Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19505-0

Social music discovery: an ethical recommendation system based on friend's preferred songs

Marco Furini[1] · **Francesca Fragnelli**[2]

Received: 29 June 2023 / Revised: 10 May 2024 / Accepted: 26 May 2024

Abstract
Music recommendation systems have become ubiquitous in today's world, but they raise ethical concerns related to bias, discrimination, and lack of transparency. To address these issues, we propose a recommendation system that combines content-based and collaborative filtering approaches within three different recommendation algorithms. These algorithms create playlists that mimic the user's listening habits while identifying similar tracks within the listening histories of the user's friends. To evaluate the effectiveness of our system, we asked ten participants to rate a total of ninety playlists. The results showed high satisfaction among participants with the playlists generated by two of the proposed recommendation algorithms. Specifically, participants who preferred to stay within their musical comfort zone appreciated one specific recommendation algorithm, while those who were willing to explore new music tended appreciated the other recommendation algorithm. In summary, by leveraging the user's social connections, our proposed system provides a more transparent and ethical approach to music recommendations. It provides a personalized and enjoyable music discovery experience that considers the nuances of individual musical taste and preferences. These findings suggest the potential impact of our proposal in addressing ethical concerns and enhancing user satisfaction in music recommendation services.

Keywords Music recommendation algorithm · Ethical concerns · Social connections

1 Introduction

Discovering music through friends is a fascinating phenomenon that has been observed and studied by researchers in the field of science, human and social psychology [14, 18]. Friend-based music recommendations offer a unique and compelling approach, primarily because they are rooted in personal relationships and social connections. This contrasts with traditional methods that rely on impersonal data and metrics to suggest music [7, 17].

✉ Marco Furini
marco.furini@unimore.it

1 University of Modena and Reggio Emilia, Reggio Emilia, Italy

2 Reggio Emilia, Italy

Published online: 11 June 2024

Discovering music through friends can have a profound impact on our emotional well-being, as it can create a sense of social bonding and shared experiences [36]. Studies suggest that when people discover new artists or songs through their friends, they are more likely to associate that music with positive emotions and memories of the shared experience, leading to potential positive effects on mental health and well-being [31]. Consequently, the phenomenon of discovering music through friends represents a crucial area of investigation, offering insights into the complex interplay between social relationships and musical preferences [8]. However, with the advent of technology and music platforms, borrowing cassette tapes or CDs from friends has become a thing of the past. Music platforms now provide users with access to music anytime, anywhere, and on any internet-enabled device, serving as constant companions for daily activities. Leveraging artificial intelligence and data science algorithms, these platforms personalize the user experience, aiming to create a 'lean-back' experience where listeners can enjoy music without actively choosing it [32].

The shift from physical music collections to digital streaming platforms has not only changed the way we listen to music, but it has also revolutionized the way music is recommended to us. There are two main types of recommendation systems used by music platforms: content-based and collaborative filtering [9, 30]. Content-based filtering [10, 35] suggests music based on the characteristics of the songs, like genre, tempo, and instrumentation, as well as the user's past listening behavior. In contrast, collaborative filtering [1, 24] suggests music based on the behavior of similar users. This approach is based on the idea that users who have similar music preferences will enjoy similar songs. Therefore, the system recommends music that has been enjoyed by other users with comparable tastes.

While recommendation systems have become increasingly popular, they also give rise to numerous ethical concerns [21, 25, 34]. One significant concern is the potential for bias and the perpetuation of stereotypes and discrimination if these systems are trained on data that reflects existing social inequalities such as race or gender. This issue highlights the importance of ensuring fairness in the design and implementation of recommendation algorithms. Another worrisome aspect is the susceptibility of recommendation systems to manipulation by malicious actors with ulterior motives. These actors may exploit the algorithms to promote specific content or ideas, potentially leading to the spread of misinformation or the amplification of harmful ideologies. This poses a threat to the integrity and reliability of the recommendations provided to users. Moreover, the opacity of recommendation systems further complicates the ethical landscape.

Our idea is to address ethical concerns about music recommendation systems by going back to the basics of discovering new music through borrowing cassette tapes or CDs from friends. In this paper, we propose an innovative recommendation mechanism for streaming platforms that leverages a user's social connections to create a highly personalized and ordered playlist of songs that match the user's established musical preferences. This creates a unique and enjoyable music discovery experience that enhances user engagement with the platform. Our approach is a hybrid of content-based and collaborative filtering. We use analytical song features such as acousticness, loudness, and liveness to transform each song in the user's listening history into a point in a multidimensional space. We then use clustering techniques to understand the user's musical taste and behavior patterns during specific periods of time, such as Monday mornings or Saturday nights. We use this knowledge to search for similar tracks within the listening histories of the user's friends and, for each period, we propose three different playlists that closely mimic the user's listening habits, providing a highly personalized music discovery experience.

