference representations (Zhang et al., 2026). These approaches demonstrate that semantics can facilitate knowledge transfer across heterogeneous item domains, but still operate within a unified cross-domain recommendation setting where multi-domain information is available to a common recommendation pipeline. In contrast, our setting considers autonomous recommender systems whose users, items, behavioral data, and recommendation models are independently maintained. FedSG therefore uses shared semantics as a cross-system knowledge carrier, enabling item and preference knowledge to be coordinated without requiring cross-system entity correspondence or a unified recommendation model.

A.2 FEDERATED RECOMMENDATION

Federated recommendation broadly studies collaborative recommendation under decentralized data ownership, with a substantial body of work treating individual users or devices as federated clients for personalized recommendation. Our setting instead concerns collaboration across independently operated recommendation environments. We therefore focus on federated cross-domain and cross-system methods that are more closely related to autonomous recommender collaboration. For exposition, we organize representative approaches according to their dominant coordination carrier: user-side preference knowledge, item-side content or semantic knowledge, and model-side knowledge encoded by parameters, representations, or predictions. These categories are not mutually exclusive, but highlight the primary interface through which knowledge is transferred.

User-side knowledge transfer. A major line of federated cross-domain recommendation abstracts local behavior into transferable user preference knowledge. FedCDR separates a private personalization module from a cross-domain transfer module and introduces personalized update and aggregation strategies (Wu et al., 2022). P2FCDR maps cross-domain user embeddings through an orthogonal transformation, protects them with local differential privacy, and uses gated fusion to select

useful transferred information (Chen et al., 2023). FPPDM++ models probabilistic preference distributions and exploits similar-user structure through compactness co-clustering (Liu et al., 2023), while PPCDR coordinates global and domain-specific preferences through federated graph learning (Tian et al., 2024). FUPM further derives comprehensive user preferences from interactions, reviews, and potential positive items and transfers them through privacy-preserving preference prototypes (Wang et al., 2025a). These methods demonstrate the value of transferring behavior-derived preference knowledge, but commonly rely on user correspondence or predefined preference representation spaces across domains.

Item-side knowledge transfer. Another line of work coordinates knowledge associated with item characteristics. FREIB addresses label skew by explicitly modeling item bias and aligning rating-related feature prototypes across participants (Wang et al., 2025c). Content-based approaches instead exploit textual information as a transferable reference across domains. PFCR learns federated content representations from item descriptions and adapts them through prompt tuning (Guo et al., 2024a), while FFMSR considers fully non-overlapping user and item spaces and coordinates multi-layer semantic representations derived from item text (Lu et al., 2025). These studies show that item content and semantics can provide transferable knowledge across heterogeneous item spaces, but their coordination is primarily centered on item-side information.

Model-side knowledge transfer. Federated recommendation also coordinates knowledge through model parameters, latent representations, or predictions. Fed-CLR combines federated learning with intra- and inter-model contrastive objectives (Wang et al., 2025b). FedDCSR disentangles domain-shared and domain-exclusive sequence representations Zhang et al., while FedHCDR separates shared and domain-specific user signals through hypergraph modeling (Zhang et al., 2024a;b). FedPDD transfers recommendation knowledge through double distillation rather than iterative parameter aggregation (Wan et al., 2024), and FedCSR combines model- and sequence-level contrastive learning without requiring cross-platform user alignment (Zheng et al., 2025). Despite relaxing different data-sharing or alignment assumptions, these approaches still coordinate knowledge through predefined model, representation, or prediction interfaces.

Overall, existing federated recommendation methods coordinate knowledge through predefined preference, item, or model interfaces. Our work instead establishes shared semantics as the collaboration interface and grounds them with local behavioral feedback, allowing different forms of recommendation knowledge to be coordinated within the same semantic space.

B LLM PROMPTS FOR SEMANTIC GROUNDING

We employ LLMs at three stages of the semantic grounding pipeline: (1) constructing a shared hierarchical semantic schema, (2) associating individual items with fine-grained descriptors in that schema, and (3) constructing semantic preference profiles and re-expressing them in the item context of another autonomous recommender system. For readability, repetitive instructions and exhaustive descriptor sets are abbreviated with "..." below.

B.1 Prompt for Shared Semantic Schema Construction

Stage I: Shared Hierarchical Semantic Schema Construction

Role. Act as an expert in multi-domain feature classification.

Input Domains. I have a dataset containing items from Video Games, Electronics, Cell Phones and Accessories, and Sports and Outdoors.

Task. Analyze item features across all four domains and construct a shared hierarchical semantic schema. The high-level concepts should capture common feature dimensions, and each concept should contain fine-grained descriptors that can be applied to items in all four domains.

Requirements.

- Each semantic concept must contain multiple possible descriptor values.
- The schema will provide shared semantic anchors for a multi-domain recommender system.
- Include only feature dimensions that are meaningful and applicable across all four domains.
- Provide no explanation; return only the schema in the format below.

