Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Socially-Motivated Music Recommendation

Ben Lacker*, **Samuel F. Way**

Spotify
benl@spotify.com

Abstract

Extensive literature spanning psychology, sociology, and musicology has sought to understand the motivations for why people listen to music, including both individually and socially motivated reasons. Music's social functions, while present throughout the world, may be particularly important in collectivist societies, but music recommender systems generally target individualistic functions of music listening. In this study, we explore how a recommender system focused on social motivations for music listening might work by addressing a particular motivation: the desire to listen to music that is trending in one's community. We frame a recommendation task suited to this desire and propose a corresponding evaluation metric to address the timeliness of recommendations. Using listening data from Spotify, we construct a simple, heuristic-based approach to introduce and explore this recommendation task. Analyzing the effectiveness of this approach, we discuss what we believe is an overlooked trade-off between the precision and timeliness of recommendations, as well as considerations for modeling users' musical communities. Finally, we highlight key cultural differences in the effectiveness of this approach, underscoring the importance of incorporating a diverse cultural perspective in the development and evaluation of recommender systems.

Introduction

Music is commonly understood to be a universal human activity, appearing across all known societies (Mehr et al. 2019). This universality encompasses a wide range of musical behaviors, servicing a variety of functions. The act of listening to music itself has been the focus of a vast literature spanning psychology, neuroscience, philosophy, sociology, musicology, anthropology, and even computer science, all sharing a common emphasis on the reasons why people listen to music and how they develop the preferences that they do. While these motivations have been investigated and modeled in heterogeneous ways, existing research points to a few common underlying functions for music listening: self-awareness, arousal and mood regulation, and social relatedness (Schäfer et al. 2013). This is to say, music's enjoyment is both a personal and social endeavor.

The social reasons behind music listening have been extensively studied. Indeed, the field of sociomusicology devotes itself to an understanding of the social dimensions of human musical behavior (Dowd 2007). Within this broader category, a multitude of specific social functions of music can be articulated, such as demonstrating belonging to a social group, feeling connected to friends, and learning about one's social environment (Schäfer et al. 2013). These social functions are often studied in the context of shared musical experiences, examples of which include attending a concert together or listening to music with family (Boer and Abubakar 2014). However, studies have also shown that social benefits can arise from listening to music by oneself, an idea known as "social surrogacy" (Hargreaves and North 1999; Groarke and Hogan 2016). Schäfer and Eerola found evidence for a number of different mechanisms through which individual music listening fulfills social needs, including providing a feeling of company (i.e., "music makes me feel less lonely"), reminiscence, and shared experiences (Schäfer and Eerola 2020). It is the latter phenomenon that we wish to advance in this work: the way individual music listening can contribute to shared social experiences and, correspondingly, an opportunity for recommender systems to assist individuals in this function.

Despite music's clear social value, algorithmic systems have tended to adopt particularly individualistic approaches to its recommendation. This focus arises at least in part from algorithmic systems' role in *personalizing* end-user experiences tied to individual accounts, which develop increasingly rich characterizations of that individual's musical tastes and serve recommendations based upon them. What's more, while the techniques that facilitate algorithmic recommendations often draw inspiration from social processes, they seem to ignore important elements that impart social value. Collaborative filtering, for example, serves to identify music that a person may enjoy because other persons with similar musical tastes also seem to enjoy it. However, this common arrangement makes two perhaps very individualistic assumptions: (1) that the person receiving the recommendation needn't care *who* those other people are and (2) that it needn't matter *when* those other people enjoy it, and, in particular, if they happen to currently be enjoying it.

But existing literature suggests that these factors may be especially important for recommendations to service music's social functions. Specifically, studies show that social proof (Cialdini and Goldstein 2004; Amblee and Bui 2011) can greatly alter individual consumer decisions and, collectively, what becomes popular in cultural markets like music (Salganik, Dodds, and Watts 2006). In addition, these factors may be especially important in collectivist societies (Boer and Fischer 2012) where cohesion and greater alignment with the needs and desires of the collective may encourage music listening to be particularly social. This importance may not be unique to collectivist settings, however, as other studies looking at cross-cultural differences emphasize that music's social dimensions seem to be universally important (Saarikallio et al. 2021), meaning that systems designed to service these needs should have broad impact, if especially so in collectivist societies.

Drawing upon substantial research evidencing the value of shared musical preferences for social bonding and friendship formation in offline environments (Knobloch, Vorderer, and Zillmann 2000; Lonsdale and North 2009; Boer et al. 2011; Soley and Spelke 2016; Selfhout et al. 2009), as well as the aforementioned "social surrogacy" functions of listening, we propose that, in the context of individual listening, one expression of the desire for shared experience is the choice to deliberately listen to music that is popular within one's community. This type of listening (as distinct from individually-motivated listening aimed at regulating mood or achieving self-awareness) allows individuals to lay the groundwork for social bonding through shared musical familiarity. But, crucially, algorithmic recommender systems have historically overlooked this function. Accordingly, our work here focuses on this specific musical goal and explores how a music recommender system might help someone achieve it. To do so, we organize our study around the following research questions:

- RQ1: How might we recommend music that fulfills the social needs connected with individual music listening?
- RQ2: How might we evaluate the success of such a recommender?
- RQ3: How does the effectiveness of this approach vary across cultures?

Our study is arranged as follows: we frame a new recommendation task explicitly designed to fulfill the desire for shared social listening experiences. We begin with a discussion of related works from relevant literature. We then introduce a data set collected from a large-scale commercial music streaming service and outline a simple, heuristic approach to this recommendation task. We discuss important considerations for this example system's evaluation and introduce suitable evaluation metrics to emphasize social utility. Finally, we discuss our findings, making cross-cultural comparisons, and conclude with some reflections and suggestions to invite future research.

Related Work

While many algorithmic recommender systems tend to target individually-motivated goals, past studies have explored two broad themes for socially-motivated recommendation: group recommendations and community-based techniques for generating recommendations.

Group recommendation, as described in Masthoff's 2011 overview (Masthoff 2011), addresses the problem of recommending items to a group of users who will consume the items together. These approaches are useful in a variety of social scenarios, such as a fitness center that seeks to play music that satisfies the taste of all those present. Group recommendation can also be useful when recommending to an individual, for example by aggregating the tastes of virtual group members along with those of the individual, as in the case of a parent who would like their child to watch TV programs that include educational content that matches the parent's preferences. While these individual applications represent a small niche within group recommendations, they do resemble the social surrogacy scenario that constitutes the focus of our work.

