EXPLORING THE RELATION BETWEEN NOVELTY ASPECTS AND PREFERENCES IN MUSIC LISTENING

ABSTRACT

Discovering new songs or artists is a central aspect of music consumption. To assist users in this task, several mechanisms have been proposed to enable recommendation systems to incorporate novelty recommendation. This study takes a complementary approach to these efforts investigating how the preference of listeners varies according to different dimensions of novelty. We use historical data from a set of listeners to investigate how familiarity and mainstreamness of novel artists relate to how relevant the listener found the novelty to be. The results of this analysis point to distinct groups of users according to preferences. Our results analyze such groups and derive implications for technologies enabling new music discovery.

1. INTRODUCTION

Discovering new songs or artists that one likes is a central aspect of music consumption. To a higher or lesser degree, all listeners look for novelty over time. Due to the immense collections presently available, there is a high potential for tools that assist listeners in finding novel and relevant items.

Both commercial systems and research efforts have made progress in addressing the task of incorporating novelty in music recommendation.

In this work we explore how considering multiple dimensions of novelty can be useful for mechanisms assisting music discovery. Our hypothesis is that considering other aspects of novel items can improve our understanding about what novelties would be preferred by music listeners. To test this hypothesis, we examine how two dimensions of novelty in artists discovered by listeners over a period – familiarity and mainstreamness – relate to the preferences of these listeners to the artists.

On the one hand, our analysis suggests that there is no overall correlation between familiarity or mainstreamness and novelty relevance in our sample. On the other hand, when considered individually, most listeners do have a significant correlation between either familiarity or mainstreamness of novel artists and the relevance they see in these artists.

This result leads us to cluster the listeners we study according to their preferences and listening behavior. Studying the resulting clusters reveals eight archetypical profiles that explain how different groups have preferences for the different aspects of novelty we consider.

Finally, a comparison of subjects' preferences for novel artists and artists previously points that there is a significant difference in the preferences of listeners when considering novelty or known artists.

Our results can inform the development of music discovery mechanisms, including recommendation systems.

2. RELATED WORK

Several studies point to the relevance of considering novelty in music recommendation systems. In general, different efforts agree that it is necessary to evaluate such systems not only by their accuracy and recall, but also by how many new and relevant items are introduced to a user [3,4]. In materializing this view, two major challenges have been tackled in the literature so far: how to build recommendation systems that introduce novelty adequately, and how to evaluate these systems [2].

On the first front, multiple algorithms have been put forward. An approach used by multiple of these algorithms is to define classes of items and to recommend items diversifying among the classes already seen by a user in the past [5, 8]. Other approaches include randomly selecting items that are above a certain similarity threshold with respect to the user [7] and recommending items that, in a network of items, are well connected to items previously seen and highly distant items.

The challenge of evaluating such algorithms relies on the fact that traditional metrics and methods do not account for unexpected recommendations or for items absent from typical testing data. To address this issue, Vargas et al. [6] put forward a formal framework to define diversity and novelty metrics. The authors make a distinction between three concepts related to novelty and diversification: choice (a user picks an item), discovery (a user is introduced to an item) and relevance (a user likes an item). In Vargas et al.'s framework, our study examines the effect of familiarity and mainstreamness of novel items picked by users on their relevance.

Somehow related to evaluation, some authors have also found that listeners are not always prone to the introduction of novelty. Three factors that affect the interest in novelty put forward by An et al. [1] are the listener's convictions, emotional state and social context.

An overall regularity in previous studies on novelty for music discovery is that they have considered novelty as a unidimensional property of items. Our study complements previous efforts by exploring how considering multiple dimensions of novelty in items – besides the fact that users do not know them – can help us further understand listener's preferences for novel items. This knowledge can be combined with previous approaches to inform the design of music recommendation or discovery services.

3. DATA COLLECTED

Two datasets are used in our analysis: the first contains the subjects whose behavior will be analyzed, together with their historical listening habits, and the second concerns metadata about the artists listened by the chosen subjects. All data was collected from Last.fm [1] through the service's public API.

3.1 Subject data

Our goal is to identify and collect data about a sample of Last.fm users which has been exposed to novel artists during a period of interest, which we set as the 52 weeks between March 2012 and March 2013 and dub as the experiment period.