During the experimental phase, we asked ten participants (age 18-30) to rate the playlists created for them using our proposed recommendation system. Since a person's musical tastes and listening habits can vary from moment to moment (what one likes on Monday morning may be different from what is appreciated on Saturday night), the evaluation process considered three different times of day (9am, 6pm, and 9pm) and produced three different playlists for each time period and user (one playlist for each recommendation algorithm). We asked participants to rank the suggested playlists and provide reasons in a free-text box. The study found that one recommendation algorithm was strongly preferred by participants who prefer to stay within their musical comfort zone. Participants who were more open to exploring new music tended to prefer a different algorithm, while the third proposed algorithm was not well received.

These findings highlight the importance of considering users' musical preferences and openness to exploring new music when designing and implementing recommendation systems. In summary, this study contributes to the understanding of the complex world of music recommendations, emphasizing the significance of personalization and user preference in the design of ethical recommendation systems.

The remainder of the paper is organized as follows. In section 2 we present recent studies in the field of music recommendation systems; in section 3 we describe details of our proposal and a possible implementation is shown in section 4. Section 5 shows results of a user evaluation study. Discussions and limitation of the proposed idea are outlined in section 6, whereas conclusions are drawn in section 7.

2 Related work

Our proposed music recommendation mechanism takes into account the music listening habits of individuals who are socially closest to us, recognizing the potential value of incorporating personal information to enhance the effectiveness of recommendation algorithms. This approach has been supported by a growing body of research, which we will explore further. In particular, we will discuss recent studies that have leveraged personal information, such as emotions or activities, to improve music recommendation algorithms. Additionally, we will delve into the ethical considerations surrounding these algorithms, as they play a crucial role in shaping the responsible development and deployment of music recommendation systems.

2.1 Recommendation Algorithms that use personal information

Recommendation algorithms have revolutionized music consumption, providing users with personalized playlists. Recent studies suggest that incorporating personal information such as current activity and mood state can significantly improve the performance of these systems [11]. Several approaches have emerged that leverage personal information in music recommendation. For instance, *Lin et al.* [26] extracted a list of favorite artists from a user's Facebook profile and recommended songs based on both the novelty of the singers and their similarity to the user's preferences. Singers similarity was measured through the Last.fm API. *Moscato et al.* [29] proposed an approach that identifies a user's personality traits, moods, and emotions from their social behavior, embedding personality and mood in a content-based filtering system for improved accuracy. *Kim et al.* [22] developed a smartphone-based system that recognizes human activities and recommends music based on an ensemble of dynamic and sequence classification methods, using a deep residual bidirectional gated recurrent neural network for activity recognition and tempo-oriented classification for music indexing.

Alvarez et al. [2] proposed a context and emotion-aware system that recommends and plays Spotify songs. The system consists of a location-based mobile app, an emotional wearable, and a recommendation service that uses AI techniques and emotionally-annotated music to predict the next song. *Tahmasebi et al.* [38] proposed a hybrid recommendation method that uses user demographic data to alleviate cold start problems in recommender systems. The method enriches the neighborhood set of users with rating data, improving the accuracy of recommendations. *Andjelkovic et al.* [3] developed a music-artists recommender system based on mood-based filtering that allows users to explore a music collection by musical mood dimensions.

The preceding studies have established that the incorporation of personal information is a key factor in improving the performances of recommendation systems. Building on this body of research, our proposal seeks to further enhance recommendation systems by integrating personal data, such as the music preferences of a user's friends. By leveraging this information, we can offer more tailored and relevant recommendations to users, thereby enhancing their overall listening experience.

2.2 Ethical concerns towards recommendation algorithms

Discussions about bias, fairness, and transparency in the development of music recommendation systems are still in their early stages [37]. These algorithms raise ethical concerns, in part because it can be challenging to understand how recommendations are generated and what underlying models represent [15, 28]. Three main concerns have emerged. **Gender Fairness** has been investigated in several studies [12, 16, 27], revealing significant discrepancies in algorithm performance between male and female user groups. **Group Fairness** is another concern, as research has identified potential cultural bias problems where recommendation algorithms may systematically and unfairly discriminate against certain individuals or groups [19, 23]. **Unscrupulous Behavior** has also been observed. For instance, some studies have shown a bias towards popular songs over less known tracks, which can advantage certain artists or record labels [5, 13, 33].

The recommendation algorithm proposed in this paper is transparent because it is based on the user's listening history and that of their social contacts. The recommended songs are those listened to by the user's social contacts, thereby minimizing discrimination towards gender, groups, or favoritism towards popular songs. By incorporating personal information from the user's social network, the algorithm aims to provide more accurate and relevant recommendations while maintaining transparency and fairness. Users can understand how the recommendations are generated and have control over the information shared with the algorithm. This approach can help to address some of the ethical concerns associated with music recommendation systems, such as bias and lack of transparency.