Community detection, a well-established research area within network science (Fortunato and Hric 2016), has been applied to socially-minded recommender systems in numerous ways (Gasparetti, Sansonetti, and Micarelli 2021). In one example, certain users are dynamically promoted to "experts", and the behavior of these experts is then leveraged to recommend novel items (Lee and Lee 2014). Another approach detects user communities based on a graph of co-interest in movies, then uses this graph as the basis for recommendations (Fatemi and Tokarchuk 2013). Another study expounds on the difference between "social recommendation" and "trust-based recommendation", both of which require explicit data on users' social connections (Ma et al. 2011).

Generalizing beyond these two themes, existing work on group recommendations and community-based recommendations highlight the variety of ways that social connections can inform both recommendation problems and solutions. The idea of aggregating the taste of virtual group members, as well as the idea that social trust is an important criterion for recommendation informs this work.

As a limitation to this space, in recent years, researchers have pointed out assumptions that have tended to focus recommender systems research to a narrow conception of individual user-item relevance. A survey on multi-stakeholder recommendation points out these assumptions and where they break down. Within this framework, the authors mention group recommendations and success metrics besides relevance (Abdollahpouri et al. 2020). Another position paper argues that the convention of treating recommendation as a relevance-prediction problem has hobbled the field; the authors argue for alternative approaches to evaluation of recommender systems, including "beyond-accuracy metrics" and user research (Jannach and Bauer 2020). A more recent survey extends these ideas further, pointing out the numerous stakeholders, objectives, and time-scales that recommender systems must serve (Jannach 2022). In this work, we articulate a specific recommendation objective and evaluation metric that challenge the classic framing around user-item relevance, one that may hold special significance within cultures that tend to be less well represented in research in the general sense (Henrich, Heine, and Norenzayan 2010).

Cultures are commonly contrasted according to their degree of individualistic versus collectivistic tendencies. While the underlying ideas stretch back centuries, the contemporary practice of quantifying this cultural dimension for systematic analysis often builds upon Geert Hofstede's 1984 book *Culture's Consequences* (Hofstede 1984), which provided a common (however criticized (Baskerville 2003)) framework to quantify several cultural dimensions, including individualism. Harry Triandis's book *Individualism and Collectivism* explored these specific aspects of culture in greater depth (Triandis 2018), and a more recent review of the vast literature on these concepts suggests that the core elements of individualism are independence and uniqueness, and that those of collectivism are duty to in-group and maintaining harmony (Oyserman, Coon, and Kemmelmeier 2002). Another review proposes distinguishing between relational collectivism, based on groups of individuals with personal relationships, and group collectivism, based on common membership in an abstract group (Brewer and Chen 2007).

The ways people articulate and fulfill their common, basic needs vary between collectivist and individualist cultures (Itai et al. 2008). Empirical studies suggest that this variation extends to the way people express needs related to music listening. A body of work by Boer found that people in collectivist cultures report using music for social diversion more frequently than do those in individualist cultures (Boer and Fischer 2012); that the emphasis placed on the social functions of music differed between cultures (Boer et al. 2012); and that while shared music rituals contribute to group cohesion across cultures, "musical family rituals affect emotional well-being particularly in more traditional/collectivistic contexts" (Boer and Abubakar 2014). More recent work on motivations, emotions, and mechanisms related to music listening found differences in musical emotions between individualist and collectivist cultures; this study also suggests differences in motivations for music listening may be country-specific (Juslin et al. 2016).

In addition to differences in the motivations for music listening, there are cultural differences in the ways people pursue these motivations. Drawing on data from before the era of streaming music, one study found increased radio ownership per capita in more individualist countries, noting that while "in collectivist cultures one radio per family is enough, in individualist cultures everyone wants his or her own radio" (Mooij 2003). Several studies of music streaming data identify specific manifestations of these cultural differences. Notably, the listening of people in collectivist cultures coalesces around a smaller number of artists, gravitates towards more mainstream music, and exhibits less diversity, compared to that of people in individualist cultures (Ferwerda et al. 2016; Ferwerda and Schedl 2016). This is consistent with the greater emphasis on harmony and reduced emphasis on uniqueness that characterize collectivist cultures. Other studies have found that a country's individualism is connected to openness to watching music videos from other cultures (Park et al. 2017), diversity of musical taste (Park et al. 2015), and the appeal of K-Pop (Baek 2015).

The studies above articulate a number of cultural differences in music listening motivations and behaviors. We argue that these differences imply culturally conditioned expectations of music recommender systems, and that for a music recommender to fully satisfy its users, it should provide for the functions and listening behaviors that are most salient in each user's culture. Yet, as a summary of the state of the art of music recommender systems points out, "little effort has been made to analyze cultural differences and patterns of music consumption behavior, which is, as we believe, a crucial step to build culture-aware [music recommender systems]." (Schedl et al. 2018)

A handful of studies have addressed the role of cultural differences in music recommender systems specifically. A cross-cultural study of user trust found that the cultural dimension of masculinity was strongly correlated with users' preferences for certain ways of presenting and explaining movie recommendations (Berkovsky et al. 2018). Other existing research consists of a study that analyzes how cultural differences affect evaluation of recommendation interfaces (Chen and Pu 2014), a study that examines preferences for items from different cultures by users in a single country (Tang, Winoto, and Ye 2011), a social-theoretical discussion of attitudes towards adoption of mobile recommender systems (Choi et al. 2014), and a series of experiments on how values of individualism and collectivism relate to users' preferences for different methods of contextualizing product recommendations (Tian 2020). Bauer and Schedl offer what we believe to be the only existing study of considering cultural differences in determining the content (as opposed to the presentation) of music recommendation, finding that incorporating country-specific "mainstreamness" attributes improves rating prediction accuracy (Bauer and Schedl 2019).

Broader Impact and Ethical Considerations

This work introduces a recommendation problem motivated by social functions of music listening and proposes a suitable evaluation metric. To illustrate these points, we describe a method for recommending music that is trending within a community of users. Here we discuss the potential impact of this work, were it to be implemented in a real-world context.

The benefits of such a socially-motivated recommendation system lie primarily in its ability to fulfill a neglected musical need that is particularly salient in collectivist cultures. As such, this work offers the possibility of counteracting a historical bias in the field of recommender systems towards individualistic behaviors. The risks of this approach include: accentuating popularity bias, by drawing further attention to music that is already popular; and stereotyping users, by incorrectly modeling the communities to which they belong.

