Advanced Sciences and Technology Journal

ASTJ vol. 2, no. 1 (2025), P 1017
10.21608/astj.2025.342474.1017
https://astj.journals.ekb.eg/

A Comprehensive Review of Music Recommendation Systems

Ologia Eskander, Ragy Eskander, Noha Hussen, Iyad Abuhadrous*

Software Engineering Department, Faculty of Engineering and Technology, Egyptian Chinese University, Cairo, Egypt
**iyad.mohammed@ecu.edu.eg*

ARTICLE INFO

Article history:
Received 9 December 2024
Revised 9 January 2025
Accepted 20 January 2025
Available online 25 January 2025

Handling Editor:
Prof. Dr. Mohamed Talaat Moustafa

Keywords:
Music Recommendation Systems (MRS),
Hybrid Recommendation Systems,
Collaborative Filtering (CF),
Content-Based Filtering (CBF),
Graph-Based Models

ABSTRACT

Music recommendation systems (MRS) play a crucial role in navigating extensive music libraries, helping users discover content that aligns with their preferences while addressing challenges such as decision fatigue and overload. This paper explores the evolution of MRS, emphasizing the limitations of traditional approaches like Collaborative Filtering and Content-Based Filtering, which struggle with issues such as cold-start problems, data sparsity, and popularity bias. Hybrid systems, which integrate these methodologies, have emerged as a robust solution, offering improved accuracy, diversity, and personalization. The analysis focuses on advanced hybrid techniques, including graph-based models, multimodal data integration, and artificial intelligence methods such as deep embeddings and adversarial learning. These innovations address critical challenges, including the semantic gap and scalability, while promoting fairness and diversity through metrics that extend beyond accuracy. Furthermore, emerging trends, such as socially motivated frameworks and context-aware recommendations, are examined for their potential to redefine user engagement and enhance the overall recommendation experience. The findings underline the scalability and robustness of hybrid systems, particularly graph-based methodologies, as the future of MRS. However, significant challenges remain, including the optimization of computational efficiency and the creation of equitable recommendation ecosystems. This study concludes by identifying future directions, including real-time adaptability, multimodal integration, and the development of fairness-aware frameworks. These insights underscore the need for continued innovation to meet evolving user needs and technological advancements in the field of music recommendation systems.

1. Introduction

Music recommendation systems (MRS) play a pivotal role in how listeners discover and engage with music in an era defined by vast, ever-growing digital catalogs. With millions of tracks available at any given moment, users often face decision fatigue, emphasizing the importance of efficient recommendation systems that cater to diverse musical tastes and contexts.

Music has long been recognized as an integral part of daily life, offering a means of communication, emotional expression, and self-reflection. Studies indicate that people listen to music more frequently than other forms of media, such as watching television or reading books. This ubiquity has spurred significant advancements in music information retrieval (MIR) techniques, which have evolved to address challenges such as genre classification, artist identification, and instrument recognition. Among these developments, music recommendation systems have emerged as essential tools to help users filter and discover songs tailored to their preferences, raising both opportunities and challenges in understanding and modeling user tastes.

Over the years, MRS technologies have transitioned from simple rule-based approaches to more sophisticated systems leveraging machine learning. Collaborative Filtering (CF) and Content-Based Filtering (CBF) are the most prominent traditional methods. CF, which identifies shared user preferences based on interaction data, is effective but limited by cold-start issues, data sparsity, and popularity bias. Conversely, CBF uses metadata and audio features to recommend tracks but struggles with the semantic gap and lacks robust personalization. These shortcomings have led to the development of hybrid systems that integrate the strengths of both approaches [1].

Emerging trends in MRS address these challenges using advanced methodologies, including graph-based models, multimodal data integration, and socially-aware frameworks [2]. These innovations promise systems that are not only accurate but also diverse, novel, and user-centric. Recent advancements, such as graph neural networks (GNNs) [3], transformer-based architectures, and adversarial learning, underscore the increasing sophistication of MRS technologies [4]. Additionally, emotion-based and context-aware models, which consider mood and contextual factors, are gaining traction, offering further personalization capabilities. However, these approaches remain in the early stages, highlighting the need for continued research and development.

This review aims to analyze the current state of music recommendation systems, focusing on hybrid approaches and emerging technologies. By evaluating their effectiveness, limitations, and impact on key metrics such as diversity and personalization, this paper seeks to clarify knowledge gaps and identify promising research directions. The objectives are threefold:

1.To explore the strengths and limitations of CF and CBF, highlighting how hybrid methods address their shortcomings.

2.To evaluate cutting-edge techniques, including graph-based systems, deep learning, and multimodal integration, for improving recommendation quality.

3.To discuss challenges like scalability, computational efficiency, and fairness, and propose future directions that align with user-centric and equitable recommendation objectives.

Many listeners, particularly those with niche tastes, find existing systems inadequate for discovery and personalization. This study advocates for more inclusive, adaptive systems that balance computational efficiency with meaningful user engagement.

2. Methodology

The primary objective of this systematic review is to analyse existing research on music recommendation systems (MRS) with a specific focus on hybrid approaches and emerging trends. This review aims to consolidate current knowledge, identify key advancements, and propose future research directions. By critically examining diverse methodologies, from collaborative filtering to multimodal systems, we seek to provide a comprehensive understanding of MRS and their implications for accuracy, diversity, and personalization.

2.1 Adherence to PRISMA

This review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor and transparency. PRISMA's structured framework includes defining the research objectives, developing a search strategy, selecting studies based on inclusion and exclusion criteria, extracting data, and synthesizing results comprehensively.

2.2 Research Scope and Keywords

The review investigates music recommendation systems, particularly focusing on hybrid approaches integrating collaborative and content-based filtering techniques. Keywords central to this search included: "music recommendation systems," "collaborative filtering", "content-based filtering", "hybrid recommendation systems", “Features Extraction”, and "Music Classification".