3 Our proposal

The primary aim of this study is to develop an innovative recommendation system that effectively utilizes a user's listening history and social connections to generate personalized playlists that accurately reflect the user's music preferences. Additionally, the system aims to identify similar tracks within the listening histories of the user's friends, enhancing the personalized playlist creation process. To accomplish this objective, we have identified three main goals: (i) to leverage the user's social circle to identify new music options; (ii) to provide tailored music recommendations that closely align with the user's individual musical tastes; and (iii) to organize these recommendations in a sequential order that corresponds to the user's established listening habits.

The intended outcome of our proposed system is to offer a curated sequence of songs that are mathematically similar to the user's established preferences. Furthermore, we aim to curate playlists specifically tailored to different time periods, such as Monday or Tuesday afternoons, to facilitate the discovery of new music appropriate for those moments. This is crucial because musical preferences can vary depending on the time of day and the activities being undertaken. For instance, the type of music one prefers in the afternoon may differ from that enjoyed in the morning, or the music chosen for a Saturday morning workout might not align with preferences during a Tuesday afternoon work session. The playlists are also ordered, as this is crucial for effectively mimicking user listening habits. Unlike a random order, which lacks structure and coherence, an ordered playlist allows for personalized experiences tailored to individual preferences. For example, some users may prefer to start their listening sessions with low-beat songs to ease into the mood, while others may prefer high-beat songs to jump-start their energy. By arranging songs in a specific order, we can replicate these preferences and create playlists that resonate with users on a deeper level. This level of personalization is only achievable through an ordered list of songs that aligns with the user's listening habits and enhances their overall music enjoyment.

Four different steps (see, Fig. 1) are necessary to build our system.

- **Step 1**. Enrich and clean the user's listening history;
- **Step 2**. Produce track lists by splitting the user's listening history according to specific time periods;
- **Step 3**. For each track list, use clustering techniques to identify groups of songs with similar characteristics. Then, consider the largest cluster and identify the most representative track (the so-called, reference track);
- **Step 4**. For each reference track of Step 3, search through the listening histories of the user's friends to find the most similar song to recommend.

3.1 Step 1: enriching and cleaning the user's listening history

The user's listening history is typically composed of a collection of songs, each described by various elements including title, author, genre, length, play date, and time. Once the listening history is acquired, it becomes necessary to enrich it by incorporating analytical values for individual songs and filtering out the ones that were not favored by the user. The selection of analytical values to characterize each song in the listening history depends on the

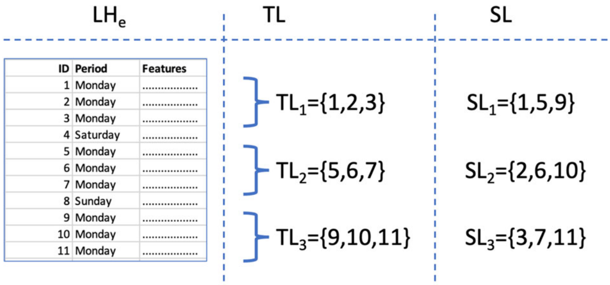

Fig. 1 The four steps necessary to build our ethical recommendation system

available data. These values encompass a range of factors, such as audio features (e.g., tempo, pitch, rhythm, harmony) and metadata (e.g., song popularity, release date, artist information), providing a comprehensive description of the songs. Similarly, the methodology employed to identify the songs that the user did not appreciate can vary based on the available data. Common approaches involve leveraging user feedback, such as ratings or thumbs up/down, or assessing whether the playback duration significantly deviates from the song's length. These techniques aid in distinguishing songs that may not align with the user's preferences and facilitate the refinement of the listening history.

3.2 Step 2: grouping songs played in a specific period

The enriched and filtered listening history (LH_e) produced at step 1 is further analyzed and transformed into N different track lists, each corresponding to a specific time period. Then, we select the i-th song from each track list to create a new song list SL_i. Specifically, SL_i contains only the i-th songs from each track list of the considered period, such that:

$$SL_i = TL_1(i), TL_2(i), TL_3(i), ..., TL_N(i). \tag{1}$$

If one of the track lists does not contain an i-th song, we exclude it from the song list creation process. For example, suppose we have four track lists, TL_1, TL_2, TL_3, and TL_4, and we want to create a playlist that includes the 3rd song from each history. However, let's say that TL_3 only has two songs, so there is no 3rd song to include. In this case, the new song list would only include the 3rd songs from TL_1, TL_2, and TL_4, and would not include any songs from TL_3.

At the end of this process, we have Z different song lists, where Z is the length of the longest listening history. Each of these lists represents a unique combination of songs that the user has listened to across different time periods and will be used to generate personalized music recommendations.

Figure 2 illustrates this process. Suppose our goal is to produce a playlist to be listened to on Monday. We transform the user's listening history, LH_e, into three track lists: TL_1 (songs played on the first available Monday in LH_e), TL_2 (songs played on the second available Monday in LH_e), and TL_3 (songs played on the third available Monday in LH_e). We then build the song lists as follows: SL_1 consists of the first song in each track list, SL_2 consists of the second song in each track list, and SL_3 consists of the third song in each track list. At the end of the process, SL_1 represents songs that the user has listened to as the first song on Mondays, SL_2 represents songs that the user has listened to as the second song on Mondays, and SL_3 represents songs that the user has listened to as the third song on Mondays (Table 1).