We see great potential for this work to motivate further research into cultural differences around music recommendations. In addition to the specific social function of music listening treated in this paper, many other musical behaviors vary between cultures; future work should investigate these variations and how they might inform music recommendations. We propose to mitigate the risk of popularity bias by emphasizing that the approach described here should function as part of a broader, multi-faceted recommendation system; it is designed to complement, rather than replace, more traditional recommendations that optimize for metrics like relevance, novelty, and serendipity. In order to reduce the risk of incorrectly assigning users to communities, future research should focus on developing and regularizing more personalized approaches to modeling musical community.

Recommendation Task

How might a recommender system help a user accomplish the task of finding the music that is currently gaining popularity among their friends, family, or community? While standard collaborative filtering approaches (see (Koren, Rendle, and Bell 2021)) certainly offer a solution, their ability to meet this specific social need depends on the similarity of a user's musical preferences and those of other community members. We argue, as above, that the desire to listen to music that's popular within one's community includes cases when the music does not necessarily align with that user's individual tastes and, also, that servicing this desire introduces a dimension of timeliness to the recommendation task.

If we assume that people will learn about this music *eventually* through real-life interactions, what role can a recommender system play in meeting this need? We assert that recommendations should help users discover this music sooner. Unlike conventional recommender systems, the goal for this task is not to change *what* music people listen to; a recommender system designed to meet this social need should help accelerate *when* people become aware of music that is trending within their communities.

We also assume that this social need is particularly relevant when it comes to newly-released music. In the context of online media, "trending" refers to concentrated, ephemeral interest in a particular topic. When it comes to music, the most common subjects of these short bursts of interest are new releases, with occasional exceptions (e.g., the 2022 popularity of Kate Bush's 1985 song "Running Up That Hill"). For this reason, in this paper we limit our study to newly-released music, under the assumption that the social dynamics of interest will be most salient here.

Data

In order to recommend music that's trending within a user's community, two basic components are necessary: a model of a user communities, and a method of determining what music is trending within those communities. In this section we describe a data set constructed to study these tasks, consisting of records of the consumption of newly-released music on Spotify, a large commercial music streaming service.

Users

Our data set comprises the listening records of a sample of 100,000 users, with 5000 users sampled from each of 20 countries: Argentina, Austria, Brazil, Chile, Colombia, Egypt, Germany, India, Indonesia, Japan, Mexico, New Zealand, Portugal, Saudi Arabia, South Africa, Spain, Taiwan, Thailand, Turkey, and the USA. We consider only those countries with a total user population larger than the desired sample size of 5000. This set of countries was selected to more generally resemble the global distribution of individualism and collectivism (based on Hofstede's most recent data on these countries' cultural indices (Hofstede, Hofstede, and Minkov 2010)) than would a uniform random sample of users on Spotify, a service that has operated the longest in what are generally considered to be more individualistic countries. Further, we included only users who had been using Spotify for at least one year, as new users of the service experience different recommendations and exhibit different behaviors from longer-tenured users (Way, Garcia-Gathright, and Cramer 2020). Finally, only users who had listened to at least one newly-released song during the sample period were included; this ruled out 30% of users in the sample countries.

This sample of users differs from Spotify's overall user base in important ways. Notably, our sampled users have all listened to a newly-released song during a specific period, which shapes the demographics of the studied population. In particular, compared to the overall Spotify user population, users in our sample are slightly younger: our sample has a mean age of 30.3, while the overall population has a mean age of 30.9. Our sample users also tend to be more active Spotify users: the median number of song streams per week for users in our sample is 114, while for the overall user population it is 66.

While these differences suggest our sample will not generalize exactly to the overall population, we assert that this is desirable. The proposed recommendation task targets people who wish to engage with music that is trending within their communities and do so specifically on Spotify. Further, a bias towards a younger population aligns well with the observation that social listening holds special value among young people (Selfhout et al. 2009).

Songs

We consider sample users' consumption of music over the months of June and July in 2022, focusing on songs released during the same period. Users in the sample streamed a total of 142,241 unique songs that were released during the sample period. Subsequent analyses in this section draw upon all these songs, however, 48% of songs were streamed by only a single user in the sample, and the experiments described in the following section discard all the latter songs.

As seen in Figure 1, compared to the music streamed by the entire population of the sampled countries, the songs in our sample under-index on the genres favored by large countries (e.g. rock, which is popular in the U.S., and desi pop, which is popular in India) and over-index on other genres that tend to be popular in a greater number of countries (e.g. trap latino, reggaeton).

How does the popularity of a new song spread within a community in the days after its release? We consider a user's listening to a song as an implicit indication of awareness. While we acknowledge cases where a user may listen to a song without paying meaningful attention to it (e.g. a playlist put on in the background), modeling the extent of these cases is outside the scope of this work. In order to gain intuition about the dynamics of community popularity of newly released songs, we examine a single simplified example of a

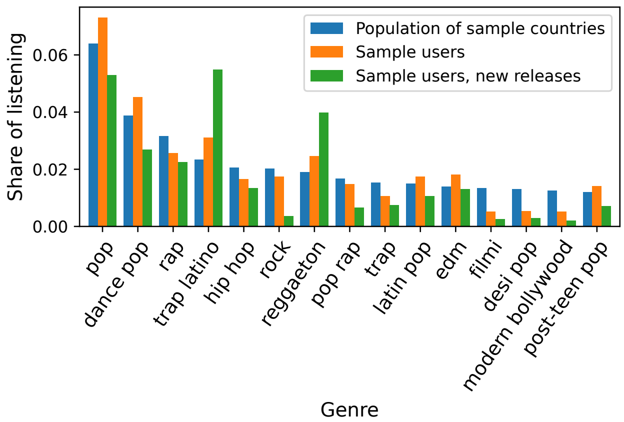

Figure 1: Share of listening by genre among (1) all listeners in the sampled countries, (2) all sampled listeners, and (3) newly released music streamed by sampled listeners. The differences in genre shares between these three groups illustrate the properties of our sample data.