Output Format.

```
Common feature taxonomy = {
  "class name":  str,
  "feature value":  list[str]
}
Return:  list[Common feature taxonomy]
```

B.2 Prompt for Item-to-Descriptor Association

Stage II: Item-to-Descriptor Association

Role. Act as a professional recommendation-system semantic analyst.

Input. I will provide the features of an item from Video Games, Electronics, Cell Phones and Accessories, or Sports and Outdoors. The available features may include the item's title, category, and brand.

Task: Associate the Item with Shared Semantic Descriptors.
Using the "Classification Schema" below, identify the descriptors relevant to the item. For **each** category in the schema, select at least one descriptor based on the provided item features and your world knowledge. Use only the descriptor labels defined in the schema.

Classification Schema.

Brand Positioning:
Premium/Flagship; Mainstream/Mid-range; Value/Budget; Specialized/Niche; Generic.

Product Type Classification:
Core Device; Hardware Component; Accessory/Peripheral; Software/Digital Content; Protection/Storage; Consumable.

Target Demographic:
Children/Youth; Adults; Professional/Expert; Casual/Beginner; Unisex/All-ages; Family.

Connectivity Technology:
Wireless; Bluetooth; Wired; USB; Wi-Fi; NFC; Cellular/LTE; Mechanical/None.

...

Aesthetic Style:
Minimalist/Modern; Industrial/Professional; Sporty/High-visibility; Gaming/RGB; Classic/Retro.

Package Configuration:
Single Item; Bundle; Kit; Set; Multi-pack; Replacement Part; Accessory Only.

Association Rules.

1. Interpret the item using both the provided information and your world knowledge.
2. For **each** category in the Classification Schema, select one or more relevant descriptors; no category may be left empty.
3. Preserve the category names and descriptor labels exactly as given, and return only the category-to-descriptor assignments.

B.3 Prompt for Cross-Context Preference Re-expression

Stage III: Cross-Context User Preference Re-expression

System Prompt

Role. You are a senior consumer psychologist and cross-domain recommender expert. Your goal is to construct a user's semantic preference profile from behavior observed in a source system and re-express that profile in the item context of a target system without introducing artificial biases.

Task 1: Underlying Preference Profile.
Write a professional profile of approximately 50 words that summarizes the user's underlying preference characteristics and decision-making logic. Describe the user rather than the specific products purchased in the source system.

Task 2: Cross-Context Re-expression in the Target System.
Re-express the same preference characteristics in the recommendation context of System p (Video Games & Electronics). Generate six pseudo-reviews, each paired with a plausible item title: three for Video Games and three for Electronics.
Preserve the preference characteristics inferred from the source system, including quality expectations, price sensitivity, evaluative standards, and critical depth. Do not create direct functional correspondences between source-system and target-system items. Instead, preserve the continuity of the user's evaluation logic across the two recommendation contexts.

Constraints.

- The `user_summary` must be professional and insightful.
- Each generated review must preserve the user's expression style in the source system.
- Each interaction must contain the corresponding domain, a plausible item title, and a review consistent with the inferred preference profile.

Output JSON.

```
{
"user_id": user_id,
"user_summary": "...",
"Video_interactions": [
{"domain": "Video_Games", "item_title": "...",
"review": "..."}, ...
],
"Electronics_interactions": [
{"domain": "Electronics", "item_title": "...",
"review": "..."}, ...
]
}
```

User Input Template

Input: Source System q (Cell Phones and Accessories & Sports and Outdoors)

`User {user_id}'s interaction history:`
{user interaction history}

C Training Algorithm

Algorithm 1 summarizes the complete training procedure of FedSG. Each communication round alternates between local recommendation learning and server-side semantic coordination. The participating systems first update their local recommenders and semantic components using their own behavioral data, after which the server performs item-level semantic complementation and preference-level mapping coordination. The resulting semantic knowledge is then returned to the corresponding systems for the next round of local learning. Throughout this process, raw behavioral data and complete recommendation models remain local.

Algorithm 1 Federated Semantic Grounding for Autonomous Recommender Collaboration

Require: Autonomous recommenders $\mathcal{R}^p$ and $\mathcal{R}^q$ with local datasets $\mathcal{D}^p$ and $\mathcal{D}^q$; shared semantic schema $\mathcal{C}$; federation hyperparameters $\tau_p, \tau_q, \rho_p, \rho_q$; communication rounds R
Ensure: Collaboratively enhanced local recommenders F^{p*} and F^{q*}

Semantic Preparation
Initialize local recommenders and semantic components for both systems
Establish item-to-anchor associations and local/cross-context preference profiles
for $r = 1$ to R **do**
Local Learning at Autonomous Recommenders
  for all $s \in \{p, q\}$ **in parallel do**
    Construct item representations according to Eq. (10)
    Construct user representations according to Eq. (11)
    Update the local recommender and trainable semantic components according to Eq. (1)
    Upload descriptor semantics, support statistics, and preference mappings to the server
  end for
Semantic Coordination at the Server
  Aggregate descriptor semantics according to Eq. (5)
  Perform selective semantic complementation according to Eq. (6)
  Coordinate cross-context preference mappings according to Eq. (9)
Collaborative Knowledge Redistribution
  Return the complemented item semantics and coordinated preference mappings to the corresponding systems
end for
return F^{p*} and F^{q*}

Table 5: Statistics of the three collaboration scenarios constructed from the Amazon Review data.