3.3 Step 3: understand the characteristics of the songs played in a specific period

After creating the song lists SL_1, SL_2, ..., SL_M, each list is analyzed to understand the analytical characteristics of the songs. This process involves clustering, which groups together songs with similar characteristics, where each characteristic corresponds to a different audio feature, such as tempo, pitch, or timbre. The distance between two songs is calculated based on the differences between their audio features. The number of these features depends on the available data and/or on the music streaming platforms.

To perform clustering, we might use algorithms such as k-means clustering or hierarchical clustering. The choice of clustering algorithm depends on several factors, including the size and complexity of the dataset, the number of clusters desired, and the specific characteristics

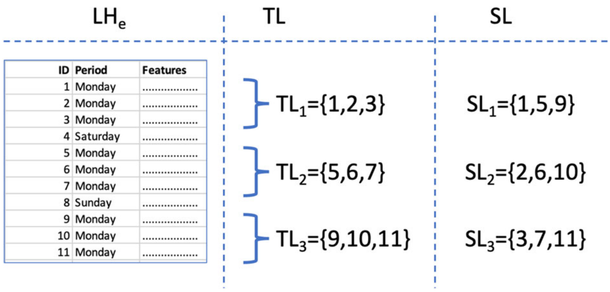

Fig. 2 Step2: From the user's listening history to track lists (songs played within a specific period) to Song lists (songs played in a specific position)

of the data. These algorithms work by iteratively grouping together data points that are close to each other in the feature space, until all of the data points have been assigned to a cluster.

Once the clustering process is complete, each song in the set SL_i is assigned to one of the K cluster, such as C_1^i, C_2^i, C_3^i, and so on. These clusters represent groups of songs that have similar audio features and can be used to identify patterns or trends in the music preferences of the listener.

The largest cluster, denoted as $\overline{C_m^i}$, summarizes the listening habits for the i-th song of the selected time period in quantitative terms. This cluster represents the set of songs that are most similar to the i-th song in terms of audio features and provides insight into the user's music preferences during that time period.

3.4 Step 4: playlist Production of the recommended songs

Each $\overline{C_*^i}$ produced in step 3 contains a reference song, called the reference track, which is used as a key to search for a similar track to recommend to the considered user (U1). Starting from the reference track, we search for a song that, within the listening histories of U1's friends, has the most similar quantitative characteristics in terms of audio features. This involves comparing the audio features of the reference track to those of other songs in the listening histories of U1's friends and selecting the song with the most similar features as the recommendation.

Table 1 Notation used in this study

Notation	Description
LH	Listening History
LH_e	Enriched and filtered listening history
TL_i	Track list of period i
SL_i	List of all the $i-th$ songs played in the specific periods
M	Length of the longest track list
K	Number of cluster
c_i^j	$j-th$ cluster of all the $i-th$ songs
$\bar{c}_k^i$	the largest cluster among the ones of the $i-th$ songs
$song_i$	i-th song recommended to user

Let $song_i$ be the song with the lowest distance. This song is recommended to U1 for the selected time period and for the i-th position in the playlist. To avoid including duplicate songs in the final recommendation list, the algorithm has been refined to suggest the second most similar song to the reference track if the first one is already present in the list of recommendations.

At the end of step 4, we have a sequence of K recommended songs: $song_1$, $song_2$, ..., $song_M$, that are ordered according to U1's musical habits and suggested based U1's friends. This personalized playlist provides U1 with songs that are mathematically similar to their established musical preferences, enhancing the music discovery experience within the streaming platform.

4 Proposal implementation

To evaluate our proposal, we utilize the listening history provided by Spotify, which is a JSON-formatted file that contains all the songs played within the past 365 days. Each song is described by various data points, such as those shown in Table 2, which are included in the listening history file.

It is important to note that our proposal is not limited to Spotify, and can be applied to any other music streaming service as the key requirement is the availability of the user's listening history, and the possibility to describe a song with features in a multi-dimensional space.

4.1 Step 1: enriching and cleaning the user's listening history

We enrich the user's listening history through the features listed in Table 3. Since Spotify API may not be used to train machine learning or AI model, we used different sources to get the additional song's features such as [6] and Kaggle.com.

To filter out tracks that the user did not enjoy, we utilized the duration of each song. Given the absence of direct feedback in the available listening histories, we made the assumption that if a song had been played for more than 30% of its total length, the user found it enjoyable. Conversely, if the playout time was less than 30%, we inferred that the user did not like the song and thus removed it from their listening history.

