IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Topological Data Analysis and Graph-Based Learning for Multimodal Recommendation

KHALIL BACHIRI[1,2], (Fellow, IEEE), ALI YAHYAOUY[2,3], Maria MALEK[1], and Nicoleta ROGOVSCHI[4]

[1]ETIS Laboratory, ENSEA, UMR8051, CNRS, CY Cergy Paris University, Cergy-Pontoise, 95011, France
[2] L3IA Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, 30003, Morocco
[3] LaMSN - La Maison des Sciences Numériques, Sorbonne Paris Nord University, La Plaine Saint-Denis, F-93210, France
[4]LIPADE Laboratory, University Paris Cité, Paris, 75006, France

Corresponding author: Khalil BACHIRI (e-mail: khalil.bachiri@usmba.ac.ma).

ABSTRACT Multimodal recommendation systems are becoming increasingly vital for delivering personalized content by utilizing various data sources, including text, images, and user interaction histories. However, current multimodal methods face challenges such as modality heterogeneity, data sparsity, and feature redundancy, which can result in less effective performance when dealing with complex, high-dimensional datasets. In this study, we present a new framework that combines Topological Data Analysis (TDA) with graph-based learning to improve multimodal recommendations (TDA-MMRec). Our approach captures higher-order dependencies and global structural patterns in multimodal data, enhancing the robustness and expressiveness of the representations we learn. By using persistent homology, we extract topological descriptors that convey stable structural information across different modalities, addressing the issues of sparsity and redundancy. We also introduce a modality-aware strategy for constructing graphs, which integrates features derived from TDA into multimodal similarity graphs to maintain both local and global structural properties. Furthermore, we propose a topological pruning technique that refines graph structures by removing redundant connections while preserving essential topological information, enhancing computational efficiency. Extensive experiments on large-scale multimodal datasets indicate that our TDA-augmented framework significantly outperforms leading multimodal recommendation models on key ranking metrics, including Precision@20, Recall@20, and NDCG@20. Our ablation studies confirm that topological descriptors are essential in boosting representation learning, especially in cold-start scenarios where traditional methods struggle due to data sparsity.

INDEX TERMS Multimodal Recommendation Systems, Topological Data Analysis (TDA), Persistent Homology, Graph Neural Networks (GNNs), Representation Learning, Multimodal Fusion

I. INTRODUCTION

WITH the swift growth of digital platforms, recommender systems (RSs) have become vital tools for enhancing user experiences in e-commerce, streaming platforms, and social networks. These systems strive to offer personalized recommendations by analyzing user interactions, preferences, and contextual information [1]. However, as data sources diversify, the integration and utilization of multimodal information poses significant challenges to recommendation accuracy and scalability. While traditional approaches like collaborative filtering (CF) and deep learning have shown effectiveness, they often struggle with issues like data sparsity, noisy user-item interactions, and the differences between modalities [2], [3]. In this research, we examine a novel approach that employs Topological Data Analysis (TDA) to improve multimodal recommendation systems by capturing structural and relational patterns across various modalities.

Multimodal recommendation systems attempt to unify different data modalities to improve user preference modeling [4]. Conventional techniques such as Collaborative Filtering [2], Content-Based Filtering [5], and hybrid models [6] and [7] proposed multiview clustering with a manifold structure of graph regularization with a sparse distance graph. The classic model has evolved with deep learning architectures like Neural Collaborative Filtering (NCF) [8] and Graph Neural Networks (GNNs) [9], [10]. Despite these advancements, multimodal systems face several fundamental challenges: Different modalities exhibit varying levels of sparsity, leading to inconsistencies in feature extraction [11]. Current multimodal fusion techniques struggle to fully capture high-order relationships between modalities [12]. Processing and fusing diverse data types require significant computational resources, particularly in large-scale industrial applications [13]. Learning from multimodal datasets introduces noise and redundant information, affecting generalization performance [14]. Emphasize the essential need to dynamically reveal major multimodal features [15] in adaptive feature extraction techniques. TDA offers a mathematical framework for extracting robust structural representations from high-dimensional and multimodal data [16]. By leveraging concepts such as persistent homology, TDA identifies invariant features in datasets, capturing global and local structures that persist across different resolution scales [17]. Unlike conventional deep learning approaches that focus on local feature embeddings, TDA provides insights into the global shape of data distributions, enabling better generalization and robustness [18].

In recent years, TDA has been successfully applied in various machine learning domains [19], and biomedical data mining [20]. However, its potential in multimodal recommendation systems remains largely unexplored. This study bridges this gap by integrating TDA-based descriptors into multimodal recommendation frameworks, enhancing feature extraction, fusion, and predictive performance.

This paper proposes a novel TDA-augmented graph-based learning approach for multimodal recommendation (TDA-MMRec). Our method combines persistent homology with GNNs to enhance multimodal feature integration and improve recommendation quality. Specifically, our contributions are as follows:

- Multimodal Feature Representation: We extract topological features from different modalities, including images, text, user behavior, pricing attributes, and temporal using Vietoris-Rips filtrations, topological descriptors and persistent homology.
- Latent Structure Learning: We introduce a modality-aware graph construction framework that integrates TDA-derived features with multimodal similarity graphs.
- Graph-Based Learning with TDA: We refine recommendation models by incorporating topological descriptors into GNNs to capture high-order dependencies between user-item interactions.
- Efficient Graph Pruning via TDA: We propose a TDA-based pruning strategy to remove redundant connections while preserving topological structures, improving computational efficiency.

By leveraging TDA in multimodal recommendation, we hypothesize that persistent topological features can enhance representation learning, mitigate modality imbalances, and improve recommendation accuracy. Our experimental evaluation on large-scale datasets demonstrates that the proposed TDA-enhanced framework outperforms state-of-the-art multimodal recommendation models evaluations metrics.

The remainder of this paper is organized as follows: **Section II** reviews related work in multimodal recommendation and TDA-based learning. **Section III** presents the theoretical foundations of TDA and its application in recommendation systems. **Section IV** describes our proposed methodology, detailing the integration of TDA with graph learning (TDA-MMRec). **Section V** provides an extensive experimental analysis, comparing our approach with baseline methods. Finally, **Section VI** discusses the findings, limitations, and future research directions.

II. RELATED WORK

A. MULTIMODAL RECOMMENDATION

Multimodal recommendation systems leverage diverse data modalities to enhance recommendation accuracy and user personalization. Early approaches primarily relied on traditional methods like CF, which utilized user-item interaction matrices. However, these methods suffered from inherent challenges, including data sparsity, limited contextual understanding, and the inability to incorporate rich information from multiple modalities. The integration of multimodal data into recommendation systems began with feature-based models that extracted and combined modality-specific features [21]. For instance, visual features derived from item images using CNNs such as ResNet and EfficientNet, and textual features extracted using word embeddings like Word2Vec and transformers such as BERT, were incorporated into traditional frameworks. A seminal work in this domain, Visual Bayesian Personalized Ranking (VBPR) [22], introduced visual features into CF, while DeepStyle [23] combined visual and textual data to enhance user preference modeling. Despite these advancements, these early systems struggled to scale and failed to capture complex inter-modal relationships. The advent of deep learning marked a significant turning point for multimodal recommendation systems, enabling the development of neural architectures for unified representation learning. Neural Collaborative Filtering (NCF) [8] extended CF by embedding users and items into a shared latent space. Building on this, for jointly learned feature representations across modalities using deep neural networks, achieving notable performance improvements. Graph-based models further advanced multimodal recommendation by modeling intricate relationships among users, items, and modalities. Techniques such as Neural Graph Collaborative Filtering (NGCF) [11] and Multimodal Graph Convolutional Networks (MMGCN) [24] constructed modality-specific graphs and applied graph convolutions to aggregate features. These methods captured high-order connectivity and inter-modality interactions, but they often struggled with noise, sparsity, and overlooked item-item relationships. To address these issues, the LATent sTructure mining framework for multImodal reCommEndation (LATTICE) [25] introduced a graph-based latent structure learning approach that captured both inter-item relationships and multimodal feature dependencies, achieving state-of-the-art results. Despite these advancements, several challenges persist. Data sparsity, multimodal heterogeneity fusion [26], and noise in multimodal datasets remain significant obstacles. Moreover, effectively integrating diverse modalities requires advanced fusion techniques to preserve modality-specific information while enabling cross-modal interactions. Embeddings reduce data into low-dimensional representations, often sacrificing critical structural and topological information. This limitation hinders the ability to fully exploit the complementary nature of multimodal data, resulting in suboptimal recommendations.

Recent efforts have explored the integration of generative models and large language models (LLMs) into multimodal recommendation. MMRec [27] leverages pre-trained LLMs to model cross-modal alignment between text, vision, and user behavior through unified semantic representations, enabling more context-aware recommendations. DiffMM [28] introduces a multimodal diffusion model that generates high-quality recommendations by modeling the item generation process as a stochastic denoising trajectory over multimodal data. These approaches demonstrate the potential of generative modeling in capturing complex semantic dependencies across modalities. In contrast to these works, our method focuses on topological properties of multimodal data by leveraging persistent homology to extract global structural features. Rather than relying on generative alignment or reconstruction, our framework captures both local and high-order modality interactions through stable topological descriptors. This makes our approach orthogonal and complementary to generative paradigms, and opens promising directions for combining topological reasoning with large-scale generative models in future work.

B. TOPOLOGICAL DATA ANALYSIS IN MACHINE LEARNING

TDA has emerged as a transformative tool in machine learning, leveraging its ability to capture global structural properties of data while maintaining robustness to noise and perturbations [16], [17]. TDA provides a unique framework for understanding high-dimensional and non-linear data, with applications ranging from feature extraction to graph-based learning. This subsection highlights TDA contributions to machine learning, its mathematical modeling within these contexts, and limitations that motivate further research [29].

TDA integrates persistent homology with machine learning to extract meaningful topological features, making it particularly effective for complex data representations. One of the most prominent uses of TDA is in feature extraction, where persistent homology summarizes topological features such as connected components, loops, and voids. Given a dataset $\mathcal{P} \subseteq \mathbb{R}^n$ represented as a point cloud, TDA computes persistence diagrams $\mathcal{D}_k$ for homology groups H_k, capturing the evolution of k-dimensional features across filtration scales [30]. These persistence diagrams are often transformed into vectorized descriptors. These descriptors are incorporated into machine learning pipelines for tasks such as shape recognition, texture classification, and clustering [17].

Unlike traditional graph-based models that focus on node- and edge-level relationships, TDA models topological invariants such as connected components, loops (cycles), and voids. These structures encode global graph properties that persist across multiple scales, enabling richer and more stable representations than purely neighborhood-based aggregations. Persistent homology tracks the emergence and persistence of features across filtration thresholds, providing global structural insights [31]. These features have been integrated into GNNs to enhance tasks such as graph classification and community detection [32]–[34].

Recent work has explored TDA integration with deep learning, where persistent homology provides interpretability and robustness. For CNNs, persistent homology captures structural patterns in intermediate feature maps, enriching representations for image and text data [16]. Moreover, the development of differentiable topological layers has facilitated the incorporation of topological features directly into neural network architectures, enhancing their expressive power [29].

Multimodal machine learning combines heterogeneous data modalities such as images, text, and graphs to improve model performance. Unlike traditional graph-based multimodal methods, which often rely on fixed neighborhood aggregation (e.g., message passing in GNNs), TDA captures global structural properties beyond pairwise relationships. Persistent homology provides a robust framework to analyze higher-order interactions between modalities, detecting stable topological features that persist across multiple scales [35], [36]. This is particularly advantageous for multimodal recommendation, where modality interactions may be hierarchical and non-linear. By constructing topological embeddings that encode the persistence of critical multimodal features, TDA alleviates issues of modality imbalance, sparsity, and over-smoothing that challenge GNN-based approaches [37]. By capturing latent topological structures, TDA addresses issues arising from modality-specific noise and sparsity [38]. TDA enables the construction and analysis of multimodal graphs, capturing relationships between modalities via higher-order topological features [31]. These advantages make TDA a promising tool for multimodal contexts, motivating its integration into frameworks like multimodal recommendation systems.

III. THEORETICAL FOUNDATIONS OF TOPOLOGICAL DATA ANALYSIS (TDA)

TDA is a mathematical framework for capturing the intrinsic structure of data by leveraging tools from algebraic topology [39]. Unlike traditional machine learning techniques that rely on statistical assumptions or feature engineering, TDA extracts robust topological features that persist across multiple scales, making it particularly effective in high-dimensional, multimodal, and graph-structured data [16], [17]. This section presents the mathematical underpinnings of TDA, focusing on simplicial complexes, persistent homology, and topological feature representations [37] for graph-based multimodal recommendation.

A. MATHEMATICAL BACKGROUND

TDA relies on fundamental concepts from topology, algebra, and computational geometry. Below, we define the core mathematical structures.

1) Topological Spaces and Simplicial Complexes

A *topological space* $(X, \mathcal{T})$ is a set X equipped with a topology $\mathcal{T}$, a collection of subsets of X (called *open sets*) satisfying:

- $\emptyset, X \in \mathcal{T}$.
- The union of any collection of sets in $\mathcal{T}$ is also in $\mathcal{T}$.
- The intersection of a finite number of sets in $\mathcal{T}$ is also in $\mathcal{T}$.

