Spectral Content Ranking Method Based on Heterogeneous Hypergraph Models

P. A. Reshetarov*, E. V. Khvorostukhina‡

A content ranking method based on heterogeneous hypergraph models is proposed. A 5-partite hypergraph is constructed on music listening data. Latent behavioral clusters, serving as proxy representations of genre and mood characteristics, are extracted from the interaction matrix using SVD and the k-means method. Ranking is performed by a generalized spectral operator. In experiments on the Yambda-50M dataset, an NDCG@10 value of 0.469 is achieved, outperforming PureSVD by 7.1 % and ItemKNN by 23.6 %.

Keywords: hypergraph, recommender systems, spectral methods, music recommendation, Yambda.

1. Introduction

Hypergraph models describe relationships of arbitrary arity and find applications in recommender systems [1]. The ability of a hyperedge to combine an arbitrary number of entities makes it possible to model complex interactions in content consumption without loss of context. Previously, the authors proposed a model based on a 4-partite hypergraph (user, artist, genre, mood) on Last.fm data [2]. However, that approach operated on macro-level artist characteristics and did not distinguish individual tracks. In this work, the construction is generalized to a 5-partite listening hypergraph (hereafter LH), where tracks are explicitly included as independent vertices. The experimental study is conducted on the open Yambda dataset [3].

The goal of this work is to develop and evaluate a track ranking method based on the spectral operator of the LH. Let five pairwise disjoint base vertex sets be given: users V_U, tracks V_T, artists V_A, and two sets of latent behavioral clusters V_G and V_M serving as proxy representations of genre and mood attributes, along with listening counts n_{ut} ($u \in V_U$, $t \in V_T$). Integration of additional information channels is permitted. For each user, a ranked list of unheard tracks must be generated in decreasing order of predicted interest.

* *Reshetarov Platon Aleksandrovich* — Master's student, SSTU, Dep. of Information and Communication Systems and Software Engineering, e-mail: grecc10@bk.ru

‡ *Khvorostukhina Ekaterina Vladimirovna* — PhD, Associate professor, SSTU, Dep. of Inform. and Commun. Systems and Software Engineering; SSU, Dep. of Geometry, e-mail: katyanew2007@rambler.ru, ORCID: 0000-0002-2775-5732

Published under CC-BY 4.0

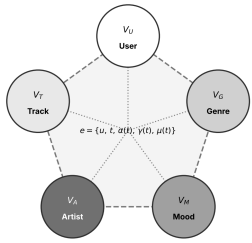


Figure 1: 5-partite listening hypergraph

2. Proposed approach

The listening hypergraph (LH) is defined as a 5-partite hypergraph $\mathcal{H} = (V, E)$ with vertex set $V = V_U \sqcup V_T \sqcup V_A \sqcup V_G \sqcup V_M$. Each hyperedge $e = \{u, t, \alpha(t), \gamma(t), \mu(t)\} \in E$ is formed when $n_{ut} > 0$ and contains exactly one vertex of each type, where $\alpha\colon V_T \to V_A$, $\gamma\colon V_T \to V_G$, $\mu\colon V_T \to V_M$ are mappings that associate a track with its artist and two latent behavioral clusters, respectively. Removing the track vertex (reverting to the previously proposed [2] 4-partite hypergraph) makes tracks with identical triples (α, γ, μ) indistinguishable. In the dataset used, approximately 70 % of tracks fall into this situation. Explicit inclusion of the track in the hyperedge structure fully eliminates this drawback. An example of the LH structure is shown in Figure 1.

Latent feature extraction. Since the Yambda dataset [3] does not contain explicit genre or mood annotations, the mappings γ and μ are constructed from behavioral data in four steps: 1) a sparse interaction matrix $L \in \mathbb{R}^{|V_U| \times |V_T|}$ with entries $L_{ut} = \ln(1 + n_{ut})$ is formed from training interactions only; 2) truncated SVD is computed: $L \approx U_r \Sigma_r V_r^\top$ with rank $r = 64$ to obtain dense latent profiles; 3) the rows of $V_r \in \mathbb{R}^{|V_T| \times r}$ are ℓ_2-normalized to produce track feature vectors; 4) `MiniBatchKMeans` with $k = 20$ and $k = 15$ (with different random seeds) determines the clusters γ and μ. The two clusterings capture distinct latent groupings of tracks based on co-listening patterns; the labels "genre" and "mood" are conventional designations reflecting the intended proxy role of these clusters, not verified semantic categories.

For each type of relationship in the LH structure, a local incidence matrix $H_\tau \in \{0, 1\}^{n \times m_\tau}$ is defined, where $n = |V_T|$ and m_τ is the cardinality of the vertex set of type τ. The approach permits integration of additional information channels beyond the original LH structure. The following blocks are considered: H_U — the transposed binary user–track interaction matrix constructed from training data, H_A — the track–artist membership matrix, H_G and H_M — matrices encoding track membership in genre and mood clusters.

