KISS MIR: Keep It Semantic and Social Music Information Retrieval

Amna Dridi and Mouna Kacimi

Faculty of Computer Science, Free University of Bozen-Bolzano, Piazza Domenicani 3, Bolzano, Italy

Keywords: Music Information Retrieval, Music-context, User-context, Personalization, Social Information.

Abstract: While content-based approaches for music information retrieval (MIR) have been heavily investigated, user-centric approaches are still in their early stage. Existing user-centric approaches use either music-context or user-context to personalize the search. However, none of them give the possibility to the user to choose the suitable context for his needs. In this paper we propose KISS MIR, a versatile approach for music information retrieval. It consists in combining both music-context and user-context to rank search results. The core contribution of this work is the investigation of different types of contexts derived from social networks. We distinguish semantic and social information and use them to build semantic and social profiles for music and users. The different contexts and profiles can be combined and personalized by the user. We have assessed the quality of our model using a real dataset from Last.fm. The results show that the use of user-context to rank search results is two times better than the use of music-context. More importantly, the combination of semantic and social information is crucial for satisfying user needs.

1 INTRODUCTION

With the increasing volume of digital music available on the world wide web, developing information retrieval (IR) techniques for music is challenging. The main reason is the wide variety of ways music is produced, represented, and used (Smiraglia, 2001; Casey et al., 2008). However, despite the high potential of systems presented at ISMIR and similar venues, the development of music IR (MIR) systems that fit user music taste is still in its early stages. For instance, most of existing MIR approaches are *content-based* considering low-level features of music (audio signal). These methods are capable to define music structure, to identify a piece from noisy recording, or to compute similarities between music pieces. However, they cannot capture semantic information which is essential to many users of music(Knees et al., 2013). Moreover, they do not take into account user preferences which can highly improve and facilitate access to music. In fact, several recent surveys on MIR (Kaminskas and Ricci, 2012; Schedl and Flexer, 2012; Weigl and Guastavino, 2011; Y.Song, 2012) emphasized the limitations of content-based MIR systems and suggest new MIR directions which are *user-centric*.

Existing user-centric MIR approaches aim at improving user access to music following two main strategies. The first one consists in enriching *music-context* using annotations (Li and Ogihara, 2006; Saari et al., 2013; Sanden and zhang, 2011), and the second one exploits *user-context* for a personalized music retrieval (Hoachi et al., 2003; Boland and Murray-Smith, 2014). These approaches aim at satisfying user needs imposing a specific type of context. However, user needs are not predictable and should not be bound to a specific setting. Thus, there is a need for providing users the possibility to interact with the music retrieval system and choose the suitable context for his needs.

In this paper, we propose KISS MIR a novel approach for music retrieval that enriches user search experience by combining both music-context and user-context. To this end, we exploit social networks as the main source for context information. We categorize the information provided in such networks into *semantic* and *social* information. Semantic information is reflected by tagging actions of users while social information represents user activities and relations in the network. The context of music and user can be either semantic, social, or both. Based on these types of context, we propose a ranking model for music tracks that helps selecting the right context for each query and user need. The versatility of our approach provides a flexible access to music. The main contributions of this paper are as follows:

Dridi, A. and Kacimi, M..
KISS MIR: Keep It Semantic and Social Music Information Retrieval.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 433-439
ISBN: 978-989-758-158-8

1. We propose to model context-information as semantic or social to investigate their role in satisfying user needs.
2. We define user profile and music profile based on semantic and social information. Semantic information reflects music description and user interests while social information reflects music popularity and user behaviour.
3. We propose a personalized ranking model that combine both music-context and user-context which can be semantic, social, or both. The model allows the user to choose the most suitable setting for his needs.
4. we evaluate our model on real world dataset from Last.fm[1] and show that user-context that combines both semantic and social information outperforms all other settings.

The remainder of this paper is organized as follows: Section 2 reviews the related work, Section 3 describes the framework of KISS MIR, Section 4 introduces user and music profiles, Section 5 describes our scoring model, Section 6 presents our experiment setting and results, and Section 7 concludes the paper.