Scenario	System	Local Domain	Users	Items	Interactions	Sparsity
Scenario 1	System p	Electronics	3,809	2,044	37,708	99.5157%
		Video	3,809	2,124	41,057	99.4925%
	System q	Phones	2,713	1,520	18,277	99.5568%
		Sports	2,713	1,865	20,965	99.5857%
Scenario 2	System p	Home	703	771	5,685	98.9511%
		Kindle	703	844	5,308	99.1054%
	System q	Books	9,503	12,144	185,142	99.8396%
		Tools	9,503	5,407	87,832	99.8291%
Scenario 3	System p	Movies	8,537	10,211	268,413	99.6921%
		CDs	8,537	7,450	153,000	99.7594%
	System q	Office	1,637	866	12,693	99.1046%
		Arts	1,637	815	12,013	99.0996%

D Detailed Experimental Setup

D.1 Datasets

We construct three collaboration scenarios from the Amazon Review data, each involving two autonomous recommender systems operating over distinct local item domains. To reflect the broader service scope of an autonomous recommender, each system is instantiated with two product domains. The scenarios use different domain combinations, yielding diverse recommendation environments and varying degrees of cross-system domain heterogeneity. This design allows us to examine collaboration across different cross-system domain configurations. Detailed domain compositions and data statistics are summarized in Table 5.

D.2 BASELINES

We use C2DSR and SyNCRec as two representative recommendation backbones to examine whether the effectiveness of FedSG is consistent across different underlying architectures. For each backbone, its independently trained version serves as the basic reference, and the compared methods are organized into four groups. **(1) Independent Recommendation** includes the original backbone trained separately in each system, and semantic enhanced backbone, named Sem-C2DSR and Sem-SyNCRec, which further incorporates the same local semantic components used by FedSG without cross-system semantic coordination. **(2) Conventional Federated Learning** includes FedAvg, FedAtt, and FedMoE, representing different model-aggregation strategies applied to the participating systems and providing a direct reference for whether conventional model-level federation can support effective collaboration among autonomous recommenders. **(3) Semantic-Augmented Federated Learning** includes Sem-FedAvg, Sem-FedAtt, and Sem-FedMoE, which combine the same local semantic components with the corresponding federated aggregation strategies, allowing us to examine whether semantic augmentation alone can make conventional model federation effective across systems. **(4) Federated Semantic Recommendation** includes FFMSR, which performs federated coordination over semantic representations and therefore provides the closest comparison to our semantic-oriented collaboration paradigm.

D.3 IMPLEMENTATION DETAILS

For data preprocessing, we treat all observed interactions as positive feedback, order them chronologically, remove duplicate interactions, and filter items and users with fewer than 10 and 3 interactions, respectively. For the cross-domain sequential backbones, users appearing in both local domains are retained within each system, while no user correspondence is assumed across systems. We adopt a leave-one-out protocol, using the last interaction for testing, the second-to-last for validation, and the remaining interactions for training. For evaluation, each positive test item is ranked together with 99 sampled negatives, and performance is measured by MRR, HR@K, and NDCG@K, where $K = 10$. For model configuration, C2DSR uses a hidden dimension of 64 and a batch size of 256, while SyNCRec uses a hidden dimension of 128 and a batch size of 128; methods instantiated with the same backbone follow the same backbone settings. The semantic schema and cross-context preference profiles are constructed offline using an LLM and encoded with a pretrained text encoder. All method-specific hyperparameters are selected according to validation performance.

E ADDITIONAL RESULTS WITH SYNCREC

To complement the results based on C2DSR, we further report the complete results with SyNCRec for Scenarios 2 and 3 in Table 6. The same baseline groups and evaluation protocol are used. The results are consistent with those reported in the main text: FedSG achieves the strongest overall performance across the two scenarios and different local domains, while the semantic-enhanced baselines generally outperform their counterparts without semantic augmentation. These results further confirm that the effectiveness of FedSG is not specific to a particular recommendation backbone.

F ADDITIONAL RESULTS UNDER HETEROGENEOUS RECOMMENDER CONFIGURATIONS

Table 7 reports the additional results on Scenarios 2 and 3 for the two heterogeneous collaboration settings introduced in Section 4.3. The results are consistent with those observed on Scenario 1: FedSG improves over independent training when the participating systems use different cross-domain backbones, and also remains effective when a cross-domain recommender collaborates with single-domain recommenders. These results further support the applicability of semantic collaboration across heterogeneous local recommender configurations.

G FULL HYPERPARAMETER SENSITIVITY ANALYSIS

We provide the complete sensitivity analysis of the four hyperparameters studied in Section 4.5: the semantic complementation criterion τ_p, the preference coordination coefficient ρ_p, the number

Table 6: Additional recommendation performance on Scenarios 2 and 3 with SyNCRec as the backbone. "(a)"–"(d)" denote independent recommendation, conventional federated learning, semantic-augmented federated learning, and federated semantic recommendation, respectively. The best results are highlighted in bold, and the second-best results are underlined.