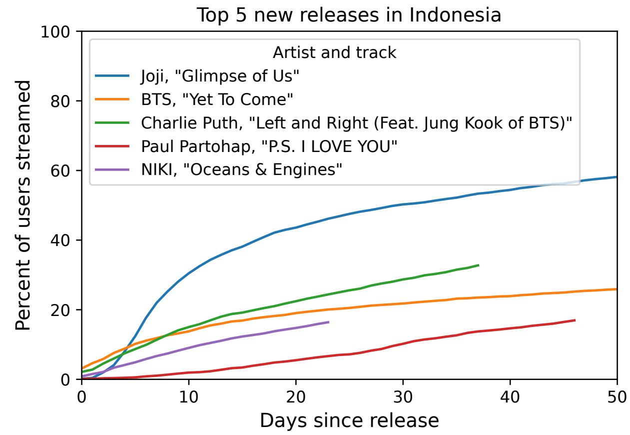

Figure 2: Spread of new song awareness in Indonesia. Each line represents one of the five most popular songs released during the sample period in Indonesia. The y-axis represents the percent of sample users in Indonesia who had listened to the song by the day since its release, given by the x-axis. Some songs gain popularity quickly, while others take longer but ultimately achieve greater popularity. Many users do not show awareness of popular songs until over a month after their release.

community: users in our sample who reside in Indonesia. In Figure 2 we see the percentage of Indonesian users in our sample who have listened to a song at least once by day x after its release. The gradual spread of awareness, as well as the slightly different paths taken by different songs is apparent. The most popular song depicted in this figure, for example, had been streamed by over half the sample users in Indonesia by the end of the sample period, while most other songs had been streamed by less than 20%.

The spread of song awareness varies among countries in our sample. Analyzing the average, over the top ten songs in each country in our sample, of the percentage of users who have listened to a song at least once by day x after its release, in Figure 3, we can see that in some countries the values are much higher than in others. This suggests that users in different countries display different levels of cohesion in terms of musical behavior. Some countries evince more homogeneous listening, while others seems to spread their listening out among a broader set of songs. We'll keep these differences in mind as we explore different ways of modeling groups of users.

These analyses help illustrate the opportunity for a socially-motivated approach to recommendation. Looking at the spread of awareness of new songs, it is apparent that some users' awareness lags weeks or even months behind the earliest listeners. For example, for the top song in Indonesia from our sample, one week after its release, only half the users in this group who would eventually listen to it had done so. We see an opportunity for a recommendation system to help the remaining half of these users, users who eventually show interest in this song, but who take longer to discover it on their own. How might we approach making recommendations to help these users?

Methods

We seek to accelerate people's awareness of music that is trending within their communities. There are many ways to approach this recommendation task. In this paper we don't seek to explore all the possibilities or to determine the best. Further, we acknowledge the possibility that, given time, a more traditional recommender system based on collaborative filtering could meet this need to some extent (see Discussion for more details). Such an exhaustive comparison of approaches is outside the scope of this work. Our goal is to call attention to a form of recommendation task, to introduce a new evaluation metric suited to this task, and to inspire additional research in these directions. We present a simple heuristic approach to making socially-motivated recommendations, with the aim of illustrating this task and the associated evaluation metric.

We approach this recommendation task by introducing the following components. First, we introduce a method for identifying and recommending newly-released music that's trending within a community. Then we introduce a way to model user communities based on demographic and listening data. Next, we describe two evaluation metrics that capture the important elements of our task. Finally, we use these components to run a series of offline experiments, comparing the effectiveness of this approach under different parameter values.

Making Recommendations

Within a community of users, how might we identify trending, new music for the purpose of recommendation? Following the aforementioned idea of the gradual spread of song

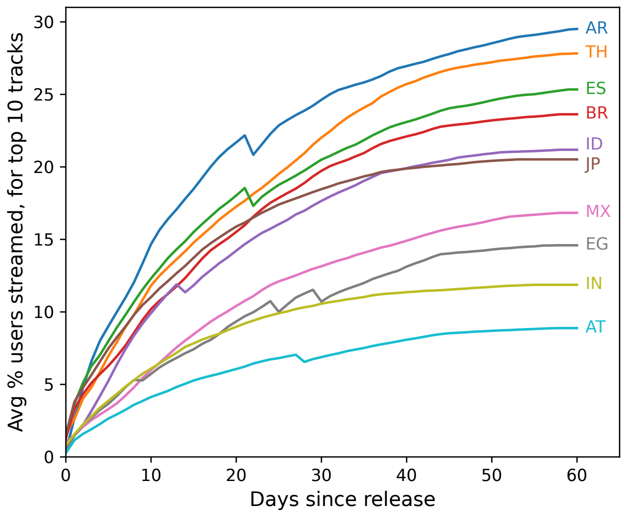

Figure 3: Average new song awareness of top 10 songs by country. Each line represents a single country from the sample data. The y-axis represents the percent of users in the country who have streamed a song by the day since the song's release given by the x-axis, averaged across the ten most popular songs in the country released during the sample period. The ten countries shown here were chosen at random from the twenty in the sample data. The share of users that listen to the most popular new music is considerably higher in some countries than in others.

awareness, we focus on the fraction of the group's users that has listened to the song as an indicator of the song's popularity within that community.

We implement a heuristic approach to making recommendations based on this quantity. We choose a number of days after a song's release, which we call $days\_waited$, at which to evaluate a song's popularity within a community. We choose a minimum fraction of a community's members that must have listened to a song for it to be recommendable, which we call $frac\_users\_required$. There is an intuitive relationship between the two recommendation parameters: the higher $days\_waited$, the higher $frac\_users\_required$ must be to achieve a comparable level of sensitivity in identifying trending songs. We therefore report $frac\_users\_required$ as the percentage of users required per day. Given a group of users and a newly released song, if the fraction of the group's users that have listened to the song by $days\_waited$ after the song's release exceeds $frac\_users\_required \cdot days\_waited$, we recommend the song to all users in the group who have not yet listened to the song.

Modeling Communities

Next, we seek to model user communities that are relevant to the recommendation task, which is to say those communities where members care about each others' musical preferences. Prior research suggests these groups may include users' families, friends, or classmates (Soley and Spelke 2016; Boer et al. 2012; Boer and Abubakar 2014). Such groups could be modeled either with explicit information on users' social ties or with implicit connections derived from listening behavior and other attributes. For the sake of simplicity and generalization, we focus on the latter.

We consider a variety of user attributes when modeling communities, which can broadly be categorized as demographic data and listening data. Demographic data includes a user's self-reported age and gender, the country in which they've registered their Spotify account, and the length of their tenure as Spotify users. We divide numerical attributes, such as age and Spotify tenure, into discrete buckets. Listening data includes the language in which the user has listened to the most music (for vocal music), the platform (e.g. mobile, desktop, TV) on which the user listens most frequently, as well as two measures of genre preference: affinity, which represents the number times a user has streamed music in a genre; and completeness, which captures the fraction of a genre's music the user is familiar with. For each measure of genre preference, we consider a user's highest-scoring genre, a categorical attribute. All user attribute data was collected the day before the start of the sample period.