2.3 Search and Selection Process

2.3.1 Initial Manuscript Selection

A broad search was conducted across multiple academic databases, including IEEE Xplore, ScienceDirect, SpringerLink, Semantic Scholar, and Google Scholar. Manuscripts were initially screened based on titles and abstracts to determine their relevance to music recommendation systems and their advancements.

2.3.2 Manuscript Identification

The initial search yielded 81 relevant studies, refined through database cross-referencing to eliminate duplicates. After preliminary filtering, 38 studies focusing on advanced methodologies, such as graph neural networks and multimodal approaches, were shortlisted.

2.3.3 Rigorous Review and Inclusion Criteria

- Inclusion Criteria:

  (a) Focus on hybrid MRS integrating collaborative and content-based filtering.

  (b) Studies published between 2010 and 2024 to ensure contemporary relevance.

  (c) Methodologically sound papers with clear findings and innovative contributions.

- Exclusion Criteria:

(a) Irrelevant focus, such as single-method MRS.

(b) Methodological weaknesses or incomplete data representation.

(c) Studies not published in peer-reviewed journals or conference proceedings.

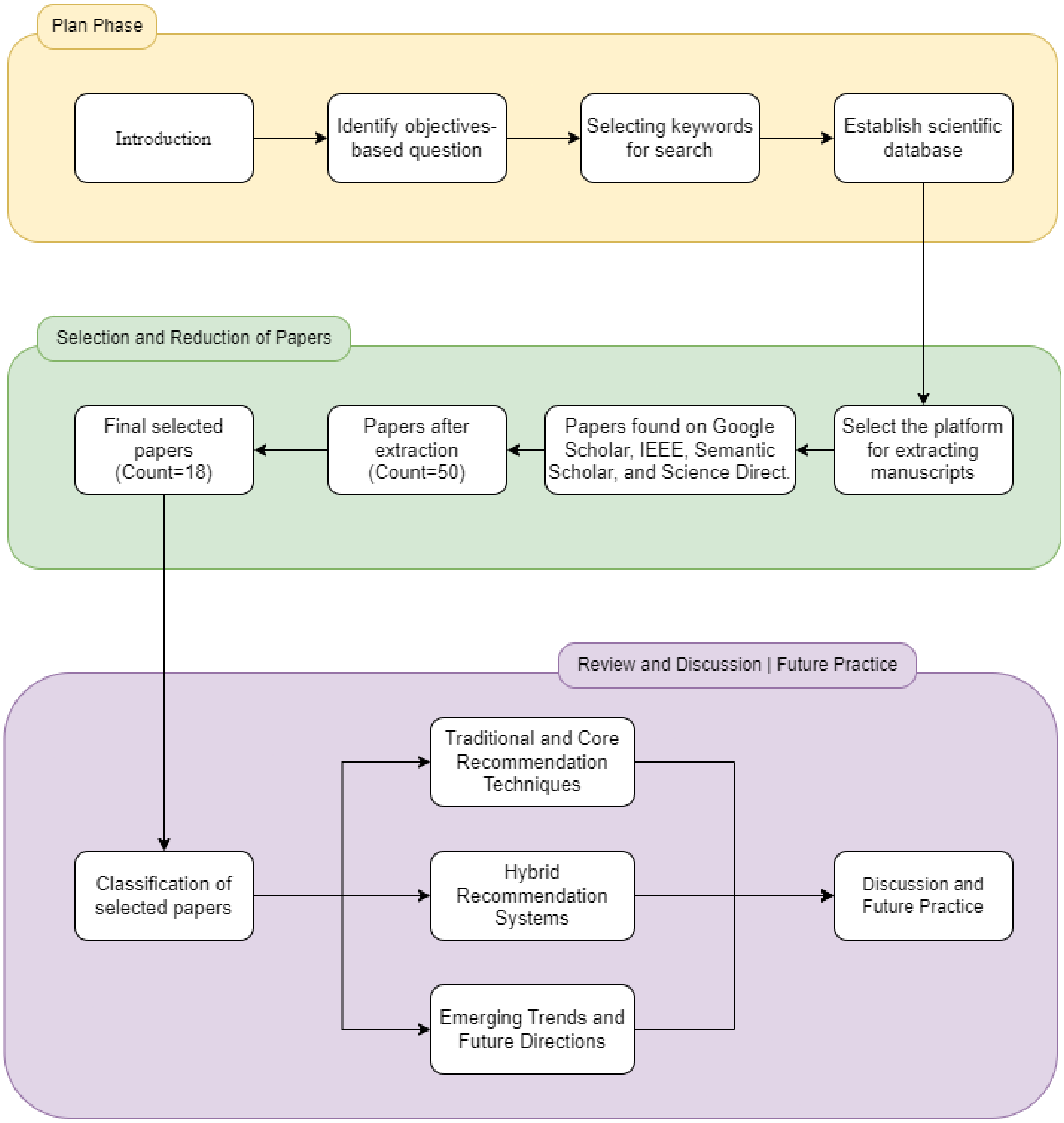

Fig. 1: Research framework overview

2.4 Data Extraction Process

2.4.1 Collaborative Data Extraction

Our systematic review collected detailed information from every study using a systematic extraction approach. This included categorizing the type of music recommendation, Music Classification, Features Extraction, Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Recommendation Systems. In order to create Table 1, where difficult and complex articles were evaluated collectively, teamwork and the help of AI tools was essential.

