MUSIC RECOMMENDATION DIVERSITY: A TENTATIVE FRAMEWORK AND PRELIMINARY RESULTS

Lorenzo Porcaro[1], Carlos Castillo [2], Emilia Gomez[1,3]
[1]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[2] Web Science and Social Computing Group, Universitat Pompeu Fabra, Barcelona, Spain
[3]Joint Research Centre, European Commission, Seville, Spain
{lorenzo.porcaro,carlos.castillo,emilia.gomez}@upf.edu

ABSTRACT

Music recommendations are increasingly part of the listening experience of people all over the world, especially in the context of streaming services. In this scenario, recommender systems' role is to help users in finding music that can fit their interests and tastes. However, Western-centric perspectives in systems' design are often subject to criticisms because of their power of reinforcing already existing cultural bias and therefore potentially impacting negatively on the music distribution mechanisms. In our research proposal, we aim to address the problem of assessing the impact of music recommendation diversity, or the lack thereof. This requires 1) the formalization of a working definition of diversity in the music field 2) the development of evaluation practices for estimating diversity in the context of music recommender systems 3) the observation of emerging impact due to music recommendation diversity 4) the proposal of countermeasures for mitigating negative or reinforcing positive impact observed. Basing on already known consequences of information technologies in political, economic and social areas, our goal is to understand the cultural impact that music recommender systems can have on our society.

1. INTRODUCTION

Our age has been dominated by the advent of streaming services, and music, or more generally audio-visual content, is a cultural product for which the enjoyment and production have been extensively influenced and re-shaped in the last century. As Benjamin at the beginning of the 20th century underlined in his "*Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit*" ("The Work of Art in the Age of Mechanical Reproduction"), the dual identity of music, which influences and is influenced by society, is strictly defined by the historical circumstances, and fundamental are the relationships between the work of art, the medium and human perception [2].

The widespread use and easy accessibility of streaming platforms is an important achievement for today's music listeners, which are one click away from the possibility to enjoy their favourite music and to discover new artists. In this panorama, recommender systems play a key role as bridge between users and the large amount of content accessible in digital repositories. Indeed, to browse the large catalogue of tracks in streaming platforms is a task that may reveal truly complicated for a user [10]. Music recommender systems represent a fraction of the wide family of recommender systems, often used in streaming services for tasks such as playlists continuation or the creation of personalized music radios [24]. Consequently, most of the users' listening behaviours are nowadays subject to different layers of algorithmic-generated interactions, accomplishment that can bring along with it opaque social, economic and cultural issues [12, 33].

In our work, we are interested in assessing how the diversity, which characterizes music from different parts of the world, is represented within recommender systems, and what can be the consequences of diversity in music recommendations, or the lack thereof. We argue that to preserve the richness of musical cultures is undoubtedly a crucial mission of modern technologies, and recommender systems have a critical role in that because of their impact on music distribution [8].

The document is structured as follows. Section 2 provides an overview of the previous work on Recommender Systems (RS) and its relationships with the concept of diversity, and also how this concept has been approached in the Music Information Retrieval (MIR) field. Afterwards, in Section 3 we present the research goals identified in our research, followed by preliminary outcomes in Section 4. Finally, conclusions and future work are discussed in Section 5.

2. RELATED WORK

2.1 Diversity in Recommender Systems

First concerns about the lack of user-centric perspectives in the design of recommendation systems, and related evaluation metrics, start to emerge around two decades ago [19]. In particular, the focus on improving the systems' per-

formance looking exclusively at the accuracy started to be criticized, and alternative evaluation frameworks took hold, introducing the so-called *beyond-accuracy* metrics. In the literature, the most established of these metrics are diversity, serendipity, novelty, and coverage, and we refer to [11] for an extended survey and analysis of these metrics. Following, we focus on diversity, the central topic of our work.

The need to adopt diversification strategies, i.e."*to identify a list of items that are dissimilar with each other, but nonetheless relevant to the user's interests*" [34], is often related to counteract the so-called *portfolio effect*, a situation where recommended items are highly similar to the ones targeted by the user, described by Burke in [4]. At the same time, it justified the emergence of new evaluation metrics, such as the Intra-List Similarity presented by Ziegler et al. in [36], where the items of a list are compared between each other for estimating the general degree of the list dissimilarity. However, when trying to improve the performance of recommender systems in terms of diversity metrics, the accuracy of these systems started to be challenged. This led to the emergence of the *diversity-accuracy dilemma*, hence how to balance the trade-off between diversity and accuracy. Several techniques have been proposed for solving this trade-off, for instance using directed random walks [16], or hybrid algorithms [35].