Starting with the profile of the first author, a snowball sampling procedure was conducted on the Last.fm friendship network until 100,000 users were identified. Over this initial data, users with at least 100 song executions in at least 30 different weeks of the experiment period were retained (approx. 5.500 users). These represent reasonably active users which likely inform much of their music listening to Last.fm.

Next, data about artists knowledge prior to the experiment period and listening habits during the experiment period were collected. The former is collected as the set of 200 artists most listened by the user since he or she joined Last.fm. The data collected about the listening habits during the experiment period is the union of the sets of 100 most listened artists by the user in each week between March 1st 2012 and March 1st 2013. Using this data we select in our sample only users that were exposed to a significant number of novel artists in the experiment period. For that, we keep in our sample only users who, during the experiment period, listened 10 times songs from each of 10 artists not present in their previous artist knowledge. Finally, we removed from our data users who had more than 1000 song executions from a same artist in the experiment period, which were decided to be outliers.

Taken together, this selection process leads us to a sample of 3,191 users who (i) have active highly listening habits, (ii) are likely to inform a large portion of these habits to the data source we use, and (iii) have experienced a number of novelties in their listening that enable us to investigate the relation between novelty characteristics and user preferences.

[1] www.lastfm.com

3.2 Artist data

For each artist present in subjects' listening data, we collected artist popularity and tag data from Last.fm. Popularity data was collected as the number of users who listened the artist as informed by Last.fm on March 10th 2013. A tag is a label given by a user to the artist, ranging, in Last.fm, from music genres (eg. rock, samba) to mood (eg. sad, lively) and other contextual metadata (eg. summer, lovesong). Each tag has a popularity reflecting how often users have assigned the tag to a given artist. After collecting tag data from all artists in subjects listening data, we filter out tags with tag popularity lower than 0.15 of the most popular tag for the artist. Furthermore, tags denoting personal circunstances, such as seen live, favorite and albums I own were manually removed.

4. CHARACTERIZING LISTENER PROFILES AND NOVELTIES

To examine how different characteristics of novel items affect their relevance for listeners, we resort on three constructs we now describe in turn: a model of a listener profile, a set of dimensions on which we consider novel items may vary, and a definition of relevance.

4.1 Novel Items

For our experiment, we consider as items artists, and an artist is novel for a subject if this artist is absent from the subject's previous artist knowledge and was listened to at least five times in a week by this subject during the first half of the experiment period. The first half of the experiment period is using for identifying novel artists listened by the subject, while the whole experiment period is used to evaluate how much attention the subject payed to these novel items (a process detailed in Section 4.4).

With this definition, we identify 245,407 novel items in our data, with a subject discovering on average (resp. standard deviation) 76.9 (71) novel artists during our measurement. For our analysis, we consider two dimensions along which novel items vary: similarity to previous profile and popularity.

4.2 Listener Profiles

A model of a listener profile is necessary to describe a subject's artist knowledge and listening habits prior to being exposed to a novel item. We model each listener profile as a set of clusters of artists in his or her listening history prior to the experiment period. These clusters are obtained applying the DBScan clustering algorithm to the set of artists known to the listener that are not listened for the first time in the experiment period. For the execution of the algorithm, the distance between two artists is defined as the inverse of the cossine similarity of the tags assigned to each artist. Moreover, the two parameters of the DBScan algorithm – the minimum number of points for a cluster to be considered and the maximum distance between two points in a cluster – were empirically defined as