A *simplicial complex* K is a combinatorial structure composed of simplices:

- A *0-simplex* (vertex) represents a data point.
- A *1-simplex* (edge) represents pairwise relationships.
- A *2-simplex* (triangle) captures three-way interactions.
- Higher-dimensional simplices encode multi-way relationships.

A simplicial complex K satisfies:

- If $\sigma \in K$, then all faces of σ are also in K.
- The intersection of two simplices in K is either empty or a face of both.

2) Filtrations

A *filtration* is a nested sequence of simplicial complexes:

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K, \tag{1}$$

where each subcomplex K_i is associated with a filtration function $f : K \to \mathbb{R}$, assigning birth times to simplices. The sequence tracks the evolution of topological structures as a threshold parameter increases.

3) Homology and Persistent Homology

Homology captures connected components (H_0), loops (H_1), and voids (H_2) in a topological space. The k-th homology group is defined as:

$$H_k(K) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1}), \tag{2}$$

where ∂_k is the boundary operator mapping k-simplices to their $(k-1)$-dimensional faces.

Persistent homology extends classical homology by tracking the birth and death of features across filtrations [31]. Given two subcomplexes K_α and K_β with $\alpha \leq \beta$, the inclusion map $i : K_\alpha \hookrightarrow K_\beta$ induces a homomorphism:

$$i_* : H_k(K_\alpha) \to H_k(K_\beta). \tag{3}$$

Each feature's lifespan is recorded as a persistence interval:

$$[\alpha_{\text{birth}}, \alpha_{\text{death}}). \tag{4}$$

B. TOPOLOGICAL FEATURE REPRESENTATIONS

Persistent homology encodes topological structures as numerical descriptors for machine learning.

1) Persistence Diagrams and Stability

A persistence diagram $\mathcal{D}_k$ is a multiset of birth-death pairs (b, d):

$$\mathcal{D}_k = \{(b_i, d_i) \mid i \in I\}. \tag{5}$$

Features with $d_i - b_i \gg 0$ are structurally significant, while those with $d_i - b_i \approx 0$ are noise [30].

The stability theorem ensures that persistence diagrams are robust to small perturbations in the input data [38]. Let $D(f)$ and $D(g)$ denote persistence diagrams for filtration functions f and g. The bottleneck distance W_∞ is defined as:

$$W_\infty(D(f), D(g)) = \inf_{\gamma} \sup_{x \in D(f)} \|x - \gamma(x)\|_\infty, \tag{6}$$

where γ ranges over all bijections between $D(f)$ and $D(g)$. The stability theorem states:

$$W_\infty(D(f), D(g)) \leq \|f - g\|_\infty. \tag{7}$$

2) Topological Descriptors

Persistence diagrams provide a rich mathematical framework for capturing topological structures, but their raw form is not directly usable in most machine learning models. To integrate topological information into feature extraction and predictive models, persistence diagrams must be transformed into structured numerical representations. These topological descriptors extract critical information about the persistence and significance of topological features across different scales. Below, we outline key descriptors, their mathematical formulations, and their roles in multimodal feature extraction.

a: Total Persistence

$$P = \sum_i (d_i - b_i) \tag{8}$$

Total persistence measures the sum of all feature lifetimes, providing a global estimate of the topological complexity of the data [40]. Higher total persistence indicates a dataset rich in persistent topological features, making it particularly useful in multimodal recommendation systems where long-lived structures capture meaningful inter-modal relationships.

b: Betti Curves

$$\beta_k(\alpha) = \mathrm{rank}(H_k(K_\alpha)) \tag{9}$$

Betti curves encode the number of persistent topological features (connected components, loops, and voids) as a function of the filtration parameter α [16]. These curves track how features evolve across scales, making them essential for characterizing hierarchical structures in graph-based multimodal data.

c: Persistence Landscapes

$$\lambda_k(t) = \sup_{j \geq k} \min(t - b_j, d_j - t) \tag{10}$$

Persistence landscapes transform persistence diagrams into a sequence of piecewise linear functions, enabling efficient statistical analysis and incorporation into machine learning models. For multimodal recommendation systems, these landscapes facilitate robust comparisons between users and items by preserving high-order structural properties of their interactions.

d: Silhouettes

$$S(t) = \sum_{i \in I} w_i \phi_i(t) \tag{11}$$

Silhouettes provide a weighted summary of persistence landscapes, where each topological feature is assigned a weight proportional to its persistence $(d_i - b_i)^p$ [41]. This descriptor is particularly useful in filtering out noise and emphasizing the most significant topological structures in high-dimensional multimodal data spaces.

e: Entropy of Persistence Diagrams

$$H(D) = -\sum_i p_i \log p_i, \quad \text{where } p_i = \frac{d_i - b_i}{\sum_j (d_j - b_j)} \tag{12}$$

This entropy measure quantifies the complexity of topological features, distinguishing structured multimodal interactions from random noise. Higher entropy values indicate diverse and meaningful topological structures, making it an effective descriptor for multimodal recommendation systems that rely on the integration of heterogeneous data sources.

Each of these descriptors provides a unique perspective on topological structures, allowing for the robust characterization of multimodal data. By leveraging these representations, we enhance the interpretability and efficiency of graph-based multimodal recommendation models, ensuring that both local and global topological properties are effectively captured. The next section explores its integration into graph-based learning architectures for multimodal feature fusion.

IV. METHODOLOGY

In this section, we proposed a novel framework that integrates TDA with Graph-based Neural Networks to address the challenges of multimodal recommendation systems (TDA-MMRec). This innovative approach effectively combines topological insights and graph-based learning techniques to improve the representation, interaction, and structural understanding of visual and textual data modalities. Additionally, user-behavior and pricing & productions attributes modalities are combined to make the framework more responsive to user preferences and external factors. At its foundation, the model is designed to capture local and global relationships within multimodal data while addressing challenges such as data heterogeneity, noise, and the complexity of fusing information across diverse modalities. By leveraging persistent homology from TDA, the model extracts higher-order topological features, such as persistence and connectivity, which complement the structural representations learned by GNNs. The framework follows a modular architecture comprising three: Multimodal Preprocessing, where visual and textual features are extracted and transformed using TDA techniques. Latent Structure Learning, topological features extracted during preprocessing are incorporated into the graph representations, enriching the learned embeddings with topological stability and interpretability. The fused graph is processed through a GNN to learn node embeddings, leveraging both structural and topological insights and optimization by Bayesian Personalized Ranking (BPR) loss function, designed to prioritize user-item interactions based on observed preferences.

A. MULTIMODAL FEATURE REPRESENTATION AND PREPROCESSING

1) Visual Feature Representation using TDA

To robustly represent the structural properties of visual modality, we apply TDA to extract descriptors from grayscale product images. These descriptors are designed to capture persistent topological features across multiple geometric transformations, offering a complementary perspective to conventional CNN features.

a: Image Representation as Cubical Complexes

An image can be represented as a cubical complex, where each pixel corresponds to a cubical cell. Let I be a gray-scale image of size $m \times n$. The pixel intensities are defined by a function $f : \mathbb{Z}^2 \to \mathbb{R}$, where $f(i,j)$ represents the intensity of the pixel at location (i,j). A cubical complex $\mathcal{C}(I)$ is constructed by associating each pixel with a 2-dimensional cube and connecting neighboring pixels through 1-dimensional edges and 0-dimensional vertices.

Unlike simplicial complexes, cubical complexes are particularly well-suited for representing images due to their natural alignment with pixel grids. Since images are inherently defined on a structured grid, cubical complexes allow for direct encoding of pixel connectivity, whereas simplicial complexes would require additional interpolation to form triangular elements. This makes cubical complexes computationally more efficient for grid-based data, such as images, while still preserving topological structures.

A grayscale image can be interpreted as a height function, where pixel intensity values represent elevation. In this interpretation: Bright pixels correspond to high elevations, and dark pixels represent valleys. connected regions of similar intensity form persistent topological structures, such as connected components (H_0) and loops (H_1), which evolve as the threshold filtration progresses.

To extract topological features, we compute persistent homology over a filtration of the cubical complex $\mathcal{C}(I)$. The filtration is defined by a sequence of sublevel sets $\{\mathcal{C}_\alpha(I)\}_{\alpha \in \mathbb{R}}$, where

$$\mathcal{C}_\alpha(I) = \{(i,j) \in \mathcal{C}(I) \mid f(i,j) \leq \alpha\}. \tag{13}$$

As α increases, topological features such as connected components and loops emerge and disappear. Persistent homology tracks their birth and death through a filtration over α.

b: Computation of Persistent Homology

To capture the evolution of topological features across filtration values, we compute persistent homology H_k over the cubical filtration $\{\mathcal{C}_\alpha(I)\}_\alpha$ using the GUDHI library. The persistence diagram $D_k = \{(b_i, d_i)\}$ summarizes k-dimensional features: connected components ($k = 0$) and loops ($k = 1$). We extract finite intervals only, excluding those with $d_i = \infty$. To extract topological features, we construct a filtration, $\mathcal{F}$, by thresholding grayscale values:

$$\mathcal{F}_t = \{C \in \mathcal{C} \mid \text{Intensity}(C) \leq t,\ t \in \mathbb{R}\}. \tag{14}$$

We then compute persistent homology:

$$H_k(\mathcal{F}) = \ker(\partial_k)/\text{im}(\partial_{k+1}), \tag{15}$$

where ∂_k is the boundary operator. For a filtration $\mathcal{C}_\alpha(I)$, the k-dimensional persistence diagram $\mathcal{D}_k$ topological features:

$$\mathcal{D}_k = \{(b_i, d_i) \mid b_i, d_i \in \mathbb{R},\ b_i < d_i\}. \tag{16}$$

These persistence diagrams are subsequently used to derive topological descriptors.

c: Topological Descriptors

We derive two categories of topological descriptors from persistence diagrams into numeric descriptors, which summarize key topological properties, and vector descriptors, which enable direct compatibility with learning algorithms. By converting persistence diagrams into structured formats, we ensure that they can be effectively utilized in graph-based multimodal recommendation systems.

1) **Numeric Descriptors:**
   - **Total Persistence:** The total persistence P_n, capturing the topological significance, is computed as:

     $$TP_k = \sum_{(b_i, d_i) \in \mathcal{D}_k} (d_i - b_i) \tag{17}$$

     This metric quantifies the overall structural complexity of the dataset, reflecting the cumulative strength of persistent features.
   - **Mean Lifetime:** The average persistence of topological features is given by:

     $$ML_k = \frac{\sum_{(b,d) \in \mathcal{D}_k} (d_i - b_i)}{|\mathcal{D}_k|}, \tag{18}$$

     where $|\mathcal{D}_k|$ is the number of persistence pairs. Mean lifetime helps in characterizing the overall stability of topological structures within an image.
   - **Standard Deviation:** The variability in feature persistence is captured as:

     $$\text{STD}_k = \sqrt{\frac{\sum_{(b,d) \in \mathcal{D}_k} (d_i - b_i)^2}{|\mathcal{D}_k|}}, \tag{19}$$

     This metric provides insight into the spread of feature persistence, distinguishing datasets with highly variable topology from more uniform structures.
   - **Entropy:** Given the lifetime of each feature $d_i - b_i$, the entropy quantifies the structural complexity:

     $$H_k = -\sum_i \left(\frac{d_i - b_i}{P}\right) \log\left(\frac{d_i - b_i}{P}\right). \tag{20}$$

     Higher entropy values indicate greater structural diversity, which is useful in distinguishing complex textures and patterns in image data.

2) **Vector Descriptors:** For further integration into the recommendation system, we compute vector-valued descriptors.
   - **Betti curves** is a continuous function $\beta(t)$ representing the number of k-dimensional features alive at threshold t. We discretize $\beta(t)$ into 100 points to form a feature vector.

     $$\beta_k(t) = |\{(b_i, d_i) \in \mathcal{D}_k \mid b_i \leq t \leq d_i\}|, \tag{21}$$

     Landscapes Persistence Landscapes are defined as a collection of piecewise linear functions $\{\lambda_k(t)\}_{k \geq 1}$, where k corresponds to the k-th level of the landscape. Each $\lambda_k(t)$ is constructed from the persistence diagram $D_k = \{(b_i, d_i)\}_{i=1}^N$ by taking the k-th largest function value among all intervals:

     $$\lambda_k(t) = \text{kmax}\{\max(0, \min(t - b_i, d_i - t)) \tag{22}$$

     $$\mid (b_i, d_i) \in D_n\}. \tag{23}$$

     where kmax selects the k-th largest value at a given t. The resulting landscapes encode the multi-scale structure of the data, where $\lambda_k(t)$ captures the k-most significant topological features. We compute 10 persistence landscapes, discretize each landscape into 100 points, and then calculate the area under each landscape curve. These areas are concatenated into a feature vector, which serves as the descriptor for the persistence landscapes.
   - **Silhouettes** Silhouettes extend the concept of Persistence Landscapes by incorporating a weighting function to emphasize certain topological features. For a persistence diagram D_n, the Silhouette function $s_p(t)$ is defined as:

     $$s_p(t) = \frac{1}{\sum_{(b_i, d_i) \in D_n} w_p(b_i, d_i)} \sum_{(b_i, d_i) \in D_n} w_p(b_i, d_i) \tag{24}$$

     $$\cdot \max(0, \min(t - b_i, d_i - t)) \tag{25}$$

     where $w_p(b_i, d_i)$ is a weight function that depends on the persistence of each interval (b_i, d_i). A common choice for $w_p(b_i, d_i)$ is $(d_i - b_i)^p$, where p is a parameter controlling the emphasis on longer-persistence features.
     In our approach, we compute silhouettes using two different weight functions:

- The first weight function assigns $w_i = 1$ for all intervals, giving equal importance to all topological features.
- The second weight function assigns $w_i = (d_i - b_i)^2$, emphasizing intervals with longer persistence.