This approach outlines the research framework, starting from the planning phase and proceeding through the application of inclusion and exclusion criteria for the literature review, followed by the review and discussion of findings. Figure 1 provides a visual summary of the process, including the scope for future work, illustrating the systematic methodology employed in the study to ensure the relevance and timeliness of the reviewed literature

Table 1. Overview of Selected Manuscripts

Title	*Ref*	*Author*	*Publication Year*	*Item Type*	*Explanation/Conclusion*
Socially-Motivated Music Recommendation	[5]	Lacker, Benjamin; Way, Samuel F.	2024	journal Article	This study explores social motivations for music listening, focusing on community trends and the trade-off between precision and timeliness in recommendations. It highlights cultural differences, user attributes, and the role of implicit feedback.
Collaborative Music Similarity and Recommendation	[6]	Knees, Peter; Schedl, Markus	2016	book Section	The paper explores music recommender systems, focusing on collaborative filtering, content-based methods, and hybrid approaches. It addresses challenges such as cold-start issues, rating biases, and overfitting. The use of topic modelling, graph-based models, and metric learning enhances recommendation accuracy.
Music Recommendation	[7]	Celma, Òscar	2010	book Section	The paper discusses music recommendation challenges, including user profiling, collaborative filtering, and the music information plane. It highlights the role of cultural information, social tagging, and privacy concerns, emphasizing the need for effective user profile representation.

Content-Aware Collaborative Music Recommendation Using Pre-Trained Neural Networks	[8]	Liang, Dawen; Zhan, Minshu; Ellis, D.	2015	conference Paper	The paper presents a content-aware recommendation system combining neural networks and collaborative filtering, excelling in cold-start problems but facing challenges with scalability and feature extraction.
A Survey of Audio-Based Music Classification and Annotation	[9]	Fu, Zhouyu; Lu, Guojun; Ting, Kai Ming; Zhang, Dengsheng	2011	journal Article	The survey reviews audio-based music classification, highlighting challenges, task-specific issues, and gaps in automated systems, excluding symbolic classification.
Music encoding and deep learning for music transcription and classification based on visually represented audio features	[10]	Achkar, Charbel El	2023	thesis	The paper focuses on music genre classification and transcription, addressing the lack of XML support for oriental music. It introduces MusicPatternOWL for knowledge extraction, proposes deep learning techniques, and presents the MEI2JSON converter, which outperforms existing methods.
Research on pattern recognition of different music types in the context of AI with the help of multimedia information processing	[11]	Sun, Wei; Sundarasekar, Revathi	2023	journal Article	The paper introduces IMuCo, an AI composer that creates soundtracks based on images, focusing on moods rather than specific emotions. It explores AI's role in music composition, the need for low-cost evaluation methods, and aims to enhance existing music classification and generation methods.
Hybrid music recommendation with graph neural networks	[1]	Bevec, Matej; Tkalčič, Marko; Pesek, Matevž	2024	journal Article	The paper introduces PinSage, a graph-based hybrid music recommendation system addressing data sparsity with improved accuracy and flexibility. While effective against collaborative filtering baselines, it highlights challenges like popularity bias and limitations of offline evaluation metrics.
Categorization of songs using spectral clustering	[12]	Blomkvist, Linus Below; Darke, Felix	2021	journal Article	The paper explores data analysis in a digital context, focusing on spectral clustering and graph Laplacian properties to categorize 50,704 songs. It highlights challenges like high computational costs and emphasizes machine learning for predicting song skips based on user behaviour.

2.4.2 Collaborative Filtering (CF)

Collaborative Filtering (CF) remains one of the most widely used methods in music recommendation systems (MRS). It operates by leveraging user interaction data, relying on the assumption that similar users prefer similar items. This domain-independence makes CF highly versatile, as it requires no content information beyond user-item interaction matrices. Studies highlight its success in terms of recommendation accuracy, making it a dominant approach in various recommender systems [6], [7].

2.5 Strengths

The primary strength of CF lies in its domain independence. CF algorithms, whether user-based or item-based, depend exclusively on user behavior data, bypassing the need for metadata or audio content. This characteristic is particularly advantageous in domains with well-established user bases, allowing CF systems to adapt to user preferences with little manual intervention [6].

Additionally, CF systems excel at identifying shared preferences among user communities, leveraging patterns from collective behavior to generate recommendations [2], [5].

2.6 Limitation

Collaborative Filtering (CF) is not without its challenges. A significant issue is data sparsity, especially prevalent in large music catalogs where the majority of items have minimal interaction data. This sparsity undermines CF’s reliability in providing recommendations for lesser-known tracks and artists. Even for well-established datasets, CF systems suffer from popularity bias, where frequently consumed items dominate recommendations. This bias leads to over-representation of popular music, sidelining niche or emerging artists and creating inequities in exposure [13], [14], [15].

Mechanisms of Popularity Bias

- Data Representation: Popular items are over-represented in the training data, leading to their disproportionate recommendation [16], [17].
- User Preferences: CF systems often cater to users with mainstream tastes, while those preferring less popular items receive less relevant recommendations [18].
- Personality Influence: Users with less extroverted personalities tend to experience greater unfairness due to CF's

tendency to reinforce popular items [19].

These factors significantly impact user experience, particularly for listeners who seek variety and discovery. Mitigating popularity bias is essential to create fairer, more inclusive recommendation systems[18].

2.7 Enhancements

To address these shortcomings, enhancements to CF have been explored. One strategy involves incorporating implicit feedback, such as play counts, skips, and song completions, rather than relying solely on explicit ratings. Implicit feedback provides a richer interaction dataset, alleviating sparsity issues and improving system robustness [6], [20].

Popularity correction terms have also been proposed to counteract biases in similarity measures that arise from item popularity disparities. These adjustments aim to create a balanced recommendation landscape, ensuring lesser-known tracks gain visibility alongside mainstream items. Moreover, pre-trained models and neural architectures have begun to integrate implicit features from interaction logs, further enhancing recommendation quality [5], [8].

2.8 Beyond Accuracy Metrics

Emerging research in CF also emphasizes the importance of beyond-accuracy objectives, such as diversity, novelty, and personalization. These metrics evaluate a system's ability to provide unique and user-specific recommendations, thereby enhancing user satisfaction and engagement [14]. Addressing popularity bias directly contributes to these beyond-accuracy goals.