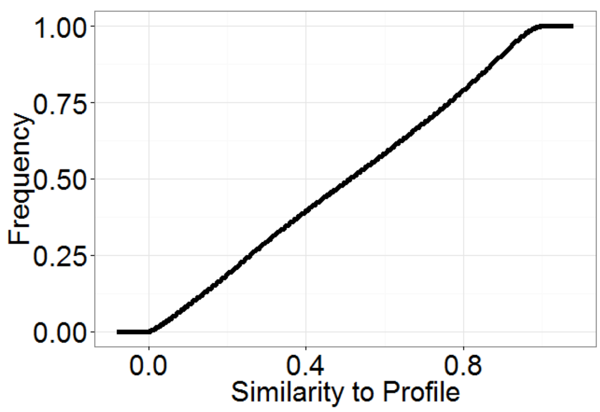


Figure 1: Distribution of the similarity of all novel artists

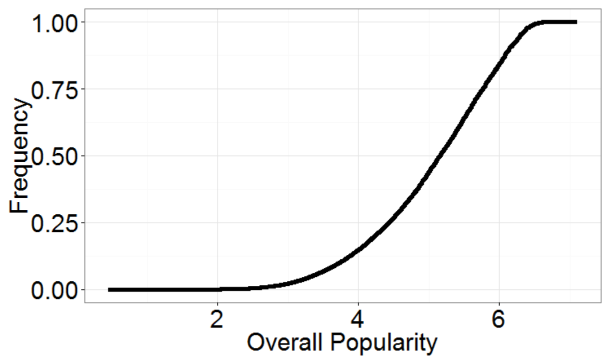


Figure 2: Distribution of the overall popularity log of all novel artists

3 and 0.875 (resp.) after consulting an online community of music afficiionados about which of several clustering solutions better described members' taste profiles. Profiles obtained through this process for our subjects have an average (resp. standard deviation) of 5.3 (2.9) clusters found among the artists present in the subject's listening history.

4.3 Characteristics of novelties

4.3.1 Similarity to previous profile

The similarity of an item to the previous profile of a subject is the distance between this item and the listener profile. This dimension models the intuition that for a jazz fan, novel heavy metal artists are likely to sound less familiar than a novel artists from a subgenre of jazz. Formally, the similarity of an artist and a listener profile is defined as highest cossine similarity among those calculated using the artist's tags and those of the centroids of the artists clusters in the listener profile. Figure 1 displays the distribution of the similarity of all novel artists in our experiment to their listener profiles.

4.3.2 Overall Popularity

The second dimension we consider for novel artists found during the experiment period is the overall popularity of the artist in Last.fm. We expect the overall popularity of an artist to represent how familiar the artist is to a subject before the subject devotes attention to listening this artist. This familiarity likely comes from other medias, such as mentions in news, social networks or television. Our hypothesis is that some users may have a preferences for mainstream novel items. Because the artists popularity distribution observed is highly skewed, our analysis uses the log of this popularity in all results reported. Figure 2 displays the distribution of the overall popularity log of all novel artists in our experiment.

4.4 Preferences for novel items

We measure subjects' preferences for different types of novel artists gauging how much or how often each novel artist was listened to during the experiment period. For this analysis, an artist/item is said more relevant than another if the former was listened to more times or more frequently than the latter.

To put forward this approach there are two metrics for item relevance in our analysis. Let δ be the period between the first time a subject listened to a novel artist and the end of our experiment period, then (i) the total attention a subject gives to an artist is the total number of executions of songs of this artist during δ divided by the number of weeks in δ; and (ii) the period of attention of a subject to a novel artist is the fraction of number of weeks in δ in which the subject listened at least once to the artist. Because novel artists are found over a period of time, some of these artist have a smaller time window in our experiment period in which they can be listened to. We tackle this issue with two measures. First, taking δ in consideration limits the bias this may introduce in the counting of song executions or weeks of attention. Second, as mentioned in Section 4.1, only novelties discovered in the first half of our experiment period are considered in the analysis. This gives each novel item at least six months of observation in our traces, what limits possible biases caused by too short observations.

Figure 3 shows the distribution of the total attention and period of attention for each novelty found by the subjects in our experiment.

5. PREFERENCES FOR NOVELTY CHARACTERISTICS

Our central research question is to understand how considering novelty as multidimensional can enhance our understanding about listener preferences. To address this question, we first evaluate whether there is a correlation between novelty characteristics – similarity to previous profile and overall popularity – and relevance – total attention and attention period.

Table 1 shows that analyzing all novel items together, there seems to be no relevant correlation between novelty characteristics and relevance that is valid for all subjects. On the other hand, analyzing the correlation coefficient between novelty characteristics and relevance for the novel items of each subject individually, there is a different pattern. For each pair of variables we examine, there is a significant correlation (above 0.15 or below -0.15) between