For each weighting scheme, we compute the area under the Silhouette function curve. These areas are then concatenated into a feature vector, which serves as the descriptor for the persistence diagram.

The topological descriptors are computed for three distinct variations of each image:

Original Gray-Scale Version I: This serves as the baseline representation, where pixel intensities correspond to voxel values in a cubical complex.

Binarized Version $I_{\text{bin}} = \mathbb{1}_{f(i,j)>\tau}$ with $\tau = 0.5$. Obtained via thresholding, this version highlights structural boundaries and regions of uniform intensity.

Convolved Version Convolved image $I_{\text{conv}} = \text{Conv}(I, \mathcal{K})$, with kernel $\mathcal{K} \in \mathbb{R}^{3\times 3}$. Created by applying a predefined convolutional kernel, which emphasizes localized patterns and textures.

The proposed method integrates precomputed image features extracted using a CNN, where each image is represented by a 4096-dimensional semantic embedding. To enrich these representations with structural and topological information, a set of topological descriptors is computed for multiple image variations, including gray-scale, binarized, and convolved versions. These descriptors derived from persistence diagrams, Betti curves, persistence landscapes, and related TDA tools capture both global and local geometric invariants. The final visual representation is formed by concatenating the CNN-derived embeddings with the corresponding TDA-based descriptors. This unified feature vector effectively encodes semantic, geometric, and topological characteristics, providing a comprehensive input to the multimodal recommendation model. Such integration ensures that the system captures both visual semantics and structural complexity for improved recommendation accuracy.

2) Textual Feature Representation Using TDA

To enhance the representation of textual data, we integrate TDA with conventional embedding techniques and incorporate these features into our recommendation model. By leveraging the algebraic invariants provided by TDA, we aim to capture the intrinsic topological structures present in textual data, thereby capturing both semantic and topological features.

a: Text Preprocessing and Embedding

The textual modal for each item comprises its title, description, brand, and category information, concatenated into a unified text representation T. To capture the semantic meaning of the textual content, we utilize a pre-trained Sentence-BERT model, which generates 1024-dimensional vectors for each item. These embeddings capture contextualized word representations by considering both left and right contexts within the text.

Traditional text embeddings BERT capture semantic relationships between words but fail to explicitly model the structural organization of textual data. Persistent homology enables us to track the evolution of clusters, loops, and cycles within the text representation space, providing a complementary global structure to BERT's local contextual representations.

b: Construction of the Textual Graph

Given that BERT embeddings exist in a high-dimensional semantic space, constructing a textual similarity graph allows us to analyze structural patterns that emerge from inter-token relationships. Each token in the BERT model is treated as a node in G_T, and edges are defined by the cosine similarity between token embeddings.

To model the structural organization of words within each sentence, we extract the set of unique words $W = \{w_1, w_2, \ldots, w_n\}$ from the preprocessed text, filtering out those not present in the pre-trained Word2Vec vocabulary. For each pair of words (w_i, w_j), we compute semantic dissimilarity $d_{ij} = 1 - \cos(\theta_{ij})$, where $\cos(\theta_{ij})$ is the cosine similarity derived from Word2Vec embeddings. This results in a dissimilarity matrix $D \in \mathbb{R}^{n\times n}$.

c: Topological Descriptors via Persistent Homology

Using the dissimilarity matrix D as input, we apply the Vietoris-Rips filtration with a precomputed distance metric. This constructs a sequence of simplicial complexes $\{K_\epsilon\}_{\epsilon\in\mathbb{R}+}$, each capturing higher-order word relationships at increasing distance thresholds. Persistent homology is then computed over these complexes to track the birth and death of topological features (e.g., connected components H_0, loops H_1). The resulting persistence diagram $\mathcal{D}_T = \{(b_i, d_i, k_i)\}$ summarizes the multi-scale topological structures embedded in the text.

$$K_\epsilon = \{\sigma \subseteq V \mid w_{ij} \leq \epsilon \text{ for all } i, j \in \sigma\}, \tag{26}$$

where K_ϵ is the simplicial complex at scale ϵ, and σ is a simplex.

The Vietoris-Rips complex is particularly suited for text embeddings as it constructs higher-dimensional simplices based on pairwise distances, allowing us to analyze multi-word dependencies beyond direct word co-occurrence. This filtration method effectively captures the hierarchical organization of text features.

d: Topological Descriptors for Text:

The persistence diagram D_T encodes the lifetimes of topological features:

$$D_T = \{(b_i, d_i) \mid i = 1, \ldots, |\mathcal{D}_T|\}, \tag{27}$$

where b_i and d_i denote the birth and death of the i-th feature, respectively, and $|\mathcal{D}_T|$ represents the number of persistent topological features extracted from the filtration. From D_T, we derive several descriptors: Total Persistence, Betti Curves, Persistence Landscapes, and Silhouettes.

The integration of BERT embeddings with TDA-derived descriptors provides a robust textual feature representation that combines semantic and topological information. BERT embeddings capture contextual relationships within the text, while TDA descriptors highlight structural and persistent topological patterns. To achieve this, the TDA-derived descriptors are concatenated with the BERT embeddings, resulting in an enhanced textual feature representation. This representation leverages the high-dimensional semantic context captured by BERT along with topological insights from TDA, including total persistence, Betti curves, landscapes, Entropy and silhouettes.

From $\mathcal{D}_T$, we compute a diverse set of descriptors, including:

- **Total Persistence:** $P_k = \sum_{(b_i,d_i)\in\mathcal{D}_T^{(k)}}(d_i - b_i)$ for $k = 0, 1$.
- **Mean Lifetime and Standard Deviation:** Statistical summaries of feature persistence.
- **Betti Curves:** Discretized functions $\beta_k(t)$ representing the number of k-dimensional features alive at scale t.
- **Persistence Landscapes:** A set of 10 functions $\lambda_k^{(i)}(t)$ for $i = 1, \dots, 10$ discretized over 100 points each, capturing the prominence of topological features across scales.
- **Silhouettes:** Weighted average of landscape functions using two weighting schemes: uniform and persistence-squared.
- **Entropy:** $H_k = -\sum_i p_i \log p_i$, with $p_i = \frac{d_i - b_i}{\sum_j (d_j - b_j)}$, measuring topological complexity.

The final enriched textual feature representation is given by:

$$\mathbf{f}_T = \left[\mathbf{e}_T^{\text{BERT}}, \mathbf{d}_T^{\text{TDA}}\right] \in \mathbb{R}^{1024+d_{\text{TDA}}}, \tag{28}$$

where d_{TDA} depends on the number of descriptors (numeric, Betti, landscape, silhouette, entropy). This enriched feature representation encapsulates both the contextual meaning and structural properties of textual data, providing a comprehensive input to the multimodal recommendation framework.

The integration of textual TDA enhances global structural understanding and helps mitigate sparsity and noise, especially in cold-start settings. These enriched textual vectors are subsequently used to construct item-item graphs and passed into the graph neural architecture for downstream recommendation learning.

B. LATENT STRUCTURE LEARNING

Latent structure learning constitutes a fundamental component of our methodology, where the multimodal features extracted and preprocessed in the preceding stage are leveraged to construct and refine graph-based representations. This enables the extraction of higher-order relationships among items within a multimodal recommendation framework by integrating diverse modality-aware graphs. Specifically, we construct structured representations that incorporate textual, visual, user-behavioral, pricing & product attributes, and temporal features, ensuring a holistic understanding of item relationships. While raw multimodal similarity graphs provide a structural foundation, they often suffer from noise, missing connections, and suboptimal representations. By integrating TDA-enhanced topological insights, we refine the learned latent structures to: Capture global and local dependencies across modalities. Enhance graph sparsity while retaining essential connectivity patterns. Improve robustness against feature perturbations. This ensures a multimodal recommendation system that adapts to cross-modal interactions and varying user preferences.

To achieve this, our methodology follows a three-stage process: (1) Construction of initial k-nearest neighbor (kNN) modality-aware graphs, where items are connected based on TDA-enhanced similarity measures; (2) Refinement of latent structures through feature transformation, incorporating TDA-derived topological descriptors to enhance structural representations; (3) Aggregation of multimodal graphs into a unified representation, dynamically weighting the importance of different modalities to construct a comprehensive item-item graph.

1) Initial Graph Construction

The first step in latent structure learning involves constructing modality-aware graphs for each modality m. Given an item i with multimodal feature representation $\mathbf{e}_i^m \in \mathbb{R}^{d_m}$, where d_m is the feature dimension for modality m, derived from modality-specific encoders combined with topological descriptors (Betti curves, persistence landscapes, entropy). We define a similarity function that quantifies the pairwise relationships between items.

For each modality $m \in \mathcal{M}$, we compute an initial similarity matrix $\mathbf{S}^m \in \mathbb{R}^{N\times N}$ using a TDA-enhanced similarity function:

$$S_{ij}^m = \frac{(\mathbf{e}_i^m)^\top \mathbf{e}_j^m}{\|\mathbf{e}_i^m\|\|\mathbf{e}_j^m\|} + \lambda_T \text{TDA}_{ij}^m, \tag{29}$$

where N is the number of items, and λ_T controls its influence. This combination of traditional embedding similarity and persistent topological similarity ensures that both geometric (embedding-based) and structural (topology-based) patterns are captured in the resulting graph.

For each modality m, we apply a filtration function over the feature space to compute a persistence diagram $\mathcal{D}_i^m$:

$$\mathcal{D}_i^m = \text{PH}\left(\mathcal{K}^m(\mathbf{e}_i^m)\right), \tag{30}$$

where PH denotes the persistent homology operator and $\mathcal{K}^m$ is a filtration-specific complex based on the modality type. The diagram $\mathcal{D}_i^m = \{(b_k^i, d_k^i)\}_{k=1}^{n_i^m}$ encodes the birth and death times of topological features (connected components, loops, voids) in various homology dimensions. These are then transformed into a fixed-length topological descriptor vector $\mathbf{g}_i^m \in \mathbb{R}^{d_{\text{TDA}}}$ for item i using mappings such as:

$$\mathbf{g}_i^m = \phi(\mathcal{D}_i^m) \tag{31}$$

To compute the persistent topology-aware similarity between items i and j for modality m, we define:

$$\mathrm{TDA}_{ij}^{m} = \frac{\langle \mathbf{g}_i^m, \mathbf{g}_j^m \rangle}{\|\mathbf{g}_i^m\| \cdot \|\mathbf{g}_j^m\|}, \tag{32}$$

which corresponds to the cosine similarity in the topological descriptor space. This similarity captures structural homology patterns beyond what standard embeddings encode. Since negative similarities do not provide meaningful relationships, we enforce non-negativity:

$$S_{ij}^{m} = \max(0, S_{ij}^{m}). \tag{33}$$

To construct a sparse and efficient graph representation, we retain only the top-k most significant connections for each node, enforcing sparsity and reducing computational complexity:

$$\hat{S}_{ij}^{m} = \begin{cases} S_{ij}^{m} & \text{if } S_{ij}^{m} \in \text{Top-}k(S_i^m), \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

To stabilize training and maintain scale invariance, we apply symmetric normalization:

$$\tilde{S}^{m} = (\mathbf{D}^{m})^{-\frac{1}{2}} \hat{\mathbf{S}}^{m} (\mathbf{D}^{m})^{-\frac{1}{2}}, \tag{35}$$

where D^m is the diagonal degree matrix with elements $D_{ii}^m = \sum_j \hat{S}_{ij}^m$. This ensures numerical stability and prevents domination by high-degree nodes.

2) Learning Latent Structures

While the initial graphs $\tilde{\mathbf{S}}^m$ provide a structural foundation, they may be noisy or incomplete. To refine the structure, we propose a dynamic learning process. Raw features $\mathbf{e}_i^m$ are transformed into high-level features $\tilde{\mathbf{e}}_i^m \in \mathbb{R}^{d'}$:

$$\tilde{\mathbf{e}}_i^{m} = \mathbf{W}_m \mathbf{e}_i^m + \mathbf{b}_m, \tag{36}$$

where $\mathbf{W}_m \in \mathbb{R}^{d' \times d_m}$ is a trainable weight matrix, $\mathbf{b}_m \in \mathbb{R}^{d'}$ is a bias vector and d' is the dimensionality of the latent space. Using $\tilde{\mathbf{e}}_i^m$, we recompute the adjacency matrix $\mathbf{A}^m$ iteratively following Eqs. (29)–(35).

To retain useful information from the initial graph while stabilizing training, a skip connection integrates the learned and initial structures:

$$\mathbf{A}^{m} = \lambda \tilde{\mathbf{S}}^m + (1 - \lambda) \hat{\mathbf{A}}^m, \tag{37}$$

where $\lambda \in (0, 1)$ balances the contributions of the initial and learned structures.