The multiple facets of diversity have implied the urgency for considering a wide conceptualization and different measurements, therefore different approaches also in the recommender systems literature have tried to tackle the problem of tuning the degree of diversity of recommendation lists. Among the others, in our research line we see links with works on temporal diversity [15], intent-oriented diversity [32], genre diversity [31], and multi-attribute diversity [6]. For an extended review of the approaches to diversity in recommender systems, we refer to [13].

2.2 Diversity in Music Information Retrieval

Our understanding of music diversity is partly based on the analysis of the semiology of music done by Molino in [20]. Indeed, what the author argues is that historical circumstances lead to a constant process of *aesthetic revolution*, making difficult to appear processes of standardization of the music languages. Even if subject to an increasing influence of the mass industry, music symbolic evolution tends to preserve its nature of constant diversification process. In a parallel direction, when analyzing the relationship between Western Music and other musical cultures ("*its Other*"), Born and Hesmondhalgh in [3] underline that 1) the mutual influence that different musical cultures have on each other, and 2) the development of these cultures in a system with specific socio-economic characteristics, are two processes strictly related.

In the field of MIR, diversity has often been examined in relation to musical tastes, hence to the aesthetic domain i.e. the more diverse is the music that you like, the more diverse are your musical taste. From another perspective, diversity, intended as differences in cultural music traditions, has also been at the centre of attention when dealing with information technologies built for extracting knowledge from the poietic domain i.e. the analysis of creative processes. In this direction, the work started by Serra in 2011 in the context of the project CompMusic [1] is an outstanding example of the need for including diverse approaches when tackling the multicultural reality of the music of the world [26].

Music recommender systems research has focused on the problem of understanding and representing diversity at two levels. At an individual level, it has been shown how personal traits and listening behaviours influence users' need for diversified recommendations [18, 29]. From another perspective, by means of cross-country analysis relationships between diversity and musical preferences have been investigated, aiming at evidencing how cultural differences in geographical regions can also be reflected in the listening experiences [7, 17]. These research lines reflect two of the future research directions of music recommender systems research identified by Schedl et al. in [24]: *Psychologically-inspired music recommendation* and *Culture-aware music recommendation*.

3. RESEARCH GOALS

The work of diversity assessment designed is structured in four main goals. In the beginning, the main attention is posed on the development of theoretical instruments for defining and evaluating diversity in the area researched. Subsequently, the focus is shifted in understanding the consequences of music recommendation diversity from a human-centric perspective, as described in the next sections.

3.1 Develop a Framework for Defining and Evaluating Music Recommendation Diversity

The first research goal identified is the creation of a framework for defining and evaluating music recommendation diversity, specific to the MIR and RS fields. A framework that must assure: 1) a solid theoretical background; 2) the development of analysis tools (toolkit, software, etc.). Given the not univocal nature of diversity, as discussed by Stirling in [27], our starting point is to understand how diversity has been approached in different literatures.

3.2 Assess Music Recommendation Diversity

The next step is to research how music recommender systems behave in different scenarios, focusing on the two classic dimensions of recommender systems: user and item. By making explicit the questions that we hypothesize, can we affirm that a recommendation list is more diverse than another one? Can we compare recommendation lists created by different systems, and affirm which system embeds more diversity? Can we compare recommendation lists created for different users, and stating if a list is more diverse than the other? Undoubtedly, the pitfalls contained

[1] https://compmusic.upf.edu

in these questions are several. At the end of this phase, we aim at being able to compare different music recommender systems and to evaluate how tuning systems settings can influence the outcome diversity.

3.3 Understand the Consequences of Music Recommendation Diversity

What we are most interested in is to understand what might be the societal impacts of using recommendation technologies in the context of cultural development. Starting from already-known negative effects on society (such as filter bubbles [21], echo chambers [28], and cyberbalkanization [30]), our objectives are: 1) to understand if consequences found in other areas can be reproduced while analyzing music recommender systems; 2) if others consequences can be found, specifically related to music field.

3.4 Propose Countermeasures for Tuning Music Recommendation Diversity

The final part of this work will target the consequences and impacts found, proposing novel methods for contrasting the counter effects of these technologies on humans. It is difficult to identify *a priori* the techniques to be developed, considering the wide range of scenarios that might derive from the research planned. However, we aim at including a complete spectrum of outcomes, from negative ones, where recommender systems are proven to damage human beings, to positive ones, where on the contrary the use of this technology can improve the well-being and can be used for the social good.