Scenario 2: Home/Kindle–Books/Tools													
Category	Model	Home			Kindle			Books			Tools		
		MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
(a)	SyNCRec	0.1367	0.2480	0.1466	0.0925	0.2093	0.1030	0.1986	0.4284	0.2350	0.1342	0.2694	0.1484
	Sem-SyNCRec	0.1598	0.3033	0.1774	0.1607	0.3488	0.1855	0.2835	0.6064	0.3452	0.1416	0.3211	0.1646
(b)	FedAvg	0.1313	0.2418	0.1406	0.0885	0.1628	0.0884	0.2017	0.4190	0.2353	0.1272	0.2636	0.1416
	FedAtt	0.1162	0.2234	0.1252	0.0798	0.1535	0.0803	0.2007	0.4313	0.2377	0.1287	0.2589	0.1418
	FedMoE	0.0974	0.2111	0.1084	0.0589	0.0884	0.0491	0.1656	0.3707	0.1944	0.1160	0.2546	0.1301
(c)	Sem-FedAvg	0.1592	0.3279	0.1845	0.1669	0.3442	0.1976	0.2850	0.6097	0.3472	0.1365	0.3116	0.1586
	Sem-FedAtt	0.1743	0.3381	0.1966	0.1567	0.3395	0.1815	0.2735	0.6050	0.3370	0.1512	0.3277	0.1716
	Sem-FedMoE	0.1815	0.3525	0.2056	0.1636	0.3070	0.1791	0.2811	0.6082	0.3441	0.1498	0.3316	0.1740
(d)	FFMSR	0.1778	0.3361	0.2008	0.1448	0.2744	0.1583	0.2487	0.5575	0.3054	0.1410	0.3192	0.1639
	FedSG	**0.1869**	**0.3648**	**0.2135**	**0.1996**	**0.3581**	**0.2190**	**0.3052**	**0.6231**	**0.3663**	**0.1538**	**0.3382**	**0.1745**
Scenario 3: Movies/CDs–Office/Arts													
Category	Model	Movies			CDs			Office			Arts		
		MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10	MRR	H@10	N@10
(a)	SyNCRec	0.1522	0.2906	0.1671	0.1265	0.2622	0.1403	0.1123	0.2129	0.1194	0.0954	0.1947	0.1016
	Sem-SyNCRec	0.2545	0.5213	0.3018	0.2597	0.5520	0.3127	0.1476	0.3146	0.1683	0.1648	0.3496	0.1837
(b)	FedAvg	0.1487	0.3035	0.1682	0.1255	0.2668	0.1407	0.0989	0.2088	0.1080	0.0775	0.1672	0.0820
	FedAtt	0.1405	0.2977	0.1604	0.1116	0.2378	0.1233	0.1032	0.2157	0.1119	0.0792	0.1738	0.0849
	FedMoE	0.1418	0.2960	0.1609	0.1133	0.2501	0.1267	0.1075	0.2102	0.1146	0.0830	0.1606	0.0837
(c)	Sem-FedAvg	0.2700	0.5276	0.3153	0.2781	0.5575	0.3205	0.1660	0.3173	0.1858	0.1784	0.3388	0.1985
	Sem-FedAtt	0.2592	0.5188	0.3047	0.2731	0.5542	0.3239	0.1640	0.3214	0.1835	0.1547	0.3575	0.1852
	Sem-FedMoE	0.2497	0.5041	0.2936	0.2640	0.5477	0.3151	0.1722	0.3187	0.1886	0.1682	0.3443	0.1918
(d)	FFMSR	0.2367	0.4828	0.2778	0.2579	0.5351	0.3071	0.1527	0.2926	0.1664	0.1547	0.3476	0.1815
	FedSG	**0.2762**	**0.5415**	**0.3243**	**0.2786**	**0.5630**	**0.3315**	**0.1760**	**0.3407**	**0.1913**	**0.1835**	**0.3643**	**0.2025**

of shared semantic concepts K, and the number of cross-context preference examples. The first two directly control the extent of item- and preference-level cross-system coordination, while the latter two determine the breadth and contextual detail of the semantic knowledge used by FedSG. We vary one parameter at a time while keeping the others fixed, and evaluate performance using HR@10 on the four local domains of each collaboration scenario. For τ_p, system-specific criterion values are considered because descriptor-support distributions differ across systems; ρ_p is varied over $\{0.1, 0.3, 0.5, 0.7, 0.9, 0.99\}$. Because the shared semantic schema is induced separately for each collaboration scenario, its total number of concepts differs across scenarios. We therefore vary K from 1 to 10 for Scenarios 1 and 2, and from 1 to 9 for Scenario 3, using the cumulative concept order of each schema. The number of cross-context preference examples is varied over $\{0, 2, 4, 6\}$. Hyperparameter values used in the main experiments are selected according to validation performance. Each point in the sensitivity curves corresponds to one recorded run evaluated at the validation-selected checkpoint.