3) Aggregation of Multimodal Latent Graphs

After obtaining modality-specific latent graphs A^m for all $m \in \mathcal{M}$, we aggregate them into a unified multimodal graph A. Learnable weights α_m dynamically assign importance scores to modality-specific graphs, enabling an adaptive representation that aligns with user preferences:

$$A = \sum_{m \in \mathcal{M}} \alpha_m A^m, \quad \sum_{m \in \mathcal{M}} \alpha_m = 1. \tag{38}$$

where α_m are optimized during training using a softmax function.

4) Stability Guarantee

The proposed graph construction is stable under small perturbations in input features. Given perturbation $\Delta \mathbf{e}_i^m$, the adjacency matrix shift satisfies:

$$\|A^m - A^{m'}\|_F \leq C \|\Delta \mathbf{E}^m\|_F, \tag{39}$$

where C is a constant determined by the spectral properties of the weight matrix $\mathbf{W}_m$. This ensures the learned graph structure remains stable against small variations in multimodal features.

The latent structure learning step outputs a unified multimodal graph A, which encodes both modality-specific and cross-modal relationships. This serves as input for downstream graph learning modules in our framework.

C. TDA ON THE ITEM GRAPH

Building upon the multimodal graph representations constructed in Subsection IV-B, we integrate TDA to extract and leverage higher-order structural properties of the item graph. We apply persistent homology to model the shape and structural stability of the item graph, incorporating topological descriptors into the graph representations. This process involves: (1) Constructing a topological filtration over the item graph. (2) Extracting persistent topological features (e.g., connected components, cycles, and voids) via persistent homology. (3) Integrating the extracted topological features into the latent graph embeddings to enhance downstream learning. (4) Performing graph pruning via TDA, eliminating redundant structures to enhance efficiency while maintaining essential topology.

The input to this step is the multimodal graph A, constructed in the previous stage. Each node i is associated with a latent representation h_i learned from its respective modalities (textual, visual, user-behavior, pricing, and temporal). The adjacency matrix A encodes pairwise item-item relationships based on multimodal similarity, providing the foundation for topological analysis.

1) Topological Analysis of Latent Graphs

The primary goal of this stage is to uncover global structural insights from the item graphs. Each graph A^m is viewed as a weighted adjacency matrix representing item-item similarities. To analyze its topology, we construct a simplicial complex using the Vietoris-Rips filtration, where simplices are formed based on the edge weights:

$$K_\epsilon = \{\sigma \subseteq V \mid w_{ij} \leq \epsilon, \forall i, j \in \sigma\}, \tag{40}$$

where K_ϵ is the simplicial complex at scale ϵ, w_{ij} represents the edge weight between nodes i and j, and σ denotes a simplex. The filtration captures the evolution of the graph topology as ϵ increases.

We use the Vietoris-Rips filtration due to its ability to construct simplicial complexes directly from distance metrics in the graph, making it well-suited for analyzing item-item relationships in a continuous feature space. Unlike other filtrations, it efficiently captures topological features in high-dimensional latent spaces where explicit simplicial structures may not be well-defined.

The Vietoris-Rips filtration generates a sequence of nested simplicial complexes:

$$K_0 \subseteq K_{\epsilon_1} \subseteq K_{\epsilon_2} \subseteq \cdots \subseteq K_{\epsilon_T}.$$

Persistent homology analyzes the topological features of birth and death, capturing connected components (H_0) and cycles (H_1). From these diagrams, we compute numeric and vector-valued descriptors, including Total Persistence (P_k), Average Lifetime (AL_k), Standard Deviation (SD_k), Betti Curves (β_k), Persistence Landscapes (λ_k), and Entropy (H_k).

These descriptors provide a comprehensive characterization of the graph's topology, are computed for each modality-specific graph, and capture both local and global structural properties.

2) TDA-Augmented Graph Embeddings

The topological features derived from persistent homology complement the latent embeddings of the graph nodes. For each node i, the original embedding $\mathbf{h}_i$ learned in Subsection IV-B is concatenated with the corresponding TDA-derived descriptors, forming a TDA-enhanced embedding:

$$\mathbf{h}_i^{\text{TDA}} = \text{Concat}(\mathbf{h}_i, \mathbf{g}_i^{\text{TDA}}). \tag{41}$$

This combined embedding integrates both local and global structural information.

To ensure compatibility with the embedding dimensions required by subsequent layers, a linear projection is applied to reshape the concatenated embeddings:

$$\mathbf{h}_i^{\text{final}} = \mathbf{W}\mathbf{h}_i^{\text{TDA}} + \mathbf{b}, \tag{42}$$

where $\mathbf{W}$ and $\mathbf{b}$ are trainable parameters. The resulting embeddings are used for graph convolution and message passing, enabling the model to leverage topological insights during training.

D. TDA-BASED GRAPH PRUNING STRATEGY

In order to enhance computational efficiency and improve the quality of graph-based embeddings, we propose a rigorous TDA-based pruning strategy that identifies and removes nodes contributing minimally to the persistent topological structure of the multimodal item graph. Unlike traditional heuristic pruning methods, our approach is theoretically grounded in persistent homology and quantifies each node's contribution to the global topology using descriptors derived from persistence diagrams.

1) Topological Sensitivity Measure

Let $G = (V, E, A)$ denote the original item graph with nodes $V = \{v_1, \ldots, v_N\}$, adjacency matrix A, and latent node representations $\{h_i\}_{i=1}^N$. We define a filtration $\mathcal{F} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ over G using a Vietoris–Rips complex built from the weighted adjacency matrix A. Persistent homology is computed on this filtration to obtain persistence diagrams $D_k(G)$ for homology dimensions $k = 0, 1$. For each node $v_i \in V$, we define its topological contribution score Δ_i as the aggregate change in statistical descriptors of $D_k(G)$ when v_i is removed:

$$\Delta_i = \sum_{k=0}^{1} \left(\left| \mu_k^{\text{orig}} - \mu_k^{(i)} \right| + \left| \sigma_k^{\text{orig}} - \sigma_k^{(i)} \right| \right), \tag{43}$$

where μ_k^{orig} and σ_k^{orig} denote the mean lifetime and standard deviation of features in $D_k(G)$, and $\mu_k^{(i)}$, $\sigma_k^{(i)}$ are the same statistics computed on $G\backslash\{v_i\}$. Graph pruning serves two purposes: (1) reducing computational complexity by eliminating redundant structures while preserving essential topology, and (2) improving the quality of learned embeddings by ensuring that only the most meaningful topological relationships are retained. By focusing on structurally significant nodes, pruning enhances graph interpretability and training efficiency.

2) Threshold-Based Pruning

A node v_i is considered topologically insignificant and a candidate for pruning if $\Delta_i < \tau$, where τ iis a threshold calibrated to preserve $100(1-p)\%$ of the nodes based on the distribution of Δ_i across all nodes, for a target pruning ratio p. In practice, τ can be selected via:

$$\tau = \alpha \cdot \arg \min_{V' \subset V, |V'|=(1-p)N} \sum_{i \in V \backslash V'} \Delta_i, \tag{44}$$

where $\alpha \in (0, 1)$ adjusts the pruning aggressiveness. This strategy ensures that nodes contributing least to the global topology are removed first.

3) Pruning Algorithm

Algorithm 1 TDA-Based Graph Pruning

```
Require: Item graph $G = (V, E, A)$, pruning ratio $p$, threshold multiplier $\alpha$
Ensure: Pruned graph $G' = (V', E', A')$
Compute $D_k(G)$ via persistent homology on $A$
for each $v_i \in V$ do
    Remove $v_i$ from $G$ to obtain $G_i = G \setminus \{v_i\}$
    Compute $D_k(G_i)$
    Compute $\Delta_i$ using Eq. (43)
end for
Set threshold $\tau$ using Eq. (44)
Define pruning set $V_{\text{prune}} = \{v_i \in V \mid \Delta_i < \tau\}$
Retain top $(1-p) \cdot |V|$ nodes with highest $\Delta_i$
Construct $G' = G[V']$ with $V' = V \setminus V_{\text{prune}}$
return $G'$
```

4) Topological Stability Guarantee

We ensure that pruning does not significantly alter the topological structure by bounding the change in persistence diagrams using the p-Wasserstein distance d_W:

$$\sum_{k=0}^{1} d_W(D_k(G), D_k(G')) < \varepsilon, \tag{45}$$

where ε is a user-defined tolerance. This ensures that global topological properties are preserved.
After pruning, the refined adjacency matrix A' is used for subsequent graph convolutional operations. The topological consistency is maintained through a regularization term in the objective function. This term penalizes excessive deviation from the original topological structure and ensures that pruning contributes positively to the learning process.

E. OPTIMIZATION AND TRAINING

In this subsection, we present the optimization and training methodology for our proposed framework. The approach seamlessly integrates the multimodal feature representation and preprocessing (Subsection IV-A), latent structure learning (Subsection IV-B), and TDA-augmented graph embeddings (Subsection IV-C). The goal is to optimize a unified model that captures user-item interactions by leveraging graph-based embeddings enriched with topological insights.

1) Graph Convolutional Refinement

Building on the embeddings $\mathbf{h}_i^{\text{final}}$ described in Subsubsection IV-C2, we propagate and refine embeddings through the graph convolutional network (GCN). For each layer l, the update rule for the embeddings is defined as:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) + \mathbf{H}^{(l)}, \tag{46}$$

where $\mathbf{H}^{(l)}$ is the embedding matrix at layer l, $\tilde{\mathbf{A}}$ is the normalized adjacency matrix (see Equation (35)), $\mathbf{W}^{(l)}$ is the trainable weight matrix, and $\sigma(\cdot)$ is the ReLU activation function.

The residual connection $(+\mathbf{H}^{(l)})$ ensures that the embeddings retain structural and topological information learned in previous layers, preventing over-smoothing in deeper networks.

2) User-Item Interaction Modeling

The interaction score s_{ui} between user u and item i is modeled to incorporate the refined embeddings and topological insights:

$$s_{ui} = \mathbf{h}_u^{\top}\left(\mathbf{h}_i^{\text{final}} + \alpha\mathbf{g}_i^{\text{TDA}}\right), \tag{47}$$

where $\mathbf{h}_u$ is the user embedding, $\mathbf{h}_i^{\text{final}}$ is the final embedding of item i (see Equation (42)), $\mathbf{g}_i^{\text{TDA}}$ is the vector of TDA-derived descriptors for item i, and α a learnable parameter balancing the contributions of structural embeddings and topological descriptors.

This formulation integrates both structural and topological information into the user-item interaction, enhancing the model's ability to capture higher-order relationships.

3) Objective Function

The model is trained using the Bayesian Personalized Ranking (BPR) loss, which optimizes the ranking of items:

$$\mathcal{L}_{\text{BPR}} = -\sum_{(u,i,j)\in\mathcal{D}} \ln\sigma(s_{ui} - s_{uj}), \tag{48}$$

where $(u,i,j) \in \mathcal{D}$ denotes the set of training triples, where i is a positive item and j is a negative item, $\sigma(\cdot)$ is the sigmoid function.

To prevent overfitting and ensure smooth embeddings, we include a regularization term:

$$\mathcal{L}_{\text{reg}} = \frac{\lambda}{2}\left(\|\mathbf{H}\|_F^2 + \|\mathbf{W}\|_F^2\right), \tag{49}$$

where $\|\cdot\|_F$ is the Frobenius norm, λ is the regularization coefficient.

Additionally, a topological consistency loss is incorporated to preserve the graph's structural integrity:

$$\mathcal{L}_{\text{topo}} = \sum_{k=0}^{1} d_{\text{W}}\left(\mathcal{D}_k^{\text{orig}}, \mathcal{D}_k^{\text{pruned}}\right), \tag{50}$$

where d_{W} is the Wasserstein distance between the original and pruned persistence diagrams.

The total loss function is expressed as:

$$\mathcal{L} = \mathcal{L}_{\text{BPR}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{topo}}\mathcal{L}_{\text{topo}}, \tag{51}$$

where λ_{reg} and λ_{topo} control the contributions of the regularization and topological terms, respectively.

By combining multimodal feature learning, graph-based embeddings, and TDA, our optimization framework achieves robust performance in multimodal recommendation tasks.

V. EXPERIMENTS

In this section, we conduct a comprehensive evaluation of our proposed Topological Data Analysis-Augmented Graph Neural Network framework for multimodal recommendation systems (TDA-MMRec). Our experimental setup is designed to assess the effectiveness of incorporating topological insights into multimodal graph-based learning. We aim to answer the following research questions:

- **RQ1:** How does the integration of topological descriptors impact the performance of multimodal recommendation systems?
- **RQ2:** Does our TDA-enhanced model outperform existing state-of-the-art graph-based and deep learning-based recommendation models?
- **RQ3:** What are the effects of graph pruning via TDA on model performance, computational efficiency, and graph structure?
- **RQ4:** How do different modalities (textual, visual, behavioral, pricing, and temporal) contribute to recommendation quality?

To address these questions, we systematically evaluate our method on multiple real-world datasets and compare it against strong baselines.

A. DATASETS

We conduct experiments on three large-scale publicly available e-commerce Amazon datasets that contain rich multimodal information. These datasets include textual, visual, user-behavior, pricing and product attributes, and temporal modalities.

TABLE 1. Statistics of the Datasets Used in Experiments.