2.8.1 Impact on User Satisfaction

- Familiarity vs. Discovery: Techniques promoting less popular items enhance user satisfaction by fostering a sense of discovery, despite initial unfamiliarity[14].
- Diversity and Fairness: Reducing popularity bias improves recommendation diversity, providing a more equitable experience for niche listeners and underrepresented artists[21].

Balancing familiarity with discovery remains a challenge, as users who prefer mainstream music may experience reduced satisfaction when popularity bias is mitigated [13].

2.8.2 Content-Based Filtering (CBF)

Content-Based Filtering (CBF) leverages item-specific attributes to generate recommendations by identifying similarities between items and the user’s interaction history. In music recommendation systems (MRS), "content" can range from metadata, such as artist or genre, to more intricate audio features extracted from the music itself [7], [22].

2.9 Traditional Approaches

Traditional CBF methods often rely on high-level metadata, such as genre tags, artist names, and album information. These descriptors are intuitive and computationally inexpensive but lack granularity. For instance, simple genre tags often fail to reflect nuanced differences in user preferences, leading to subpar recommendations [23].

Alternatively, low-level features derived from raw audio signals—such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral features, and rhythmic patterns—are employed to compute item similarity [9], [12]. These methods are more granular but face significant challenges, particularly the "semantic gap," where the extracted features do not align with human-perceived musical qualities like emotion or style. The semantic gap has been a persistent issue, as highlighted by efforts to bridge it using advanced machine learning models [24].

2.10 Limitations

CBF has notable limitations. Metadata-based methods often suffer from incomplete or inaccurate datasets, reducing the reliability of recommendations. Low-level feature extraction struggles with the semantic gap, limiting the ability to represent music in a way that aligns with user preferences [11], [12], [23].

Furthermore, pure CBF systems overlook user interaction data, resulting in recommendations that lack the personalization achieved by CF [1].

2.11 Advances in Music Information Retrieval (MIR)

Recent advancements in Music Information Retrieval (MIR) have pushed the boundaries of CBF by employing deep learning and psychological models. Deep learning techniques, such as Convolutional Neural Networks (CNNs) and Transformer-based architectures, have been instrumental in extracting meaningful representations of audio signals [4], [10], [25]. These models aim to bridge the semantic gap by learning perceptually relevant features directly from raw audio, as shown in recent work on pattern recognition and adversarial learning frameworks, and by using a recently popular technique—latent Dirichlet allocation (LDA)—and its application to listening sessions in order to recommend items or playlists [6].

In addition, the Arousal-Valence-Depth (AVD) model, derived from music psychology, offers a compact yet meaningful way to characterize user preferences. By focusing on dimensions like energy, emotional tone, and complexity, the AVD framework integrates well with collaborative filtering systems to address cold-start problems, demonstrating significant improvements in recommendation quality [26]. Furthermore, techniques that use end-to-end learning combining audio features with textual metadata and user feedback to incorporate both audio and interaction data can bridge the semantic gap and improve recommendation quality [27], [28], [29].

2.12 Beyond Audio Features

Emerging trends in CBF emphasize multimodal content integration, combining audio features with other modalities like text (e.g., lyrics), images (e.g., album art), and user-generated content (e.g., tags and reviews). This multimodal approach enhances recommendation diversity and user engagement, presenting a holistic view of musical content [2], [6].

Multimodal methods increasingly leverage advancements in AI, such as generative adversarial networks (GANs)[11] and pre-trained neural networks, to enrich data representations and improve recommendation quality. Future research is expected to focus on integrating richer contextual data, including social and cultural dimensions, to further personalize recommendations [4].

Hybrid Recommendation Systems

Hybrid recommendation systems combine Collaborative Filtering (CF) and Content-Based Filtering (CBF) to address the limitations of each. CF relies on user interaction data to recommend items, while CBF analyzes content attributes. Integrating these two approaches enhances recommendation quality, particularly by mitigating the cold-start problem and addressing sparsity and popularity biases [1], [8].

Recent advancements in hybrid systems demonstrate their ability to leverage graph-based methods, multimodal data, and pre-trained neural networks. These innovations expand the scope of hybrid systems, making them adaptable to various user needs and diverse music collections [1], [7].

2.13 Motivations for Hybrid Approaches

The motivations for hybrid systems stem from the limitations inherent in pure CF and CBF. CF struggles with data sparsity, particularly for niche or long-tail content, and exhibits a strong bias toward popular items. Conversely, CBF, while capable of handling new items, often fails to capture the depth of user preferences due to the semantic gap in feature extraction. Hybrid systems address these challenges by leveraging both user interaction and content data, creating more accurate, diverse, and personalized recommendations [18], [30].

Hybrid approaches are also crucial for mitigating popularity bias, which affects user satisfaction and engagement. By balancing exposure between popular and niche content, hybrid systems foster a fairer recommendation landscape, ensuring that users discover new music while still aligning with their tastes [17], [31].

2.14 Graph-Based Methods

Graph-based approaches have become a cornerstone of modern hybrid systems. Graph Neural Networks (GNNs) offer a scalable and flexible framework for combining CF and CBF data. These models represent users, items, and their interactions as nodes and edges within a graph structure, enabling the system to learn complex relationships between entities [31].

PinSage, one of the most notable GNN-based systems, exemplifies the power of this approach. Originally designed for large-scale recommendation tasks, PinSage leverages random walks to construct localized graph representations, capturing both interaction data and content-based features such as deep audio embeddings. This capability significantly enhances the scalability and accuracy of hybrid systems in music recommendation [1], [3], [8].

