

ACE: Semantically-Grounded Graph Alignment via Affective Contrastive Learning

Potito Aghilar*
Politecnico di Bari
Bari, Italy

Sabino Roccotelli*
Politecnico di Bari
Bari, Italy

Vito Walter Anelli*
Politecnico di Bari
Bari, Italy

Alejandro Bellogín
Universidad Autónoma de Madrid
Madrid, Spain

Michelantonio Trizio
Wideverse
Bari, Italy

Tommaso Di Noia
Politecnico di Bari
Bari, Italy

Abstract

Graph Contrastive Learning (GCL) methods for recommendation learn representations by propagating signals over user-item interaction graphs. However, modeling these graphs with homogeneous edges can lead to *semantic-structural misalignment*, where information is exchanged between structurally adjacent but semantically dissimilar items, adversely affecting retrieval quality. Existing solutions typically rely on auxiliary encoders or additional supervision, increasing model complexity and training cost. We propose **ACE** (Affective Contrastive Embeddings), a framework that improves representation alignment by incorporating affective semantics into contrastive learning. ACE encodes the affective dimensions of *valence* and *arousal* as a topological prior, encouraging consistency between learned embeddings and an affective semantic space distilled from large language models. To operationalize this alignment, we introduce a *Semantically Weighted Noise Contrastive Estimation* (SW-NCE) loss that modulates contrastive gradients according to users' affective preferences. Experiments on Amazon, Last.fm, and iTunes demonstrate that ACE consistently improves top- K retrieval performance over 11 baselines while reducing computational overhead. These results indicate that affective geometric alignment is an effective and efficient mechanism for enhancing graph-based retrieval models.

CCS Concepts

• Information systems → Recommender systems.

Keywords

Graph Contrastive Learning; Recommender Systems

ACM Reference Format:

Potito Aghilar, Sabino Roccotelli, Vito Walter Anelli, Alejandro Bellogín, Michelantonio Trizio, and Tommaso Di Noia. 2026. ACE: Semantically-Grounded Graph Alignment via Affective Contrastive Learning. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '26)*, July 20–24, 2026, Melbourne, VIC, Australia. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3805712.3809526>

*Corresponding authors. These authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGIR '26, Melbourne, VIC, Australia*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2599-9/2026/07

<https://doi.org/10.1145/3805712.3809526>

1 Introduction

Music consumption is an inherently emotional activity, distinct from the utilitarian dynamics of e-commerce or news recommendation. Recent research in user modeling [1, 32] suggests that users do not merely consume content but consume moods, transitioning from high-energy workout playlists to melancholic introspection. Consequently, the quality of a recommendation depends as much on semantic coherence, aligning with the user's emotional trajectory, as on collaborative probability. Algorithmic approaches to recommendation relying on a Graph Neural Network architecture have shown their superior performance by exploiting the bipartite graph nature of user-item interaction signals [36]. Among these approaches, those based on Graph Contrastive Learning (GCL), such as Wu et al. [41] and Yu et al. [46], are the most promising in terms of accuracy in computing the recommendation list.

Unfortunately, the dominant paradigm in GCL operates under a “structural reductionist” assumption [12]. State-of-the-art models such as UltraGCN [22] and XSimGCL [45] achieve efficiency by simplifying message passing but remain “emotionally deaf,” treating the interaction graph as a static topology of identical binary edges. This leads to *Semantic Incongruence*: the structural proximity enforced by the collaborative objective often contradicts the semantic coherence required for user satisfaction [37]. Conversely, feature-based models and Context-Aware Recommender Systems capture attribute-level semantics [5], but fail to model high-order collaborative signals in the graph structure [36]. A growing trend attempts to bridge this gap using Large Language Models (LLMs) [55], but they often suffer from collaborative blindness, possessing world knowledge yet lacking access to the interaction graph [50]. The prevailing solution of *architectural stacking* adds computationally heavy auxiliary encoders or attention mechanisms on top of a GCN backbone [48]. We argue that this approach is theoretically redundant, attempting to learn via complex architecture what should be encoded via geometry.

We propose **Affective Contrastive Embeddings (ACE)**, a framework that incorporates the psychological dimensions of *valence* and *arousal* [28] not as auxiliary features, but as a topological prior in contrastive representation learning. ACE enforces *Semantic Isomorphism*, a geometric constraint requiring the learned latent topology to align with an affective manifold distilled from large language models. To realize this objective, we introduce a *Semantically Weighted Noise Contrastive Estimation* loss (SW-NCE), which modulates representation learning by shifting neighborhood aggregation from the forward pass to the backward pass. We interpret the resulting optimization dynamics through the lens of *anisotropic diffusion* [9],

enabling ACE to selectively attenuate gradient propagation across semantically dissimilar neighborhoods without relying on explicit graph convolution operations. Furthermore, ACE addresses user heterogeneity via *Adaptive Sensitivity Coefficients*, which allow the model to differentiate between users with stable affective preferences and those exhibiting more diverse interaction behavior. Extensive experiments across Amazon Digital Music [23], Last.fm [32], and iTunes [16] confirm that ACE outperforms 11 state-of-the-art baselines. By resolving semantic constraints via the loss function rather than the architecture, ACE operates as a “0-layer” model during inference, offering a sustainable Green AI solution that leverages LLMs as *Semantic Oracles* to guide latent geometry without burdening deployment. Our code is available in the [ACE Github repository](#).

2 Related Work

Modern collaborative filtering is defined by a fundamental tension between *structural learning* and *semantic understanding* [34, 47]. While Graph Contrastive Learning dominates the former, and Context-Aware RS the latter, current efforts to reconcile them typically culminate in *architectural stacking*, layering auxiliary encoders atop graph backbones [40]. This section reviews these paradigms and situates ACE as a geometric unification of the two.

Graph Contrastive Learning extracts self-supervised signals by maximizing mutual information between views of the interaction graph. Recent advances have moved away from computationally expensive structural augmentations (such as edge dropout) toward theoretical simplifications that generate views via lightweight embedding perturbations [46].

Stochastic View Generation and Signal Denoising. Early frameworks like SGL [41] established the paradigm by generating views via node/edge dropout. Building on the simplified linearity of LightGCN [19], subsequent research prioritized efficiency. UltraGCN [22] approximated infinite-layer convolution via constraint losses, while SimGCL [46] and XSimGCL [45] demonstrated that complex graph augmentations are unnecessary, replacing them with noise injection in the embedding space. LightGCL [8] further refined this by utilizing SVD to distinguish global collaborative signals from local noise.

Theoretical Unification of Optimization and Aggregation. Recent literature shifts towards theoretical unification and dynamic modeling. LightGCN++ [21] and LightGODE [49] address over-smoothing via improved sampling and continuous ODE dynamics, respectively, while MixRec [51] employs “Mixup” strategies to synthesize virtual interactions for sparse settings. A significant paradigm shift has been the theoretical unification of contrastive learning and neighborhood aggregation. EGCF [52] and ESIGCF [44] propose embedding-less architectures to reduce memory overhead. Most notably, LightCCF [53] proved that the gradient descent of the InfoNCE loss is mathematically equivalent to graph convolution. However, while this unifies structure and optimization, it treats the underlying graph as semantically neutral, failing to distinguish between structural connectivity and affective coherence.

Semantic Incongruence in Structural Topologies. Despite these advances, GCL methods remain “emotionally deaf.” By optimizing strictly for structural proximity, they enforce a topology where divergent thematic categories, such as a *melancholic breakup ballad*

and a *high-energy party anthem*, are treated as equivalent positive samples, leading to *semantic incongruence*. ACE addresses this limitation not by introducing additional layers, but by enforcing *Semantic Isomorphism* directly within the loss landscape.

Emotion-aware recommender systems align recommendations with psychological priors [30], typically grounding representation, such as Valence and Arousal, in Russell’s Circumplex Model [28]. While early methods relied on manual tagging, recent approaches use LLMs to extract affective features from multimodal content [32].

Feature-Based Contextual Fusion. The prevailing integration strategy treats affective context as an auxiliary feature vector concatenated with user and item embeddings. Shallow architectures like FM [26] and AFM [42] model second-order interactions, while deep architectures such as Wide&Deep [10], DeepFM [17], and DCNv2 [33] employ cross-networks to capture high-order feature intersections.

The reviewed methods suffer from *architectural stacking*, attempting to learn semantic alignment using complex feature-interaction modules rather than geometric constraints. This decouples the semantic manifold from the collaborative graph structure.

3 Preliminaries

3.1 Notation and Problem Definition

Let $\mathcal{U} = \{u_1, \dots, u_M\}$ and $\mathcal{I} = \{i_1, \dots, i_N\}$ denote the sets of users and items, respectively. The interactions are represented by the binary matrix $\mathbf{R} \in \{0, 1\}^{M \times N}$, where $r_{ui} = 1$ indicates an observed interaction. We associate each entity with learnable embedding vectors $\mathbf{e}_u, \mathbf{e}_i \in \mathcal{Z} \subset \mathbb{R}^d$, where $\mathcal{Z}$ is the latent space of embedding dimension d .

3.1.1 The Semantic Manifold Hypothesis. Standard collaborative filtering operates under a “structural reductionist” assumption [12], treating items as atomic nodes. We challenge this by assuming that every item exists not only in the discrete graph topology but also in a continuous semantic space.