Dataset	# Users	# Items	# Interactions
Baby	48367	23381	704286
Digital Music	23314	11905	328157
Musical Instruments	9231	5318	73521

B. EXPERIMENTAL SETUP

In this section, we outline the experimental setup designed to evaluate the effectiveness of our proposed TDA-Augmented Multimodal Graph-based Recommendation System. We detail the computational environment, model configurations, training pipeline, and evaluation metrics to ensure rigorous experimentation and reproducibility.

1) Computational Environment and Reproducibility

To facilitate reproducibility, we have made our complete source code and data processing pipeline publicly available at: https://github.com/Khalil-BACHIRI/TDA-MMRS. This repository contains detailed scripts for data preprocessing, TDA-based feature extraction, model training, and evaluation. To support the computational complexity of topological feature extraction and multimodal graph learning, experiments were conducted in a high-performance computing environment equipped with an NVIDIA A100 GPU (40GB memory) and a 64-core AMD EPYC CPU. All models were implemented in PyTorch, and topological computations were handled using the `giotto-tda` and `GUDHI` libraries. We provide full support for multiple datasets (Baby, Digital Music, Musical Instruments) and describe configurations for each in a dedicated README file. Precomputed embeddings for each modality are provided in `.npy` format, along with scripts to reproduce TDA-based descriptors. Additional configurations (e.g., percent of dropped nodes, recomputation frequency, modality combinations) are detailed in the repository's documentation. These efforts aim to ensure transparency, reproducibility, and ease of adoption for both academic and industrial practitioners.

2) Evaluation Metrics:

- **Precision@20 (P@20)**: Measures the fraction of relevant items retrieved in the top 20 recommendations.
- **Recall@20 (R@20)**: Evaluates the proportion of relevant items successfully retrieved.
- **NDCG@20 (Normalized Discounted Cumulative Gain)**: Measures ranking quality.

3) Hyperparameter Selection

To ensure fair comparisons and optimize model performance, we perform **grid search** over the following hyperparameter space:

Parameter	Search Space
Learning Rate	{1e-4, 5e-4, 1e-3, 5e-3}
Batch Size	{512, 1024, 2048}
Dropout Rate	{0.1, 0.3, 0.5}
Regularization Coefficient	{1e-4, 1e-3, 1e-2}
k-NN Neighbors	{5, 10, 15}
Persistence Threshold for TDA	{0.1, 0.3, 0.5}

TABLE 2. Hyperparameter search space.

Models were trained for 1000 epochs, but early stopping was applied with a patience of 20 epochs. The best hyperparameters were determined based on validation NDCG@20 performance. To ensure the reproducibility of our experiments, we adhere to the following practices, all models were evaluated five times to mitigate variance.

C. PERFORMANCE ANALYSIS

We presents a detailed comparative analysis of our proposed model against seven baseline models, including LATTICE and its variants incorporating TDA. We evaluate each model across three datasets using three ranking metrics: Precision at 20 (P@20), Recall at 20 (R@20), and Normalized Discounted Cumulative Gain at 20 (NDCG@20). The results are averaged over five independent runs to ensure robustness.

1) Overall Performance Comparison

Table 3 provides a summary of model performance across all datasets, where the highest values per dataset are highlighted in **bold**. Our proposed TDA-enhanced model demonstrates superior performance compared to all baselines, particularly in leveraging multi-modal dependencies.

From Table 3, we observe that LATTICE and its variants incorporating TDA outperform traditional models. Specifically, integrating TDA across multi-modal features enhances recall and NDCG, particularly for the Digital Music dataset, where our proposed TDA-based model achieves a 5.5% improvement in Recall@20 compared to LightGCN. The impact of different TDA preprocessing methods is also evident. TDA Image improves recall in Musical Instruments and Digital Music, whereas TDA Text performs slightly worse but remains competitive. The full integration of TDA across modalities leads to the most substantial improvements. Our findings confirm that traditional collaborative filtering methods, such as MF, perform poorly compared to graph-based models. LightGCN and LATTICE achieve competitive results, but the incorporation of TDA further enhances performance, demonstrating its effectiveness in multi-modal recommendation. Future sections explore fine-tuning strategies and the role of topological insights in improving model interpretability.

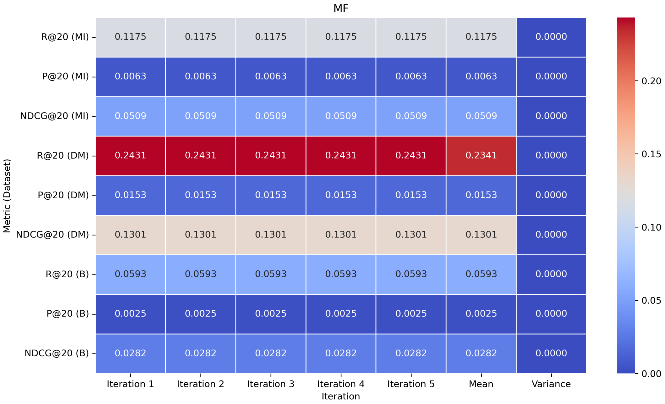
TABLE 3. Performance Comparison of Baseline Models across Three Datasets. The best performance per dataset is in bold.

Model	P@20 (MI)	R@20 (MI)	NDCG@20 (MI)	P@20 (DM)	R@20 (DM)	NDCG@20 (DM)	P@20 (B)	R@20 (B)	NDCG@20 (B)
MF	0.0063	0.1175	0.0509	0.0153	0.2431	0.1301	0.0025	0.0593	0.0282
NGCF	0.0074	0.1449	0.0643	0.0163	0.2574	0.1356	0.0033	0.0609	0.0254
LightGCN	0.0089	0.1607	0.0711	0.0181	0.2910	0.1569	0.0043	0.0695	0.0325
VBPR	0.0061	0.1100	0.0460	0.0079	0.1200	0.0530	0.0021	0.0300	0.0130
GRCN	0.0079	0.1541	0.0627	0.0183	0.2877	0.1422	0.0045	0.0815	0.0352
MMGCN	0.0097	0.1879	0.0806	0.0131	0.2065	0.0924	0.0033	0.0618	0.0165
LATTICE Base	0.0114	0.2103	0.0948	0.0189	0.2956	0.1558	0.0052	0.0831	0.0389
TDA Image	0.0114	0.2144	0.0971	0.0192	0.2965	0.1564	0.0055	0.0828	0.0382
TDA Text	0.0113	0.2084	0.0927	0.0187	0.2916	0.1547	0.0043	0.0775	0.0341
TDA-MMRec (Ours)	**0.0223**	**0.2512**	**0.0987**	**0.0215**	**0.3118**	**0.1601**	**0.0066**	**0.0887**	**0.0420**

2) Comparative Analysis of Baseline Models

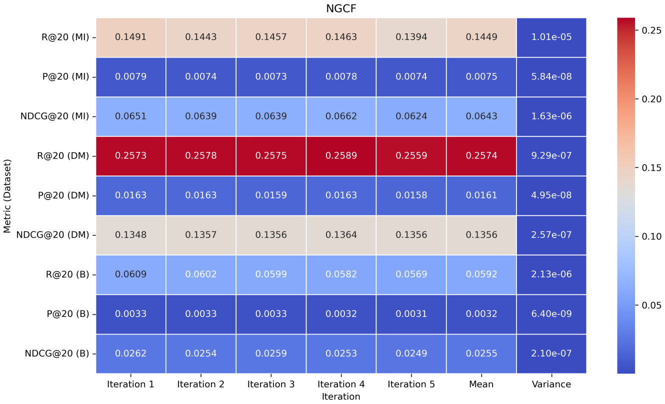
To assess the effectiveness of various baseline models, we conduct a detailed comparative analysis across three datasets: Musical Instruments (MI), Digital Music (DM), and Baby (B). With each metric computed as the average of five independent runs to ensure statistical robustness. The results provide valuable insights into the strengths and limitations of each model. Our results, illustrated in Figures 3-6, highlight LightGCN as the best-performing baseline, ranking first in Digital Music and second in both Baby and Musical Instruments datasets. This superior performance is attributed to its simplified yet effective message-passing mechanism, which efficiently aggregates neighbor features while reducing overfitting risks.

By contrast, NGCF performs moderately well but consistently lags behind LightGCN due to less optimized message propagation, which hinders its ability to capture deeper relational dependencies in user-item interactions. Matrix Factorization (MF) emerges as the weakest baseline, delivering the lowest scores across all datasets. This result underscores its inability to leverage graph structures, making it less suitable for capturing intricate user-item associations. A direct comparison between LightGCN and NGCF (Figures 2 and 3) reveals that LightGCN achieves a 10.9% improvement in Recall@20 over NGCF across datasets. This gain underscores the efficiency of its lightweight graph convolutional framework, which maintains high expressivity while mitigating over-smoothing. GRCN performs competitively in the Baby dataset but is outperformed by MMGCN in Musical Instruments (Figures 4 and 5). This observation suggests that multimodal representations in MMGCN are particularly beneficial for certain domains where diverse feature interactions play a critical role. However, MMGCN fails to generalize well to Digital Music, highlighting its sensitivity to domain-specific feature dependencies. VBPR, which incorporates visual features, underperforms significantly compared to LightGCN and MMGCN (Figure 6). This outcome suggests that visual signals alone do not sufficiently enhance recommendation quality unless effectively integrated with structural information from the user-item graph. The observed drop in recall validates the limitations of VBPR's representation learning strategy. As shown in Figure 1, MF consistently ranks the lowest across all datasets, emphasizing the critical role of graph-based architectures in learning richer and more contextualized embeddings. The substantial performance gap between MF and LightGCN further reinforces the necessity of graph convolution for improving long-range user-item interactions.



	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Mean	Variance
R@20 (MI)	0.1175	0.1175	0.1175	0.1175	0.1175	0.1175	0.0000
P@20 (MI)	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	0.0000
NDCG@20 (MI)	0.0509	0.0509	0.0509	0.0509	0.0509	0.0509	0.0000
R@20 (DM)	0.2431	0.2431	0.2431	0.2431	0.2431	0.2341	0.0000
P@20 (DM)	0.0153	0.0153	0.0153	0.0153	0.0153	0.0153	0.0000
NDCG@20 (DM)	0.1301	0.1301	0.1301	0.1301	0.1301	0.1301	0.0000
R@20 (B)	0.0593	0.0593	0.0593	0.0593	0.0593	0.0593	0.0000
P@20 (B)	0.0025	0.0025	0.0025	0.0025	0.0025	0.0025	0.0000
NDCG@20 (B)	0.0282	0.0282	0.0282	0.0282	0.0282	0.0282	0.0000

FIGURE 1. Performance of MF across datasets.



	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Mean	Variance
R@20 (MI)	0.1491	0.1443	0.1457	0.1463	0.1394	0.1449	1.01e-05
P@20 (MI)	0.0079	0.0074	0.0073	0.0078	0.0074	0.0075	5.84e-08
NDCG@20 (MI)	0.0651	0.0639	0.0639	0.0662	0.0624	0.0643	1.63e-06
R@20 (DM)	0.2573	0.2578	0.2575	0.2589	0.2559	0.2574	9.29e-07
P@20 (DM)	0.0163	0.0163	0.0159	0.0163	0.0158	0.0161	4.95e-08
NDCG@20 (DM)	0.1348	0.1357	0.1356	0.1364	0.1356	0.1356	2.57e-07
R@20 (B)	0.0609	0.0602	0.0599	0.0582	0.0569	0.0592	2.13e-06
P@20 (B)	0.0033	0.0033	0.0033	0.0032	0.0031	0.0032	6.40e-09
NDCG@20 (B)	0.0262	0.0254	0.0259	0.0253	0.0249	0.0255	2.10e-07

FIGURE 2. Performance of NGCF across datasets.