G.1 Semantic Complementation Criterion

Figures 3a–3c report the sensitivity to the semantic complementation criterion τ_p. Recall that τ_p determines which locally under-supported semantic descriptors are complemented with coordinated cross-system semantics. Across all three scenarios, performance varies non-monotonically with the criterion, and the preferred range differs substantially across domains. In Scenario 1, intermediate criteria generally provide the strongest results, whereas broader complementation eventually reduces performance in all four domains. Scenario 2 exhibits greater variation: different domains favor criteria ranging from the lower to the upper-middle portion of the sweep, indicating that the appropriate extent of complementation depends on the local semantic-support distribution. The same pattern becomes particularly clear in Scenario 3, where Movies and CDs reach their best performance at

Table 7: Additional domain-level results under heterogeneous local recommender configurations on Scenarios 2 and 3. Panel (a) considers two cross-domain recommenders with different backbones. Panel (b) considers a cross-domain recommender collaborating separately with the two single-domain recommenders. The single-domain results are reported individually, while the FedSG result for the cross-domain recommender is averaged over the two corresponding collaboration runs.

(a) Different Cross-domain Recommenders

Scenario	Method	Cross-domain (C2DSR)				Cross-domain (SyNCRec)			
		Domain	MRR	H@10	N@10	Domain	MRR	H@10	N@10
Scenario 2	Independent	Home	0.1924	0.3176	0.2078	Books	0.2835	0.6064	0.3452
	FedSG		**0.2074**	**0.3504**	**0.2260**		**0.2945**	**0.6076**	**0.3546**
	Independent	Kindle	0.1747	0.3070	0.1886	Tools	0.1416	0.3211	0.1646
	FedSG		**0.1893**	**0.3442**	**0.2008**		**0.1583**	**0.3285**	**0.1796**
Scenario 3	Independent	Movies	0.1971	0.3553	0.2171	Office	0.1476	0.3146	0.1683
	FedSG		**0.2116**	**0.4117**	**0.2422**		**0.1629**	**0.3277**	**0.1779**
	Independent	CDs	0.1727	0.3214	0.1900	Arts	0.1648	0.3496	0.1947
	FedSG		**0.2014**	**0.4118**	**0.2338**		**0.1654**	**0.3674**	**0.2046**

(b) Cross-domain and Single-domain Recommenders

Scenario	Method	Cross-domain (C2DSR)				Single-domain (SASRec)			
		Domain	MRR	H@10	N@10	Domain	MRR	H@10	N@10
Scenario 2	Independent	Home	0.1924	0.3176	0.2078	Books	0.3138	0.5320	0.3512
	FedSG		**0.2075**	**0.3269**	**0.2208**		**0.3177**	**0.5597**	**0.3610**
	Independent	Kindle	0.1747	0.3070	0.1886	Tools	0.1854	0.3331	0.2039
	FedSG		**0.1768**	**0.3093**	**0.1905**		**0.1876**	**0.3355**	**0.2053**
Scenario 3	Independent	Movies	0.1971	0.3553	0.2171	Office	0.1963	0.3274	0.2119
	FedSG		**0.2269**	**0.4119**	**0.2541**		**0.2013**	**0.3280**	**0.2153**
	Independent	CDs	0.1727	0.3214	0.1900	Arts	0.2041	0.3647	0.2258
	FedSG		**0.2205**	**0.4267**	**0.2520**		**0.2074**	**0.3830**	**0.2323**

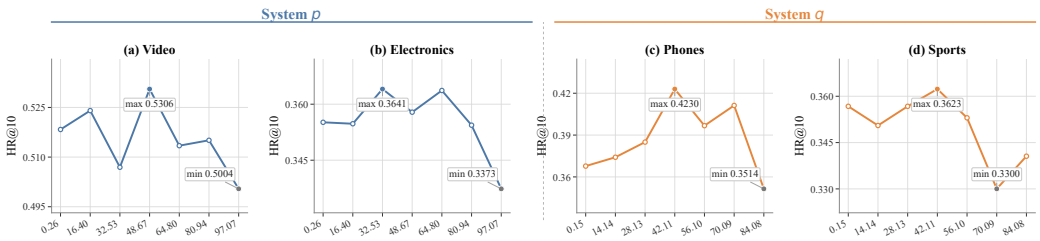
relatively moderate criteria but deteriorate substantially when the criterion approaches its largest value. Office and Arts exhibit smaller but similarly non-monotonic responses.

These results provide two observations. First, increasing the scope of semantic complementation is not uniformly beneficial: descriptors with insufficient local support can benefit from cross-system knowledge, whereas extending coordination to a larger portion of the semantic space may cease to provide useful complementary information. Second, the preferred criterion is domain-dependent, reflecting differences in how strongly individual semantic descriptors are represented in local item spaces. This behavior supports the support-aware selective design in Eq. (8), rather than applying semantic coordination uniformly to all descriptors.