The task of inferring the social communities that are meaningful to people's music listening presents an opportunity for research that exceeds the scope of this work. Here, we use a simple approach to segmenting users into inferred communities in order to demonstrate the value of this recommendation task and motivate further research. We segment users into inferred communities using subsets of the attributes above. For example, one segmentation yields segments of users who share the same top listening language and age bucket; another segmentation yields segments of users who reside in the same country and use the same type of device to use Spotify.

Evaluation

We use song plays to evaluate our recommender. A form of implicit feedback, plays are a less reliable signal of user interest in a song than an explicit interaction such as rating (Hu, Koren, and Volinsky 2008). Implicit feedback, however, is far more abundant than explicit feedback, and represents the most fundamental behavior in the context of a streaming music platform. This abundance and universality make song plays an attractive and widely used signal for evaluation (McFee et al. 2012), particularly when comparing cross-cultural behavior.

When making music recommendations based on social motivations, two qualities are important: whether the user would be interested in listening to the song, and whether the song provides social value. Following a common practice in recommender systems literature, we measure the former by computing the precision of our system's recommendations compared to a user's implicit feedback in held-out data. Concretely, if we recommend a song to a user five days after its release, we consider the recommendation successful if the user listened to that song at any point between the day of the recommendation and the end of the sample period.

The second quality of interest, the social value of a recommendation, is multifaceted. In this work we focus on the social value derived from listening to the same music as one's community, which implies that the listening must occur around the same time. We thus use a measure of timeliness to evaluate the social value provided by our recommendations.

To measure how timely our recommendations are, we consider the difference between when a user found a new song on their own and when we would have recommended it. Once again, we consider a user's organic listening to a song as an indication that they would engage with the same song had it been recommended to them earlier. We assume a user's first stream of the song during the sample period represents the moment of organic discovery (e.g., through word of mouth, hearing it on TV, or even through another recommendation algorithm).

Counting the number of days between the moment of organic discovery and the moment of recommendation (as determined by the parameter $days\_waited$), we obtain the number of "days saved." For example, if a user first listened to a song 7 days after it was released, but our approach would have recommended it 3 days after its release, we can determine that the approach "saved" the user 4 days of not knowing about the music. On the other hand, if we recommend a song 7 days after its release to a user who would have already streamed it on the 2nd day after release, the value of "days saved" is -5.

As formulated, an average timeliness score that is significantly greater than zero suggests an approach that can add real value to users. Ultimately that value will need to be weighed in light of the corresponding precision score in a subjective fashion: can the approach save users enough time while also being precise enough to avoid surfacing too many irrelevant recommendations?

Experiments

We conduct a series of offline experiments to evaluate the effectiveness of the proposed approach to socially-motivated recommendation and to examine how its performance responds to three variables: the attributes we use to segment users, and the values of $days\_waited$ and $frac\_users\_required$. For each experiment, we choose a subset of the demographic and listening attributes described above and segment the sample users into groups using these attributes. For each group, we generate recommendations by applying the approach described above to each song in our sample. We experiment with $days\_waited$ in [1, 30] and $frac\_users\_required$ in [0.5%, 1%, 2%, 3%, 4%] (per day) and report the aggregate precision and average timeliness values for each experiment.

To provide context for the precision values achieved in these experiments, we report the results of a baseline, where we disregard the temporal aspects of the sample data and generate recommendations as follows. We segment users into groups using the approach described above. Within each segment, we hold out a random 10% of users for evaluation. For each segment, we rank the songs in our sample data by the number of users in the remaining 90% who streamed them. We recommend the top k songs to the 10% of users set aside for evaluation; we consider the recommendation successful if the user had listened to the song during the sample period. We experiment with this baseline for k in [1, 10].

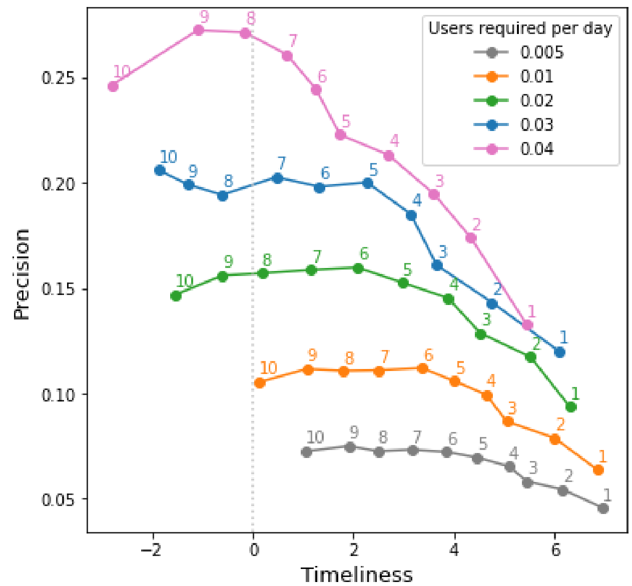

Figure 4: Precision (y-axis) vs timeliness (x-axis). Each point represents a single experiment, with each line representing a particular value of $frac\_users\_required$ (per day), and the points along each line representing particular values of $days\_waited$. Higher values of $days\_waited$ and $frac\_users\_required$ result in greater precision at the expense of timeliness.

Results

How effectively does the proposed system make music recommendations that satisfy users' interests, as measured through precision, and provide social value, as measured through timeliness? In offline experiments, our approach yielded precision values that ranged from 0.05 to 0.27, compared to the baseline approach, which yielded a maximum precision value of 0.04, with $k = 1$. In this section, we examine three aspects of these results. First, we analyze how the recommendation parameters $days\_waited$ and $frac\_users\_required$ affect recommendation quality and interact with each other. Next, we look at the effectiveness of different attributes for segmenting users to model their social communities. Finally, we consider how the performance of our approach varies along cultural dimensions.

Recommendation Parameters

To analyze the dynamics of $frac\_users\_required$ and $days\_waited$, we focus on the series of experiments using a fixed segmentation of users according to the best-performing set of attributes (reported in the following section): top genre completeness, country, and age bucket.