DEFINITION 1 (THE SEMANTIC MANIFOLD). Let $\mathcal{S} \subset \mathbb{R}^h$, with h being the dimensionality of the semantic space, be a continuous, differentiable manifold where the distance metric $d_{\mathcal{S}}(\cdot, \cdot)$ corresponds to semantic dissimilarity. We assume the existence of a mapping function $\Phi: \mathcal{I} \rightarrow \mathcal{S}$ that projects an item i to a coordinate vector $\mathbf{c}_i \in \mathcal{S}$.

3.1.2 Ensemble Semantic Distillation. To operationalize Φ , we treat the Large Language Model (LLM) as a *Semantic Oracle* $\Phi_{\text{LLM}}(\cdot)$ [39]. Since generative models are inherently stochastic, a single prompt may yield a hallucinated representation, we propose a consensus-based extraction mechanism to construct a robust manifold.

DEFINITION 2 (ENSEMBLE SEMANTIC DISTILLATION). We define the robust semantic coordinate $\mathbf{c}_i$ as the expectation of the oracle’s projection $\Phi_{\text{LLM}}(\cdot)$ over a prompt space $\mathcal{P}$, where the different $p \in \mathcal{P}$ contain the system instruction and a random portion of the permuted item list. We define **Ensemble Semantic Distillation** as the Monte Carlo approximation of this expectation using a “council” of Q distinct prompts $\{p_1, \dots, p_Q\}$:

$$\mathbf{c}_i = \mathbb{E}_{p \sim \mathcal{P}} [\Phi_{\text{LLM}}(i, p)] \approx \frac{1}{Q} \sum_{q=1}^Q \Phi_{\text{LLM}}(i, p_q), \quad (1)$$

where $\Phi_{LLM}(i, p_q)$ is the output generated under prompt context p_q .

Ensemble Semantic Distillation is introduced since the expectation $\mathbb{E}_{p \sim \mathcal{P}}$ is computationally infeasible (due to the number of possible permutations). It acts as a noise-cancellation mechanism [38], distilling a stable semantic consensus.

Affective Manifold Instantiation. While ACE is generic to any semantic domain, in this study, we instantiate $\mathcal{S}$ as the *Affective Manifold*. Here, $\mathbf{c}_i = [v_i, a_i] \in \mathbb{R}^2$ represents the Valence (positivity) and the Arousal (energy) of the item. This 2D continuous space allows us to model complex emotional transitions (e.g., from “Sad” to “Happy”) as geometric trajectories rather than discrete category jumps.

3.1.3 User Anchor and Semantic Incongruence. Having established the item manifold, we define the user’s position within it, and we formalize the fundamental limitation of existing GCL models.

DEFINITION 3 (USER AFFECTIVE ANCHOR). We define a user’s semantic position $\mathbf{c}_u \in \mathcal{S}$ as the centroid of their historical interactions $\mathcal{N}_u$:

$$\mathbf{c}_u = \frac{1}{|\mathcal{N}_u|} \sum_{i \in \mathcal{N}_u} \mathbf{c}_i. \quad (2)$$

This vector represents the user’s “affective comfort zone” and serves as the geometric anchor for our contrastive alignment.

DEFINITION 4 (SEMANTIC INCONGRUENCE). Given a user-item pair (u, i) with an observed interaction $r_{ui} = 1$, Semantic Incongruence occurs when the structural proximity in the interaction graph contradicts the semantic distance in the manifold $\mathcal{S}$. Formally, a pair is incongruent if:

$$\text{sim}(\mathbf{e}_u, \mathbf{e}_i) > \xi \quad \text{AND} \quad d_{\mathcal{S}}(\mathbf{c}_u, \mathbf{c}_i) > \epsilon, \quad (3)$$

where $\text{sim}(\cdot, \cdot)$ is the structural similarity, and ξ, ϵ are thresholds for structural similarity and semantic distance, respectively.

Existing models only optimize for the former while ignoring the latter, leading to “emotionally deaf” representations.

3.2 Contrastive Learning as Aggregation

State-of-the-art collaborative filtering models, such as LightGCN, rely on explicit graph convolution layers to aggregate neighborhood information [19], typically following the rule $\mathbf{e}_u^{(k+1)} = \sum_{i \in \mathcal{N}_u} \alpha_{ui} \mathbf{e}_i^{(k)}$ where α_{ui} represents the graph normalization coefficient $\frac{1}{\sqrt{|\mathcal{N}_u|} \sqrt{|\mathcal{N}_i|}}$, with $|\mathcal{N}_i|$ being the number of the historical interactions of the neighbor v , and k being the convolution layer. While effective, this operation is computationally expensive and prone to over-smoothing. Explicit message passing has been shown to be theoretically redundant when using contrastive objectives [53].

Consider the standard InfoNCE loss for a user-item pair (u, i) [41]:

$$\mathcal{L}_{CL}(u, i) = -\log \frac{\exp(\text{sim}(\mathbf{e}_u, \mathbf{e}_i)/\tau)}{\sum_{j \in \mathcal{B}} \exp(\text{sim}(\mathbf{e}_u, \mathbf{e}_j)/\tau)} \triangleq -\log \frac{\exp(\mathbf{e}_u^\top \mathbf{e}_i/\tau)}{\sum_{j \in \mathcal{B}} \exp(\mathbf{e}_u^\top \mathbf{e}_j/\tau)}, \quad (4)$$

where $\mathcal{B}$ is a batch of items, and τ is the temperature. Analyzing the gradient of this loss with respect to the user embedding $\mathbf{e}_u$ reveals the implicit optimization:

$$\nabla_{\mathbf{e}_u} \mathcal{L}_{CL} \propto \underbrace{(p_{ui} - 1)\mathbf{e}_i}_{\text{Pull Positive}} + \sum_{j \in \mathcal{B}^-} \underbrace{p_{uj}\mathbf{e}_j}_{\text{Push Negatives}}, \quad (5)$$

where p_{ui} (p_{uj}) is the softmax probability, namely the predicted probability of item i (j) being the positive match for u , and $\mathcal{B}^-$ is the set of negative items in the optimization step. Since $p_{ui} \leq 1$ the term $(p_{ui} - 1)$ is negative. In gradient descent, this pulls $\mathbf{e}_u$ towards $\mathbf{e}_i$. Conversely, the term p_{uj} is positive. In gradient descent, this pushes $\mathbf{e}_u$ away from the negative embeddings $\mathbf{e}_j$. This implies that optimizing the InfoNCE loss is mathematically equivalent to performing neighborhood aggregation, as it minimizes the distance between a user and their interacted items in the latent space.

Semantic Blindness. Standard formulations perform indiscriminate aggregation. The objective pulls the user embedding towards all interacted items with equal force, modulated only by the dot product, regardless of semantic consistency. If a user interacts with an item inconsistent with their usual preferences (e.g., out of curiosity), blindly aggregating such items introduces “semantic noise.” This observation motivates our proposed method, which seeks to filter these noisy aggregations via semantic weighting.

4 ACE: Affective Contrastive Embeddings

In this section, we present **ACE** (**A**ffective **C**ontrastive **E**mbeddings). Building upon the theoretical insight that contrastive learning inherently performs neighborhood aggregation (Section 3.2), we propose a framework that replaces explicit graph convolutional layers with a geometry-aware contrastive objective. ACE does not treat affective context as auxiliary features to be processed, but as a topological prior that guides the gradient flow. We move the burden of context integration from a mere architectural level to the optimization level. This design is governed by a fundamental geometric principle we term *Semantic Isomorphism*.

DEFINITION 5 (SEMANTIC ISOMORPHISM). We define *Semantic Isomorphism* as the requirement that the learned latent space $\mathcal{Z}$ must be homeomorphic to the local geometry of $\mathcal{S}$ (defined in 3.1). Formally, assuming local convexity, for any user u and two items i, j , the optimization objective loss $\mathcal{L}$ over θ parameters must satisfy the monotonicity condition:

$$d_{\mathcal{S}}(\mathbf{c}_u, \mathbf{c}_i) < d_{\mathcal{S}}(\mathbf{c}_u, \mathbf{c}_j) \implies \|\nabla_{\theta} \mathcal{L}(u, i)\| > \|\nabla_{\theta} \mathcal{L}(u, j)\|, \quad (6)$$

This constraint ensures that the magnitude of the gradient flow is inversely proportional to the semantic distance. Consequently, the model effectively dampens the learning signal derived from “emotional noise” (interactions that are structurally positive but semantically incongruent with the user’s emotional state), while accelerating convergence for emotionally resonant items. To enforce this isomorphism without incurring the computational cost of heavy auxiliary encoders, ACE implements four coupled mechanisms: Latent Space Injection (Sec. 4.1) initializes the latent space geometry using affective priors; Adaptive Semantically-Grounded Modeling (Sec. 4.2) introduces two separate user-specific mood sensitivity coefficients; Semantically-Aware Ranking with Penalty (Sec. 4.3) integrates mood sensitivity in the scoring function; Semantically-Weighted Noise Contrastive Estimation (NCE) loss (Sec. 4.4) modulates the gradient flow via a mood-sensitive weighting function.