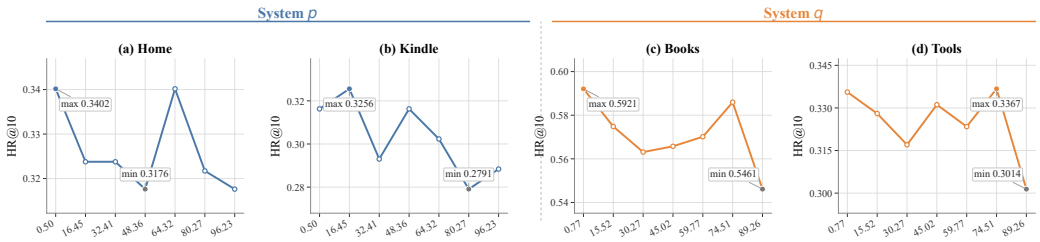
G.2 Preference Coordination Coefficient

Figures 4a–4c examine the preference coordination coefficient ρ_p, which balances the existing cross-context mapping with the locally learned preference mapping contributed by the collaborating system. The results show no universal coordination ratio across domains or scenarios. In Scenario 1, the four domains favor noticeably different settings, ranging from relatively small to relatively large values of ρ_p. Scenario 2 shows the same behavior: Home and Books achieve their strongest results at $\rho_p = 0.7$, whereas Kindle prefers a smaller value and Tools performs similarly at multiple intermediate settings. Scenario 3 provides an even clearer contrast. Movies and CDs favor substantial retention of the existing cross-context mapping, with both performing best around $\rho_p = 0.9$, while Office and Arts prefer more balanced coordination.

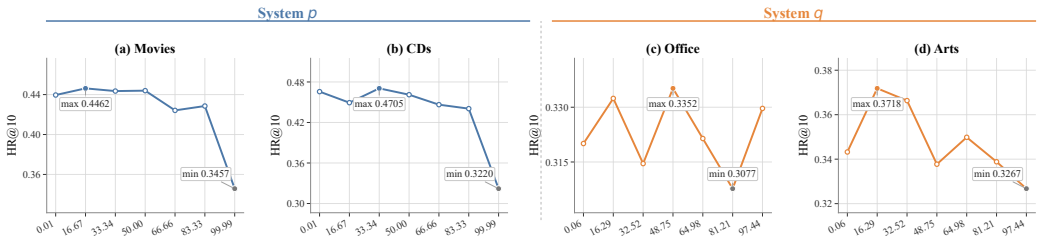
The consistent variation in the preferred ρ_p indicates that the relative usefulness of retained cross-context knowledge and newly introduced local interpretation depends on the recommendation en-



(a) Scenario 1: Video/Electronics–Phones/Sports.



(b) Scenario 2: Home/Kindle–Books/Tools.



(c) Scenario 3: Movies/CDs–Office/Arts.

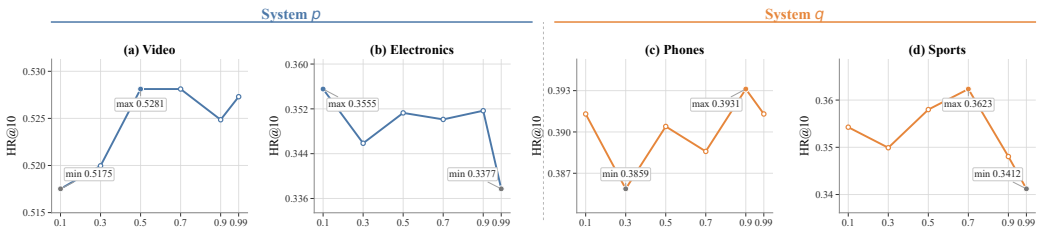
Figure 3: Sensitivity of FedSG to the semantic complementation criterion τ_p across the three collaboration scenarios. Performance is measured by HR@10.

vironment. Importantly, performance does not exhibit a common monotonic relationship with ρ_p: neither retaining almost all existing knowledge nor aggressively incorporating the collaborating system's mapping is uniformly optimal. This motivates selecting the coordination strength according to validation performance rather than fixing a global coefficient across systems and domains.

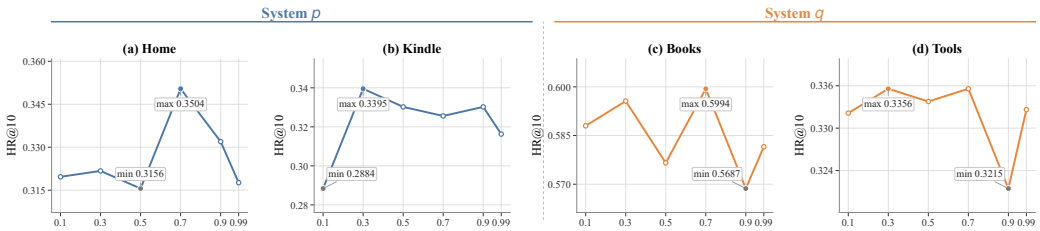
G.3 Number of Shared Semantic Concepts

Figures 5a–5c analyze the number of shared semantic concepts K. Increasing K expands the semantic schema available for characterizing items and therefore increases the breadth of the shared semantic space. Across the three scenarios, a richer schema is often beneficial compared with very small values of K, but the improvement is clearly non-monotonic. In Scenario 1, different domains reach their strongest performance with different schema sizes, showing that the useful semantic capacity is not uniform even within the same pair of systems. Scenario 2 exhibits a similar pattern, with the preferred K ranging across intermediate schema sizes and particularly large fluctuations for Kindle. Scenario 3 further highlights this domain dependence: Movies and CDs benefit from relatively broad semantic coverage, while Office peaks at a smaller schema size and subsequently declines, whereas Arts continues to benefit from additional concepts.