Other graph-based techniques integrate co-occurrence graphs, such as playlist-song graphs, where nodes represent tracks and edges reflect their co-occurrence in user playlists. This method effectively captures implicit user preferences and has been successfully applied in platforms like Spotify to improve recommendation diversity and catalog coverage [6], [23]. Integrating GNNs with multimodal data sources—such as audio features, textual metadata, and user feedback—has proven effective in improving recommendation diversity and addressing popularity bias [32].

2.15 Applications and Results

Hybrid systems employing graph-based and multimodal methods have consistently demonstrated superior performance across accuracy and beyond-accuracy metrics. For instance, systems integrating deep learning models with graph-based frameworks outperformed traditional CF and CBF methods in precision, diversity, and novelty evaluations. These systems are particularly adept at promoting long-tail content, ensuring that niche tracks and lesser-known artists receive exposure [14], [33].

User-Centric and Item-Centric Techniques

- User-Centric Approaches: These techniques focus on promoting lesser-known items that align with individual user preferences, enhancing satisfaction by fostering discovery [14].
- Item-Centric Approaches: Strategies like negative sample mixing (e.g., the PopMix algorithm) help balance the exposure of popular and underrepresented items while maintaining recommendation accuracy [20].

Hybrid systems that incorporate domain-aware methodologies (e.g., Graph Neural Networks) can reflect intrinsic similarities between songs, improving recommendations for less mainstream content [31]. Techniques optimizing embedding spaces for cultural inclusivity further ensure that diverse artists gain visibility, promoting fairness across different demographics [17].

2.16 Beyond-Accuracy Metrics

Hybrid systems are particularly effective in achieving beyond-accuracy objectives, such as diversity, novelty, and personalization. By integrating multiple data sources and learning frameworks, these systems provide recommendations that cater to unique user preferences while encouraging exploratory behavior.

2.16.1 Fairness and Inclusivity

- Addressing popularity bias helps ensure that niche interests are represented, leading to a more equitable recommendation landscape [18].
- Techniques that optimize embedding spaces for cultural inclusivity further promote fairness by increasing

visibility for underrepresented artists [17].

2.16.2 Impact on User Loyalty

- By diversifying recommendations, hybrid systems enhance user engagement and loyalty, as users discover new music aligned with their evolving tastes [14].
- However, balancing novelty with familiarity remains critical, as some users prefer popular tracks, and mitigating popularity bias too aggressively may reduce their satisfaction [31].

2.16.3 Emerging Trends in Music Recommendation

Emerging trends in music recommendation systems reflect the dynamic evolution of technology, user expectations, and content diversity. These advancements emphasize the integration of cutting-edge machine learning techniques, multimodal data fusion, and socially-driven frameworks to enhance user experience and engagement [11], [34].

Socially-motivated recommendation systems focus on the communal and cultural aspects of music consumption. These systems aim to capture the influence of community trends, cross-cultural variations, and the social bonding facilitated by music. Studies have highlighted the role of music in group dynamics and cultural identity, suggesting that socially-aware systems could improve engagement by tailoring recommendations to shared p references and trends [5].

Approaches incorporating community-driven data have demonstrated higher precision in collectivist societies, where music recommendations align with group preferences. For example, systems that analyze playlist-sharing behaviors and social interactions identify emergent trends within communities, enabling timely and relevant recommendations [2].

Future directions include enhancing these models by integrating explicit social signals, such as likes and shares, alongside implicit signals like co-listening behaviors. This can help refine socially-motivated recommendation algorithms, making them more adaptable to diverse cultural contexts [6], [9].

Artificial intelligence has revolutionized music recommendation systems by introducing sophisticated architectures like transformers, generative adversarial networks (GANs) [11], and deep embedding models. These methods capture intricate relationships between songs and users, improving the system's ability to recommend nuanced and contextually relevant music. Transformers, in particular, have been employed to model sequential listening behaviors, capturing temporal patterns that traditional methods overlook [6], [8].

Adversarial learning frameworks, such as ALI (Adversarially Learned Inference), further refine these embeddings, improving clustering performance for music categorization tasks. These models are particularly effective in addressing challenges like cold starts and sparsity by generating plausible representations for less-explored data [4]. Research incorporating cross-modal variational auto-encoders (CMVAEs) has also demonstrated success in aligning video content with background music, enriching the user experience [34].

2.16.4 Multimodal Integration

The integration of multimodal data is redefining the scope of music recommendation systems. By combining audio features, textual data (e.g., lyrics or user reviews), visual content (e.g., album art), and user interaction logs, these systems offer a holistic understanding of musical preferences. This fusion not only enhances the diversity and relevance of recommendations but also captures the multifaceted nature of music perception [2], [6].

Moreover, the use of contextual data, such as activity logs, implicit feedback such as skips and playtime, allows systems to adapt recommendations to real-time user contexts.

2.16.5 Expanding Beyond Accuracy Metrics

Modern music recommendation research increasingly emphasizes metrics beyond traditional accuracy, such as diversity, novelty, and serendipity. Hybrid models and multimodal systems have demonstrated their ability to excel in these dimensions, providing a richer and more exploratory user experience. For instance, graph-based hybrid models balance precision with novelty, exposing users to less familiar yet relevant tracks while maintaining high engagement levels [1], [8].

Furthermore, efforts to improve fairness and reduce popularity bias are gaining traction. These initiatives aim to ensure equitable exposure for niche and emerging artists, addressing systemic imbalances in music ecosystems. By integrating fairness objectives into hybrid systems, researchers are paving the way for inclusive and sustainable recommendation platforms [5], [22].

Emerging trends in music recommendation underscore the potential of socially-aware, AI-driven, and multimodal systems to revolutionize the listening experience. These advancements highlight the need for continued innovation to meet the growing complexity of user preferences and the diversity of musical content.