4.1 Latent Space Injection

The standard approach in Context-Aware Recommender Systems involves concatenating auxiliary features to user and item embeddings

or processing them via attention heads [33, 42]. We argue that this complexity is unnecessary. We treat affective context as a geometric prior, encoding semantic signals into the embedding initialization. We define a linear projection $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}^d$ that maps the semantic coordinates $\mathbf{c} \in \mathcal{S}$ of Valence and Arousal into the latent space:

$$\phi(\mathbf{c}) = \mathbf{W}_{va}\mathbf{c} + \mathbf{b}_{va}, \quad (7)$$

where $\mathbf{W}_{va}$ and $\mathbf{b}_{va}$ are trainable parameters. We inject this signal into the user and item embeddings to obtain context-enriched representations $\mathbf{z}_u$ and $\mathbf{z}_i$:

$$\mathbf{z}_u = \mathbf{e}_u + \delta\phi(\mathbf{c}_u), \quad \mathbf{z}_i = \mathbf{e}_i + \delta\phi(\mathbf{c}_i), \quad (8)$$

where δ modulates the representation enrichment. This operation acts as a *Semantic Positional Encoding* [56]. Rather than increasing the dimensionality of the input, it shifts the embeddings in the vector space based on their emotional characteristics before any collaborative learning occurs. This enforces a geometric initialization where entities with similar affective profiles are naturally organized, providing a starting point for the contrastive alignment.

4.2 Adaptive Semantically-Grounded Modeling

A rigid application of affective priors assumes that all users are equally driven by mood. This assumption is flawed; while some users strictly adhere to emotional anchor (e.g., maintaining a “sad” mood), others are eclectic explorers whose interactions span the emotional spectrum [27]. Enforcing strict Valence-Arousal alignment for the latter group would be detrimental. To address this, we introduce learnable user-specific weights that allow the model to automatically distinguish between “mood-sensitive” and “mood-agnostic” users. We define two scalar parameters for each user u :

- (1) λ_u^α : **Score-Oriented Mood Sensitivity**. This parameter governs the recommendation logic by controlling the semantic penalty applied to candidate items based on their distance from the user’s affective anchor. The model dynamically decides whether to enforce strict adherence to the user’s current mood or to permit broader exploration of the semantic manifold for eclectic users.
- (2) λ_u^γ : **Aggregation Mood Sensitivity**. This parameter modulates the Semantically-Weighted NCE loss. It determines the user-specific sensitivity to semantic dissonance, controlling how aggressively the model dampens the gradient flow from interactions that are structurally positive but affectively incongruent. This effectively allows the model to learn a personalized filter for the user’s interaction history.

The parameters are learned end-to-end, allowing ACE to dynamically adjust its reliance on semantic priors based on the user’s behavior.

4.3 Semantically-Aware Ranking with Penalty

To generate rankings that are both accurate and semantically coherent, we employ a Bayesian Personalized Ranking (BPR) loss augmented with an explicit semantic penalty. The predicted score $\hat{y}_{ui}$ is given by:

$$\hat{y}_{ui} = \text{sim}(\mathbf{z}_u, \mathbf{z}_i) - \alpha \cdot \lambda_u^\alpha \cdot d_S(\mathbf{c}_u, \mathbf{c}_i) + \beta \cdot b_i \triangleq \mathbf{z}_u^\top \mathbf{z}_i - \alpha \cdot \lambda_u^\alpha \cdot \|\mathbf{c}_u - \mathbf{c}_i\|_2 + \beta \cdot b_i. \quad (9)$$

This scoring function comprises three terms: the latent collaborative similarity ($\mathbf{z}_u^\top \mathbf{z}_i$), an explicit re-ranking penalty weighted by the user’s mood sensitivity weight λ_u^α , and an item bias term b_i . Here,

the hyperparameter α acts as a global semantic parameter, determining the severity of the penalty for affective incongruence, while β modulates the influence of intrinsic item popularity bias on the final ranking. This ensures that for mood-sensitive users, items with high emotional distance are explicitly demoted in the ranking list.

4.4 Semantically-Weighted NCE Loss

Standard Graph Contrastive Learning [41] performs indiscriminate aggregation, as discussed in Sec. 3.2, pulling user embeddings towards all positive items with equal force. We propose *Semantically-Weighted Noise Contrastive Estimation (NCE) loss*, a mechanism that performs soft-filtering of the interaction graph via the loss function. We first compute a confidence weight w_{ui} for each interaction, representing its semantic incoherence:

$$w_{ui} = \gamma \cdot \lambda_u^\gamma \cdot d_S(\mathbf{c}_u, \mathbf{c}_i) \triangleq \gamma \cdot \lambda_u^\gamma \cdot \|\mathbf{c}_u - \mathbf{c}_i\|_2. \quad (10)$$

Here, $d_S(\mathbf{c}_u, \mathbf{c}_i)$ is the Euclidean distance in the affective manifold $\mathcal{S}$, and γ is a global hyperparameter. As the emotional distance increases, w_{ui} increases, signaling a likely noise interaction. We integrate w_{ui} into the InfoNCE objective by modulating an adaptive semantic margin for the positive pair, obtaining a *Semantically-Weighted NCE loss*:

$$\mathcal{L}_{SWNCE}(u, i) = -\log \frac{\exp((\text{sim}(\mathbf{z}_u, \mathbf{z}_i) + w_{ui})/\tau)}{\sum_{j \in \mathcal{B}} \exp((\text{sim}(\mathbf{z}_u, \mathbf{z}_j) + w_{uj})/\tau)}, \quad (11)$$

where $\text{sim}(\cdot, \cdot)$ is a dot product and $\mathcal{B}$ represents the set of users and items in a batch. This formulation implements *Semantic Gradient Weighting* via an adaptive margin mechanism. When w_{ui} is high (indicating semantic incongruence for a positive interaction), the similarity score is artificially inflated. Consequently, the magnitude of the gradient $\nabla \mathcal{L}$ is reduced for that specific pair, allowing the model to “learn less” from emotionally dissonant interactions and perform denoising via gradient descent. This additive formulation also naturally handles negative samples. For a semantically distant negative item j , a large w_{uj} increases its contribution to the denominator’s partition function. This upward shift in the negative’s score strengthens the repulsive gradient, pushing the dissonant item further from the user’s representation. This design choice enforces a stricter geometric constraint than simple loss re-weighting. As established in margin-based representation learning literature [13, 31], multiplicative loss weights merely scale the gradient step size without altering the optimal target distance between embeddings. An additive margin, instead, explicitly shifts the decision boundary, forcing the latent space to physically reorganize to satisfy the semantic topological prior. This aligns with Riemannian Optimization principles, where the step size is large along semantically consistent manifolds and dampened across semantic faults [14, 24].

4.5 Optimization and Complexity Analysis

The model is trained end-to-end by minimizing the joint loss:

$$\mathcal{L} = \mathcal{L}_{BPR} + \omega_{SWNCE} \cdot \mathbb{E}_{(u, i) \sim \mathcal{R}} [\mathcal{L}_{SWNCE}(u, i)] + \lambda_{reg} \|\Theta\|_2^2, \quad (12)$$

where ω_{SWNCE} balances the representation learning and the ranking task, and λ_{reg} is the regularization coefficient.

Time Complexity. Standard GCN-based models (e.g., LightGCN, SGL) rely on explicit message passing layers to aggregate neighborhood information, incurring a computational cost of $O(L \cdot |E| \cdot d)$ per

epoch, where L is the number of layers, $|E|$ is the number of edges, and d is the embedding dimension. However, recent theoretical advancements by LightCCF [53] have shown that the gradient descent of the contrastive loss function is mathematically equivalent to performing graph convolution, suggesting that explicit stacking of GCN layers is theoretically redundant, often leading to over-smoothing without additional gain. Leveraging this insight, ACE adopts the “0-layer” paradigm. By relying on the *Semantically-Weighted NCE* loss to handle structural aggregation during the backward pass, we eliminate the need for explicit graph convolutions during the forward pass. Consequently, forward propagation in ACE is reduced to a simple lookup operation followed by latent space injection (Eq. 8). For a user u and item i , the inference score is computed as the dot product of their semantically-enriched embeddings, penalized by the user’s specific sensitivity gate λ_u^α (Eq. 9). This reduces the inference time complexity to $O(d)$ per pair, matching the speed of Matrix Factorization. Crucially, the training complexity per epoch is reduced to $O(|E| \cdot d)$, removing the multiplicative factor L and the overhead of sparse matrix multiplications required by GCNs.

Space Complexity. A critical bottleneck in scaling GCNs to industrial datasets is the memory requirement for storing the sparse adjacency matrix on the GPU. Standard models such as LightGCN and SGL have a space complexity of $O((M+N)d+|E|)$, where storing the graph structure $|E|$ often dominates memory usage for large-scale sparse graphs. In contrast, ACE operates as an *embedding-less GCN* during the forward pass. Since the structural information is encoded implicitly via the contrastive loss sampling rather than explicit matrix multiplication, ACE does not require loading the adjacency matrix into VRAM. The space complexity of ACE is strictly $O((M+N)d)$, allocated solely for user and item embeddings and lightweight auxiliary vectors (e.g., sensitivity coefficients λ_u). By decoupling space complexity from graph density $|E|$, ACE achieves linear scalability with respect to the user-item population. This facilitates deployment on industrial-scale datasets, effectively circumventing the memory bottlenecks that typically constrain traditional GCNs.