2 RELATED WORK

Existing user-centric MIR approaches can be divided into two main classes. The first class of approaches exploits music-context to improve user access to music. Concretely, they rely on music annotation with semantic labels. Some approaches annotate music with emotions and they rely on emotion detection based on music content (Li and Ogihara, 2006; Huron, 2000; Kaminskas and Ricci, 2011; Braunhofer et al., 2013). While some other approaches combine human-annotated tags with music content for emotion detection (Lee and Neal, 2007; Saari et al., 2013; Y.Song et al., 2013; Lamere, 2008; Feng et al., 2003), multimodal music similarity (Zhang et al., 2009), artist descriptions (Pohle et al., 2007), music pieces characterisation (Knees et al., 2007), verifying the quality of music tag annotation via association analysis (Arjannikov et al., 2013), or via multi-label classification (Sanden and zhang, 2011) .

The second class of approaches exploits user-context to build user profile for personalized MIR. For instance, (Hoachi et al., 2003) propose to build user profile based on what he likes and what he hates. Moreover, (Celma et al., 2005) use Friend Of A Friend (FOAF) documents to define user profile. While, in another work, (Herrara, 2009) suggest that user profile can be categorised to three domains: *demographic*, *geographic* and *psychographic*. Additionally, (Chen and Chen, 2001) derive user profile from his access history and (Boland and Murray-Smith, 2014) capture the change of user profile over time.

In our work, we aim at enriching the user search experience, by combining both music-context and user-context in the retrieval process. The most similar approaches to our are music recommender systems that help users to filter and discover music according to their tastes (Bugaychenko and Dzuba, 2013; Celma et al., 2005; Chen and Chen, 2001; Chedrawy and Abidi, 2009). While these approaches provide music recommendation that matches user profile, we are providing a music search system with personalized ranking involving the user in the center of the retrieval process.

3 KISS MIR FRAMEWORK

KISS MIR consists in combining both *music-context* and *user-context* in the music retrieval process. To extract these contexts, we exploit social networks as a prominent and rich source for information about both music properties and user activities. The first step towards this goal is to understand (1) which kind of information can we find in such networks and (2) how can we use it to extract music and user contexts. To this end, we distinguish the following entities as main components of the information provided by social networks:

1. **Users:** represent the participants to a social network.
2. **Music Tracks:** represent the content shared by users in a social network
3. **Descriptions:** represent *tags* or *annotations* provided by users to describe music tracks. Tags can also be exploited to indicate user preferences.
4. **Reactions:** represent user feedback reflected by different actions (*comment, like, dislike, favourite, etc*). Reactions capture user interests and the popularity of music tracks. In some cases, they also categorize interests as negative or positive (*like, dislike, etc.*)
5. **Communities:** represent sets of users who are interconnected. Users can be linked based on different criteria such as friendship, location, be-

[1] www.lastfm.fr

haviour, or simply belonging to the same social network.

We model social networks as a graph that represents the entities mentioned above. Figure 1 shows the graph model consisting of four types of nodes *Users*, *Music tracks*, *Descriptions* and *Reactions*. Users provide music tracks, describe them, and react to them. Communities can be implicit, and thus they are not represented in the graph.

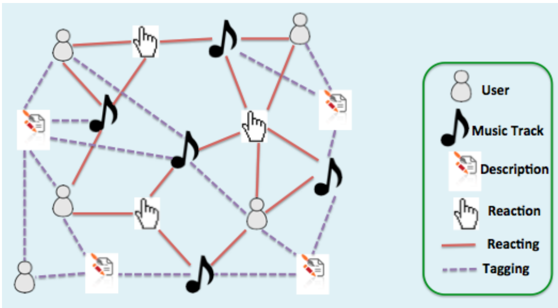

Figure 1: Social Music Graph.

4 USER AND MUSIC PROFILES

The different entities of social networks provide two types of information defined as follows:

Semantic Information. Semantic information reflects meanings, roles, and properties. This type of information is provided by descriptions where users tag music tracks with semantic labels. Descriptions about music include genre, singer, title, *etc*. They also give labels to user interests such as Rock and Pop.

Social Information. Social information reflects activities in social networks. These activities are represented by the links between entities in the social graph shown in Figure 1. They include reacting and tagging actions of users towards music tracks.