4.2 Step 2: grouping songs played in a specific period

Grouping songs played during a specific period is a straightforward process that doesn't require any particular implementation choice. Let's assume that the chosen period is Monday from 6:00pm to 10:00pm. Thanks to the timecode present in the 365-day listening history, it's possible to identify all the songs played on Mondays from 6:00pm to 10:00pm. All the first songs played on different Mondays (6:00pm-10:00pm) will form the list with the first songs;

Table 2 Listening History: Features that describe each played song

Name	Description
artistName	Name of the artist of the song
trackName	Title of the song
endTime	Day and time when the song playback ended
msPlayed	Duration in milliseconds of the song playback

Table 3 Additional features used to describe a song

Feature name	Description
Uri	Unique song identifier
Duration	Total song duration in milliseconds
Danceability (0.0 ... 1.0)	Describes how suitable a track is for dancing and is based on a combination of musical elements, including tempo, rhythm, stability, beat strength, and overall regularity. It goes to not danceable (0.0) to very danceable (1.0)
Energy (0.0 ... 1.0)	A perceptual measure of the intensity and activity of a track; typically, energetic tracks are fast, loud, and noisy
Key (-1 ... 11)	The musical key of the track
Loudness (-60dB ... 0dB)	The relative volume of the track expressed in decibels (dB)
Mode (0 ... 1)	The mode (major or minor) of a track, or the type of scale from which its melodic content is derived;
Speechiness (0.0 ... 1.0)	Presence of spoken words in a track; the more the track is vocal, the closer the attribute value is to the maximum value.
Acousticness (0.0 ... 1.0)	How likely a track is to be acoustic.
Instrumentalness (0.0 ... 1.0)	Predicts whether a track contains vocals; the closer the instrumental value is to 1.0, the greater the likelihood that the track does not contain vocal content.
Liveness (0.0 ... 1.0)	Detects the presence of an audience in the audio recording; higher values represent a greater likelihood that the track was performed live.
Valence (0.0 ... 1.0)	Describes the positive musical emotion conveyed by a track; tracks with high valence are more positive (e.g., happy, cheerful, euphoric), while tracks with low valence are more negative (e.g., sad, depressed, angry)
Tempo	The estimated overall tempo of a track measured in bpm; in music terminology, tempo is the speed or pace of a given track and is directly derived from the average duration of the beat.

all the second songs played will form the list with the second songs. The process continues until N different lists are created. Likely, since not all Mondays N songs have been played, the lists may have different sizes. In this implementation, to avoid clustering problems, we decide to not consider lists with less than 5 songs. Therefore, this step outputs M different lists, with $Z \leq N$.

4.3 Step 3: understand the characteristics of the songs played in a specific period

When performing clustering analysis, data normalization is a crucial first step. The goal of clustering algorithms is to group together data points that are more similar to each other than to those in other groups. However, different features of the data may have different scales or units of measurement, which can cause certain features to dominate the clustering process and others to be overlooked. For instance, as reported in Table 3, *loudness* and *tempo* might dominate all the other features. Normalizing the data involves transforming the features to a common scale, such as between 0 and 1, or standardizing them to have a mean of 0 and a standard deviation of 1.

Once the data has been normalized, we can move on to the clustering phase. For this project, we chose to use the K-means algorithm, which is a widely-used, simple, and high-performing clustering algorithm. Given the number of clusters as input, the algorithm works by calculating the centroid of each cluster. To avoid any bias in the calculation, we employed the silhouette algorithm to find the most accurate number of clusters. The silhouette algorithm measures how similar data points are to their own cluster compared to other clusters, and calculates a silhouette score for each data point, which ranges from -1 to 1. A score of 1 indicates that the data point is well-matched to its cluster, while a score of -1 indicates that it is more similar to a neighboring cluster. The average silhouette score is then calculated for each clustering configuration, and the number of clusters that produces the highest average score is considered to be the optimal number of clusters for the dataset. By using the silhouette algorithm, we were able to ensure that the resulting clusters are meaningful and accurate, and that the clustering analysis is optimized for our specific dataset.

The clustering process produced multiple clusters for every list of Step 2. To identify the most representative cluster for each list, we selected the largest cluster. As a result, this step generates a total of M largest clusters, which are the most meaningful and accurate groupings of data points based on our chosen clustering algorithm and number of clusters.

4.4 Step 4: playlist Production of the recommended songs

Within each largest cluster, it is possible to identify the representative song in different ways. Since there is no method better than another, in this implementation, we propose three heuristics (see, Fig. 3):

- **Centroid** as the reference track. The advantage of this method lies in the fact that the centroid is the point that best describes the characteristics of all the songs in the cluster itself; the disadvantage is that this point might not correspond to an actual song, as it is a fictional point created by the K-means algorithm.
- **Closest** track to its centroid as the reference track. The advantage in this case is that the reference song corresponds to a real track appreciated by the user; the disadvantage is that this is not an effective midpoint compared to all the other tracks in the cluster.
- **Random** track as the reference track. The advantage of this method is that it is a simple and easy-to-implement approach that does not require any additional calculations or processing; the disadvantage is that this approach might not be as precise or sophisticated as the other methods proposed.