4.6 ACE as Anisotropic Manifold Diffusion

To understand the advantage of ACE over standard GCL, we analyze the gradient dynamics of the Semantically-Weighted NCE loss (Eq. 12). Standard GCNs perform *isotropic smoothing* [9]: they aggregate information from all neighbors $\mathcal{N}_u$ equally (normalized by degree), effectively treating the graph as a heat conductive surface with uniform thermal conductivity. This leads to the well-known “oversmoothing” problem where distinct user interests are washed out. In contrast, the gradient of the ACE objective with respect to the user embedding $\mathbf{e}_u$ can be derived as:

$$\nabla_{\mathbf{e}_u} \mathcal{L}_{ACE} \propto \sum_{i \in \mathcal{N}_u} \underbrace{(1 - p_{ui}(w_{ui}))}_{\text{Conductance } C_{ui}} \cdot \mathbf{e}_i, \quad (13)$$

where p_{ui} is the prediction probability modulated by w_{ui} (defined in Equation (10)). This formulation follows *Anisotropic Diffusion* [25]. Here, the term $C_{ui} = (1 - p_{ui})$ acts as the *conductance* coefficient. When the semantic distance between a user’s anchor and an item is high, w_{ui} increases, artificially inflating the logit value. This forces $p_{ui} \rightarrow 1$. Consequently, the conductance term $C_{ui} \rightarrow 0$, effectively halting the diffusion of information. Thus, ACE does not merely

Table 1: Statistical summary of the experimental datasets.

Dataset	# Users	# Items	# Interactions	Sparsity
Amazon DM	100,952	70,519	130,434	99.9982%
Last.fm	52,829	815,631	2,200,763	99.9949%
SiTunes	30	298	1850	79.3065%

“weight” samples; it dynamically alters the topology of the latent space during optimization, creating “semantic cuts” in the graph where structural edges exist but semantic coherence is lost. This explains why ACE avoids the collaborative blindness of standard GCNs without requiring explicit attention layers.

5 Experimental Setup

To validate the effectiveness of ACE, we conduct extensive experiments to answer the following research questions:

- **RQ1:** Does ACE outperform state-of-the-art graph-based and feature-based baselines in terms of recommendation accuracy, and how does it navigate the trade-off between accuracy and diversity (popularity bias)?
- **RQ2:** How effective are the *Adaptive Sensitivity Coefficients* in capturing user heterogeneity, and can they autonomously balance the trade-off between semantic consistency (for mood-driven users) and structural exploration (for eclectic users)?
- **RQ3:** Does ACE achieve a superior efficiency profile regarding training carbon footprint, and memory scalability by operating as a “0-layer” model?
- **RQ4:** Do Large Language Models act as reliable *Semantic Oracles*? Does LLM-distilled affective manifold exhibit *geometric isomorphism* with real human emotional perception?

5.1 Datasets

To evaluate the robustness of ACE we utilize three music benchmarks: Amazon Digital Music, Last.fm, and SiTunes. The statistics for these datasets are summarized in Table 1.

Amazon Digital Music. We utilize the Amazon Review Digital Music Dataset [23]. We retain the original non-binarized ratings (1–5 stars) to enable explicit feedback models to leverage the full granularity of user preferences, while implicit feedback baselines can internally binarize interactions using the standard threshold.

Last.fm. We adopt the Last.fm dataset introduced in Wang et al. [32]. We deliberately discard the user-generated tags, which often suffer from semantic propagation errors. Instead, we rely on our distilled affective priors to test the model’s capacity to correct structural noise.

SiTunes. To substantiate our semantic manifold hypothesis with psychometric ground truth, we incorporate SiTunes [16], a situational music recommendation dataset containing rich physiological and psychological signals. Unlike massive-scale datasets where affect is inferred, it provides human-annotated Valence-Arousal labels, used in RQ4 to validate the geometric isomorphism of the learned manifold.

A contribution of this work is the construction of the *Affective Manifold S* via the *Ensemble Semantic Distillation* process defined in Definition 2. To operationalize this, we employ a state-of-the-art

Prompt: "For each song in the list below (track id, track name by artist), provide the most precise estimated value of valence (musical positiveness) and arousal (energy/intensity) score on a scale from 0.0 to 1.0:
 198392, New Storms for Older Lovers by La Dispute
 . . .
 40904, They Dance Alone by Sting"

Figure 1: Ensemble Semantic Distillation prompt. We aggregate K Monte Carlo LLM queries to derive a stable Valence-Arousal centroid c_i from metadata.

Foundation Model, Gemini 2.5 flash lite [29], as the *Semantic Oracle*, configured with a fixed temperature and top-p at 0 and 0.95, respectively. For every unique track, we extract continuous Valence and Arousal coordinates $c_i \in \mathbb{R}^2$. Figure 1 reports the prompt used to retrieve the Valence and Arousal values. To mitigate the inherent stochasticity of generative models, we set the Monte Carlo ensemble size to $Q=10$, querying the oracle with batched items shuffled with seeds ranging from 42 to 51, eliciting Valence and Arousal scores on a continuous scale of $[0,1]$. The semantic coordinates c_i are computed as the centroid of these responses. This *offline* extraction, with the coordinates frozen, ensures that the LLM’s computational cost is decoupled from the recommender system’s training and inference.

5.2 Baselines

To evaluate ACE, we compare it, using the WarpRec framework [6], against 11 state-of-the-art baselines, categorized into *Graph-based* and *Feature-Based*, to highlight the trade-off between modeling high-order collaborative signals and integrating semantic context.

Graph-Based Models. These models prioritize topological signal propagation. We include LightGCN [19] as the foundational linear backbone. Representing Graph Contrastive Learning, we select SGL [41] and LightGCL [8], which generate views via structural dropout and SVD-based global signal extraction, respectively. Focusing on computational efficiency, we include UltraGCN [22] and XSimGCL [45]; these methods approximate infinite-layer convolution via constraint losses and embedding noise injection, thereby bypassing explicit message passing. Finally, we incorporate recent theoretical advancements: EGCF [52], which utilizes an embedding-less architecture for memory scalability, and LightCCF [53], which unifies contrastive optimization with neighborhood aggregation.

Feature-Based Models To validate the hypothesis that “structure alone is insufficient,” we compare against feature-based architectures: FM [26], AFM [42], Wide&Deep [10], and DCNv2 [33]. These models explicitly ingest the exact same LLM-derived Valence-Arousal vectors as ACE. However, they process this context via *architectural stacking*—treating affect as an auxiliary feature interaction—rather than as a geometric prior, allowing us to directly test the efficacy of topological alignment over feature concatenation.

5.3 Evaluation Metrics

To provide a holistic assessment of ACE, we move beyond simple accuracy maximization [4] to test the geometric and semantic properties of the learned latent space. We report standard top-K ranking

metrics, NDCG@K and Recall@K (with $K=10$), as our primary benchmark for collaborative utility. However, we argue that accuracy alone is insufficient to validate semantic alignment; a model can easily maximize Recall by recommending structurally dominant “pop” songs while remaining “emotionally deaf” to the user’s specific mood. To detect this scenario, we analyze the distributional properties of the recommendations. We report the Gini Index to quantify the inequality of recommendation frequency (where a lower score indicates mitigation of the “Matthew Effect”) and Expected Popularity Complement (EPC) to measure the model’s ability to surface semantically relevant but structurally sparse items from the long tail.

5.4 Implementation Details

Data Splitting. To simulate real-world scenarios, we employ a temporal holdout strategy. For each user, interactions are sorted by timestamp; the most recent 10% of interactions are reserved for the test set.

Hyperparameter Settings. To ensure reproducibility and fair comparison, all models are implemented in PyTorch and trained on a single NVIDIA A100 GPU. We optimize using the Adam optimizer with a batch size of 4096. We employ an early stopping strategy with both patience and grace period set to 20 epochs, monitoring $nDCG@10$ with 100 negative samples in evaluation. For the fixed parameters, the regularization coefficient and dropout rate are set to 10^{-4} and 0.1, respectively, while ω_{NA} is fixed at 1. The optimal hyperparameters are identified via grid search strategies: the embedding size is selected from $\{64, 128, 256, 512\}$, and the learning rate is tuned within $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. We ablate the coefficients α, β, γ , and δ over the set $\{0,1\}$, while the temperature parameter τ is selected from $\{0.15, 0.22, 0.4\}$. For baselines, we utilize the optimal hyperparameters tuned via the same grid search protocol.

6 Results and Discussion

In this section, we present a comprehensive evaluation of ACE. We benchmark against state-of-the-art graph-based and feature-based baselines (**RQ1**). We then deconstruct the model’s internal mechanics, investigating whether the *Adaptive Sensitivity Coefficients* successfully capture user heterogeneity by balancing semantic consistency for mood-driven users against structural exploration for eclectic ones (**RQ2**). Subsequently, we validate the efficiency claims of our “0-layer” inference paradigm, analyzing the training carbon footprint to determine if ACE qualifies as a sustainable Green AI solution (**RQ3**). Finally, we assess the reliability of the LLM-derived affective manifold by quantifying its *Geometric Isomorphism* with human emotional perception using psychometric ground truth (**RQ4**).

6.1 RQ1: Structural-Semantic Alignment and Recommendation Efficacy

To rigorously validate the hypothesis that *Semantic Isomorphism* provides a superior inductive bias compared to architectural complexity, we benchmark ACE against 11 state-of-the-art baselines. Table 2 summarizes the performance across three datasets. The empirical results substantiate three critical findings regarding the interplay between latent geometry and collaborative filtering.