4.1 User Profile

Based on semantic and social information described above, we define two types of user profiles: *User Semantic Profile* and *User Social Profile* defined as follows:

4.1.1 User Semantic Profile

User semantic profile consists in the set of descriptions (tags) used by a user u to annotate music tracks of his choice. We denote this set by $T_u = \{t_1, t_2, ..., t_k\}$. In social networks tags represent a strong indicator of users interests.

4.1.2 User Social Profile

User social profile consists in a set of users that have a relationship with user u, namely the community of u. We denote this set by $C_u = \{u_1, u_2, ..., u_{n1}\}$ where u_i is a user i linked to user u.

4.2 Music Profile

Similarly to user profile, we define two types of profiles for music tracks: *Music Semantic Profile* and *Music Social Profile* defined as follows:

4.2.1 Music Semantic Profile

Music semantic profile consists in a set of features that represent a music track m. The set of features is predefined and has a fixed length k. Music track features include genre, singer, year, and a set of keywords representing its content such as lyrics or title. We represent the semantic profile of music track m as vector of k features $F_m = \{f_1, f_2, ..., f_k\}$. The values of these features are mainly provided by tags.

4.2.2 Music Social Profile

Music social profile consists in the set of users who reacted to the music track m. We denote this set by $U_m = \{u_1, u_2, ..., u_{n2}\}$ where the u_i is a user who reacted to music m. The number of users who reacted to music track m indicates its popularity in the social network.

5 SCORING MODEL

When users search for music, they typically do not have a clear idea about what they are looking for. Thus, a keyword search is not a suitable scenario unless users have precise information about a given music track such as title or singer. To allow a wider access to music, we choose a query-by-example paradigm where users can find similar music tracks to what they are interested in. Given a user u, and a query Q=m, where m is a music track, we search for music tracks that are relevant to the query Q and match the interests of user u. A set of music results $\{m_1, \ldots, m_{n3}\}$ is returned to the user where the score of each result m_i is given as following:

$$S(Q, u, m_i) = \lambda S_{music}(m_i, Q) + (1 - \lambda) S_{user}(m_i, u) \tag{1}$$

where $S_{music}(m_i, Q)$ denotes the *music relevance score* of m_i to Q and $S_{user}(m_i, u)$ denotes the *user relevance score* of m_i to u . The parameter λ controls the amount of personalization ($0 \leq \lambda \leq 1$). Setting $\lambda = 1$ means that we aim at finding what matches the query music track and setting $\lambda = 0$ means that we aim at finding what matches user interests. Values in between combine the two components with different degrees.

5.1 Music Relevance Score

The music relevance score, of a music track m given a query Q, indicates to which degree m matches Q. Considering the query-by-example paradigm, we compute the music relevance score using a similarity measure between m and Q:

$$S_{music}(m, Q) = Similarity(m, Q) \quad (2)$$

The similarity measure can take different forms such as *cosine similarity*, *Jaccard similarity*, or *Euclidean distance*. In this paper, we use *Jaccard distance* to compute the similarity between two music tracks m and Q:

$$Similarity(m, Q) = \frac{|F_m \cap F_Q|}{|Fm \cup F_Q|} \quad (3)$$

where F_m and F_Q represent the set of features of music tracks m and Q respectively. These features are extracted from the semantic music profiles of m and Q.

5.2 User Relevance Score

The community of a user have a big influence on the music he listens to. This means that for a music track to be liked by the user, it needs to be popular in his community. Thus, we exploit the social profile of the user to compute the *user relevance score* of music tracks m in the community of u as follows:

$$S_{user}(m, u) = \frac{|U_m \cap C_u)|}{|Users|} \quad (4)$$

where: U_m is the set of users who reacted to music track m, C_u is the set of users in the community of user u. and $|Users|$ is the total number of users in the network. It is important to mention that we focus, in this paper, only on positive reactions. So, the more reactions there are, the more popular the music track is.

Additionally to the social profile of the user, we can enhance the user relevance score by taking into account his semantic profile. This means that among the users of the community of user u, we just target those who have similar interests as u. Recall that user interests are reflected by tags. In this case, we would limit the community of user u only to users who have similar tagging actions as user u. We say that two users u_i and u_j have similar tagging behaviour if the size intersection of their tag sets T_{u1} and T_{u2} exceeds a certain threshold. The threshold setting depends on the social network and the amount of data it contains.