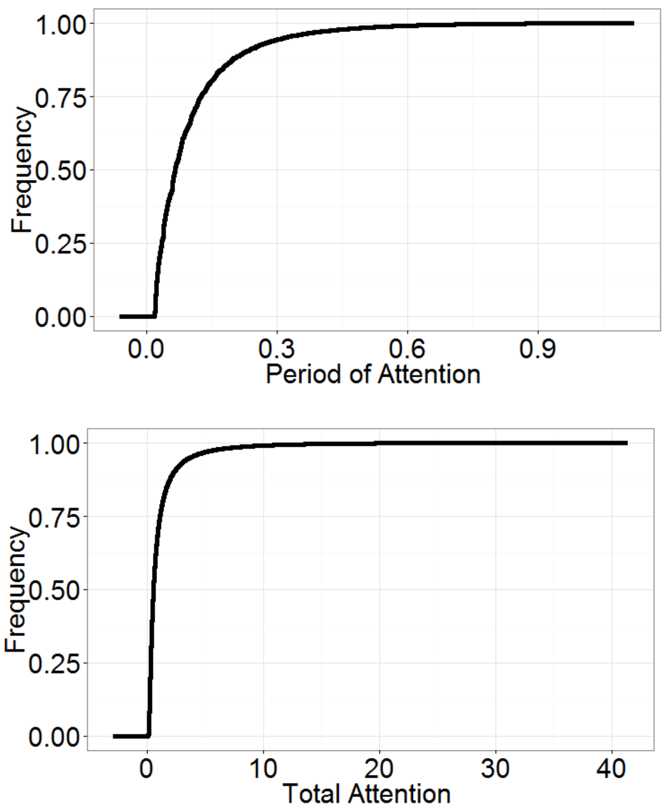


Figure 3: Distribution of the total attention and period of attention for each novelty

Dimentions	PoA	StP	OP
Total Attention (TA)	0.75	0.03	0.05
Period of Attention (PoA)	1.0	0.04	0.07
Similarity to Profile (StP)	-	1.0	0.01
Overall Popularity (OP)	-	-	1.0

Table 1: Correlation (Spearman's coefficient) between novelty characteristics and relevance, analyzing all novel items together.

the variables for approximately one third of all subjects 5. Moreover, 60% of all subjects have in their data a significant correlation between at least one of the novelty characteristics and a measure of relevance.

Taken together, our results point that while there is no overall common behavior regarding subjects' preferences for different types of novelty, most users have some preference for a type of novelty in their listening behavior.

6. LISTENER GROUPS WITH RESPECT TO NOVELTY

The analysis of the correlation between novelty types and preferences indicates that there are different types of listener in our data. In this section we apply clustering analysis to delve further in the identification and analysis of such groups. For that, we use only the set of users for which at least one of the correlations analyzed in the previous section is not in the [-0.15; 0.15] interval.

Our analysis uses the Ward hierarchical clustering method considering normalized versions of two dimensions to describe each subject's preferences for novelty aspects and two dimensions that measure the subject's usual musical taste. The dimensions related to preferences are (a) the correlation between number of song executions for novel artists and their similarity to previous profile and (b) the correlation between number of song executions for novel artists and overall popularity. To represent usual tastes, we use (c) the average popularity of the artists listened by the subject, and (d) the number of clusters present in the subject's listening profile. Upon examination, the correlations with period of attentions for novel artists were discarded as redundant dimensions for the clustering analysis.

Applying this approach leads us to a solution with 8 clusters whose centroids are shown in Figure 4. Analyzing these centroids, we labelled the groups as follows:

1. *Underground*: Listeners who usually listen to artists out of the mainstream, have a relatively homogeneous listening profile, and display no marked preferences in the dimensions we consider.

2. *Niche radical*: Listeners that habitually focus on non-mainstream artists, and have a clear preference for novelty very close to their profile and also out of the mainstream.

3. *Focused mainstreamer*: Subjects with marked preference for mainstream artists and novelty from the mainstream, but that listen relatively few artist groups.

4. *Ecletic mainstreamer*: Listeners with high preference for mainstream artists and novelty while maintaining a diverse listening profile.

5. *Explorer*: Listeners with noted preference for novelty dissimilar to their previous profile.

6. *Risk averse*: Listeners that strongly value novelty that is familiar; also tend to listen to mainstream artists.