Table 2: Top-10 ranking performance. Bold and underlined denote best and second-best results. † and * indicate statistical significance over the specific baseline and all baselines, respectively (paired t-test, $p < 0.05$), under Bonferroni correction.

Model	Amazon Digital Music				Last.fm				iTunes			
	nDCG↑	Recall↑	EPC↑	Gini↓	nDCG↑	Recall↑	EPC↑	Gini↓	nDCG↑	Recall↑	EPC↑	Gini↓
FM [26]	0.3395	0.5152	0.0753	0.8636	0.6085	0.6915	0.3215	0.8355	0.4558†	0.5023	0.0410	0.8189
Wide&Deep [10]	0.3155	0.4842	0.0699	0.8929	0.5360	0.6323	0.2846	0.8202	0.4731	0.5200	0.0545	0.8225
AFM [42]	0.3234	0.4954	0.0717	0.8784	0.5121	0.6117	0.2722	0.8813	0.4734	0.5020	0.0464	0.8305
DCNv2 [33]	0.3073	0.4798	0.0681	0.7678	0.5251	0.6259	0.2786	0.8543	0.4694	0.5045	0.0429	0.8164
LightGCN [19]	0.3037	0.3703	0.0673	<u>0.5829</u>	0.7659	0.8308	0.3993	0.8349	0.5370	0.5932	0.0519	0.7998
UltraGCN [22]	<u>0.4276</u>	<u>0.5766</u>	<u>0.0947</u>	0.8386	0.7815	0.8259	0.4070	0.8947	0.4305	0.4224†	0.0501	0.8590
SGL [41]	0.3141	0.3787	0.0698	0.5870	0.8217	0.8569	0.4261	0.7656	0.3593†	0.3470†	0.0189	0.5741
LightGCL [8]	0.3132	0.3887	0.0695	0.6471	0.7666	0.8146	0.3999	0.7801	0.3920†	0.4180†	0.0269	0.6733
XSimGCL [45]	0.3180	0.3863	0.0706	0.6154	0.7909	0.8164	0.4116	0.7459	0.3736†	0.3622	0.0273	0.5191
EGCF [52]	0.3112	0.3928	0.0691	0.6999	0.8367	<u>0.8661</u>	0.4328	0.7899	0.6079	0.6620	0.0553	0.7508
LightCCF [53]	0.2643	0.3240	0.0587	0.5546	<u>0.8383</u>	0.8617	<u>0.4331</u>	<u>0.7617</u>	0.2566†	0.2157†	0.0117	<u>0.5452</u>
ACE (Ours)	0.4928*	0.6364*	0.1093*	0.8088	0.8601*	0.8961*	0.4433*	0.8040	<u>0.6025</u>	<u>0.6400</u>	0.0567	0.7361

6.1.1 Overcoming the Limits of Architectural Stacking. ACE consistently outperforms both structure-centric (e.g., LightGCN, UltraGCN) and context-centric (e.g., Wide&Deep, DCNv2) paradigms on the large-scale Amazon and Last.fm datasets. On Amazon Digital Music, ACE achieves a statistically significant improvement in nDCG@10 over the strongest baseline, UltraGCN.

This result offers a profound critique of the *architectural stacking* status quo: context-aware baselines ingest the exact same LLM-derived Valence-Arousal vectors as ACE, yet they fail to match the performance of graph-based models. This confirms the phenomenon of *Collaborative Blindness*: treating affective context merely as a feature interaction fails to leverage the high-order connectivity of the user-item graph. Conversely, while recent efficient Graph Contrastive Learning models like LightCCF and EGCF successfully simplify the graph topology, they hit a performance ceiling by remaining “emotionally deaf”—optimizing for structural co-occurrence without semantic regularization. ACE bridges this gap not by increasing architectural depth, but by utilizing the affective manifold to modulate gradient flow, effectively achieving the structural reach of GCNs with the semantic precision of context-aware models.

6.1.2 Navigating the Accuracy-Diversity. A non-trivial outcome of our analysis is the relationship between accuracy (nDCG) and novelty (EPC). In standard Collaborative Filtering, optimizing for retrieval accuracy typically exacerbates popularity bias, resulting in lower EPC and a higher Gini coefficient [54]. ACE circumvents this trade-off: on the Amazon dataset, it achieves the highest accuracy (nDCG) concomitant with the highest novelty (EPC) and a significantly reduced Gini coefficient compared to the runner-up UltraGCN. We attribute this to the *Semantically-Weighted NCE* loss. Popular items often accumulate interactions due to exposure bias rather than genuine user preference (e.g., a “Sad” user clicking a “Happy” viral track). By penalizing these “emotional noise” interactions—structurally adjacent but semantically dissonant—ACE implicitly down-weights false-positive popular items during the backward pass. This filtering mechanism allows the model to surface niche, long-tail items semantically congruent with the user’s mood, demonstrating that high-fidelity recommendation requires geometric alignment rather than indiscriminate structural aggregation.

6.1.3 Robustness to Sparsity and Data Scale. The performance differential across datasets highlights the utility of semantic priors under varying data densities. On the massive, extremely sparse Amazon dataset, ACE’s semantic injection is critical for connectivity, yielding substantial gains. Conversely, on the small-scale, dense iTunes dataset, the structural signal is nearly complete (exhibiting a sparsity of 79.3% compared to ~99.9% for the other datasets). In this highly connected regime, purely structural models such as EGCF naturally excel, and ACE demonstrates comparable efficacy. This behavior aligns with established findings in context-aware and hybrid recommendation literature [18, 35]: when the interaction graph is dense and fully observed, the collaborative signal is largely saturated, thereby reducing the marginal utility of an explicit semantic or topological prior. Nevertheless, ACE maintains robust performance in this environment, confirming that the integration of semantic constraints does not degrade the underlying collaborative signal. More importantly, this result underscores that ACE is most impactful in sparse, real-world scenarios, where structural information alone is insufficient and the semantic prior becomes crucial for accurate retrieval.

6.1.4 Efficiency as a Geometric Byproduct. Cross-referencing these accuracy gains with the carbon footprint analysis in Figure 3, ACE operates within the Green AI quadrant. Unlike *architectural stacking* methods that achieve marginal accuracy gains through exponential parameter growth, ACE’s performance is driven by the *quality* of the optimization signal rather than the *quantity* of parameters. By operating as a “0-layer” model during inference, ACE confirms that aligning the latent space with the affective manifold is not only effective but computationally sustainable.

6.2 RQ2: Adaptive Sensitivity and the Semantic-Structural Trade-off

A core premise of ACE is that not all users are equally driven by affective coherence; some are “mood-sensitive” while others are “eclectic explorers.” This section studies whether the *Adaptive Sensitivity Coefficients* (λ_u^a and λ_u^v) successfully capture this heterogeneity, analyzing this at both the macroscopic (platform) and microscopic (user) levels.

6.2.1 Macroscopic Analysis: Platform-Specific Affective Cultures. Figure 2 visualizes the density distribution of the learned

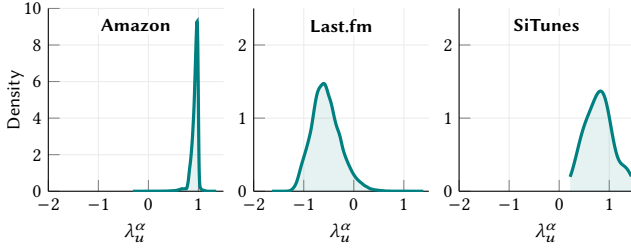


Figure 2: Distribution of learned user-specific Adaptive Sensitivity Coefficient (λ_u^α). Distinct distributions validate the necessity of user-level adaptation.

score-oriented sensitivity coefficient (λ_u^α) across the three datasets. The distinct morphological differences in these distributions reveal that ACE captures the underlying “affective culture” of each platform. On Amazon, the distribution is extremely narrow and peaked around $\lambda \approx 1.0$, reflecting the utilitarian nature of e-commerce where users typically rate items strictly aligned with a specific intent, necessitating high semantic consistency. Conversely, Last.fm exhibits significantly greater variance with a heavy tail, indicating that ACE autonomously grants “semantic slack” to a larger portion of the user base, enabling structural exploration beyond strict mood adherence in a passive listening environment. iTunes shows signs of bimodality, mirroring the controlled nature of the dataset, in which subjects were induced into specific emotional states, creating distinct clusters of “compliant” versus “resistant” subjects.

6.2.2 Microscopic Analysis: The “Semantic Shield” Mechanism. To understand the dynamics of ACE, we analyze the Semantic Shield Mechanism regulated by the score-oriented sensitivity coefficient. This parameter effectively governs the permeability of the semantic filter: a high λ_u^α enforces strict adherence to the user’s affective history, while a low λ_u^α permits the model to prioritize collaborative signals over semantic consistency. We examine two distinct user archetypes from the Last.fm dataset to validate this adaptive behavior.