6 EVALUATION

6.1 Music Dataset

To evaluate our approach, we have used a dataset from *Last.fm*[2][3]. The dataset contains music tracks, users, and their activities. As there was not enough semantic information about music tracks, we have exploited Wikipedia[4]to enrich the dataset. For each music track, we have used the title to access its Wikipedia page. From the Wikipedia page, we have got information about the music track from the infobox including singer, producer, writer, year, label, and other features. Further, we have removed all tracks that do not have Wikipedia pages. Regarding user activities, the dataset contains *descriptions* corresponding to *tags*, and *reactions* corresponding to *clicks* which were the only reactions available in the dataset. Table 1 gives statistics about the resulted dataset.

Table 1: Statistical characteristics of the *Last.fm* dataset.

# Music Tracks	# Users	# Clicks	# Tags
11523	3387	642490	68335

6.2 Evaluation Methodology

We have run our experiments using 100 queries. Recall that we use a query-by-example paradigm and thus queries correspond to music tracks. We have randomly selected the 100 queries from the $top250$ most clicked music tracks. The reason of this choice is driven by the requirements of the automatic assessment of the results described below. As a further step, we proceeded with the selection the query initiators. For each query, we have selected the users who clicked on the query music track and ranked them. The rank of users was computed based on the number of clicks he has on that track and the number of clicks he has globally. The query initiator was then selected randomly among the $top20$ users.

[2]www.lastfm.fr

[3]http://www.dtic.upf.edu/∼ocelma/MusicRecommendationDataset/lastfm-1K.html

[4]www.wikipedia.org

After selecting all queries and their query initiators, we have use our model to rank the results of each query. We have tested different strategies of our model:

1. **Music-context (Semantic):** consists in returning results that match only the semantic profile of the query without considering the query initiator profile. This is achieved by setting the parameter $\lambda = 1$.
2. **Music-context (Semantic + Social):** consists in returning results that match the semantic and the social profile of the query without considering the query initiator profile. This is achieved by setting the community of the query initiator to all users in the network. Thus the S_{user} score would be independent from the query initiator reflecting only the popularity of the music track in the whole network.
3. **User-context (Social):** consists in returning results that match the social profile of the query initiator setting $\lambda = 0$. In this experiments, the community of any user is the set of all users in the network and thus the score is solely based on the popularity of the music track in the network. This setting is equivalent to **Music-context (Social)** where the result matches only the social profile of the query.
4. **User-context (Semantic + Social):** consists in taking into account both the social and the semantic profiles of the query initiator. In this setting music tracks should match the interests of the community and the query initiator. This is obtained by setting $\lambda = 0$ and restricting the community of the query initiator only to users with similar tagging behaviour.
5. **Music-context + User-context (Semantic + Social):** consists in using all the elements of our approach. We set $\lambda = 0.5$ and we use both semantic and social profiles for music tracks and users to rank results.

6.3 Assessment and Evaluation Metrics

To avoid any subjectivity in the assessment of the results, we have exploited click information to indicate whether a user likes a music track or not. So, for each returned result we check if the user has clicked on it. If he has clicked, then we set the result as relevant and give a value of 1, otherwise it is irrelevant and has a value of 0. To measure the effectiveness of our approach, we have used the precision $P@k$ which represents the fraction of retrieved music tracks that are relevant to the query considering only the top-k results. It is given by:

$$P@k = \frac{|Relevant\_Track \cap topk\_Track\_Results|}{k} \quad (5)$$

We have also used the *Mean Average Precision (MAP)* which is a widely adopted standard measure in IR given by:

$$MAP@n = \frac{\sum_{i=1}^{N} AverageP@n_i}{N} \quad (6)$$

where N is the total number of queries, n is a given position and $AverageP$ is the average precision of each query.

6.4 Results and Discussion

Tables 2 and 3 show the precision and MAP values for the different strategies of our model. It is clear from the results that user-context approaches perform the best in terms of precision and MAP values. More precisely, when the social and the semantic user profiles are both taken into account to find relevant music tracks. Compared to using only a Music-context approach based on the semantic profile of the music track, the precision@5 increases from 0.285 to 0.478 which is a substantial improvement. Similarly, the MAP@5 highly improves from 0.242 to 0.423. We note a decrease in precision and MAP for the $top10$ results which is due to the decrease of the popularity of music tracks at lower ranks.