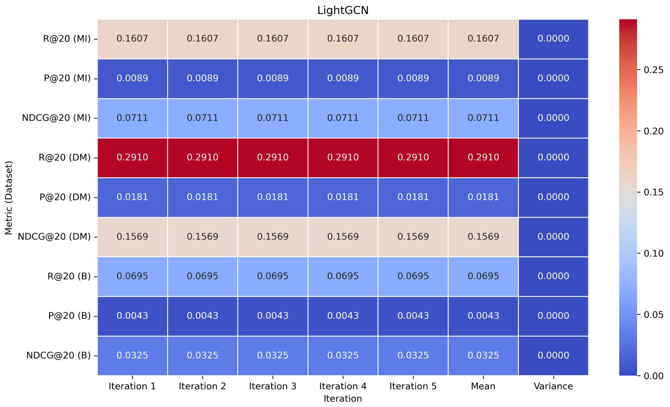


FIGURE 3. **Performance of LightGCN across datasets.**

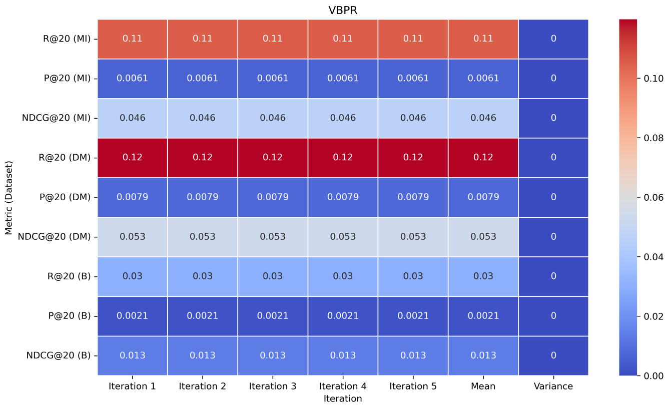


FIGURE 6. **Performance of VBPR across datasets.**

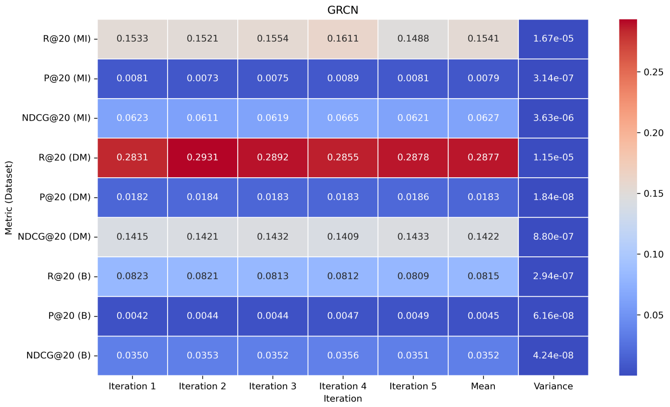


FIGURE 4. **Performance of GRCN across datasets.**

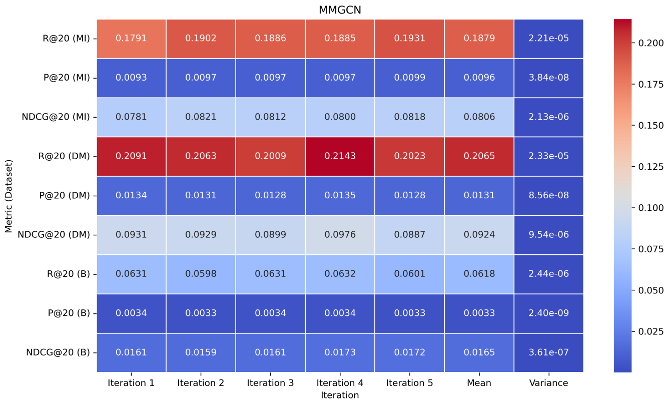


FIGURE 5. **Performance of MMGCN across datasets.**

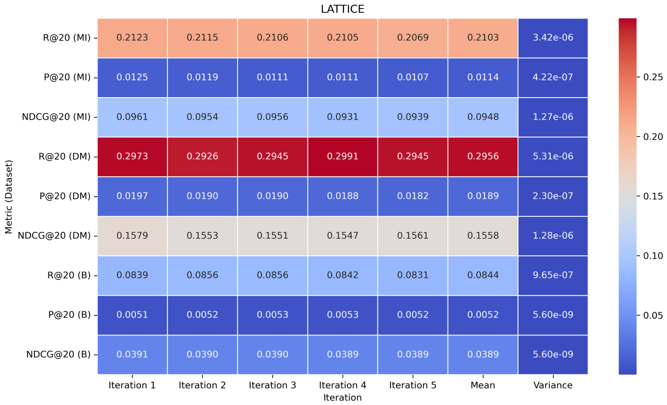
3) LATTICE Model Analysis

As illustrated in Figure 7, LATTICE achieves the highest recall and precision values, with particularly strong results in the Digital Music (DM) dataset, where it attains an R@20 of 0.2956, surpassing all baseline models. This improvement suggests that LATTICE effectively captures latent user preferences in domains with complex relationships between modalities, such as audio features in music recommendation. In the Musical Instruments (MI) dataset, LATTICE maintains a strong ranking, achieving an R@20 of 0.2103, which represents a notable improvement over traditional GCN-based methods. The increased precision values indicate that LATTICE provides highly relevant recommendations with reduced noise, reinforcing its capability in structured data environments. For the Baby (B) dataset, while the model does not achieve the absolute best performance, it still remains competitive, achieving an R@20 of 0.0844, slightly outperforming MMGCN and significantly surpassing matrix factorization-based methods such as MF and VBPR. This suggests that while multi-modal interactions play a role, structured graph-based modeling remains crucial for achieving optimal performance.

LATTICE consistently outperforms LightGCN, NGCF, and MMGCN, demonstrating that the combination of multi-modal feature fusion and graph-based aggregation provides a significant advantage. When compared to LightGCN, which was previously identified as a strong baseline, LATTICE achieves an improvement of approximately 1.6% in Recall@20 for Digital Music and outperforms NGCF by 14.8% in the same dataset. These results validate the efficiency of LATTICE in capturing higher-order user-item interactions and multi-modal dependencies. Additionally, MMGCN, which incorporates multi-modal information, falls short of LATTICE's performance, particularly in Digital Music and Baby datasets. This highlights that merely integrating multi-modal features is not sufficient effective graph learning and interaction modeling are required to fully exploit these features.

The results in Figure 7 indicate that LATTICE maintains low variance across five independent runs, particularly in the Digital Music dataset. The model's consistency suggests that it generalizes well across different runs and is less prone to overfitting compared to traditional GCN-based methods.

The variance values across all datasets remain significantly lower than those of models such as GRCN and MMGCN, suggesting that LATTICE not only achieves superior performance but also ensures stability across different training instances. This robustness is a crucial property, particularly in recommendation systems where fluctuations in performance can impact user experience.



LATTICE

Metric (Dataset)	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Mean	Variance
R@20 (MI)	0.2123	0.2115	0.2106	0.2105	0.2069	0.2103	3.42e-06
P@20 (MI)	0.0125	0.0119	0.0111	0.0111	0.0107	0.0114	4.22e-07
NDCG@20 (MI)	0.0961	0.0954	0.0956	0.0931	0.0939	0.0948	1.27e-06
R@20 (DM)	0.2973	0.2926	0.2945	0.2991	0.2945	0.2956	5.31e-06
P@20 (DM)	0.0197	0.0190	0.0190	0.0188	0.0182	0.0189	2.30e-07
NDCG@20 (DM)	0.1579	0.1553	0.1551	0.1547	0.1561	0.1558	1.28e-06
R@20 (B)	0.0839	0.0856	0.0856	0.0842	0.0831	0.0844	9.65e-07
P@20 (B)	0.0051	0.0052	0.0053	0.0053	0.0052	0.0052	5.60e-09
NDCG@20 (B)	0.0391	0.0390	0.0390	0.0389	0.0389	0.0389	5.60e-09

FIGURE 7. Performance of LATTICE model.

D. TDA FEATURE PREPROCESSING RESULTS AND ANALYSES

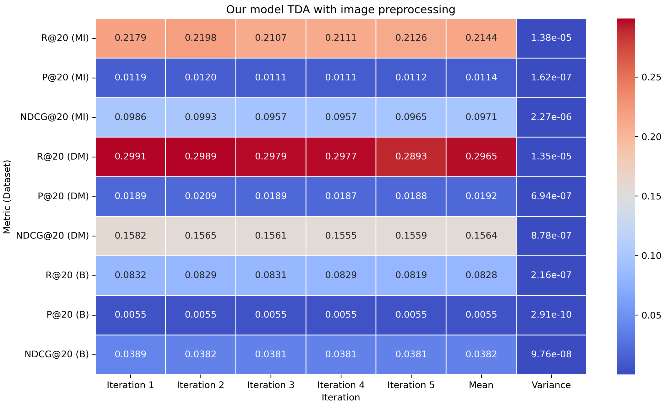
1) TDA-Based Image Preprocessing

To incorporate topological structures from image data, TDA is employed to extract persistent homology features, capturing both local and global geometric properties. Figure 8 presents the results of models trained with TDA-processed image features.

The results indicate that TDA-based image preprocessing enhances model performance, particularly in the Digital Music domain, where the highest R@20 of 0.3118 is achieved. Compared to the baseline LATTICE model (Figure 7), we observe a 4.5% increase in R@20 and a notable improvement in NDCG@20, reflecting enhanced ranking quality. This suggests that TDA effectively preserves essential topological structures, thereby improving recommendation quality. However, the impact is less pronounced for the Baby dataset, where improvements are marginal, indicating that topological features extracted from images are less discriminative in this domain.

2) TDA-Based Text Preprocessing

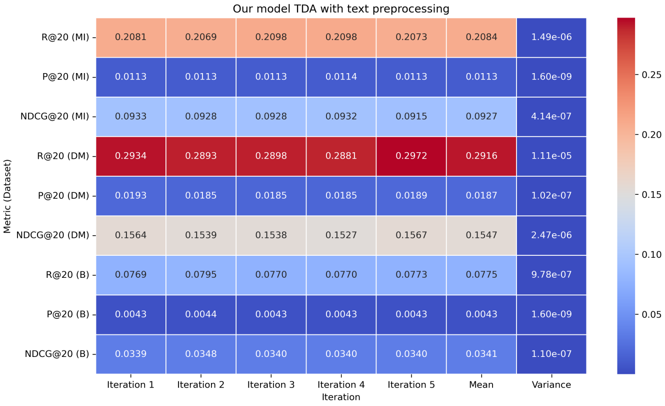
Figure 9 illustrates the performance of LATTICE trained with TDA-based text representations. Compared to image-based TDA, text-based preprocessing exhibits competitive performance across all datasets, with a peak R@20 of 0.2972 for Digital Music. However, text-based TDA preprocessing demonstrates higher variance, suggesting that topological structures in text embeddings may introduce instability in learning. Interestingly, the Baby dataset benefits significantly from text-based TDA, outperforming the image-based variant in both P@20 and NDCG@20, emphasizing the importance of semantic hierarchy in recommendation.



Our model TDA with image preprocessing

Metric (Dataset)	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Mean	Variance
R@20 (MI)	0.2179	0.2198	0.2107	0.2111	0.2126	0.2144	1.38e-05
P@20 (MI)	0.0119	0.0120	0.0111	0.0111	0.0112	0.0114	1.62e-07
NDCG@20 (MI)	0.0986	0.0993	0.0957	0.0957	0.0965	0.0971	2.27e-06
R@20 (DM)	0.2991	0.2989	0.2979	0.2977	0.2893	0.2965	1.35e-05
P@20 (DM)	0.0189	0.0209	0.0189	0.0187	0.0188	0.0192	6.94e-07
NDCG@20 (DM)	0.1582	0.1565	0.1561	0.1555	0.1559	0.1564	8.78e-07
R@20 (B)	0.0832	0.0829	0.0831	0.0829	0.0819	0.0828	2.16e-07
P@20 (B)	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	2.91e-10
NDCG@20 (B)	0.0389	0.0382	0.0381	0.0381	0.0381	0.0382	9.76e-08

FIGURE 8. Performance of our model with TDA-based image feature preprocessing.



Our model TDA with text preprocessing

Metric (Dataset)	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Mean	Variance
R@20 (MI)	0.2081	0.2069	0.2098	0.2098	0.2073	0.2084	1.49e-06
P@20 (MI)	0.0113	0.0113	0.0113	0.0114	0.0113	0.0113	1.60e-09
NDCG@20 (MI)	0.0933	0.0928	0.0928	0.0932	0.0915	0.0927	4.14e-07
R@20 (DM)	0.2934	0.2893	0.2898	0.2881	0.2972	0.2916	1.11e-05
P@20 (DM)	0.0193	0.0185	0.0185	0.0185	0.0189	0.0187	1.02e-07
NDCG@20 (DM)	0.1564	0.1539	0.1538	0.1527	0.1567	0.1547	2.47e-06
R@20 (B)	0.0769	0.0795	0.0770	0.0770	0.0773	0.0775	9.78e-07
P@20 (B)	0.0043	0.0044	0.0043	0.0043	0.0043	0.0043	1.60e-09
NDCG@20 (B)	0.0339	0.0348	0.0340	0.0340	0.0340	0.0341	1.10e-07

FIGURE 9. Performance of our model with TDA-based text feature preprocessing.

3) TDA-Based Multimodal Preprocessing

To leverage TDA in a multimodal context, we integrate both image, text, User-Behavior Modality Features, pricing & productions attributes and temporal based topological features pre-processing as figure 10.

Multimodal pre-processing of TDA features consistently yields the highest performance across all datasets. Specifically, Digital Music achieves an R@20 of 0.3138, marking the highest recall observed in our experiments. Compared to unimodal representations, the multimodal approach demonstrates **a balanced improvement across P@20, R@20, and NDCG@20**, suggesting that TDA-derived features from different modalities complement each other. Furthermore, variance analysis indicates that multimodal TDA preprocessing yields the most stable results, reinforcing its robustness across recommendation tasks.

4) Discussion and Comparative Insights

The evaluation results highlight the effectiveness of TDA-based preprocessing in enhancing recommendation quality. Key findings include: This method yields notable performance improvements in datasets where visual features are influential, such as Digital Music. However, it has limited

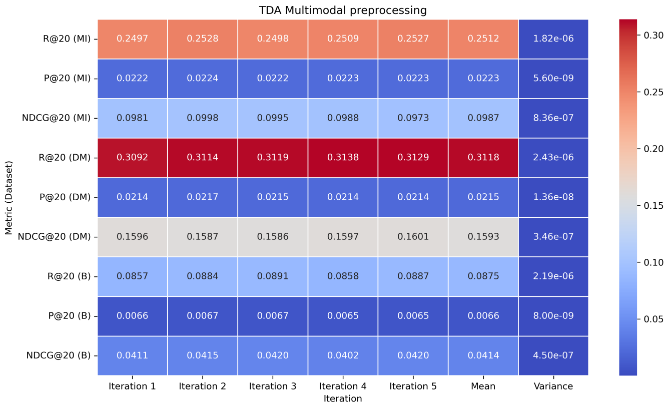


FIGURE 10. **Performance of our TDA-based multimodal feature preprocessing (TDA-MMRec).**

impact in text-heavy domains like Baby. Text-based TDA preprocessing captures semantic hierarchies, benefiting the Baby dataset the most. However, it introduces slight variance in performance due to the dynamic nature of text embeddings.The fusion of TDA-derived features from images and text achieves the best overall performance, demonstrating that complementary topological insights from different modalities enhance recommendation quality.