- **The Mood-Sensitive (High λ_u^α):** User #8616 exhibits a focused interest in high-energy Pop/Funk, resulting in a high-valence, high-arousal centroid ($V \approx 0.70, A \approx 0.68$). Recognizing this semantic homogeneity, ACE converges on a high sensitivity coefficient ($\lambda_u^\alpha \approx 1.0$). The predictions remain topologically identical to the history (Semantic Distance ≈ 0.05), ensuring that structural noise does not dilute the user’s specific energy preference.
- **The Eclectic Explorer (Low λ_u^α):** User #33161 displays complex behavior ($|N_u| = 53$), oscillating between aggressive Metal and Pop. A static semantic constraint would force recommendations toward the geometric centroid—a “semantically ambiguous” mid-valence point ($V \approx 0.35$) that satisfies neither the Metal nor the Pop mode. Instead, ACE autonomously learns a low sensitivity ($\lambda_u^\alpha \approx 0.1$), effectively performing manifold relaxation. This allows the gradient flow to bypass the semantic penalty and follow the strongest structural gradients (the Metal cluster). The resulting recommendations are structurally relevant but semantically distinct from the mean ($V \approx 0.15$), yielding a high Semantic Distance of 0.28.

Table 3: Complexity analysis (per-epoch time; GPU space). Notation: L : layers, $|E|$: interactions, d : dim, S : views, M, N : users/items, q : rank, $\tilde{n}$: avg. features, L_c : cross layers. *: memory for 0-th layer embeddings and graph structure.

Model	Reference	Time Complexity	Space Complexity*
FM	ICDM '10 [26]	$O(E \cdot \tilde{n} \cdot d)$	$O((M+N)d)$
Wide&Deep	DLRS '16 [10]	$O(E \cdot (\tilde{n} + d_{DNN}))$	$O((M+N)d + \Theta_{DNN})$
AFM	IJCAI '17 [42]	$O(E \cdot \tilde{n}^2 \cdot d)$	$O((M+N)d + \Theta_{Att})$
DCNv2	WWW '21 [33]	$O(E \cdot L_c \cdot d^2)$	$O((M+N)d + L_c \cdot d^2)$
LightGCN	SIGIR '20 [19]	$O(L \cdot E \cdot d)$	$O((M+N)d + E)$
UltraGCN	CIKM '21 [22]	$O(E \cdot d)$	$O((M+N)d)$
SGL	SIGIR '21 [41]	$O(S \cdot L \cdot E \cdot d)$	$O((M+N)d + E)$
LightGCL	ICLR '23 [8]	$O(L[E \cdot d + q(M+N)d])$	$O((M+N)d + E)$
XSimGCL	TKDE '24 [45]	$O(L \cdot E \cdot d)$	$O((M+N)d + E)$
EGCF	TOIS '25 [52]	$O(L \cdot E \cdot d)$	$O(N \cdot d + E)$
LightCCF	SIGIR '25 [53]	$O(E \cdot d)$	$O((M+N)d)$
ACE (Ours)	-	$O(E \cdot d)$	$O((M+N)d)$

6.3 RQ3: Carbon Footprint vs. Accuracy

As recommender systems scale, the tension between performance and computational cost becomes a critical bottleneck. We evaluate whether ACE qualifies as a sustainable Green AI [7] solution by analyzing its training efficiency, inference latency, and memory scalability. While the generation of semantic priors via *Ensemble Semantic Distillation* incurs a significant one-time computational cost, we argue that this approach offers a more sustainable long-term solution. By shifting complexity from a recurring training cost (i.e., deep GCN layers trained for many epochs) to a one-time, offline inference cost, ACE dramatically reduces the computational budget of the operational training and deployment loop. Our analysis therefore focuses on this operational efficiency, which is critical for large-scale systems [2, 3]. To ensure a rigorous assessment of this operational phase, we used Codecarbon [11] to track real-time emissions against state-of-the-art baselines.

6.3.1 Accuracy vs. Carbon Footprint Analysis. Figure 3 illustrates the recommendation performance (nDCG@10) against the estimated carbon emissions (kg CO₂eq/s) during training for Amazon and Last.fm, this efficiency-accuracy trade-off is consistently observed on iTunes as well. The analysis reveals a distinct stratification across the efficiency landscape:

- **The High-Emission Cluster (UltraGCN, LightCCF):** Despite theoretical claims of simplification, these models exhibit emission rates orders of magnitude higher than efficient competitors. Their failure to achieve proportional accuracy gains suggests that architectural simplification alone does not guarantee sustainability.
- **The Low-Emission Competitors (SGL, Wide&Deep):** While these models cluster in the low-emission quadrant, they suffer from a “Performance Ceiling.” They lack the geometric regularization required to maximize retrieval quality, resulting in significantly lower accuracy.
- **The ACE Advantage (Top-Left Quadrant):** ACE dominates the top-left quadrant, rendering the cost-performance trade-off obsolete. It achieves state-of-the-art accuracy with emission rates comparable to the simplest baselines. This confirms that Semantic Isomorphism provides a superior optimization path compared to architectural depth; by shifting context integration from the

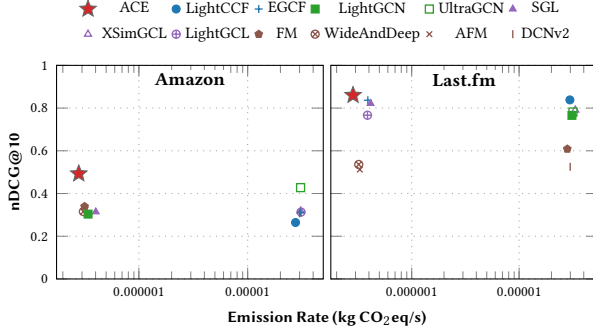


Figure 3: Recommendation performance vs. carbon footprint. ACE (red star) achieves competitive accuracy with significantly lower emission rates than baselines across Amazon and Last.fm datasets.

forward pass to the loss function, ACE eliminates high-emission training dynamics.

6.3.2 The “0-Layer” Inference Paradigm. ACE’s efficiency is theoretically grounded in its “0-layer” design (Sec. 4.5), which decouples graph reasoning from architectural depth. As defined in Table 3, unlike standard GCNs (e.g., LightGCN) that require recursive neighborhood aggregation—scaling inference cost with graph density $O(|E|)$ and demanding heavy VRAM to store sparse adjacency matrices—ACE operates as an embedding-less GCN during the forward pass. By shifting structural integration to the loss function, the trained embedding e_u inherently encapsulates the collaborative signal. This yields two critical advantages:

- **Inference Latency:** Complexity is reduced to $O(d)$ per pair (simple dot products), matching the speed of Matrix Factorization and enabling real-time deployment.
- **Memory Scalability:** Space complexity is strictly $O((M+N)d)$, allocated solely for embeddings. This eliminates the need to load the interaction graph into GPU memory, preventing Out-Of-Memory (OOM) errors and ensuring linear scalability on industrial-scale datasets.

6.4 RQ4: Semantic Oracle Validation

To validate the geometric isomorphism between the LLM-distilled manifold (S) and human perception, we utilize SiTunes, the only dataset in our suite with human-annotated Valence and Arousal (V-A) scores, to answer a critical question: *Does the LLM-derived latent geometry reflect human emotional perception sufficiently to guide the gradient descent?* We analyze the alignment between the LLM-generated coordinates and human perception through the lenses of Rank Preservation and Categorical Consistency.

6.4.1 Rank Preservation and Subjectivity Ceiling. Since ACE relies on relative distances to modulate gradient flow, rank ordering is paramount. We observe strong alignment for Arousal (Spearman’s ρ of 0.54 and Kendall’s τ of 0.38) and moderate alignment for Valence ($\rho = 0.33$ and $\tau = 0.22$). The correlations are statistically significant ($p < 0.001$). While the Valence scores are lower, they must be interpreted within the context of the “Subjectivity Ceiling” in Music Emotion

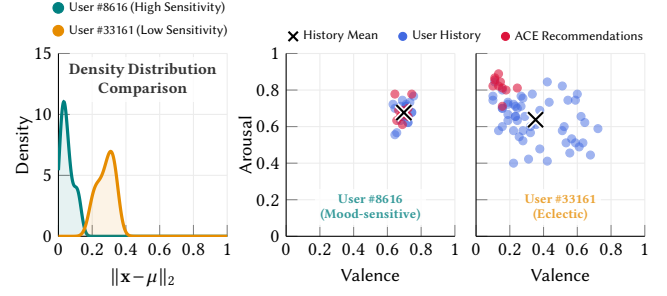


Figure 4: Qualitative analysis of Adaptive Sensitivity. Left: Density of anchor-to-item distances. Right: Recommendations relative to the history mean. User #8616 exhibits strict constraints, while User #33161 shows broader exploration.

Recognition. Literature confirms that while Arousal is acoustically objective, Valence is culturally subjective [43], with human inter-annotator agreement typically plateauing between 0.30–0.50 [15, 20]. Consequently, the LLM does not fail; rather, it successfully captures the “Conservative Consensus” of the track’s emotional profile.

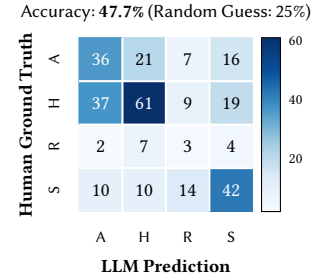


Figure 5: Confusion Matrix on SiTunes between human ground-truth and LLM-generated physiological and psychological signals. A = Angry, H = Happy, R = Relaxed, S = Sad.

6.4.2 Categorical Sanity Check. Finally, we investigate the distributional properties of the categorical signal. When discretizing the manifold into Russell’s four quadrants (Figure 5), the LLM achieves an accuracy of 47.7%, nearly doubling the random baseline (25%). The Confusion Matrix shows that errors are primarily “adjacent” (e.g., confusing “Happy” with “Relaxed”), while “diagonal” errors (e.g., confusing “Happy” with “Sad”) are rare.