4. PRELIMINARY OUTCOMES

Following the research line previously defined, exploratory experiments have been carried out, thanks to which it has been possible to achieve two initial results. On one hand, in [22] we were able to explore standard diversity measures from the Information Theory literature, applying them for a comparative analysis of playlist datasets. On the other hand, in [23] we made a first attempt of proposing new measures for evaluating the variations of recommendation lists in different scenarios.

Both studies have target items related to recommender systems framework, playlist in [22] and recommendation list in [23], and through the use of information retrieval, statistics, and mathematical modelization techniques, we have evaluated a degree of diversity, statically in the first case whereas dynamically in the second. In these studies, we centered our attention on two dimensions often considered in the MIR literature: popularity, or mainstreamness [1, 5], and semantic information [14].

In [22], characterizing and comparing four playlist datasets created in different historical and technological contexts, we notice the emergence of diverse patterns in the users' grouping choices. Similarly, by means of the comparison of analog and streaming radios, in [23] we started to tackle the limitation of specific metrics for evaluating music recommendation diversity.

5. CONCLUSIONS AND FUTURE WORK

This paper provides an initial formalization of the research work plan and first results on the study of the impact of music recommendation diversity. Until now, few studies in the MIR literature have focused on comprehending what have been the consequences of introducing music recommendation technologies in the actual cultural context. Using the terminology introduced by Selbst et al. in [25], this might cause a *Ripple-Effect Trap*, defined as a "*Failure to understand how the insertion of technology into an existing social system changes the behaviours and embedded values of the pre-existing system*". With our research, we aim at raising awareness in the MIR field for avoiding to fall in this abstraction trap while designing and implementing music recommender systems.

Parallelly, another challenge is to define what diversity can represent, and how it can be represented in the music recommendation field. In 2004, Huron in his article *Issues and Prospects in Studying Cognitive Cultural Diversity* [9], included a call for action where he states:

> We[*music researchers*] should be concerned about the loss of cultural diversity for the same reason that biologists worry about the loss of biodiversity: we don't yet know what the loss will mean, but we do know that the loss will be irreversible.

Even if the comparison between cultural diversity and biodiversity might be challenged in some aspects, the identification as a main problem of the unknown consequences of the loss of diversity is part of the motivation of this work. Indeed, we imagine that the lack of representation or the misrepresentation of diversity potentially could lead to unfair treatment in music distribution, which in the worst scenario might lead to cultural discrimination.

Apart from that, our choice to focus on diversity as sociotechnical concept is also driven by recent debates about the impact of Artificial Intelligence (AI) systems on human behaviour and the related ethical, social, economic and legal issues. Imagining recommender systems as part of a broader field which can be AI, the importance of considering diversity is proven by its inclusion within the list of seven key requirements that AI systems should meet in order to be trustworthy, proposed in the *Ethics Guidelines for Trustworthy Artificial Intelligence*, written by the High-Level Expert Group on Artificial Intelligence [2].

Finally, our motivations are also partly reflected in the less recent but fundamental debate about the importance of the preservation of cultural diversity, which in 2001 has led to the *UNESCO Universal Declaration on Cultural Diversity* [3]. In this document, it is emphasized how cultural diversity is a vehicle by which promoting pluralism, human rights, international solidarity and also, creativity.

[2] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Diversity

[3] http://portal.unesco.org/en/ev.php-URL_ID=13179&URL_DO=DO_TOPIC&URL_SECTION=201.html

6. ACKNOWLEDGMENTS

This work is partially supported by the European Commission under the TROMPA project (H2020 770376).

7. REFERENCES

[1] Christine Bauer and Markus Schedl. Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. *PLOS ONE*, 14(6):1–36, 2019.

[2] Walter Benjamin. *The Work of Art in the Age of Mechanical Reproduction*. Hannah Arendt, ed., Illuminations. London: Fontana., 1968.

[3] Georgina Born and David Hesmondhalgh. Introduction: On Difference, Representation and Appropriation in Music. In *Western Music and Its Others: Difference, Representation, and Appropriation in Music*, pages 1–58. University of California Press, 2000.

[4] Robin Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

[5] Óscar Celma. *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer-Verlag Berlin Heidelberg, 2010.

[6] Tommaso Di Noia, Jessica Rosati, Paolo Tomeo, and Eugenio Di Sciascio. Adaptive multi-attribute diversity for recommender systems. *Information Sciences*, 382-383:234–253, 2017.