Once the reference track is selected, we gather and enrich the listening histories of the user's friends to search for the song that is closest to the reference track. To compute the similarity, we measure the distance between the reference track and other songs in an 11-dimensional feature space, using the Euclidean distance metric. It's worth noting that while we utilized the Euclidean distance in our analysis, alternative distance metrics are viable options. For instance, in our experiments, we also explored the cosine distance; however, we found no significant differences compared to the distance metric employed by k-means. The song closest to the reference track becomes a candidate for recommendation. By iterating

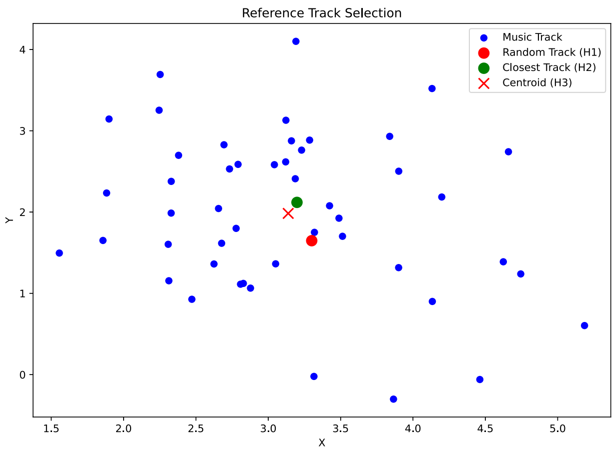

Fig. 3 Graphical explanation, on a two-dimensional space, of the three heuristics proposed to select the reference track of the largest cluster. H1 selects a random track; H2 selects the track closest to the cluster centroid; H3 selects the centroid even if does not represent a real music track

this process across all clusters, we generate an ordered playlist of M songs that the user can listen to within a specified period. The ordered playlist is tailored to the user's musical tastes and listening habits, as the selection of the i-th song in the playlist is based on the analysis of all the i-th songs found in different periods of the listening history. This process ensures that each song in the playlist is aligned with the user's preferences and listening patterns, enhancing the overall listening experience.

5 User evaluation

The algorithm was implemented as a fully functional and testable application using the Python programming language. To perform a real-world evaluation, we select participants for the evaluation process in such a way that each participant has at least one friend among the other participants. This allows us to simulate social connections and interactions, which are important factors in music listening behavior.

To evaluate the effectiveness of the algorithm with a reasonable sample size, we conducted a literature review and consulted relevant studies [4, 20] to determine an appropriate number of participants. Based on these findings, we selected a group of 10 users aged between 18 and 30 to assess the effectiveness of the algorithm. This age range was chosen to analyze and compare individuals who were not too distant, but potentially had divergent musical tastes due to generational differences. Participants were asked to download their Spotify listening histories and provide the necessary information to access the data. Thanks to the voluntary contributions of these users, a dataset with a total of 174,268 songs was obtained. At the end of the experiments, the 10 users who had provided their listening histories were recalled for a face-to-face interview aimed at evaluating the quality of the playlists generated by three different heuristics to select the most representative song of a cluster.

Ninety (90) different recommended playlists were generated for the evaluation (10 users, 3 different methods and 3 different periods):

– P1: 9:00am, to capture listening habits at the beginning of the day, during commuting to work or university;

- P2: 6:00pm, to capture listening habits at the end of both study and work activities;
- P3: 9:00pm, to capture listening habits during the evening.

Each user was asked to evaluate a total of nine playlists (3 for each period). A link where to find the playlists and the evaluation form was sent to them. To ensure that the evaluation process was not too time-consuming, the playlists were limited to 10 songs each.

The evaluation procedure followed by each user consisted of the following steps: (i) playlists listening, (ii) specify the preferred and the less preferred playlists, (iii) Explain the choice with a free text box. This free-text box was used to gather general qualitative feedback on assessing the algorithm's performance and identifying areas for improvement.

The analysis of the 10 interviews identified the preferred playlist among the three proposed. Based on the results, playlists based on heuristic H2 (track closest to the centroid) and H3 (centroid as virtual track) outperformed playlists produced with heuristic H1 (random track). Specifically, playlists H2 were the most preferred playlists, with 5 out of 10 users selecting it, followed by playlist H3, which was preferred by 4 out of 10 users. Only 1 out of 10 users chose playlist H1.

To have a more analytical understanding of the user's preferences, we assigned scores of 2, 1, and 0 to the first, second, and third preferences, respectively. Playlists H2 received a total score of 14, playlists H3 received a total score of 11, and playlists H1 received a total score of 5. The results clearly show that playlists H2 was the preferred playlist, while playlists H1 had the worst performance.