The LLM is not a perfect replicator of individual human annotation, acting instead as a “Conservative Consensus.” However, the strong CCA score and significant rank correlation confirm that it successfully distills the structural properties of the affective manifold. This provides a mathematically sound basis for using LLM priors to guide the ACE framework in the absence of manual labels.

7 Qualitative Analysis

To validate the physical manifestation of *Adaptive Sensitivity Coefficients* in the latent space, we analyze the geometric arrangement of recommendations for two distinct user archetypes (Figure 4).

The Mood-Sensitive (User #8616). This user exhibits a learned sensitivity of $\lambda_u^\alpha \approx 1.0$. The Kernel Density Estimation of anchor-to-item distances reveals a high-kurtosis peak centered near zero (Fig 4, left), while the 2D projection shows recommendations tightly clustered around the affective centroid (Fig 4, right). Geometrically, ACE “collapses” the search space, acting as a strict filter that enforces local geometric isomorphism for users requiring affective consistency.

The Eclectic Explorer (User #33161). ACE assigns a low sensitivity of $\lambda_u^\alpha \approx 0.1$. The distance distribution is platykurtic with a heavy right tail, and the spatial trajectory forms a diffuse cloud deviating from the history mean. This confirms that the “Semantic Shield” is effectively lowered, allowing the gradient to flow across the manifold and prioritize structural connectivity over semantic proximity.

8 Conclusion

We revisit the challenge of integrating affective context into collaborative filtering and propose that the solution may lie in better geometric representations rather than increased architectural complexity. We introduced **ACE** (**A**ffective **C**ontrastive **E**mbeddings), a framework that operationalizes *Semantic Isomorphism* to align the latent interaction graph with the underlying affective manifold.

Our theoretical analysis reveals that this alignment can be achieved by unifying Graph Contrastive Learning with Anisotropic Diffusion. By shifting the burden of context integration to the backward pass, ACE effectively filters “emotional noise”—such as structurally adjacent but affectively dissonant interactions—without relying on explicit graph convolutions. Our results show that this approach enables ACE to outperform complex state-of-the-art baselines while operating as a “0-layer” model during inference. This represents a promising step toward Green AI, significantly reducing the computational and carbon costs of training and deployment. Furthermore, our validation on the SiTunes dataset suggests that Large Language Models, when stabilized via *Ensemble Semantic Distillation*, can serve as effective “Semantic Oracles” that respect the nuances of human emotional perception. We hope this work encourages the community to further explore the potential of geometric priors in recommender systems, shifting the focus from the quantity of parameters to the quality of the optimization signal. While our current implementation focuses on music emotion, our framework is fundamentally domain-agnostic. Future research will explore extending these principles to higher-dimensional, non-affective semantic spaces. By mapping the manifold $\mathcal{S}$ to price-category topologies in e-commerce or topic-sentiment spaces in news recommendation, we aim to demonstrate that geometric priors can serve as a highly scalable alternative to architectural complexity across a broad spectrum of recommendation scenarios.

Acknowledgments

The authors acknowledge partial support of OVS Fashion Retail Reloaded, Lutech Digitale 4.0, Servizi Locali 2.0, Konecra New KOCS ERA, and P+ARTS.

References

- [1] Potito Aghilar, Vito Walter Anelli, Andrea Lops, Fedelucio Narducci, Azzurra Ragone, Sabino Rocchetti, and Michelantonio Trizio. 2025. Adaptive User Modeling in Visual Merchandising: Balancing Brand Identity with Operational Efficiency. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2025, New York City, NY, USA, June 16–19, 2025*. ACM, 358–360. doi:10.1145/3699682.3730976
- [2] Potito Aghilar, Vito Walter Anelli, Michelantonio Trizio, and Tommaso Di Noia. 2023. Scalable Cloud-Native Pipeline for Efficient 3D Model Reconstruction from Monocular Smartphone Images. In *Software Architecture - 17th European Conference, ECSA 2023, Istanbul, Turkey, September 18–22, 2023, Proceedings (Lecture Notes in Computer Science)*, Bedir Tekinerdogan, Catia Trubiani, Chouki Tibermacine, Patrizia Scandurra, and Carlos E. Cuesta (Eds.). Springer, 266–282. doi:10.1007/978-3-031-42592-9_18
- [3] Potito Aghilar, Vito Walter Anelli, Michelantonio Trizio, Eugenio Di Sciascio, and Tommaso Di Noia. 2025. Training-free, Identity-preserving Image Editing for Fashion Pose Alignment and Normalization. *Expert Syst. Appl.* 293 (2025), 128579. doi:10.1016/J.ESWA.2025.128579
- [4] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, and Fedelucio Narducci. 2022. User-controlled federated matrix factorization for recommender systems. *J. Intell. Inf. Syst.* 58, 2 (2022), 287–309. doi:10.1007/S10844-021-00688-Z
- [5] Vito Walter Anelli, Tommaso Di Noia, Pasquale Lops, and Eugenio Di Sciascio. 2017. Feature Factorization for Top-N Recommendation: From Item Rating to Features Relevance. In *Proceedings of the 1st Workshop on Intelligent Recommender Systems by Knowledge Transfer & Learning co-located with ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, August 27, 2017 (CEUR Workshop Proceedings)*, Yong Zheng, Weike Pan, Shaghayegh (Sherry) Sahebi, and Ignacio Fernández (Eds.). CEUR-WS.org, 16–21. <https://ceur-ws.org/Vol-1887/paper3.pdf>
- [6] Marco Avolio, Potito Aghilar, Sabino Rocchetti, Vito Walter Anelli, Chiara Malamaci, Vincenzo Paparella, Marco Valentini, Alejandro Bellogin, Michelantonio Trizio, Joseph Trotta, Antonio Ferrara, and Tommaso Di Noia. 2026. WarpRec: Unifying Academic Rigor and Industrial Scale for Responsible, Reproducible, and Efficient Recommendation. *CoRR abs/2602.17442* (2026). arXiv:2602.17442 doi:10.48550/ARXIV.2602.17442
- [7] Joeran Beel, Alan Said, Tobias Vente, and Lukas Wegmeth. 2024. Green Recommender Systems: A Call for Attention. *SIGIR Forum* 58, 2 (2024), 1–5. doi:10.1145/3722449.3722468
- [8] Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. 2023. LightGCL: Simple Yet Effective Graph Contrastive Learning for Recommendation. In *ICLR*. OpenReview.net.
- [9] Ben Chamberlain, James Rowbottom, Maria I. Gorinova, Michael M. Bronstein, Stefan Webb, and Emanuele Rossi. 2021. GRAND: Graph Neural Diffusion. In *Proceedings of the 38th International Conference on Machine Learning, ICLR 2021, 18–24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 1407–1418. <http://proceedings.mlr.press/v139/chamberlain21a.html>
- [10] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishu Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ipsir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. 2016. Wide & Deep Learning for Recommender Systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*, Alexandros Karatzoglou, Balázs Hidasi, Domonkos Tikk, Oren Sar Shalom, Hagai Roitman, Bracha Shapira, and Lior Rokach (Eds.). ACM, 7–10. doi:10.1145/2988450.2988454
- [11] Benoit Courty, Victor Schmidt, Sasha Luccioni, Goyal-Kamal, MarionCoutarel, Boris Feld, Jérémy Lecourt, LiamConnell, Amine Saboni, Inimaz, supatomic, Mathilde Léval, Luis Blanche, Alexis Cruveiller, ouminasara, Franklin Zhao, Aditya Joshi, Alexis Bogroff, Hugues de Lavoreille, Niko Laskaris, Edoardo Abati, Douglas Blank, Ziyao Wang, Armin Catovic, Marc Alencon, Michał Stęchly, Christian Bauer, Lucas Otávio N. de Araújo, JPW, and MinervaBooks. 2024. *mlco2/codecarbon: v2.4.1*. doi:10.5281/zenodo.11171501
- [12] Jie Dai, Qingshan Li, Tianyi Nong, Qipeng Bi, and Hua Chu. 2024. Joint contrastive learning of structural and semantic for graph collaborative filtering. *Neurocomputing* 586 (2024), 127547.
- [13] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019*. Computer Vision Foundation / IEEE, 4690–4699. doi:10.1109/CVPR.2019.00482
- [14] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. 2021. Sharpness-aware Minimization for Efficiently Improving Generalization. In *ICLR*. OpenReview.net.
- [15] Gilles E. Gignac and Eva T. Szodorai. 2016. Effect size guidelines for individual differences researchers. *Personality and Individual Differences* 102 (2016), 74–78. doi:10.1016/j.paid.2016.06.069
- [16] Vadim Grigorev, Jiayu Li, Weizhi Ma, Zhiyu He, Min Zhang, Yiqun Liu, Ming Yan, and Ji Zhang. 2024. SiTunes: A Situational Music Recommendation Dataset with Physiological and Psychological Signals. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval, CHIIR 2024, Sheffield, United Kingdom, March 10–14, 2024*, Paul D. Clough, Morgan Harvey, and Frank Hopfgartner (Eds.). ACM, 417–421. doi:10.1145/3627508.3638343
- [17] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction.