Table 2: Mean precision values for all queries.

	P@5	P@10
Music-context (Semantic)	0.285	0.204
Music-context (Social) / User-context (Social)	0.45	0.309
Music-context (Semantic + Social)	0.433	0.293
User-context (Semantic + Social)	**0.478**	**0.315**
Music-context + User-context (Semantic + Social)	0.450	0.293

After User-context approaches come the Music-context combined with User-context approaches showing also high precision and MAP values. Thus, whenever User-context approaches are adopted, we can increase the satisfaction of the user. Now, considering only Music-context approaches we can see a notable difference when we use the semantic profile of the music track and when we enhance it with its social profile. We can see that matching music tracks based on their social profile increases the precision@5 from

Table 3: MAP values for all queries.

	MAP@5	MAP@10
Music-context (Semantic)	0.242	0.147
Music-context (Social) / User-context (Social)	0.398	0.259
Music-context (Semantic + Social)	0.405	0.254
User-context (Semantic + Social)	**0.423**	**0.268**
Music-context + User-context (Semantic + Social)	0.421	0.258

0.285 to 0.433 which is a substantial improvement. Similarly, the MAP@5 highly improves from 0.242 to 0.405. The reason is that even though the social profile of the music track does not depend on the query initiator, it depends on other users in the network. This means that it is enough to be in the same network to influence the taste of any participant. So, this strategy is indirectly User-context which explains the high improvement in the results.

Table 4: MAP @5 for all queries.

λ \ $1-\lambda$	0.0	0.2	0.5	0.8	1.0
0.0	n/a	n/a	n/a	n/a	**0.478**
0.2	n/a	n/a	n/a	**0.448**	n/a
0.5	n/a	n/a	**0.450**	n/a	n/a
0.8	n/a	**0.450**	n/a	n/a	n/a
1.0	**0.285**	n/a	n/a	n/a	n/a

Table 5: P@5 for all queries.

λ \ $1-\lambda$	0.0	0.2	0.5	0.8	1.0
0.0	n/a	n/a	n/a	n/a	**0.423**
0.2	n/a	n/a	n/a	**0.419**	n/a
0.5	n/a	n/a	**0.421**	n/a	n/a
0.8	n/a	**0.421**	n/a	n/a	n/a
1.0	**0.242**	n/a	n/a	n/a	n/a

To analyse the impact of the parameter λ on the performance of the model, we use different values ranging from 0 to 1 for the strategy that combines all types of profiles. The results are given in tables 4 and 5. In the same line as previous results, we can see that the best results are achieved when $\lambda = 0$ which corresponds to User-context strategy. To summarise, the overall results of these experiments demonstrate that involving the user in the retrieval process can be done in different ways and all of them improves highly the satisfaction of the user compared to music-context approaches.

In these experiments, we have exploited clicks, by analogy to IR evaluation paradigm, as relevance judgments. However, actually, in the case of music, a simple click does not really reveal the relevance judgment since the user can stop the music after some seconds. To overcome this problem, in our ongoing experiments, we are currently testing our approach by introducing a graded relevance scale based on the number of clicks considering that having clicked a track several times is a much stronger indication of its relevance.

7 CONCLUSIONS AND FUTURE WORK

We have presented a user-centric approach for music retrieval which gives the opportunity to users to get involved in the search process. To this end, we exploited a variety of contexts. More specifically, we have extracted semantic and social contexts for music tracks and for users. Based on these different contexts, we proposed a ranking model that can use music-context, user-context, or both. Additionally, it can set the context to semantic, social, or both. We have investigated the different settings of our model to understand what type of information is more suitable for satisfying user needs. Our experiments have shown that user-context is by far more useful than music-context for improving the quality of music search results. More importantly, both semantic and social information should be taken into account to represent user profile. As future work, we aim at investigating different types of reactions other than clicks. Comments are definitely a valuable source for user interests and can reveal a lot more about what the user is looking for and how his taste changes over time.

REFERENCES

Arjannikov, T., Sanden, C., and Zhang, J. (2013). Verifying tag annotations through association analysis. In *Proceedings of International Society of Music Information Retrieval (ISMIR)*, pages 195–200. ACM.

Boland, D. and Murray-Smith, R. (2014). Information-theoretic measures of music listening behaviour. In *Proceedings of ISMIR*, pages 561–566. ACM.