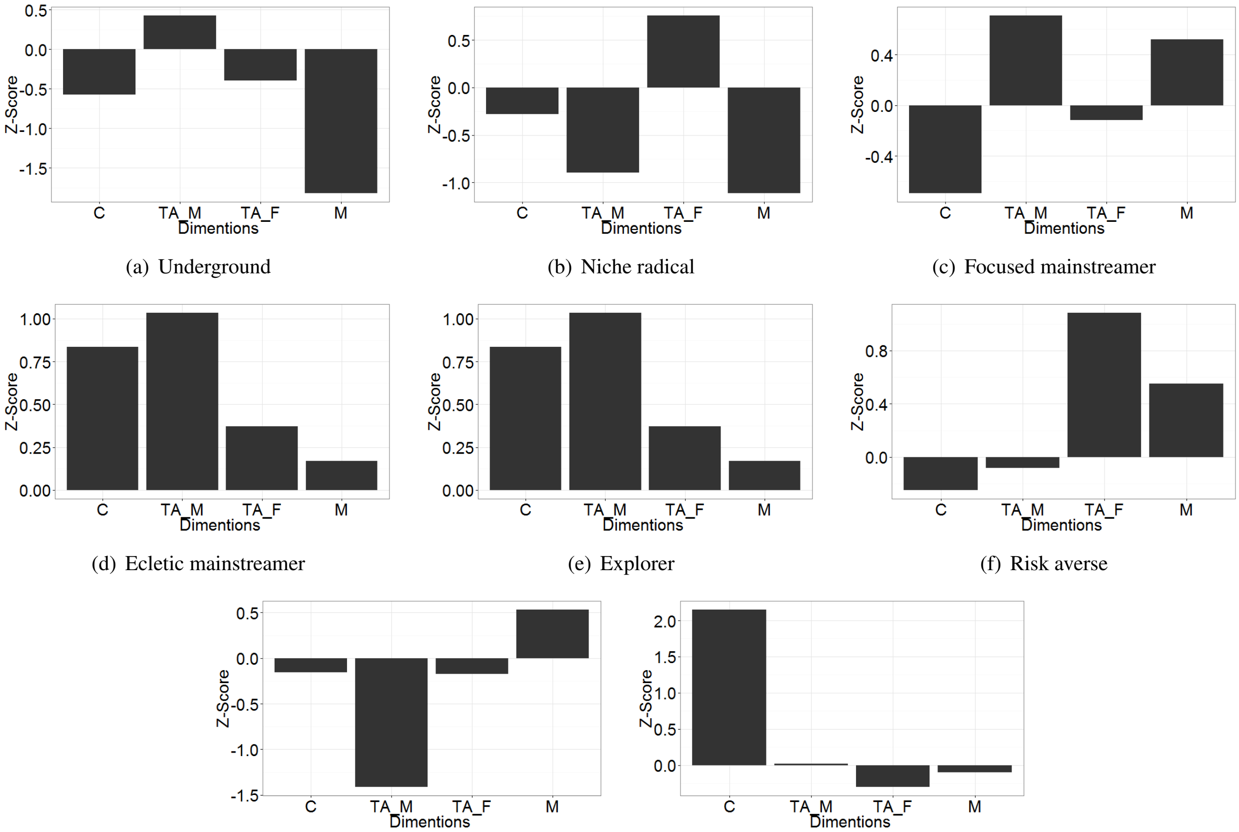
7. *Fond of surprises*: Subjects with distinct preference for novelty that is out of the mainstream in contrast with listening moderately mainstream artists.

8. *Highly eclectic*: Subjects whose listening profiles had disproportionaly high number of artist clusters, and value novelty irrespective of familiarity or mainstreamness.

7. SIMILARITY AND POPULARITY AMONG RELEVANT NOVELTY

Another question that follows from our analysis is whether different types of novelty have a different effect on subjects' listening behavior when they are relevant.

To answer this question, we consider only the novel artists with higher total attention given by a subject than



(a) Underground (b) Niche radical (c) Focused mainstreamer

(d) Ecletic mainstreamer (e) Explorer (f) Risk averse

(g) Fond of surprises (h) Highly eclectic

Figure 4: Users' group centroids. C = number of Clusters, TA_M = correlation between Total Attention and overall popularity (Mainstreamness), TA_F = correlation between Total Attention and similarity to profile (Familiarity), M = average popularity of the artists listened by the subject (Mainstreamness)

the average that same subject gave to other novelties during the experiment period. Among these novel artists, we first test whether among relevant novelties, those dicovered far from a listener's profile are found to be more relevant than those close to the profile. There is no support for this difference in our data (t-test, n = , p-value = .4). Unsurprisingly, investigating following a similar procedure, there is evidence that among the relevant novelty, the most popular artists are likely to have higher relevance: novel artissts whose overall popularity was higher than 5 had more overall more executions than the remainder (t-test, p-value ¡ .001).