3. Critical Review

3.1. Foundational Studies or Traditional Methods

Papers such as Collaborative Music Similarity (Knees et al., 2016) and A Survey of Audio-Based Music (Fu et al., 2011) established the strengths of Collaborative Filtering (CF) and Content-Based Filtering (CBF) in music recommendation. These studies emphasize CF's reliance on user interaction data and CBF's use of audio features. However, they lack deeper exploration into addressing the cold-start problem and the semantic gap, particularly in cases involving niche or less popular content. While these foundational approaches provide a baseline, they fall short in proposing solutions for improving recommendation diversity and personalization.

3.2. Hybrid and Graph-Based Systems

Research like Hybrid Music Recommendation (Bevec et al., 2024) and Content-Aware Collaborative Music (Liang et al., 2015) demonstrate the effectiveness of hybrid systems in overcoming the limitations of CF and CBF. These papers explore the integration of graph-based methods and neural networks to enhance accuracy and diversity. Despite their contributions, they do not fully address the challenges of scalability and computational efficiency in real-time applications. Furthermore, while these studies highlight improvements in accuracy, they provide limited insights into mitigating popularity bias and ensuring fairness in recommendations.

3.3. Emerging Trends and AI Advances

Studies like Motivic Pattern Classification of Music Audio (Achkar, 2023) and Socially-Motivated Music Recommendation (Lacker et al., 2024) highlight the potential of AI-driven methods, such as deep learning and socially-aware frameworks. These approaches address the need for more nuanced and culturally sensitive recommendations. However, they often lack comprehensive strategies for integrating multimodal data and balancing user-centric and item-centric objectives. Additionally, while these papers explore novel techniques, they fall short in offering scalable solutions for large-scale music recommendation systems.

3.4. Evaluation of Beyond-Accuracy Metrics

Research focusing on beyond-accuracy metrics, such as Content-Driven Music Recommendation: Evolution (Deldjoo et al., 2024) and Improving Context-Aware Music (Pichl et al., 2017), emphasizes the importance of diversity, novelty, and fairness in evaluating recommendation systems. These studies propose methods for enhancing user experience through exploratory recommendations. However, they often struggle to balance these objectives with accuracy and personalization. The trade-offs between novelty and familiarity remain underexplored, indicating a need for more comprehensive frameworks that integrate beyond-accuracy metrics without compromising core recommendation quality.

4. Discussion and Analysis

4.1. Interpreting Key Findings

The literature reveals that hybrid recommendation systems consistently outperform traditional CF and CBF approaches across multiple dimensions. While CF provides reasonable accuracy in cases with sufficient interaction data, it falters with cold-start and sparsity issues. Similarly, CBF methods, though robust for new items, lack the personalization achieved by CF. Hybrid systems bridge these gaps by integrating the strengths of both, delivering significant improvements in accuracy, diversity, and long-tail coverage[1], [7].

Graph-based methods, particularly those leveraging GNNs, emerge as a standout solution. Techniques like PinSage not only enhance recommendation accuracy but also address long-standing challenges such as the semantic gap and scalability. These methods are adept at incorporating multimodal data, further enriching recommendation quality. However, their complexity and resource requirements highlight the trade-offs inherent in advanced hybrid approaches [3].

Hybrid systems also excel in beyond-accuracy metrics, promoting diversity, novelty, and personalization. These dimensions are crucial for engaging users in exploratory behaviors and fostering fairer ecosystems where niche content is more accessible.

4.2. The Role of Beyond-Accuracy Objectives

Beyond-accuracy metrics such as diversity, novelty, and personalization are gaining prominence in evaluating recommendation systems. Hybrid systems consistently outperform their CF and CBF counterparts in these areas. Diversity ensures users receive a varied selection of recommendations, while novelty introduces them to lesser-known items, expanding their musical horizons. Personalization tailors suggestions to individual preferences, enhancing user satisfaction and engagement. The incorporation of multimodal and graph-based methods has been particularly effective in achieving these objectives, underscoring their importance in next-generation recommendation systems [1].

Balancing these objectives with accuracy remains a challenge. Systems designed to prioritize novelty or diversity often encounter trade-offs with predictive precision. Future advancements must explore methods to optimize this balance, ensuring that systems cater to both exploratory and predictive needs of users [1].

4.3. Challenges

Hybrid systems, despite their strengths, face several challenges. Scalability remains a critical concern, especially for graph-based methods that require substantial computational resources. As datasets grow larger and more complex, optimizing these systems to handle real-time recommendations without compromising performance is imperative [1], [24].

Additionally, the integration of multimodal data introduces its own set of complexities. Aligning heterogeneous data sources such as audio features, text, and images requires sophisticated preprocessing and representation techniques. Ensuring these systems remain interpretable and user-friendly adds an additional layer of complexity [23].

4.4. Future Prospects

The future of hybrid music recommendation systems lies in leveraging emerging technologies and methodologies. Multimodal integration, incorporating richer content such as video, user-generated tags, and real-time contextual data, promises to create more dynamic and engaging recommendation systems. Advanced hybridization frameworks, including counterfactual evaluation methods, will also enhance the robustness of these systems [1], [23].

Further research is needed to optimize graph-based methods for scalability, exploring novel architectures that balance computational efficiency with recommendation quality. Techniques like knowledge distillation and pruning may help reduce the resource footprint of complex GNNs, making them more viable for large-scale applications [1].

Additionally, incorporating user satisfaction measures, both explicit and implicit, into system feedback loops could revolutionize how systems adapt to individual preferences. Models that dynamically learn from user feedback in real-time will be pivotal in achieving the next level of personalization and engagement [5].

5. Conclusion and Future Directions

Hybrid recommendation systems surpass traditional CF and CBF approaches in accuracy, diversity, and user satisfaction. Their ability to integrate user interaction data with content features makes them particularly effective in addressing cold-start and sparsity issues.

Graph-based hybrid models, such as those employing GNNs, offer a scalable and nuanced approach to music recommendation. By leveraging user-item relationships and multimodal content, these methods enhance recommendation accuracy while promoting long-tail content.