To gain further insight into the participants' preferences, we analyzed the free-text box. We found out that all users agreed that playlist H2 contained songs that were most similar to their musical tastes. However, those who selected playlist H3 were motivated by the mix of songs they already knew and liked, as well as new songs they could potentially enjoy in the future. For instance, participant #3 stated, *"Playlist H2 is more similar to me, but I already know all the songs, so it wouldn't allow me to discover new music, while playlist H3 is a good mix of songs that I already like and unknown songs that I would actually listen to"*.

Overall, participants who preferred to stay within a restricted domain of songs, artists, and genres, also known as the "musical comfort zone," tended to prefer playlist H2. On the other hand, those who enjoyed exploring new sounds while maintaining a musical connection with what they listen to, were more likely to prefer playlist H3.

6 Discussion, limitation and future improvement

The proposed approach has important theoretical and practical implications, which can be summarized as follows: (i) the experiment underscores the crucial role of using advanced recommendation algorithms, such as method H2 and H3, to provide personalized and accurate recommendations to users; (ii) the finding that the algorithm performs better when the reference track is closest to the centroid highlights the potential of clustering techniques in enhancing recommendation accuracy; (iii) the study sheds light on users' preferences for different types of music recommendation algorithms, which can inform the development of more effective algorithms in the future; (iv) the results suggest that users' preferences for specific music recommendation algorithms are influenced by their musical attitudes, such as their desire to stay within their musical comfort zone or explore new sounds.

Considering the significance of friendships in music listening behavior, music platforms could potentially introduce a new feature to enhance its recommendation algorithm or provide a new user experience. For instance, they could create a section specifically dedicated to "Recommendations based on your friends," which would allow users to discover music based on what their friends are listening to. By incorporating social factors into its recommendation algorithm, music platforms could potentially enhance the relevance and accuracy of its recommendations, while also providing users with a more personalized and engaging music streaming experience.

The obtained results has some limitations: (i) the participants were recruited from a specific population (adults between 18 and 30) and this could limit the generality of the results to other populations. (ii) The user's current mood, activity, and listening environment could have influenced the participants music preferences and the effectiveness of the algorithm.

As a future improvement, we plan to incorporate a threshold to limit the number of songs already known to the user. Currently, the mechanism checks for the presence of multiple songs, but not for songs already familiar to the user. The introduction of the threshold (for instance, [9] notices a reuse percentage ranging from 14% to 31%, depending on the analyzed platform) will ensure the discovery of a percentage of new music. The implementation is quite simple: by checking the user's listening history, the system may know whether the song is known or not; once the threshold is reached, the algorithm can not suggest songs already played, but only similar songs (as it already does to prevent song duplication).

7 Conclusions

In this paper, we have proposed a recommendation system that leverages social connections to provide highly personalized music recommendations. Our system combines content-based and collaborative filtering approaches to create a playlist that closely mimics the user's listening habits, while also identifying similar tracks within the listening histories of the user's friends. The effectiveness of our proposed system was tested through an experimental phase involving ten participants. Results showed that our proposed system offers a more transparent and ethical approach to music recommendations, offering a personalized and enjoyable music discovery experience that takes into account the nuances of individual musical taste and preferences.

Funding Open access funding provided by Università degli Studi di Modena e Reggio Emilia within the CRUI-CARE Agreement.