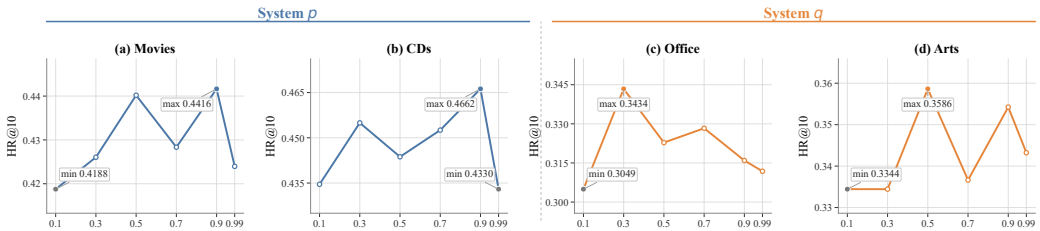
These results reveal a trade-off between semantic coverage and the inclusion of redundant or weakly relevant concepts. A small schema may omit semantic dimensions useful for relating heterogeneous items, whereas continually enlarging the schema does not guarantee additional recommendation benefit. The contrasting behavior of Office and Arts in Scenario 3 is particularly informative: additional concepts can help one local item space while providing little or even unfavorable information for another. Therefore, K should be viewed as controlling the useful semantic capacity of the shared schema rather than as a parameter for which larger values are inherently preferable.



(a) Scenario 1: Video/Electronics–Phones/Sports.



(b) Scenario 2: Home/Kindle–Books/Tools.



(c) Scenario 3: Movies/CDs–Office/Arts.

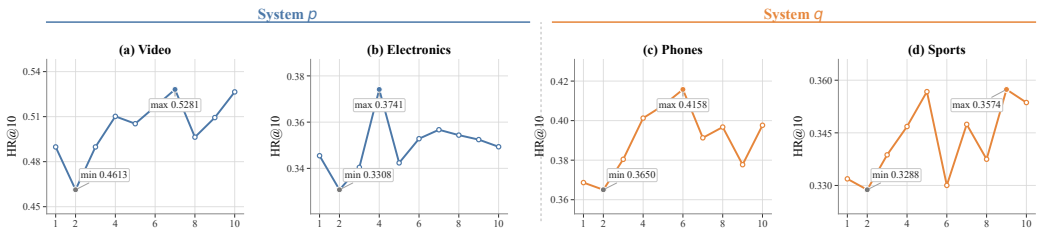
Figure 4: Sensitivity of FedSG to the preference coordination coefficient ρ_p across the three collaboration scenarios. Performance is measured by HR@10.

G.4 Number of Cross-Context Preference Examples

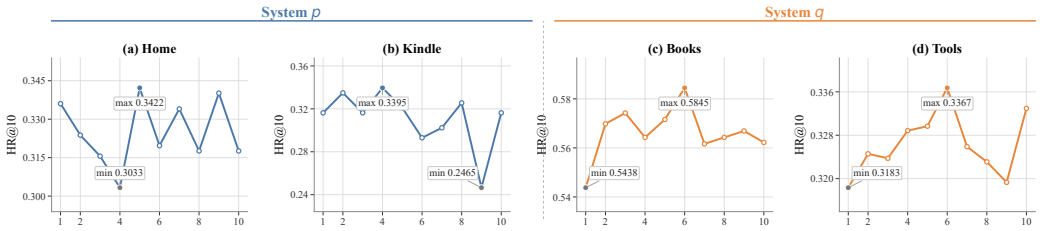
Figures 6a–6c study the amount of concrete contextual evidence included when re-expressing user preferences. We vary the number of cross-context preference examples over $\{0, 2, 4, 6\}$. Importantly, the setting with zero examples still retains the summary-level preference profile; it removes only the additional concrete examples. The experiment therefore evaluates whether, and to what extent, contextual examples provide information beyond the semantic preference summary.

The results show that such examples can be useful, but their marginal benefit is strongly domain-dependent. In Scenario 1, several domains benefit from adding examples, although their preferred counts differ. Scenario 2 shows a wider range of behaviors: Home and Tools benefit from a moderate number of examples, Kindle improves up to four examples before declining, whereas Books remains comparatively insensitive to the example count. Scenario 3 provides an especially clear contrast. Movies performs best using the summary alone, CDs changes only moderately, Office benefits substantially from a moderate number of examples, and Arts continues to improve with additional examples within the tested range.