Beyond-accuracy metrics such as diversity, novelty, and personalization are gaining prominence in evaluating recommendation systems. Hybrid systems consistently outperform their CF and CBF counterparts in these areas. Graph-based hybrid models excel in balancing the accuracy-diversity trade-off, ensuring that users are exposed to a wide range of music while maintaining high recommendation quality. Techniques focusing on fairness and serendipity are particularly effective in promoting underrepresented content and enhancing user engagement [35], [36], [37], [38].

Scalability, computational demands, and the integration of multimodal data remain key hurdles. These challenges highlight the need for further research into optimizing hybrid systems without sacrificing performance or interpretability.

The findings underscore the importance of hybrid approaches in advancing music recommendation technology. For users, these systems offer richer and more personalized experiences, introducing them to a diverse array of content while aligning with their preferences. For creators, hybrid systems provide a fairer platform, increasing the visibility of niche and emerging artists. The integration of multimodal data and beyond-accuracy objectives represents a paradigm shift toward user-centric and inclusive recommendation systems.

The next steps in hybrid recommendation research should focus on optimizing graph-based and multimodal approaches for real-world scalability. Future systems must address computational constraints while preserving their ability to deliver high-quality recommendations. Techniques such as counterfactual evaluation, advanced hybridization frameworks, and real-time learning from user feedback will be crucial in achieving this balance.

In addition, expanding the scope of multimodal integration to include richer content like social media interactions, contextual cues, and user-generated tags will further enhance personalization. Finally, developing frameworks that explicitly tackle fairness and popularity bias will be essential for creating equitable and inclusive systems that benefit all stakeholders in the music ecosystem.

This review reaffirms the potential of hybrid systems as the future of music recommendation, emphasizing the need for continued innovation and collaboration across disciplines to realize their full capabilities.

References

[1] M. Bevec, M. Tkalčič, and M. Pesek, "Hybrid music recommendation with graph neural networks," *User Model User-Adap Inter*, Aug. 2024, doi: 10.1007/s11257-024-09410-4.

[2] M. Pichl, E. Zangerle, and G. Specht, "Improving Context-Aware Music Recommender Systems: Beyond the Pre-filtering Approach," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, Bucharest Romania: ACM, Jun. 2017, pp. 201–208. doi: 10.1145/3078971.3078980.

[3] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph Convolutional Neural Networks for Web-Scale Recommender Systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London United Kingdom: ACM, Jul. 2018, pp. 974–983. doi: 10.1145/3219819.3219890.

[4] R. L. F. Cunha, E. Caldeira, and L. Fujii, "Determining Song Similarity via Machine Learning Techniques and Tagging Information," Apr. 12, 2017, *arXiv*: arXiv:1704.03844. doi: 10.48550/arXiv.1704.03844.

[5] B. Lacker and S. F. Way, "Socially-Motivated Music Recommendation," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 18, pp. 879–890, May 2024, doi: 10.1609/icwsm.v18i1.31359.

[6] P. Knees and M. Schedl, "Collaborative Music Similarity and Recommendation," in *Music Similarity and Retrieval: An Introduction to Audio- and Web-based Strategies*, P. Knees and M. Schedl, Eds., Berlin, Heidelberg: Springer, 2016, pp. 179–211. doi: 10.1007/978-3-662-49722-7_8.

[7] Ò. Celma, "Music Recommendation," in *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*, Ò. Celma, Ed., Berlin, Heidelberg: Springer, 2010, pp. 43–85. doi: 10.1007/978-3-642-13287-2_3.

[8] D. Liang, M. Zhan, and D. Ellis, "Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks," presented at the International Society for Music Information Retrieval Conference, 2015. Accessed: Oct. 24, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Content-Aware-Collaborative-Music-Recommendation-Liang-Zhan/6e06a1796bf36b2f0896a3e350a972e3160c7d66

[9] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation," *IEEE Trans. Multimedia*, vol. 13, no. 2, Art. no. 2, Apr. 2011, doi: 10.1109/TMM.2010.2098858.

[10] C. E. Achkar, "Music encoding and deep learning for music transcription and classification based on visually represented audio features," phdthesis, Université Bourgogne Franche-Comté, 2023. Accessed: Nov. 03, 2024. [Online]. Available: https://theses.hal.science/tel-04483299

[11] W. Sun and R. Sundarasekar, "Research on pattern recognition of different music types in the context of AI with the help of multimedia information processing," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Feb. 2023, doi: 10.1145/3523284.

[12] L. B. Blomkvist and F. Darke, "Categorization of songs using spectral clustering".

[13] D. Kowald, M. Schedl, and E. Lex, "The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study," vol. 12036, pp. 35–42, 2020, doi: 10.1007/978-3-030-45442-5_5.

[14] R. Ungruh, K. Dinnissen, A. Volk, M. S. Pera, and H. Hauptmann, "Putting Popularity Bias Mitigation to the Test: A User-Centric Evaluation in Music Recommenders," in *Proceedings of the 18th ACM Conference on Recommender Systems*, in RecSys '24. New York, NY, USA: Association for Computing Machinery, Oct. 2024, pp. 169–178. doi: 10.1145/3640457.3688102.

[15] A. Ferraro, "Music cold-start and long-tail recommendation: bias in deep representations," in *Proceedings of the 13th ACM Conference on Recommender Systems*, in RecSys '19. New York, NY, USA: Association for Computing Machinery, Sep. 2019, pp. 586–590. doi: 10.1145/3298689.3347052.

[16] M. Mansoury, F. Duijvestijn, and I. Mourabet, "Potential Factors Leading to Popularity Unfairness in Recommender Systems: A User-Centered Analysis," Oct. 04, 2023, *arXiv*: arXiv:2310.02961. doi: 10.48550/arXiv.2310.02961.