Braunhofer, M., Kaminskas, M., and Ricci, R. (2013). Location-aware music recommendation. 2(1):31–44.

Bugaychenko, D. and Dzuba, A. (2013). Musical recommendations and personalization in a social network. In *Proceedings of RecSys*, pages 367–370. ACM.

Casey, M., Veltkamp, R., Gosto, M., Leman, M., Rhodes, C., and Slaney, M. (2008). Content-based mir: Current directions and future challenges. In *Proceedings of IEEE*. IEEE.

Celma, O., Ramirez, M., and Herrara, P. (2005). Foafing the music: A music recommendation system based on rss feeds and user preferences. In *Proceedings of ISMIR*, pages 464–467. ACM.

Chedrawy, Z. and Abidi, S. (2009). A web recommender system for recommending predicting and personalizing music palylists. In *Proceedings of WISE*, pages 335–342. Springer.

Chen, H. and Chen, A. L. P. (2001). A music recommendation system based on music data grouping and user interests. In *Proceedings of CIKM*, pages 231–238. ACM.

Feng, Y., Zhuang, Y., and Yunhe, P. (2003). Music information retrieval by detecting mood via computational media aesthetics. In *Proceedings of WI*, pages 235–241. ACM, IEEE.

Herrara, O. C. (2009). *Music Recommendation and Discovery in the Long Tail*. PhD thesis.

Hoachi, K., Matsumoto, K., and Inoue, N. (2003). Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceedings of MULTIMEDIA*, pages 110–119. ACM.

Huron, D. (2000). Perceptual and cognitive applications in music information retrieval. In *Proceedings of ISMIR*. ACM.

Kaminskas, M. and Ricci, F. (2011). Location-adapted music recommendation using tags. In *Proceedings of UMAP*.

Kaminskas, M. and Ricci, F. (2012). Contextual music information retrieval and recommendation: State of the art and challenges.

Knees, P., Pohle, T., Schedl, M., and Widmer, G. (2007). A music search engine built upon audio-based and web-based similarity measures. In *Proceedings of SIGIR*, pages 447–454. ACM.

Knees, P., Schedl, M., and Celma, O. (2013). Hybrid music information retrieval. In *Proceedings of ISMIR*, pages 1–2. ACM.

Lamere, P. (2008). Social tagging and music information retrieval. 37(2):101–114.

Lee, H. and Neal, D. (2007). Toward web 2.0 music information retrieval: Utilizing emotion-based, user assigned descriptors. 44:1–34.

Li, T. and Ogihara, M. (2006). Toward intelligent music retrieval. In *Proceedings of MULTIMEDIA*, volume 8, pages 564–574.

Pohle, T., Shen, J., Knees, P., Schedl, M., and Widmer, G. (2007). Building an interactive next-generation artist recommender. In *CBMI*. IEEE.

Saari, P., Eerola, T., Fazekas, G., Barthet, M., Lartillot, O., and Sandlen, M. (2013). The role of audio and tags in music mood prediction. In *Proceedings of ISMIR*. ACM.

Sanden, C. and zhang, J. (2011). An empirical study of multi-label classifiers for music tag annotation. In *Proceedings of ISMIR*, pages 717–722. ACM.

Schedl, M. and Flexer, A. (2012). Putting the user in the center of music information retrieval. In *Proceedings of ISMIR*, pages 385–390. ACM.

Smiraglia, R. (2001). Musical works as information retrieval entities epistemological perspectives. In *Proceedings of ISMIR*, pages 85–91. ACM.

Weigl, D. and Guastavino, C. (2011). User studies in the music information retrieval literature. In *Proceeding of ISMIR*, pages 335–340. ACM.

Y.Song, Dixon, S., Pearce, M., and Halpern, A. (2013). Do online social tags predict perceived or induced emotional responses to music? In *Proceeding of ISMIR*, pages 89–94. ISMIR.

Y.Song, S. Dixon, M. P. (2012). A survey of music recommendation systems and future perspectives. In *Proceedings of CMMR*, pages 395–410.

Zhang, B., Shen, J., Xiang, Q., and Wang, Y. (2009). Icompositemap: a novel framework for music similarity. In *Proceedings of SIGIR*, pages 403–410. ACM.