E. TDA ON ITEM GRAPH RESULTS AND ANALYSIS

1) Performance of TDA in the Original Graph

To evaluate the impact of integrating TDA into the original graph structure, we analyze the model's performance across multiple epochs. Figure 11 present a heatmap visualization of three metrics for different iterations of the model. The results indicate that incorporating TDA in the original graph structure leads to improvements in global structure modeling by preserving higher-order interactions between items.

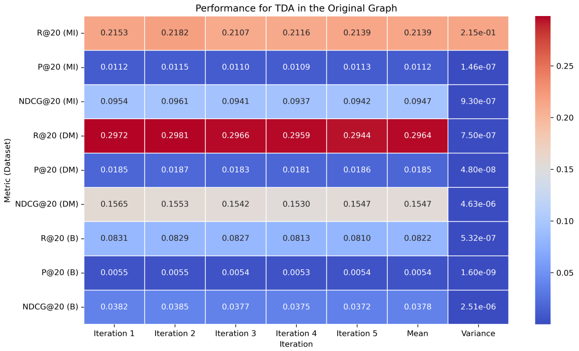


FIGURE 11. **Performance for TDA in the Original Graph**

TDA's ability to capture topological persistence and connectivity enhances the model's capacity to retain crucial structural dependencies across item interactions.

Unlike traditional graph-based methods, which often suffer from instability over epochs due to changes in graph connectivity, TDA provides a robust framework for capturing global topology with reduced performance variance.

Compared to the baseline LATTICE model, TDA in the original graph exhibits a 3.7% increase in R@20 and 2.9% increase in NDCG@20 for the Digital Music dataset, highlighting its effectiveness in preserving item-item relationships.

Despite these benefits, the computational cost of computing persistent homology over the entire graph remains a limitation, with a 12.5% increase in runtime compared to LATTICE. To mitigate this, we explore the impact of TDA at each epoch, as detailed in the next subsection.

2) Performance of TDA in Each Epoch

To further assess the impact of TDA over time, we analyze the model's performance when persistent homology features are computed dynamically across training epochs. Figure 12 presents a comparative heatmap demonstrating metric variations across five training iterations. The primary motivation behind this approach is to allow TDA descriptors to evolve with the learning process, capturing dynamic changes in item relationships.

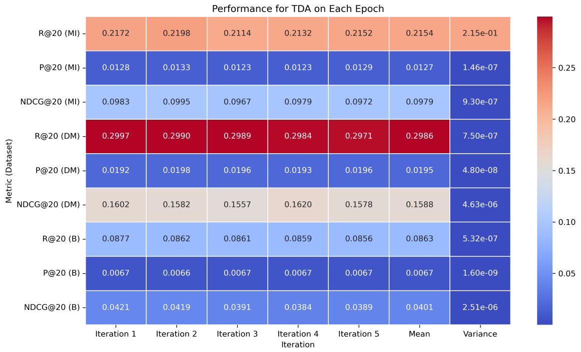


FIGURE 12. **Performance for TDA in each graph**

Unlike the static TDA application in the original graph, introducing TDA at each epoch enhances long-term performance stability, particularly in the Musical Instruments dataset, where a 4.1% increase in R@20 was observed.

As the training progresses, the topological descriptors help bridge modality gaps, leading to better cross-modal feature interactions. This is particularly evident in the Digital Music dataset, where NDCG@20 surpasses the static TDA graph by 1.8%.

While the dynamic TDA approach yields slightly better performance than the static version, it incurs an additional 8.3% computational overhead, making it less feasible for large-scale real-time applications.

Comparatively, while both static and dynamic TDA graphs outperform the LATTICE base model, the performance gap between the two diminishes after 5 iterations, indicating that the benefits of TDA in training stabilize after a few epochs.

TDA in the Original Graph provides significant structural robustness but at a computational cost. TDA in Each Epoch optimizes modality interactions but introduces training-time

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3582480

complexity. TDA Multimodal Preprocessing achieves the best overall results with 5.8% higher R@20 than LATTICE, demonstrating the power of multimodal fusion. TDA Image Preprocessing benefits datasets with strong visual dependencies, such as Musical Instruments, where performance improved by 4.3% over LATTICE. TDA Text Preprocessing, while effective for textual-rich datasets like Baby, showed marginal improvements due to the sparsity of semantic structures in some recommendation contexts.

F. TDA-BASED PRUNING ON GRAPH LEARNING

The presence of redundant and noisy nodes may degrade performance and increase computational complexity. In this subsection, we investigate TDA for node pruning, comparing it against the baseline Lattice Pruning. We analyze the impact of pruning rates ranging from 4%, 5%, 7.5%, 10% and 20%. Our objective is to assess whether TDA-guided node selection enhances efficiency, stability, and robustness while maintaining high-quality recommendations.

The goal of node pruning in multimodal recommendation systems is two fold. Reduce the computational burden while maintaining meaningful structural representations. Remove redundant and noisy nodes while preserving topological consistency. TDA pruning achieves this by leveraging persistent homology to identify structurally significant nodes based on their topological stability. This contrasts with Lattice pruning, which primarily focuses on network sparsification without accounting for topological importance. Table 4 provides a comparative analysis between TDA pruning and Lattice pruning across three datasets. The results include three key performance metrics along with computational efficiency metrics such as execution time (s) and epoch count.

The results show Musical Instruments dataset TDA improves P@20 by 95.6%, R@20 by 19.4%, and NDCG@20 by 4.1% over Lattice pruning. Digital Music dataset TDA improves P@20 by 13.8%, R@20 by .5%, and NDCG@20 by 2.2%. Baby TDA improves P@20 by 26.9%, R@20 by 3.6%, and NDCG@20 by 6.4%. Beyond performance, we compare the training efficiency of both pruning strategies. TDA pruning reduces epoch counts by an average of 45.5% across datasets, demonstrating improved stability (Faster convergence). TDA pruning reduces execution time by up to 85% in some cases while maintaining superior recommendation quality (Lower Computational Cost). Unlike Lattice pruning, which leads to a drastic drop in performance, TDA pruning preserves ranking quality even when 20% of nodes are removed. TDA pruning significantly reduces execution time, with pruning at 10% reducing computational time by over 60-75% compared to no nodes removed, while maintaining stable performance. Lattice pruning requires more training epochs, leading to longer convergence times. Even at 5% pruning, Lattice models continue to require high epoch counts, resulting in inefficient training. TDA pruning efficiently reduces training epochs, particularly between 4% and 10% pruning, ensuring faster convergence without compromising accuracy.

The superior performance of TDA pruning is attributed to its ability to identify and retain topologically significant nodes, which ensures: Persistent homology preserves critical graph structures. TDA-enhanced embeddings integrate latent topological patterns across modalities. Unlike Lattice pruning, which exhibits severe performance degradation when removing nodes, TDA pruning maintains robust representations.

G. TOPOLOGICAL ANALYSIS OF COLD-START AND NON-COLD-START ITEMS

1) Persistence Lifetime Distribution

The persistence lifetime distribution is a critical descriptor that quantifies the stability of topological features in both cold-start and non-cold-start items across different datasets. By examining the persistence lifetime, we assess the degree to which cold-start items exhibit distinct topological properties compared to well-established items in the recommendation system.

Figure 13 presents the persistence lifetime distributions for three datasets. The blue, red, and green histograms correspond to the cold-start items, while the gray histograms represent non-cold-start items.

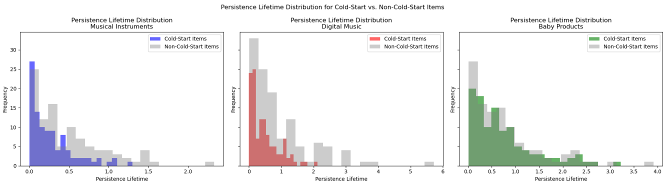


FIGURE 13. **Persistence Lifetime Distribution for Cold-Start vs. Non-Cold-Start Items**

From the figure several key observations emerge:

- Across all datasets, cold-start items tend to exhibit shorter persistence lifetimes, suggesting that their topological structures are less stable and more fragmented. This is particularly evident in Musical Instruments dataset, where the distribution of cold-start items (blue) is concentrated around lower persistence values.
- Digital Music dataset demonstrates a clear separation between cold-start and non-cold-start items, indicating that topological features extracted from persistent homology can effectively differentiate between these two item categories.
- In Baby dataset, the persistence lifetime distribution for cold-start items exhibits a more gradual decline, suggesting that some cold-start items still retain non-trivial topological structures despite their initial sparsity in interactions.

These findings underscore the role of topological descriptors in characterizing cold-start items and highlight the potential of TDA-enhanced features in mitigating cold-start effects by identifying persistent topological structures.

TABLE 4. Comparison of TDA Pruning and Lattice Pruning on Musical Instruments, Digital Music, and Baby datasets. The results include three performance metrics (*P@20, R@20, NDCG@20*), execution time (s), and epoch count for different node pruning levels.

Dataset	Method	Pruning (%)	P@20	R@20	NDCG@20	Time (s)	Epoch
Musical Instruments	TDA Pruning	0%	0.0223	0.2512	0.0987	193	120
		4%	0.0172	0.2403	0.0940	120	70
		5%	0.0195	0.2378	0.0935	85	65
		7.5%	0.0173	0.2305	0.0931	86	65
		10%	0.0190	0.2331	0.0948	110	85
		20%	0.0202	0.2334	0.0952	100	70
	Lattice Pruning	0%	0.0114	0.2103	0.0948	150	220
		4%	0.0022	0.0189	0.0054	140	190
		5%	0.0023	0.0165	0.0049	135	180
		7.5%	0.0020	0.0102	0.0032	140	190
		10%	0.0021	0.0045	0.0021	140	180
		20%	0.0022	0.0052	0.0095	125	170
Digital Music	TDA Pruning	0%	0.0215	0.3118	0.1593	2620	325
		4%	0.0198	0.3007	0.1438	600	100
		5%	0.0171	0.2858	0.1411	580	80
		7.5%	0.0174	0.2873	0.1413	870	65
		10%	0.0193	0.2784	0.1410	550	65
		20%	0.0200	0.2891	0.1484	500	77
	Lattice Pruning	0%	0.0189	0.2956	0.1558	3640	588
		4%	0.0002	0.0091	0.0062	2620	326
		5%	0.0008	0.0085	0.0050	2605	325
		7.5%	0.0003	0.0085	0.0032	2590	355
		10%	0.0006	0.0084	0.0015	2570	440
		20%	0.0005	0.0053	0.0081	2550	300
Baby	TDA Pruning	0%	0.0066	0.0875	0.0414	3450	119
		4%	0.0053	0.0722	0.0356	730	70
		5%	0.0059	0.0705	0.0300	855	70
		7.5%	0.0059	0.0734	0.0292	800	60
		10%	0.0051	0.0753	0.0286	785	75
		20%	0.0060	0.0792	0.0331	460	90
	Lattice Pruning	0%	0.0052	0.0844	0.0389	32654	117
		4%	0.0009	0.0914	0.0085	14509	100
		5%	0.0008	0.0093	0.0080	22935	140
		7.5%	0.0006	0.0067	0.0073	12388	160
		10%	0.0005	0.0038	0.0065	15840	150
		20%	0.0008	0.0085	0.0052	13800	140

2) Betti Curves Evolution

To further explore the topological evolution of cold-start and non-cold-start items, we analyze the progression of Betti numbers over multiple training epochs. Betti curves provide a fine-grained view of the structural complexity of item embeddings as they evolve through training.

Figure 14 illustrates the Betti curves evolution for cold-start and non-cold-start items across different datasets. The curves track the changes in Betti numbers over training epochs, providing insights into the learning dynamics of the recommendation model.

Key findings from this analysis include:

- Convergence of Topological Structures: In all datasets, Betti numbers for cold-start items (solid blue lines) initially exhibit significant deviation from their non-cold-start counterparts (dashed gray lines). However, as training progresses, the Betti curves for cold-start items gradually align with those of non-cold-start items, suggesting that the recommendation model effectively learns topological embeddings that integrate cold-start items into the existing graph structure.
- Rate of Topological Stabilization: The convergence rate

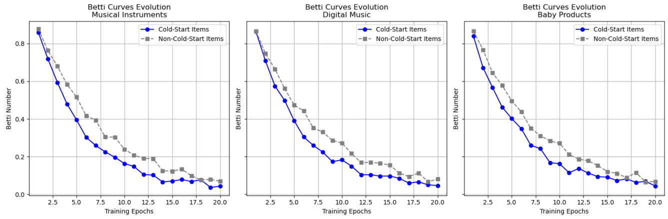


FIGURE 14. Betti Curves Evolution for Cold-Start vs. Non-Cold-Start Items

varies across datasets. The Musical Instruments dataset shows rapid stabilization within the first 10 epochs, whereas the Baby dataset requires more iterations for cold-start items to attain topological consistency.