[17] A. Moradi, N. Neophytou, and G. Farnadi, "Advancing Cultural Inclusivity: Optimizing Embedding Spaces for Balanced Music Recommendations," May 27, 2024, *arXiv*: arXiv:2405.17607. doi: 10.48550/arXiv.2405.17607.

[18] J. Prent and M. Mansoury, "Correcting for Popularity Bias in Recommender Systems via Item Loss Equalization," Oct. 07, 2024, *arXiv*: arXiv:2410.04830. doi: 10.48550/arXiv.2410.04830.

[19] E. Yalcin and A. Bilge, "Popularity bias in personality perspective: An analysis of how personality traits expose individuals to the unfair recommendation," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 9, p. e7647, 2023, doi: 10.1002/cpe.7647.

[20] M. Jangid and R. Kumar, "Enhancing user experience: a content-based recommendation approach for addressing cold start in music recommendation," *J Intell Inf Syst*, Sep. 2024, doi: 10.1007/s10844-024-00872-x.

[21] S. Song and B. Wang, “Mitigating Popularity Bias in Recommendation Systems Via Negative Sample Mixing,” 2024, *SSRN*. doi: 10.2139/ssrn.4710039.
[22] C. Hansen, “Sequential Modelling with Applications to Music Recommendation, Fact-Checking, and Speed Reading,” Sep. 11, 2021, *arXiv*: arXiv:2109.06736. doi: 10.48550/arXiv.2109.06736.
[23] Y. Deldjoo, M. Schedl, and P. Knees, “Content-driven music recommendation: Evolution, state of the art, and challenges,” *Computer Science Review*, vol. 51, p. 100618, Feb. 2024, doi: 10.1016/j.cosrev.2024.100618.
[24] Ò. Celma Herrada, P. Herrera Boyer, and X. Serra, “Bridging the Music Semantic Gap,” 2006, Accessed: Nov. 29, 2024. [Online]. Available: http://repositori.upf.edu/handle/10230/34294
[25] S. Dai, H. Zhang, and R. B. Dannenberg, “Automatic Analysis and Influence of Hierarchical Structure on Melody, Rhythm and Harmony in Popular Music,” Oct. 15, 2020, *arXiv*: arXiv:2010.07518. Accessed: Oct. 10, 2024. [Online]. Available: http://arxiv.org/abs/2010.07518
[26] P. Magron and C. Fevotte, “Leveraging the Structure of Musical Preference in Content-Aware Music Recommendation,” in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada: IEEE, Jun. 2021, pp. 581–585. doi: 10.1109/ICASSP39728.2021.9414194.
[27] S. Oramas, O. Nieto, M. Sordo, and X. Serra, “A Deep Multimodal Approach for Cold-start Music Recommendation,” *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, pp. 32–37, Aug. 2017, doi: 10.1145/3125486.3125492.
[28] J. Lee, K. Lee, J. Park, J. Park, and J. Nam, “Deep Content-User Embedding Model for Music Recommendation,” *ArXiv*, Jul. 2018, Accessed: Dec. 03, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Deep-Content-User-Embedding-Model-for-Music-Lee-Lee/b9be81d52c9918b03dc909d129c056b4f9d83e09
[29] J. Pons, O. Nieto, M. Prockup, E. Schmidt, A. Ehmann, and X. Serra, “End-to-end learning for music audio tagging at scale,” Jun. 15, 2018, *arXiv*: arXiv:1711.02520. doi: 10.48550/arXiv.1711.02520.
[30] P. Magron and C. Févotte, “Neural content-aware collaborative filtering for cold-start music recommendation,” *Data Min Knowl Disc*, vol. 36, no. 5, pp. 1971–2005, Sep. 2022, doi: 10.1007/s10618-022-00859-8.
[31] R. Salganik, F. Diaz, and G. Farnadi, “Fairness Through Domain Awareness: Mitigating Popularity Bias for Music Discovery,” in *Advances in Information Retrieval*, N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, and I. Ounis, Eds., Cham: Springer Nature Switzerland, 2024, pp. 351–368. doi: 10.1007/978-3-031-56066-8_27.
[32] P. Sathyajit and P. Thakur, “Hybrid Music Recommendation System combining Neo4j GraphDB based and ML-based Approach,” in *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Feb. 2024, pp. 428–432. doi: 10.1109/ESIC60604.2024.10481578.
[33] “Personalized Music Hybrid Recommendation Algorithms Fusing Gene Features,” *Mathematical Problems in Engineering*, vol. 2022, pp. 1–8, Apr. 2022, doi: 10.1155/2022/9209022.
[34] J. Yi, Y. Zhu, J. Xie, and Z. Chen, “Cross-Modal Variational Auto-Encoder for Content-Based Micro-Video Background Music Recommendation,” *IEEE Trans. Multimedia*, vol. 25, pp. 515–528, 2023, doi: 10.1109/TMM.2021.3128254.
[35] T. Duricic, D. Kowald, E. Lacic, and E. Lex, “Beyond-Accuracy: A Review on Diversity, Serendipity and Fairness in Recommender Systems Based on Graph Neural Networks,” Oct. 03, 2023, *arXiv*: arXiv:2310.02294. doi: 10.48550/arXiv.2310.02294.
[36] S. McNee, J. Riedl, and J. Konstan, “Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems,” 2006. Accessed: Dec. 09, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Accurate-is-not-always-good:-How-Accuracy-Metrics-McNee-Riedl/07563ba21b3508309c389ed219af141dc1bc6d51
[37] “(PDF) Being accurate is not enough: How accuracy metrics have hurt recommender systems,” in *ResearchGate*, doi: 10.1145/1125451.1125659.
[38] “A Survey on Recommendation Methods Beyond Accuracy,” *IEICE Transactions on Information and Systems*, vol. 100, no. 12, pp. 2931–2944, Dec. 2017, doi: 10.1587/TRANSINF.2017EDR0003.