[7] Bruce Ferwerda and Markus Schedl. Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In *Proceedings of the 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP) co-located with 24th Conference on User Modeling, Adaptation, and Personalization (UMAP)*, 2016.

[8] Andre Holzapfel, Bob L. Sturm, and Mark Coeckelbergh. Ethical Dimensions of Music Information Retrieval Technology. *Transactions of the International Society for Music Information Retrieval*, 1:44–55, 2018.

[9] David Huron. Issues and Prospects in Studying Cognitive Cultural Diversity. In *Proceedings of the 8th International Conference on Music Perception & Cognition*, number August, 2004.

[10] Sheena S. Iyengar and Mark R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6):995–1006, 2000.

[11] Marius Kaminskas and Derek Bridge. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, 7(1):1–42, 2016.

[12] Nedim Karakayali, Burc Kostem, and Idil Galip. Recommendation Systems as Technologies of the Self: Algorithmic Control and the Formation of Music Taste. *Theory, Culture & Society*, 35(2):3–24, 2018.

[13] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems – A survey. *Knowledge-Based Systems*, 123:154–162, 2017.

[14] Paul Lamere. Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101–114, 2008.

[15] Neal Lathia, Stephen Hailes, Licia Capra, and Xavier Amatriain. Temporal diversity in recommender systems. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'10)*, page 210, 2010.

[16] Jian Guo Liu, Kerui Shi, and Qiang Guo. Solving the accuracy-diversity dilemma via directed random walks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 85(1), 2012.

[17] Meijun Liu, Xiao Hu, and Markus Schedl. Artist Preferences and Cultural, Socio-Economic Distances Across Countries: a Big Data Perspective. In *Proceedings of the 18th International Conference on Music Information Retrieval, ISMIR 2017*, 2017.

[18] Feng Lu and Nava Tintarev. A diversity adjusting strategy with personality for music recommendation. In *Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2018)*, pages 7–14, 2018.

[19] Sean M. McNee, John Riedl, and Joseph A. Konstan. Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006)*, pages 1097 – 1101, 2006.

[20] Jean Molino and Craig Ayrey. Musical Fact and the Semiology of Music. *Music Analysis*, 9(2):105–111;113–156, 1990.

[21] Eli Parisier. *The filter bubble: What the Internet is hiding from you*. Penguin Press, New York, 2011.

[22] Lorenzo Porcaro and Emilia Gómez. 20 Years of Playlists: A Statistical Analysis on Popularity and Diversity. In *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR'19)*, 2019.

[23] Lorenzo Porcaro and Emilia Gómez. A Model for Evaluating Popularity and Semantic Information Variations in Radio Listening Sessions. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems (ImpactRS), at the 13th ACM Conference on Recommender Systems (RecSys 2019)*, 2019.

[24] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. Current Challenges and Visions in Music Recommender Systems Research. *International Journal of Multimedia Information Retrieval*, 7(2):95–116, 2018.

[25] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and Abstraction in Sociotechnical Systems. In *ACM Conference on Fairness, Accountability, and Transparency (FAT*)*, volume 1, pages 59–68, 2018.

[26] Xavier Serra. A Multicultural Approach in Music Information Research. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR'11)*, pages 151–156, 2011.

[27] Andy Stirling. A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, 4(15):707–719, 2007.

[28] Cass Sunstein. *Echo Chambers: Bush v. Gore Impeachment, and Beyond*. Princeton University Press, 2001.

[29] Nava Tintarev, Christoph Lofi, and Cynthia C.S. Liem. Sequences of Diverse Song Recommendations: An exploratory study in a commercial system. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP)*, pages 391–392, 2017.

[30] Marshall Van Alstyne and Erik Brynjolfsson. Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science*, 51(6):851–868, 2005.

[31] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. Coverage, redundancy and size-awareness in genre diversity for recommender systems. *Proceedings of the 8th ACM Conference on Recommender systems*, (October):209–216, 2014.

[32] Saúl Vargas, Pablo Castells, and David Vallet. Intent-Oriented Diversity in Recommender Systems. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'11)*, pages 1–2, 2011.

[33] Patrick Vonderau. The Spotify Effect: Digital Distribution and Financial Growth. *Television and New Media*, 20(1):3–19, 2019.

[34] Cong Yu, Laks Lakshmanan, and Sihem Amer-Yahia. It takes variety to make a world: Diversification in recommender systems. In *Proceedings of International Conference on Extending Database Technologies (EDBT) 2009*, number June 2014, page 368, 2009.

[35] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.

[36] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *WWW'15 Conference*, page 22, 2005.