Data Availability The data supporting the findings of this study were collected from various sources, including public databases/services, as well as personal and private communications with participants.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Aggarwal CC, Aggarwal CC (2016) Neighborhood-based collaborative filtering. Recommender Systems: The Textbook pp 29–70
2. Álvarez P, Zarazaga-Soria FJ, Baldassarri S (2020) Mobile music recommendations for runners based on location and emotions: the dj-running system. Pervasive Mob Comput 67:101242
3. Andjelkovic I, Parra D, O'Donovan J (2019) Moodplay: interactive music recommendation based on artists' mood similarity. Int J Human-Comput Stud 121:142–159
4. de Assunção WG, Zaina LAM (2022) Evaluating user experience in music discovery on deezer and spotify. In: Proceedings of the 21st brazilian symposium on human factors in computing systems, pp 1–11
5. Bauer C, Schedl M (2019) Global and country-specific mainstreaminess measures: definitions, analysis, and usage for improving personalized music recommendation systems. PLoS ONE 14(6):e0217389
6. Bertin-Mahieux T, Ellis DP, Whitman B, Lamere P (2011) The million song dataset. In: Proceedings of the 12th international conference on music information retrieval (ISMIR 2011)
7. Boer D, Abubakar A (2014) Music listening in families and peer groups: benefits for young people's social cohesion and emotional well-being across four cultures. Front Psychol 5:392
8. Boer D, Fischer R, Strack M, Bond MH, Lo E, Lam J (2011) How shared preferences in music create bonds between people: Values as the missing link. Pers Soc Psychol Bull 37(9):1159–1171
9. Bonnin G, Jannach D (2014) Automated generation of music playlists: survey and experiments. ACM Computing Surveys (CSUR) 47(2):1–35
10. Deldjoo Y, Schedl M, Cremonesi P, Pasi G (2020) Recommender systems leveraging multimedia content. ACM Computing Surveys (CSUR) 53(5):1–38
11. Dhelim S, Aung N, Bouras MA, Ning H, Cambria E (2022) A survey on personality-aware recommendation systems. Artif Intell Rev:1–46
12. Epps-Darling A, Cramer H, Bouyer RT (2020) Artist gender representation in music streaming. In: ISMIR, pp 248–254
13. Eriksson M, Fleischer R, Johansson A, Snickars P, Vonderau P (2019) Spotify teardown: Inside the black box of streaming music. Mit Press
14. Farnsworth PR (1958) The social psychology of music
15. Ferraro A (2019) Music cold-start and long-tail recommendation: Bias in deep representations. In: Proceedings of the 13th ACM conference on recommender systems, RecSys '19, p 586-590. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3298689.3347052
16. Ferraro A, Serra X, Bauer C (2021) Break the loop: gender imbalance in music recommenders. In: Proceedings of the 2021 conference on human information interaction and retrieval, CHIIR '21, p 249-254. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3406522.3446033
17. Furini M, Montangero M (2023) Understanding users music listening habits for time and activity sensitive customized playlists. In: 2023 IEEE 20th consumer communications & networking conference (CCNC), IEEE, pp 485–488
18. Hagen AN, Lüders M (2017) Social streaming? navigating music as personal and social. Convergence 23(6):643–659
19. Holzapfel A, Sturm B, Coeckelbergh M (2018) Ethical dimensions of music information retrieval technology. Trans Int Soc Music Inform Retrieval 1(1):44–55
20. Hosey C, Vujović L, St. Thomas B, Garcia-Gathright J, Thom J (2019) Just give me what i want: How people use and evaluate music search. In: Proceedings of the 2019 chi conference on human factors in computing systems, pp 1–12
21. Jesse M, Jannach D (2021) Digital nudging with recommender systems: survey and future directions. Comput Human Behav Rep 3:100052
22. Kim HG, Kim GY, Kim JY (2019) Music recommendation system using human activity recognition from accelerometer data. IEEE Trans Consum Electron 65(3):349–358
23. Kleinberg J, Mullainathan S, Raghavan M (2016) Inherent trade-offs in the fair determination of risk scores. arXiv:1609.05807
24. Koren Y, Rendle S, Bell R (2021) Advances in collaborative filtering. Recommender systems handbook, pp 91–142
25. Lin K, Sonboli N, Mobasher B, Burke R (2019) Crank up the volume: preference bias amplification in collaborative recommendation. arXiv:1909.06362
26. Lin N, Tsai PC, Chen YA, Chen HH (2014) Music recommendation based on artist novelty and similarity. In: 2014 IEEE 16th international workshop on multimedia signal processing (MMSP), IEEE, pp 1–6
27. Melchiorre AB, Rekabsaz N, Parada-Cabaleiro E, Brandl S, Lesota O, Schedl M (2021) Investigating gender fairness of recommendation algorithms in the music domain. Inform Process Manag 58(5):102666

28. Melchiorre, AB, Zangerle E, Schedl M (2020) Personality bias of music recommendation algorithms. In: Proceedings of the 14th ACM conference on recommender systems, RecSys ’20, p 533-538. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3383313.3412223
29. Moscato V, Picariello A, Sperli G (2020) An emotional recommender system for music. IEEE Intell Syst 36(5):57–68
30. Paul D, Kundu S (2020) A survey of music recommendation systems with a proposed music recommendation system. In: Emerging technology in modelling and graphics: proceedings of IEM Graph 2018, Springer, pp 279–285
31. Rabinowitch TC, Cross I, Burnard P (2013) Long-term musical group interaction has a positive influence on empathy in children. Psychol Music 41(4):484–498
32. Schedl M, Knees P, McFee B, Bogdanov D (2021) Music recommendation systems: techniques, use cases, and challenges. In: Recommender systems handbook, Springer, pp 927–971
33. Schedl M, Zamani H, Chen CW, Deldjoo Y, Elahi M (2018) Current challenges and visions in music recommender systems research. Int J Multimed Inform Retrieval 7:95–116
34. Seaver N (2019) Captivating algorithms: Recommender systems as traps. J Mater Cult 24(4):421–436
35. Shahbazi Z, Byun YC (2019) Product recommendation based on content-based filtering using xgboost classifier. Int J Adv Sci Technol 29:6979–6988
36. Steinbeis N, Koelsch S (2009) Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. Cereb Cortex 19(3):619–623
37. Sturm BL, Iglesias M, Ben-Tal O, Miron M, Gómez E (2019) Artificial intelligence and music: open questions of copyright law and engineering praxis. In: Arts, vol 8, MDPI, p 115
38. Tahmasebi F, Meghdadi M, Ahmadian S, Valiallahi K (2021) A hybrid recommendation system based on profile expansion technique to alleviate cold start problem. Multimed Tool Appl 80:2339–2354