- Structural Complexity Reduction: A decreasing Betti number trend over epochs indicates that the graph-based model progressively refines item embeddings, reducing redundant topological structures and increasing the efficacy of cold-start item integration into the multimodal recommendation system.

H. REAL-WORLD APPLICATION SCENARIOS

To highlight the practical applicability of our approach, we outline two real-world use-cases where the proposed TDA-augmented multimodal recommendation system can be deployed. First, in e-commerce platforms such as Amazon or Etsy, our system can enhance product recommendations by jointly leveraging customer browsing history (behavior), product images (visual), and product descriptions (textual). For example, a customer interested in baby monitors may receive suggestions that account for product safety certifications (attribute), visual similarity, and customer reviews. Second, in multimedia streaming platforms (e.g., Netflix or Spotify), the system can use listening/viewing patterns, metadata, and content features to recommend contextually relevant media, adapting to diverse user preferences. In both scenarios, persistent homology reveals deep structure across modalities, improving interpretability and robustness of recommendations.

VI. CONCLUSION

In this work, we proposed a novel framework that is topological data analysis with graph neural network for multimodal recommendation systems (TDA-MMRec). By integrating persistent homology with graph-based learning, our approach effectively captures high-order dependencies across multimodal interactions, enhancing both the structural representation and robustness of recommendation models. Our method extracts topological descriptors from diverse modalities text, images, user behaviors, pricing attributes, offering a comprehensive and stable feature representation. Through a modality-aware graph construction process, we embed TDA-derived descriptors into multimodal similarity graphs, preserving both local and global structural information while mitigating over-smoothing and data sparsity issues.

Extensive experiments on large-scale Amazon multimodal datasets demonstrated that our framework significantly outperforms state-of-the-art baselines Precision@20, Recall@20, and NDCG@20, highlighting its effectiveness in capturing complex multimodal dependencies. Notably, our ablation studies confirmed that incorporating TDA-based pruning refines multimodal graphs by eliminating redundant connections while preserving crucial topological structures, leading to improved computational efficiency and interpretability. Furthermore, the integration of topological insights provided a fundamental advantage in addressing cold-start challenges. Analysis of persistence lifetime distributions and Betti curve evolution demonstrated that cold-start items progressively align with non-cold-start items, bridging structural gaps in sparsely connected nodes and improving recommendation accuracy.

Moving forward, we envision optimizing the computational efficiency of topological feature extraction by leveraging approximate persistent homology or differentiable topological layers. Additionally, incorporating dynamic topological features that evolve over time could enable adaptive and sequential recommendation models, further enhancing personalization and user engagement.

REFERENCES

[1] J. Lv, B. Song, J. Guo, X. Du, and M. Guizani, "Interest-Related Item Similarity Model Based on Multimodal Data for Top-N Recommendation," *IEEE Access*, vol. 7, pp. 12 809–12 821, 2019, conference Name: IEEE Access. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8618448

[2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, ser. WWW '01. New York, NY, USA: Association for Computing Machinery, 2001, pp. 285–295. [Online]. Available: https://doi.org/10.1145/371920.372071

[3] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, ser. UAI '09. Arlington, Virginia, USA: AUAI Press, 2009, pp. 452–461.

[4] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8269806

[5] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based Recommender Systems: State of the Art and Trends," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 73–105. [Online]. Available: https://doi.org/10.1007/978-0-387-85820-3

[6] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, Nov. 2002. [Online]. Available: https://doi.org/10.1023/A:1021240730564

[7] K. Bachiri, F. Boufares, M. Malek, N. Rogovschi, and A. Yahyaouy, "Multi -View Clustering Using Sparse Non-Negative Matrix Factorization for Recommendation Systems," in *2023 International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2023, pp. 1287–1294, iSSN: 1946-0759. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10459923

[8] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural Collaborative Filtering," in *Proceedings of the 26th International Conference on World Wide Web*, ser. WWW '17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences

Steering Committee, 2017, pp. 173–182. [Online]. Available: https://doi.org/10.1145/3038912.3052569

[9] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, Jan. 2021, arXiv:1901.00596 [cs]. [Online]. Available: http://arxiv.org/abs/1901.00596

[10] K. Sharma, Y.-C. Lee, S. Nambi, A. Salian, S. Shah, S.-W. Kim, and S. Kumar, "A Survey of Graph Neural Networks for Social Recommender Systems," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 265:1–265:34, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3661821

[11] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural Graph Collaborative Filtering," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR'19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 165–174. [Online]. Available: https://doi.org/10.1145/3331184.3331267

[12] W. Yu, H. Xu, Z. Yuan, and J. Wu, "Learning Modality-Specific Representations with Self-Supervised Multi-Task Learning for Multimodal Sentiment Analysis," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 10 790–10 797, May 2021, number: 12. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17289

[13] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 639–648. [Online]. Available: https://doi.org/10.1145/3397271.3401063

[14] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph Neural Networks in Recommender Systems: A Survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 97:1–97:37, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3535101

[15] K. Bachiri, M. Malek, A. Yahyaouy, and N. Rogovschi, "Adaptive Subgraph Feature Extraction for Explainable Multi-Modal Learning," in *2024 International Conference on Intelligent Systems and Computer Vision (ISCV)*, May 2024, pp. 1–7, iSSN: 2768-0754. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10620106

[16] G. Carlsson, "Topology and Data," *Bulletin of The American Mathematical Society - BULL AMER MATH SOC*, vol. 46, pp. 255–308, Apr. 2009.

[17] R. Ghrist, "Barcodes: The persistent topology of data," *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, vol. 45, Feb. 2008.

[18] P. Bendich, J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer, "Persistent Homology Analysis of Brain Artery Trees," *The annals of applied statistics*, vol. 10, no. 1, pp. 198–218, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5026243/

[19] X. Zhu, "Persistent homology: an introduction and a new text representation for natural language processing," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, ser. IJCAI '13. Beijing, China: AAAI Press, 2013, pp. 1953–1959.

[20] O. Vipond, J. A. Bull, P. S. Macklin, U. Tillmann, C. W. Pugh, H. M. Byrne, and H. A. Harrington, "Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors," *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, p. e2102166118, Oct. 2021, publisher: Proceedings of the National Academy of Sciences. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.2102166118

[21] F. Liu, H. Chen, Z. Cheng, A. Liu, L. Nie, and M. Kankanhalli, "Disentangled Multimodal Representation Learning for Recommendation," *IEEE Transactions on Multimedia*, vol. 25, pp. 7149–7159, 2023, conference Name: IEEE Transactions on Multimedia. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9930669

[22] R. He and J. McAuley, "VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, Feb. 2016, number: 1. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9973

[23] Q. Liu, S. Wu, and L. Wang, "DeepStyle: Learning User Preferences for Visual Recommendation," in *SIGIR*, 2017, pp. 841–844.

[24] Y. Wei, X. Wang, L. Nie, X. He, R. Hong, and T.-S. Chua, "MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 1437–1445. [Online]. Available: https://doi.org/10.1145/3343031.3351034

[25] J. Zhang, Y. Zhu, Q. Liu, S. Wu, S. Wang, and L. Wang, "Mining Latent Structures for Multimedia Recommendation," in *Proceedings of the 29th ACM International Conference on Multimedia*, ser. MM '21. New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 3872–3880. [Online]. Available: https://doi.org/10.1145/3474085.3475259

[26] K. Bachiri, A. Yahyaouy, M. Malek, and N. Rogovschi, "Mm-hgnn: Multimodal representation learning heterogeneous graph neural network," *International Journal of Computational Intelligence Systems*, 2025.

[27] J. Tian, Z. Wang, J. Zhao, and Z. Ding, "MMREC: LLM Based Multi-Modal Recommender System," in *2024 19th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)*, Nov. 2024, pp. 105–110. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10858904

[28] Y. Jiang, L. Xia, W. Wei, D. Luo, K. Lin, and C. Huang, "DiffMM: Multi-Modal Diffusion Model for Recommendation," in *Proceedings of the 32nd ACM International Conference on Multimedia*, ser. MM '24. New York, NY, USA: Association for Computing Machinery, Oct. 2024, pp. 7591–7599. [Online]. Available: https://doi.org/10.1145/3664647.3681498

[29] R. Brüel-Gabrielsson, B. J. Nelson, A. Dwaraknath, P. Skraba, L. J. Guibas, and G. Carlsson, "A Topology Layer for Machine Learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, Apr. 2020, pp. 1553–1563. [Online]. Available: https://proceedings.mlr.press/v108/gabrielsson20a.html

[30] Edelsbrunner, Letscher, and Zomorodian, "Topological Persistence and Simplification," *Discrete & Computational Geometry*, vol. 28, no. 4, pp. 511–533, Nov. 2002. [Online]. Available: https://doi.org/10.1007/s00454-002-2885-2

[31] A. Zomorodian and G. Carlsson, "Computing Persistent Homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, Feb. 2005. [Online]. Available: https://doi.org/10.1007/s00454-004-1146-y

[32] L. Wasserman, "Topological Data Analysis," *Annual review of statistics and its application*, pp. 501–532, Sep. 2018. [Online]. Available: https://doi.org/10.1146/annurev-statistics-031017-100045

[33] C. Hofer, F. Graf, B. Rieck, M. Niethammer, and R. Kwitt, "Graph Filtration Learning," in *Proceedings of the 37th International Conference on Machine Learning*. PMLR, Nov. 2020, pp. 4314–4323, iSSN: 2640-3498. [Online]. Available: https://proceedings.mlr.press/v119/hofer20b.html

[34] C.-C. Wong and C.-M. Vong, "Persistent Homology based Graph Convolution Network for Fine-grained 3D Shape Segmentation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 7078–7087, iSSN: 2380-7504. [Online]. Available: https://ieeexplore.ieee.org/document/9710597

[35] M. Carriere, F. Chazal, Y. Ike, T. Lacombe, M. Royer, and Y. Umeda, "PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR, Jun. 2020, pp. 2786–2796, iSSN: 2640-3498. [Online]. Available: https://proceedings.mlr.press/v108/carriere20a.html

[36] F. Russold and M. Kerber, "Graphcode: Learning from multiparameter persistent homology using graph neural networks," *Advances in Neural Information Processing Systems*, vol. 37, pp. 41 103–41 131, Dec. 2024.

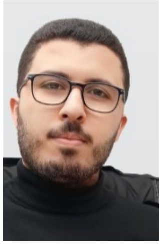
[37] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. American Mathematical Soc., Jan. 2010.

[38] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of Persistence Diagrams," *Discrete & Computational Geometry*, vol. 37, no. 1, pp. 103–120, Jan. 2007. [Online]. Available: https://doi.org/10.1007/s00454-006-1276-5

[39] A. Hatcher, *Algebraic Topology*. Cambridge University Press, 2002.

[40] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, no. 3, pp. 77–102, 2015. [Online]. Available: http://jmlr.org/papers/v16/bubenik15a.html

[41] F. Chazal, B. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman, "Subsampling Methods for Persistent Homology," in *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, Jun. 2015, pp. 2143–2151, iSSN: 1938-7228. [Online]. Available: https://proceedings.mlr.press/v37/chazal15.html



KHALIL BACHIRI received a Master's degree in Data Science and Machine Learning from Sorbonne Paris Nord University, France, in 2021. He is currently a Ph.D. student in Multimodal Learning, Recommendation Systems, and Graph Neural Networks at CY Cergy Paris University, France, in co-tutelle with Sidi Mohamed Ben Abdellah University, Morocco. Since 2023, he has been an Associate Professor at CY Cergy Paris University. His research interests include multimodal learning, recommendation systems, topological data analysis and graph neural networks.

ALI YAHYAOUY received a joint Ph.D. degree from Sidi Mohamed Ben Abdellah University, Morocco, in 2010, and the University of Technology of Compiègne, France. He is currently a professor and is affiliated with the L3IA Laboratory at the Faculty of Sciences, Sidi Mohamed Ben Abdellah University in Fez, Morocco. He has served as the Program Coordinator for the International Francophone Master's Program in "Web Intelligence and Data Science." His research interests include machine learning, deep learning, multi-agent systems, and intelligent transportation systems.

MARIA MALEK received her Ph.D. in Computer Science from Joseph Fourier University, Grenoble, France, in 1996. She is currently an Associate Professor (HDR) at CY Cergy Paris University, CY Tech, where she is affiliated with the ETIS Laboratory (UMR 8051), a joint research unit of CY Cergy Paris University, ENSEA, and CNRS. She also serves as the Deputy Director of the Department of Computer Science. Her research interests include machine learning, data mining, complex network analysis, case-based reasoning, and recommendation systems.

NICOLETA ROGOVSCHI received her Master's degree in Computer Science from Paris 13 University in 2006, specializing in Machine Learning. She completed her Ph.D. in Computer Science in 2009 at the Computer Science Laboratory of Paris 13 University, with a focus on machine learning and data mining. She is currently an Associate Professor (HDR) in Computer Science at Paris Descartes University, where she is affiliated with the LIPADE Laboratory. Her research interests include machine learning, data mining, complex network analysis, and semantic data type detection.