Additionally, through the supplementary block H_L, the external structure of track–album membership from the dataset metadata is incorporated. The generalized spectral operator of the LH Θ, aggregating incidence blocks with weights $w_\tau > 0$, extends hypergraph-based collaborative filtering [1] to spectral aggregation: $\Theta = \sum_\tau w_\tau H_\tau D_{e_\tau}^{-1} H_\tau^\top \in \mathbb{R}^{n\times n}$, where $D_{e_\tau} = \mathrm{diag}(H_\tau^\top \mathbf{1}) \in \mathbb{R}^{m_\tau \times m_\tau}$ is the diagonal hyperedge degree matrix of block τ. The entry Θ_{ij} is proportional to the intensity of co-occurrence of tracks i and j in shared contextual hyperedges. To approximate the operator without computing the dense matrix, a block matrix M is formed from vertically stacked blocks of the form $\sqrt{w_\tau}\, D_{e_\tau}^{-1/2} H_\tau^\top$, so that $\Theta = M^\top M$. Let $V_k \in \mathbb{R}^{n\times k}$ be the matrix of k right singular vectors of M corresponding to the largest singular values. For user u, a listening history vector $x_u = (\ln(1 + n_{ut}))_t \in \mathbb{R}^n$ is formed from training interactions. The final predicted score vector is computed by projection onto the constructed basis: $s_u = x_u V_k V_k^\top \in \mathbb{R}^n$, after which unheard tracks are ranked in decreasing order of $s_u[t]$. The proposed method does not involve gradient-based optimization or supervised learning; all components (SVD, clustering, spectral decomposition) are computed in a single pass over the training data. When using a single block H_U, the operator reduces to a hyperedge-degree-normalized variant of collaborative filtering; integration of additional contextual blocks (H_A, H_L, H_G, H_M) enriches the spectral basis with information about entity relationships absent from the original sparse listening matrix. The approach is implemented in Python 3 (`TruncatedSVD`, `MiniBatchKMeans`). In the experiments, the block weights are fixed at $w_U = w_A = w_L = 1.0$, $w_G = w_M = 0.5$, and the basis rank is $k = 256$.

3. Experimental results and conclusions

From the Yambda-50M dataset [3], listening events with a playback fraction $\geq$50 % were selected for 30,000 popular tracks and 2,000 active users (2,841,958 interactions, 29,973 tracks, 8,396 artists, albums extracted from metadata, density 4.7 %). The baselines are Random, Pop (by popularity), ItemKNN (cosine similarity over the full item co-occurrence matrix without neighbor truncation), and PureSVD ($k = 128$). Unique user–track interaction pairs were split globally into training, validation, and test sets in a 64/16/20 ratio by uniform random assignment, repeated for three random seeds. Top-10 metrics were computed over the full catalog excluding training items (standard deviation ≤ 0.004).

Analysis of the results (Table 1) shows that the single-block model LH(U) achieves NDCG@10 comparable to PureSVD (0.437), establishing a collabo-

Table 1: Ranking metrics on the test set

Metric	Random	Pop	ItemKNN	PureSVD	LH(U)	LH(U,A)	LH(U,A,L,G,M)
P@10	0.011	0.175	0.367	0.429	0.427	0.459	**0.461**
NDCG@10	0.011	0.185	0.379	0.437	0.437	0.467	**0.469**
HR@10	0.102	0.699	0.952	0.980	0.981	0.990	**0.993**

rative filtering baseline for the operator framework. Integration of the artist block (LH(U,A)) increases NDCG@10 by 6.9 % relative to PureSVD. The full contextual model LH(U,A,L,G,M) achieves the best NDCG@10 = 0.469 (+7.1 % over PureSVD) and HR@10 = 0.993. The most significant metric improvement is provided by adding the artist relationship structure. Thus, the 5-partite LH with a spectral operator effectively combines heterogeneous content information and outperforms classical approaches. The spectral basis constructed here can serve as input features for hypergraph neural network (HGNN) architectures, enabling end-to-end trainable models that leverage both the structural advantages of the proposed hypergraph and the representational power of deep learning. Promising research directions include joint optimization of operator block weights on the validation set and integration of the spectral operator into HGNN-based pipelines for further improvement of ranking quality.

References

[1] L. Xia, C. Huang, Y. Xu, J. Zhao, D. Yin, J. Huang, "Hypergraph contrastive collaborative filtering", *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022, 70–79. DOI: 10.1145/3477495.3532058.

[2] P. A. Reshetarov, "Description of a graph neural network for a music recommender system", *Mathematical and Computer Modelling: Proceedings of the XII Scientific Conference (Omsk, 2025)*, OSU, Omsk, 2025, 297–299, (In Russian).

[3] A. Ploshkin, V. Tytskiy, A. Pismenny, V. Baikalov, E. Taychinov, A. Permiakov, D. Burlakov, E. Krofto, N. Savushkin, "Yambda-5B – A Large-Scale Multi-modal Dataset for Ranking And Retrieval", 2025, arXiv: 2505.22238.

Received on November 9, 2026