In figure 4, we compare the timeliness and precision of the recommendations produced by different values of these two recommendation parameters. Each point represents a

Features	Precision	Timeliness
Top genre completeness	0.176	8.7
Top genre completeness + country	0.186	8.9
Top genre completeness + country + age bucket	**0.198**	**9.4**

Table 1: Comparison of user groupings based on different sets of attributes

single experiment, with lines connecting the experiments that share a value of $frac\_users\_required$. There is a clear trade-off between the two metrics. For each fixed value of $frac\_users\_required$, represented by a single line, higher $days\_waited$ values up to 10 generally yield higher precision and lower timeliness. Values above 10 yield significantly lower precision, and are thus omitted from the plot. These results correspond to the intuition that waiting longer to assess a song's trendiness allows for more accurate assessment, but it reduces the usefulness of the recommendation. The more users have already listened to a recommended song, the less time will be saved as a result of the recommendation.

Comparing different values of $frac\_users\_required$ represented by different lines in figure 4, we see that higher values yield recommendations that are more precise and somewhat less timely. This parameter can be understood as a measure of strictness; higher values lead to fewer, more effective recommendations.

Identifying User Communities

We use data describing users' demography, geography, musical taste, and language preferences to segment users into groups. We analyze the effectiveness of these attributes at modeling users' musical communities by comparing the precision and timeliness of the recommendations produced for user segments based on subsets of these attributes. In Table 1 we report the precision and timeliness of the most effective segmentations using one, two, and three attributes respectively. Results are reported for $frac\_users\_required$ =3% and $days\_waited$ =4; the relative precision and timeliness of all segmentations was consistent for values of $days\_waited$ between 1 and 7 and for values of $frac\_users\_required$ between 1% and 4%.

The most effective single attribute for segmenting users is top genre completeness. This suggest the rather simple conclusion that the groups people want to keep up with musically are those groups characterized by shared musical taste. The secondary effectiveness of country suggests the importance of geography and hints that culture indeed plays a role in people's musical communities. Future work should explore more specific ways of describing users' geography and culture, as a single country may be home to a diversity of cultures.

Third, we see that age is also a useful attribute; this echoes findings in previous work that suggest people's relationship to new music and to keeping up musically changes with age.

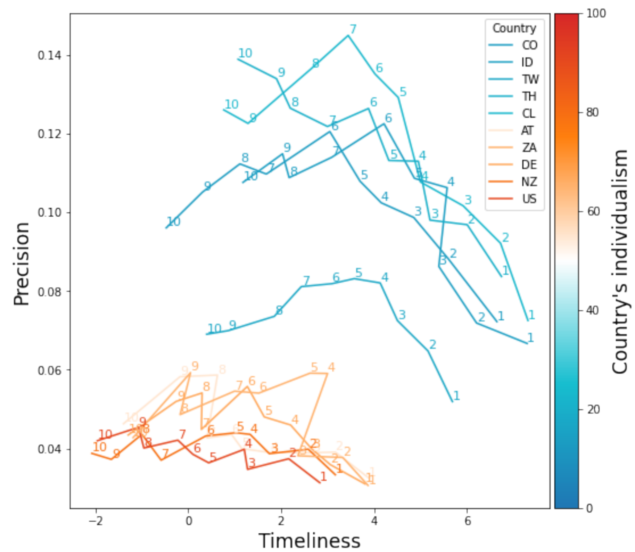

Figure 5: Experiment results aggregated by country for ten sample countries. Each line represents a single country, with each point representing a different value of $days\_waited$. Lines are colored by a country's level of individualism (Hofstede, Hofstede, and Minkov 2010). The countries displayed here consist of the five countries in the sample with the lowest levels of individualism and the five countries with the highest. Countries with low levels of individualism display higher values of both precision and timeliness.

Finally, the fact that each additional split increased the effectiveness of this approach suggests that getting more specific helps; maybe this even suggests that as we start to approximate real-world communities, the effectiveness of our approach improves.

Cultural Differences

While we have thus far reported precision and timeliness values for each experiment over our entire sample, it is important to note that precision and timeliness varied considerably between user segments. Some of this variation relates to the cultural dimension of individualism and collectivism.

In Figure 5, we plot the timeliness and precision values by country, for values of $days\_waited$ in [1,10], using the most effective attributes found for segmenting users (segmenting by top genre completeness, country, and age bucket) and a value of $frac\_users\_required$ of 1%. Each line represents the results for a single country, with each point along the line representing a single experiment with a different value of $days\_waited$. The lines are colored according to the country's level of individualism-collectivism (Hofstede, Hofstede, and Minkov 2010). Of the twenty countries in our sample, we plot the five with the highest levels of individualism and the five with the lowest.

This allows us to see that in countries characterized by lower levels of individualism (Hofstede, Hofstede, and Minkov 2010), the proposed approach to recommendation yields higher values of both timeliness and precision, across all values of $days\_waited$. While omitted from the plot for readability, the remaining ten countries in our sample follow a similar pattern. This result suggests that our approach to recommendation is more effective in more collectivist cultures, which is consistent with previous research that found that collectivism manifests in the concentration of music listening around a smaller number of artists (Ferwerda et al. 2016; Ferwerda and Schedl 2016).

Future work should further explore the types of users for which this approach to socially-motivated recommendations is most effective.

Discussion

In this work, we outline a novel recommendation task focused on socially-motivated music listening. In support of such recommendations, we propose a "timeliness" metric to quantify success for this task. To illustrate both the task and its evaluation, we explore a simple heuristic approach to making such socially-motivated recommendations, using a sample of data on consumption of newly released songs on Spotify. Furthermore, we discuss how the cultural dimension of individualism-collectivism interacts with the effectiveness and motivation behind this approach to recommendations.

Our work shows that, even with a very simple approach, recommendations based on community trends are an attractive means for identifying music that people may listen to eventually but wish to find out about sooner. Furthermore, we have a basic sense of what attributes help define the groups that are meaningful to people for this purpose; namely, these are a user's genre preferences, geography, and age.

Finally, we find that the effectiveness of our approach varies between groups. The nature of this variation suggests that social music recommendations in practice will likely operate differently in places with different cultural values. This suggests an immense opportunity for future research into how cultural differences should inform recommender systems. In particular, we argue that the greater effectiveness of socially-motivated music recommendations in collectivist cultures suggests that the suitability of a particular recommendation task may vary between cultures. This points to the possibility that the very act of framing a recommendation task may be culturally conditioned. We hope our work might inspire future research to engage with these ideas further.