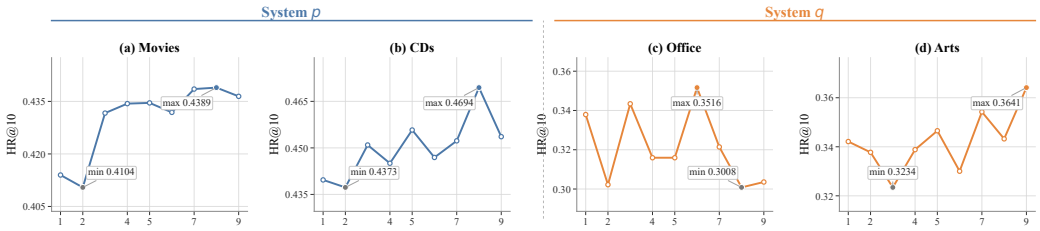
These results indicate that summary-level profiles already capture substantial transferable preference information, while concrete examples can provide complementary contextual detail when the target recommendation environment requires it. However, increasing the number of examples does not yield uniformly larger gains and may eventually provide redundant or less useful information. The preferred amount of contextual evidence should therefore be selected according to the local recommendation environment rather than fixed across systems.



(a) Scenario 1: Video/Electronics–Phones/Sports.



(b) Scenario 2: Home/Kindle–Books/Tools.

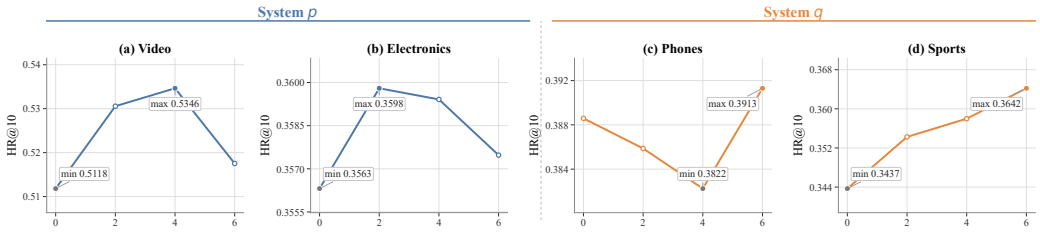


(c) Scenario 3: Movies/CDs–Office/Arts.

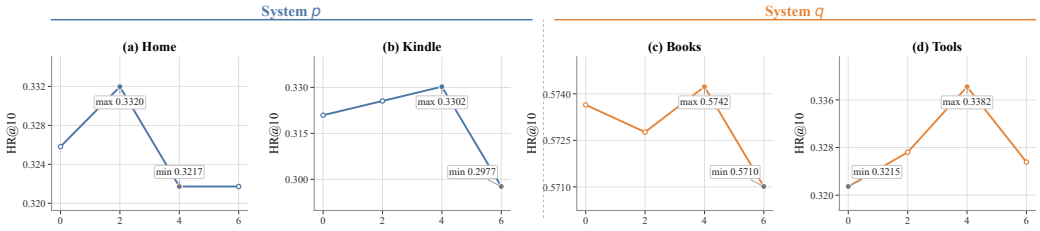
Figure 5: Sensitivity of FedSG to the number of shared semantic concepts K across the three collaboration scenarios. Performance is measured by HR@10.

G.5 Summary of Sensitivity Analysis

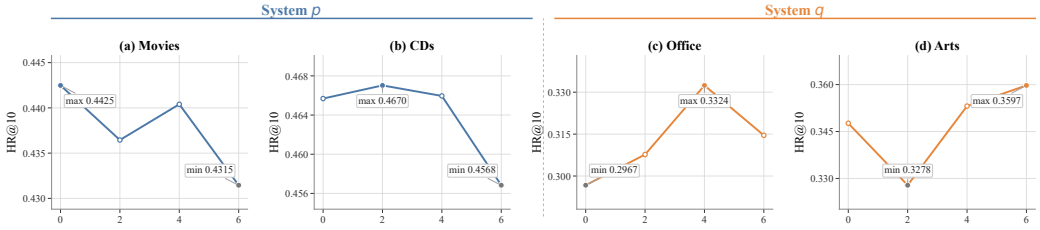
Taken together, the four analyses reveal a common property of FedSG: effective semantic grounding depends on controlling both how much cross-system knowledge is introduced and how richly that knowledge is represented. The coordination parameters τ_p and ρ_p govern the balance between local and cross-system knowledge, while K and the number of cross-context preference examples determine the breadth and contextual specificity of the semantic representation. None of these parameters exhibits a universally monotonic trend across domains, and their preferred values vary across recommendation environments. These results support validation-based parameter selection and, more importantly, show that semantic collaboration benefits from selective coordination rather than uniformly maximizing the amount of transferred semantic information.



(a) Scenario 1: Video/Electronics–Phones/Sports.



(b) Scenario 2: Home/Kindle–Books/Tools.



(c) Scenario 3: Movies/CDs–Office/Arts.

Figure 6: Sensitivity of FedSG to the number of cross-context preference examples across the three collaboration scenarios. Performance is measured by HR@10.