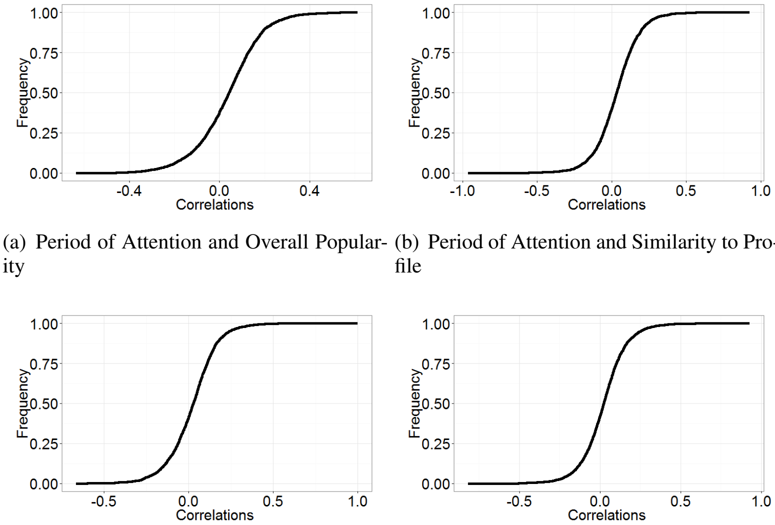
8. DISCUSSION AND IMPLICATIONS

The main objective of this study is to investigate the use of multiple dimensions of novelty to further understand listeners' preferences regarding novel artists. This understanding, in turn, is aimed at informing the design of music discovery mechanisms. In our analysis, our main finding are that:

- There is no overall correlation between familiarity of mainstreamness and relevance in the novel artists discovered by our subjects;
- Considered individually, most subjects had a clear correlation between either familiarity or mainstreamness and the relevance of novel items;
- It is possible to cluster listeners with some correlation between novelty aspects and preferences in 8 groups, which reveal how different audiences approach novelty differently;

Considered together, these results point to the need of a personalized approach in assisting a listener to discover relevant novel artists. Moreover, this personalization should be related not only to which other artists have been liked before, but should be also aware that some listeners have clear preferences regarding how mainstream or familiar novel artists are. The information that these dimensions are relevant to model listeners' behavior, and that it is highly variable among a listener population should be considered in the design of future mechanisms. On another perspective, the groups we find to describe the archetypical behaviors among our subjects can be used to develop interfaces or mechanisms that target different types of users.

There are a number of limitations on our study and directions in which future work could expand it. Notedly, considering data from a population outside Last.fm is necessary to evaluate the generalizability of our results. Also, testing whether the same behavior observed in this study would be present using other models of taste and musical



(a) Period of Attention and Overall Popularity (b) Period of Attention and Similarity to Profile

(c) Total Attention and Similarity to Profile (d) Total Attention and Similarity to Profile

Figure 5: Correlations between novelty characteristics and relevance

similarity such as content-based could be fruitful. Finally, the validation of the use of familiarity and mainstreamness as dimensions of novelty in an experiment of musical recommendation is necessary to further validate the use of multiple dimensions to address novelty.

9. REFERENCES

[1] J. An, D. Quercia, and J. Crowcroft: "This paper won't change your mind, but...: Why individuals seek diverse opinions (or why they don't)r," *Technical Report of the University of Cambridge*, 2012.

[2] Ò. Celma, P, Herrera: "A new approach to evaluating novel recommendations," *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 179–186, 2008.

[3] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, pp. 5–53, 2004.

[4] S. M. McNee, J. Riedl and J. A. Konstan "Being accurate is not enough: how accuracy metrics have hurt recommender systems," *Proceedings of CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pp. 1097–1101, 2006.

[5] M. Nakatsuji, Y. Fujiwara, A. Tanaka, T. Uchiyama, K. Fujimura, T. Ishida: "Classical music for rock fans?: novel recommendations for expanding user interests," *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 949–958, 2010.

[6] S. Vargas and P. Castells: "Rank and relevance in novelty and diversity metrics for recommender systems," *Proceedings of the fifth ACM conference on Recommender systems*, pp. 109–116, 2011.

[7] Y. C. Zhang, D. O. Séaghdha, D. Quercia and T. Jambor: "Auralist: introducing serendipity into music recommendation," *ProProceedings of the fifth ACM international conference on Web search and data mining*, pp. 13–22, 2012.

[8] C. Ziegle, S. M. McNee, J. A. Konstan, G. Lausen: "Improving recommendation lists through topic diversification," *Proceedings of the 14th international conference on World Wide Web*, pp. 22–32, 2005.