Our analysis explores a trade-off between the precision and timeliness of recommendations. While it is tempting to use recall to evaluate this task, recall supposes value in saving users *any* amount of time in discovering new music. The socially-motivated nature of this recommendation task calls for a more nuanced metric that measures *how much* time we save users explicitly. Knowing exactly how this time saved translates into value for music listeners, however, will require investigation by future studies. Our timeliness metric supposes that accelerating a listener's awareness of a song by two days is better than only one day, but it's not clear if this doubling of days saved translates into twice as much perceived value. Further, we suspect that not all days saved are equivalent (e.g., the difference between recommending at days 3 vs. 4 post-release, compared to 13 vs. 14); that it may be possible to recommend a trending song *too* soon for users who require more social proof; and that not all individuals will perceive value in the same way, even in the same cultural context. Understanding these dynamics presents an intriguing direction for future research to explore, both in the context of music specifically but also in the context of virality and social trends more generally.

A key assumption made in our analysis asserts that if a user listened to a song on one day, they would have accepted and even preferred its recommendation on an earlier day. However, as noted in our introduction and again in the previous paragraph, social proof plays an important role in shaping individual decisions and collective outcomes in cultural markets. That is to say, the popularity of a song on one day potentially contributes to whether or not its recommendation would have been successful on another day. Exactly how much it contributes and the extent to which this effect can be modeled presents potentially many studies' worth of interesting questions. For the study at hand, we resolved that by making this simplifying assumption (i.e., that recommendations would have been successful if made at an earlier time), our estimates of days saved and thus value added represent upper bounds of what is possible in practice. Further, effects like recommendations potentially going unnoticed present similar complications that we believe could limit the effectiveness of implementations in practice. With these considerations in mind, we believe our simplifying assumption is appropriate for this study, given our intentions of opening a discussion around this class of recommendation problem. Accordingly, we welcome future research into the social dynamics of music and how they shape its recommendation.

With regards to improving upon the simple approach presented here, we note a number of opportunities for future exploration. First, there is significant opportunity in learning about how best to characterize and identify relevant user groups for this recommendation task. Certainly more nuanced attributes should help, but one might also consider more sophisticated approaches to actively learn and resolve uncertainties at the individual-listener level. Intuitively, this recommendation task must strike a balance between stereotyping people by lumping them in with too large a group of people with whom they don't have enough in common and, conversely, losing information on broader cultural trends by specializing too much. Adopting – and properly regularizing – an approach at the individual level feels like the logical next step.

In the introduction, we alluded to a goal and corresponding evaluation procedure that we wish to emphasize for future work, particularly for work involving real-world experiments. That goal was to accelerate the process by which people discover socially-relevant music without changing the distribution of what people would have otherwise listened to. For randomized controlled experiments, this implies a simple but important check: the fraction, for instance, of individuals who end up listening to the recommended music should not differ significantly for those in the treatment versus control groups. Ideally, these fractions would be indistinguishable but with the treatment group reaching that fraction sooner. In the simulations presented here, precision offers some insight into how much opportunity is created for this sort influence to manifest. Understanding the extent to which that influence is realized in a real-world test, however, would inform how well recommender systems can accelerate but not alter music's discovery in practice.

Previously, we noted that traditional, collaborative-filtering-inspired approaches should, over time, service socially-motivated recommendations indirectly by gradually learning from how to surface music to similar users. Such systems, in fact, are present in the "natural" backdrop of music discovery that takes places throughout the months of our sample. We argue, however, that this passive approach is insufficient for music platforms that wish to proactively rather than reactively participate in peoples' discovery of new music, a promise that is central to the value proposition of streaming services (Webster 2021). Our results suggest, as well, that such approaches are particularly insufficient for more collectivist-leaning audiences.

A final opportunity for future research draws on our recommendation algorithm. Here, we've treated all users within a group as equal, while this is certainly not the case. Within a community there may be trendsetters, or particular users whose behavior is especially informative about the future performance of new music. An approach that focuses on individuals could also consider recommending a new song to these "trendsetting" users explicitly in order to gain more information about the predicted trendiness of a song and who may follow their lead.

Acknowledgments

The authors wish to thank Ben Carterette, Zahra Nazari, Himan Abdollahpouri, and Glenn McDonald.

References

Abdollahpouri, H.; Adomavicius, G.; Burke, R.; Guy, I.; Jannach, D.; Kamishima, T.; Krasnodebski, J.; and Pizzato, L. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1): 127–158.

Amblee, N.; and Bui, T. 2011. Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts. *International journal of electronic commerce*, 16(2): 91–114.

Baek, Y. M. 2015. Relationship between cultural distance and cross-cultural music video consumption on YouTube. *Social Science Computer Review*, 33(6): 730–748.

Baskerville, R. F. 2003. Hofstede never studied culture. *Accounting, organizations and society*, 28(1): 1–14.

Bauer, C.; and Schedl, M. 2019. Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. *PLOS One*, 14(6): e0217389.

Berkovsky, S.; Taib, R.; Hijikata, Y.; Braslavsku, P.; and Knijnenburg, B. 2018. A cross-cultural analysis of trust in recommender systems. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 285–289.

Boer, D.; and Abubakar, A. 2014. Music listening in families and peer groups: benefits for young people's social cohesion and emotional well-being across four cultures. *Frontiers in Psychology*, 5: 392.

Boer, D.; and Fischer, R. 2012. Towards a holistic model of functions of music listening across cultures: A culturally decentred qualitative approach. *Psychology of Music*, 40(2): 179–200.

Boer, D.; Fischer, R.; Strack, M.; Bond, M. H.; Lo, E.; and Lam, J. 2011. How shared preferences in music create bonds between people: Values as the missing link. *Personality and Social Psychology Bulletin*, 37(9): 1159–1171.

Boer, D.; Fischer, R.; Tekman, H. G.; Abubakar, A.; Njenga, J.; and Zenger, M. 2012. Young people's topography of musical functions: Personal, social and cultural experiences with music across genders and six societies. *International Journal of Psychology*, 47(5): 355–369.

Brewer, M. B.; and Chen, Y.-R. 2007. Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review*, 114(1): 133.

Chen, L.; and Pu, P. 2014. Experiments on user experiences with recommender interfaces. *Behaviour & Information Technology*, 33(4): 372–394.

Choi, J.; Lee, H. J.; Sajjad, F.; and Lee, H. 2014. The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change*, 86: 65–79.

Cialdini, R. B.; and Goldstein, N. J. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55: 591–621.

Dowd, T. 2007. The sociology of music. *21st century sociology: A reference handbook*, 2: 249–260.

Fatemi, M.; and Tokarchuk, L. 2013. A community based social recommender system for individuals & groups. In *2013 International Conference on Social Computing*, 351–356. IEEE.