- In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19–25, 2017*, Carles Sierra (Ed.). ijcai.org, 1725–1731. doi:10.24963/IJCAI2017/239
- [18] Ruining He and Julian J. McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12–17, 2016, Phoenix, Arizona, USA*, Dale Schuurmans and Michael P. Wellman (Eds.). AAAI Press, 144–150. doi:10.1609/AAAI.V30i1.9973
- [19] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *SIGIR*. ACM, 639–648.
- [20] Xiao Hu and J. Stephen Downie. 2010. When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands, August 9–13, 2010*, J. Stephen Downie and Remco C. Veltkamp (Eds.). International Society for Music Information Retrieval, 619–624. <http://ismir2010.ismir.net/proceedings/ismir2010-106.pdf>
- [21] Geon Lee, Kyungho Kim, and Kijung Shin. 2024. Revisiting LightGCN: Unexpected Inflexibility, Inconsistency, and A Remedy Towards Improved Recommendation. In *RecSys*. ACM, 957–962.
- [22] Kelong Mao, Jieming Zhu, Xi Xiao, Biao Lu, Zhaowei Wang, and Xiuqiang He. 2021. UltraGCN: Ultra Simplification of Graph Convolutional Networks for Recommendation. In *CIKM*. ACM, 1253–1262.
- [23] Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 188–197. doi:10.18653/v1/D19-1018
- [24] Razvan Pascanu and Yoshua Bengio. 2014. Revisiting Natural Gradient for Deep Networks. In *ICLR*.
- [25] Pietro Perona and Jitendra Malik. 1990. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 7 (1990), 629–639. doi:10.1109/34.56205
- [26] Steffen Rendle. 2010. Factorization Machines. In *ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14–17 December 2010*, Geoffrey I. Webb, Bing Liu, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu (Eds.). IEEE Computer Society, 995–1000. doi:10.1109/ICDM.2010.127
- [27] Kyle Robinson, Dan Brown, and Markus Schedl. 2020. User Insights on Diversity in Music Recommendation Lists. In *Proceedings of the 21th International Society for Music Information Retrieval Conference, ISMIR 2020, Montreal, Canada, October 11–16, 2020*, Julie Cumming, Jin Ha Lee, Brian McFee, Markus Schedl, Johanna Devaney, Cory McKay, Eva Zangerle, and Timothy de Reuse (Eds.). 446–453. <http://archives.ismir.net/ismir2020/paper/000311.pdf>
- [28] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
- [29] Gemini Team. 2023. Gemini: A Family of Highly Capable Multimodal Models. CoRR abs/2312.11805 (2023). arXiv:2312.11805 doi:10.48550/ARXIV.2312.11805
- [30] Hieu Tran, Tuan Le, Anh Do, Tram Vu, Steven Bogaerts, and Brian Howard. 2023. Emotion-Aware Music Recommendation. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7–14, 2023*, Brian Williams, Yiling Chen, and Jennifer Neville (Eds.). AAAI Press, 16087–16095. doi:10.1609/AAAI.V37i1.26911
- [31] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. 2018. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18–22, 2018*. Computer Vision Foundation / IEEE Computer Society, 5265–5274. doi:10.1109/CVPR.2018.00552
- [32] Javier Wang, Alejandro Bellogin, and Iván Cantador. 2024. Integrating sentiment features in factorization machines: Experiments on music recommender systems. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2024, Cagliari, Italy, July 1–4, 2024*. ACM, 183–188. doi:10.1145/3627043.3659561
- [33] Ruoxi Wang, Rakesh Shivanna, Derek Zhuayan Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed H. Chi. 2021. DCN V2: Improved Deep & Cross Network and Practical Lessons for Web-scale Learning to Rank Systems. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19–23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 1785–1797. doi:10.1145/3442381.3450078
- [34] Shoujin Wang, Yan Wang, Fikret Sivrikaya, Sahin Albayrak, and Vito Walter Anelli. 2023. Data science for next-generation recommender systems. *Int. J. Data Sci. Anal.* 16, 2 (2023), 135–145. doi:10.1007/S41060-023-00404-W
- [35] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4–8, 2019*, Ankur Teredesai, Vipin Kumar, Ying Li, Römer Rosales, Evimaria Terzi, and George Karypis (Eds.). ACM, 950–958. doi:10.1145/3292500.3330989
- [36] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21–25, 2019*, Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer (Eds.). ACM, 165–174. doi:10.1145/3331184.3331267
- [37] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled Graph Collaborative Filtering. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25–30, 2020*, Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 1001–1010. doi:10.1145/3397271.3401137
- [38] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-Consistency Improves Chain-of-Thought Reasoning in Language Models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1–5, 2023*. OpenReview.net. <https://openreview.net/forum?id=1PL1NIMMrw>
- [39] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 – December 9, 2022*, Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.). http://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html
- [40] Wei Wei, Chao Huang, Lianghao Xia, and Chuxu Zhang. 2023. Multi-Modal Self-Supervised Learning for Recommendation. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 – 4 May 2023*, Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben (Eds.). ACM, 790–800. doi:10.1145/3543507.3583206
- [41] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised Graph Learning for Recommendation. In *SIGIR*. ACM, 726–735.
- [42] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19–25, 2017*, Carles Sierra (Ed.). ijcai.org, 3119–3125. doi:10.24963/IJCAI2017/435
- [43] Yi-Hsuan Yang and Homer H. Chen. 2012. Machine Recognition of Music Emotion: A Review. *ACM Trans. Intell. Syst. Technol.* 3, 3 (2012), 40:1–40:30. doi:10.1145/2168752.2168754
- [44] Zhi Yang, Ruizhang Huang, Yanping Chen, Chuan Lin, and Yongbin Qin. 2025. ESIGCF: Extremely simplified but intent-enhanced graph collaborative filtering for recommendation. *Eng. Appl. Artif. Intell.* 162 (2025), 112495.
- [45] Junliang Yu, Xin Xia, Tong Chen, Lizhen Cui, Nguyen Quoc Viet Hung, and Hongzhi Yin. 2024. XSimGCL: Towards Extremely Simple Graph Contrastive Learning for Recommendation. *IEEE Trans. Knowl. Data Eng.* 36, 2 (2024), 913–926.
- [46] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. 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, Madrid, Spain, July 11 – 15, 2022*, Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai (Eds.). ACM, 1294–1303. doi:10.1145/3477495.3531937
- [47] Zheng Yuan, Fajie Yuan, Yu Song, Youhua Li, Junchen Fu, Fei Yang, Yunzhu Pan, and Yongxin Ni. 2023. Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23–27, 2023*, Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete (Eds.). ACM, 2639–2649. doi:10.1145/3539618.3591932
- [48] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, and Liang Wang. 2021. Mining Latent Structures for Multimedia Recommendation. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 – 24, 2021*, Heng Tao Shen, Yueting Zhuang, John R. Smith, Yang Yang, Pablo César, Florian Metze, and Balakrishnan Prabhakaran (Eds.). ACM, 3872–3880. doi:10.1145/3474085.3475259
- [49] Weizhi Zhang, Liangwei Yang, Zihe Song, Henry Peng Zou, Ke Xu, Liancheng Fang, and Philip S. Yu. 2024. Do We Really Need Graph Convolution During Training? Light Post-Training Graph-ODE for Efficient Recommendation. In *CIKM*. ACM, 3248–3258.
- [50] Yang Zhang, Fuli Feng, Jizhi Zhang, Keqin Bao, Qifan Wang, and Xiangnan He. 2025. CoLLM: Integrating Collaborative Embeddings Into Large Language Models for Recommendation. *IEEE Trans. Knowl. Data Eng.* 37, 5 (2025), 2329–2340. doi:10.1109/TKDE.2025.3540912
- [51] Yi Zhang and Yiwen Zhang. 2025. MixRec: Individual and Collective Mixing Empowers Data Augmentation for Recommender Systems. In *WWW*. ACM, 2198–2208.

- [52] Yi Zhang, Yiwen Zhang, Lei Sang, and Victor S. Sheng. 2025. Simplify to the Limit! Embedding-Less Graph Collaborative Filtering for Recommender Systems. *ACM Trans. Inf. Syst.* 43, 1 (2025), 22:1–22:30. doi:10.1145/3701230
- [53] Yu Zhang, Yiwen Zhang, Yi Zhang, Lei Sang, and Yun Yang. 2025. Unveiling Contrastive Learning's Capability of Neighborhood Aggregation for Collaborative Filtering. In *SIGIR*. ACM, 1985–1994.
- [54] Minghao Zhao, Le Wu, Yile Liang, Lei Chen, Jian Zhang, Qilin Deng, Kai Wang, Xudong Shen, Tangjie Lv, and Runze Wu. 2022. Investigating Accuracy-Novelty Performance for Graph-based Collaborative Filtering. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai (Eds.). ACM, 50–59. doi:10.1145/3477495.3532005
- [55] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, and Qing Li. 2024. Recommender Systems in the Era of Large Language Models (LLMs). *IEEE Trans. Knowl. Data Eng.* 36, 11 (2024), 6889–6907. doi:10.1109/TKDE.2024.3392335
- [56] Jianhang Zhu, Rongpeng Li, Xianfu Chen, Shiwen Mao, Jianjun Wu, and Zhifeng Zhao. 2024. Semantics-Enhanced Temporal Graph Networks for Content Popularity Prediction. *IEEE Trans. Mob. Comput.* 23, 8 (2024), 8478–8492. doi:10.1109/TMC.2023.3349315