Ferwerda, B.; and Schedl, M. 2016. Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-Country Analysis. In *UMAP (Extended Proceedings)*.

Ferwerda, B.; Vall, A.; Tkalcic, M.; and Schedl, M. 2016. Exploring music diversity needs across countries. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 287–288.

Fortunato, S.; and Hric, D. 2016. Community detection in networks: A user guide. *Physics Reports*, 659: 1–44.

Gasparetti, F.; Sansonetti, G.; and Micarelli, A. 2021. Community detection in social recommender systems: a survey. *Applied Intelligence*, 51(6): 3975–3995.

Groarke, J. M.; and Hogan, M. J. 2016. Enhancing wellbeing: An emerging model of the adaptive functions of music listening. *Psychology of Music*, 44(4): 769–791.

Hargreaves, D. J.; and North, A. C. 1999. The functions of music in everyday life: Redefining the social in music psychology. *Psychology of Music*, 27(1): 71–83.

Henrich, J.; Heine, S. J.; and Norenzayan, A. 2010. Most people are not WEIRD. *Nature*, 466(7302): 29–29.

Hofstede, G. 1984. *Culture's consequences: International differences in work-related values*, volume 5. sage.

Hofstede, G.; Hofstede, G. J.; and Minkov, M. 2010. Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival.

Hu, Y.; Koren, Y.; and Volinsky, C. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE international conference on data mining*, 263–272. Ieee.

Itai, I.; et al. 2008. Self actualisation: For individualistic cultures only? *International Journal on Humanistic Ideology*, 1(02): 113–139.

Jannach, D. 2022. Multi-Objective Recommender Systems: Survey and Challenges. *arXiv preprint arXiv:2210.10309*.

Jannach, D.; and Bauer, C. 2020. Escaping the mcnamara fallacy: towards more impactful recommender systems research. *AI Magazine*, 41(4): 79–95.

Juslin, P. N.; Barradas, G. T.; Ovsiannikow, M.; Limmo, J.; and Thompson, W. F. 2016. Prevalence of emotions, mechanisms, and motives in music listening: A comparison of individualist and collectivist cultures. *Psychomusicology: Music, Mind, and Brain*, 26(4): 293.

Knobloch, S.; Vorderer, P.; and Zillmann, D. 2000. The impact of music preferences on the perception of potential friends in adolescence. *Zeitschrift Fur Sozialpsychologie*, 31(1): 18–30.

Koren, Y.; Rendle, S.; and Bell, R. 2021. Advances in collaborative filtering. *Recommender systems handbook*, 91–142.

Lee, K.; and Lee, K. 2014. Using dynamically promoted experts for music recommendation. *IEEE Transactions on Multimedia*, 16(5): 1201–1210.

Lonsdale, A. J.; and North, A. C. 2009. Musical taste and ingroup favouritism. *Group Processes & Intergroup Relations*, 12(3): 319–327.

Ma, H.; Zhou, D.; Liu, C.; Lyu, M. R.; and King, I. 2011. Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 287–296.

Masthoff, J. 2011. Group recommender systems: Combining individual models. In *Recommender systems handbook*, 677–702. Springer.

McFee, B.; Bertin-Mahieux, T.; Ellis, D. P.; and Lanckriet, G. R. 2012. The million song dataset challenge. In *Proceedings of the 21st International Conference on World Wide Web*, 909–916.

Mehr, S. A.; Singh, M.; Knox, D.; Ketter, D. M.; Pickens-Jones, D.; Atwood, S.; Lucas, C.; Jacoby, N.; Egner, A. A.; Hopkins, E. J.; et al. 2019. Universality and diversity in human song. *Science*, 366(6468): eaax0868.

Mooij, M. D. 2003. Convergence and divergence in consumer behaviour: implications for global advertising. *International Journal of Advertising*, 22(2): 183–202.

Oyserman, D.; Coon, H. M.; and Kemmelmeier, M. 2002. Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128(1): 3.

Park, M.; Park, J.; Baek, Y. M.; and Macy, M. 2017. Cultural values and cross-cultural video consumption on YouTube. *PLOS One*, 12(5): e0177865.

Park, M.; Weber, I.; Naaman, M.; and Vieweg, S. 2015. Understanding musical diversity via online social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, 308–317.

Saarikallio, S.; Alluri, V.; Maksimainen, J.; and Toiviainen, P. 2021. Emotions of music listening in Finland and in India: Comparison of an individualistic and a collectivistic culture. *Psychology of Music*, 49(4): 989–1005.

Salganik, M. J.; Dodds, P. S.; and Watts, D. J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762): 854–856.

Schäfer, K.; and Eerola, T. 2020. How listening to music and engagement with other media provide a sense of belonging: an exploratory study of social surrogacy. *Psychology of Music*, 48(2): 232–251.

Schäfer, T.; Sedlmeier, P.; Städtler, C.; and Huron, D. 2013. The psychological functions of music listening. *Frontiers in Psychology*, 4: 511.

Schedl, M.; Zamani, H.; Chen, C.-W.; Deldjoo, Y.; and Elahi, M. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2): 95–116.

Selfhout, M. H.; Branje, S. J.; ter Bogt, T. F.; and Meeus, W. H. 2009. The role of music preferences in early adolescents' friendship formation and stability. *Journal of adolescence*, 32(1): 95–107.

Soley, G.; and Spelke, E. S. 2016. Shared cultural knowledge: Effects of music on young children's social preferences. *Cognition*, 148: 106–116.

Tang, T. Y.; Winoto, P.; and Ye, R. Z. 2011. Analysis of a multi-domain recommender system. In *The 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, 280–285. IEEE.

Tian, K. 2020. *Cross-cultural recommender system use and design*. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Triandis, H. C. 2018. *Individualism and collectivism*. Routledge.

Way, S. F.; Garcia-Gathright, J.; and Cramer, H. 2020. Local trends in global music streaming. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 705–714.

Webster, J. 2021. The promise of personalisation: Exploring how music streaming platforms are shaping the performance of class identities and distinction. *New Media & Society*, 14614448211027863.

Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, see Broader Impact and Ethical Considerations.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see Methods.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see Data.
   (e) Did you describe the limitations of your work? Yes, see Discussion.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see Broader Impact and Ethical Considerations.
   (g) Did you discuss any potential misuse of your work? Yes, see Broader Impact and Ethical Considerations.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see Broader Impact and Ethical Considerations.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? No, because the code and data are proprietary.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, see Methods and Experiments.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, see Evaluation.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes, see Discussion.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? NA
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA