20 YEARS OF PLAYLISTS: A STATISTICAL ANALYSIS ON POPULARITY AND DIVERSITY

Lorenzo Porcaro[1], Emilia Gomez[1,2]
[1]Music Technology Group, Universitat Pompeu Fabra, Barcelona
[2]Joint Research Centre, European Commission, Seville
{lorenzo.porcaro,emilia.gomez}@upf.edu

ABSTRACT

Grouping songs together, according to music preferences, mood or other characteristics, is an activity which reflects personal listening behaviours and tastes. In the last two decades, due to the increasing size of music catalogue accessible and to improvements of recommendation algorithms, people have been exposed to new ways for creating playlists. In this work, through the statistical analysis of more than 400K playlists from four datasets, created in different temporal and technological contexts, we aim to understand if it is possible to extract information about the evolution of humans strategies for playlist creation. We focus our analysis on two driving concepts of the Music Information Retrieval literature: popularity and diversity.

1. INTRODUCTION

The advent of the streaming era has transformed the role of music playlists, which have become central to the listening experience. The current interest by both academia and private company is illustrated by the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation, where almost 2K people registered for participating [4]. However, the interest in algorithmic-enhanced methods for playlist generation started much before, when large catalogues of digital music became available, around the beginning of this century [1, 13, 17].

In the last 20 years, Music Information Retrieval (MIR) literature has addressed several aspects of playlists, both from a user and an algorithmic perspective [8]. In this work, we focus on the analysis at scale of large playlist datasets in order to understand how humans create a playlist in different contexts. The main hypothesis of our work is that technological innovations occurring in the last two decades have affected how people are experiencing music, and thus how music pieces are grouped together.

We center our attention on two main facets considered important in the design of playlist generation systems, according to the literature. On one hand, we address the characterization of playlists in terms of popularity, which has been a focus of several studies and related to the so-called long-tail effect [3]. On the other hand, we address the semantic diversity of a playlist, considered as the contrary to the semantic similarity concept. The trade-off between similarity and diversity has been already object of analysis [6, 7], so following a similar direction we study the diversity of semantic information in playlists.

In detail, we measure the popularity of a playlist, starting from the popularity of its component tracks. The advantage of this measure relies in the possibility to compute it without looking at the content. Furthermore, using tags retrieved from *Last.fm* [1], we define a playlist diversity index based on the semantic distance between its tracks. The distance is computed using tag embeddings, a compact and meaningful representation of user-generated annotations.

This work has three main contributions. First, we propose and implement two distinct indexes for playlist dataset characterization: one related to the concept of popularity and another one to the concept of diversity. Second, we apply, study and discuss the statistical distribution of these measures to four datasets, containing more than 400K playlists created in the last 20 years. Third, we release the data and software used for the analysis in order to foster reproducible research and future work in the topic.

The paper is structured as follows. Section 2 provides an overview of previous works related to the analysis of the playlist creation processes. We then propose an analysis methodology in Section 3, which includes a description of the model, the considered datasets and the proposed statistical measures. Section 4 provides the obtained results, which are discussed in Section 5.

2. RELATED WORK

A playlist is usually described in a very broad way as "*an ordered sequence of songs meant to be listened to as a group*", a noteworthy definition in the literature [9]. Several approaches have been proposed to build computational models of the mechanisms behind playlist creation such as Combinatorial Pattern Generation [13], Gaussian Process Regression [17], Markov Chains [5], or Hypergraph Random Walks [12] among the others. Recent studies for analyzing playlists proposed context-aware algorithms, which

[1] https://www.last.fm

	AOTM	CORN	SPOT	DEEZ
Oldest playlist	1998	2010	2012	2013
Newest playlist	2011	2011	2015	2018
# playlists	100K	15K	175K	82K
Max length	60	75K	47K	400
# playlists (w/o outliers)	97K	15K	155K	74K
Max length (w/o outliers)	33	361	109	65
Avg length	19	131	27	17
# tracks	972K	75K	2,789K	277K

Table 1. Summary of the datasets.

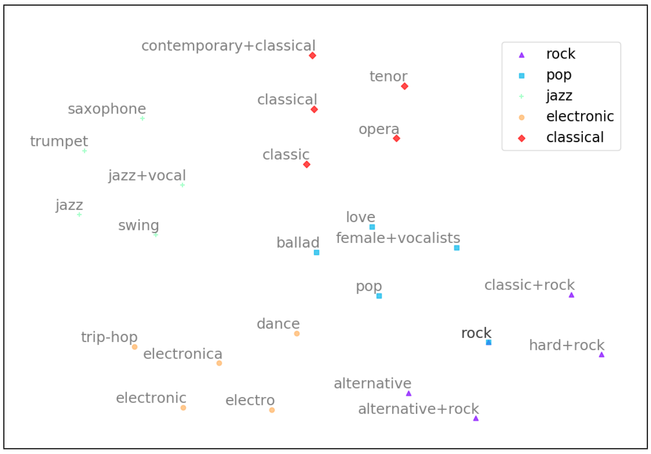

Figure 1. Top 5 similar tags found for "*rock*", "*pop*", "*jazz*", "*electronic*", "*classical*", using the tag-embeddings computed from DEEZ corpora. Distance is calculated with approximate nearest neighbor algorithm, using euclidean distance of normalized vectors, plotted using t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm.

takes into account characteristics such as the playlist title [16], and also sequence-aware algorithms, which analyze the order of songs in a playlist [20]. For an extensive review of manual, automatic and assisted techniques for playlist creation, we refer to [8].

Understanding how people create playlist is fundamental for creating computational models capable of emulating this human generation process (automatic playlist generation) [5], or predicting the most likely song to add in a given playlist (automatic playlist continuation) [4]. In addition, the intrinsic value that individuals give to a set of songs cannot be always fully explained by the analysis of the acoustic features, as tempo or tonality, emotional state or contextual information [6], and the sentence "*Making a playlist is more of an art than a science*" partly summarizes this hindrance [7]. Finally, interactive tools for supporting users during the decision-making process of playlist creation have been shown effective, but at the same time affecting human decisions. Indeed, in [11] the authors show how these kinds of tools can bias humans towards adding

AOTM	CORN	SPOT	DEEZ
Rock	Rock	Rock	Rock
Indie	Altern.	Pop	Pop
Altern.	Pop	Indie	Fem. Voc.
Pop	Jazz	Altern.	Altern.
Fem. Voc.	Fem. Voc.	Electr.	Indie
Altern. Rock	Indie	Fem. Voc.	Hip Hop
Class. Rock	Class. Rock	Hip Hop	Electr.
Indie Rock	Soul	Jazz	French

Table 2. Top tags used within each dataset (Fem. Voc.=Female Vocalist; Altern.=Alternative; Class. Rock = Classic Rock; Electr.=Electronic).

tracks more popular or more recent, in comparison to what they would independently add to the playlist. It can be considered as a consequence of the difficulty of creating models which effectively reflect human behaviors.

3. METHODOLOGY

3.1 Dataset

This study considers a total of 409K playlists (2.3M songs) from four different playlist datasets (see summary in Table 1), three of them already proposed in the literature:

1. Art of the Mix [2] (AOTM): Playlists submitted by users to the Art of the Mix website [2] .
2. Yes.com [5] (CORN): Playlists from radio stations in the United States.
3. Spotify [16] (SPOT): Playlists from Twitter's users tweeting via Spotify.
4. Deezer (DEEZ): Playlist from Deezer's users, crawled in-house.

The datasets were considered because of the heterogeneity of their nature. Indeed, they have been created using playlists from different periods, with different usage and purpose. CORN [3] provides playlists from radio stations in the United States, without restrictions on musical genres. This is the only dataset where playlists are not generated by users. AOTM [4] is the dataset containing the oldest playlists, covering a 13 years period from 1998 to 2011. It is formed by playlists submitted by users to Art of the Mix, a website where a community of playlist passionates share their creations. SPOT [5] has been composed tracking Spotify's users active in Twitter between 2012 and 2015. Similarly, DEEZ has been created using the Deezer API, selecting users playlists created between 2013 and 2018.

There are two main differences between AOTM and the other two user-generated datasets. First, SPOT and DEEZ

[2] http://www.artofthemix.org
[3] http://www.cs.cornell.edu/~shuochen/lme/data_page.html
[4] https://bmcfee.github.io/data/aotm2011.html
[5] http://dbis-nowplaying.uibk.ac.at/#playlists

Artist	Track	Tags
The Blacktop Cadence	Off track	('punk', 100), ('indie rock', 100)
The Cure	Maybe someday	('post-punk', 100), ('new wave', 92)
The Blacktop Cadence	I don't do well in social situations	('punk', 100),('indie rock', 100)
Jefferson Airplane	Today	('classic rock', 100), ('Psychedelic Rock', 98)
Husker Du	Something I learned today	('punk', 100), ('hardcore', 49)
Superchunk	Punch me harder	('indie', 100), ('college rock', 50)
Willie Bobo	Fried neck bones and some home fries	('sexy', 100), ('downtempo', 100)
Wu-tang Clan	Clan in da front	('Hip-Hop', 100), ('rap', 81)

Table 3. Example of tracks and relative tags of a playlist with low diversity ($pDI = 0.12$, *top*), and a playlist with high diversity ($pDI = 0.98$, *bottom*)

playlist creation process is embedded on particular streaming platform, while this does not apply to AOTM. Second, playlists in AOTM have been created before streaming services became intensively used worldwide [6], while DEEZ and SPOT dataset are representative of a period in which streaming technologies were already consolidated.

As pre-processing step, we filtered out playlists with less than 4 unique tracks. In addition, we excluded extremely long playlists by computing, for each dataset separately, the interquartile range (IQR) of the playlist lengths and excluding all playlists that are longer than $Q_3 + 1.5 * IQR$, where Q_3 is the 3rd quartile. Table 1 shows a summary of the dataset characteristics.

3.2 Playlist Popularity Analysis

We address the characterization of popularity by defining a popularity index for both tracks and playlists based on a set of metrics proposed in the literature.

We estimate the *Track Popularity Index* (tPI) as the track frequency within a dataset, i.e. the number of playlists in a dataset in which it occurs, as proposed in [20]. This index is normalized between 0 and 1, using min-max normalization. In order to understand how track popularity is distributed, we uniformly split each dataset into 10 different groups, according to tPI. The first one contains tracks with $tPI \in [0, 0.1)$, the second one with $tPI \in [0.1, 0.2)$, etc. We then analyze the statistical distribution of track popularity per group, using two qualitative measures proposed in the literature: *Shannon* index [18] and *Simpson* index [19]. For both indexes, 0 indicates that there is no variation in terms of popularity, while 1 indicate that popularity varies significantly within the dataset.

In addition, we define the *Playlist Popularity Index* (pPI) for a playlist p as the average of tPI for the playlist tracks t_i

$$pPI(p) = \frac{1}{len(p)} * \sum_{i=1}^{len(p)} tPI(t_i) \tag{1}$$

We compute pPI for each playlist and we then compute the $Gini$ coefficient [10] to estimate the degree of imbalance of the playlist popularity distribution for each dataset, i.e. we obtain a measure of the statistical dispersion of playlist popularity. $Gini$ coefficient is comprised between 0 and 1, where 0 express the maximum balance, which means that pPI is almost equally distributed between playlists in the dataset, while 1 represents an unbalanced situation, which means that few playlists have high pPI, while several playlists have low pPI.

In the case of playlist popularity, the choice of using *Gini* coefficient is motivated by the idea of having a value representing the influence of every playlist on the overall dataset distribution. Differently, when tracks are grouped together according to their popularity, thanks to *Shannon* and *Simpson* indexes we have an estimation of how tracks are distributed within groups.

3.3 Playlist Semantic Diversity Analysis

In order to characterize the diversity of tracks on a given playlist, we consider a semantic distance measure based on user-generated tags. For every track, we queried *Last.fm* website to retrieve its top 5 related tags. We then proceed as follows. First, we create a tag-vector representation to estimate tag distances. Second, we use a linear combination of tags-embedding distances, weighted with a tag popularity count, to obtain the distance between every two tracks of a playlist. Third, we average the distances between tracks to yield a final diversity estimation for a playlist, according to the retrieved tags.

Finally, we obtain two metrics: 1) a distance between tracks, based on tag-similarity; 2) a playlist diversity index, based on the distance variations between tracks. We provide more details on the process in the next sections.

3.3.1 Tag-embeddings

Recent developments of NLP techniques have shown how particular types of words vector representation can bring with them valuable semantic information. In our study, we select the *GloVe* [15] learning algorithm [7], an architecture that exploits the ratio of word-word co-occurrences probabilities within a corpora for generating tag embeddings.

[6] https://www.ifpi.org/downloads/GMR2016.pdf

[7] https://nlp.stanford.edu/projects/glove

We choose this representation because of its compact form, the low computation cost needed for training the model with new corpora, and the facility to compute a distance metric between embeddings. For training the model, we combine the retrieved tags from *Last.fm* to create a corpus of track tags, and we use it to generate a vector representation for each tag.

In Table 2, we report the most frequent tags for the four datasets. Most of them are shared between datasets, so we study the few not shared tags to better understand their peculiarities. As an example, *"French"* only appear in DEEZ, having Deezer been founded in France. Furthermore, we observe how *"Electr."* and *"Hip Hop"* tags only appear in SPOT and DEEZ. The rise of commercial music in these these two genres in the last decade may be reflected in their extended presence.

Figure 1 shows an example of similar tags, where the distance has been computed using the trained tag-embeddings. *"rock"* and *"pop"* clusters are quite near, as *"jazz"* and *"classical"*. Within *"electronic"* similar tags, *"dance"* is the nearest one to *"pop"*. Within the *"jazz"* cluster, there are *"trumpet"* and *"saxophone"*, two instruments often related to this genre. These are some examples of observations that can be derived, and which reflect semantic information contained in the computed vector representation.

3.3.2 Playlist Track-Tag diversity index

We define a track *tr* as linear combination of its T weighted tags:

$$tr = \sum_{i=1}^{T} w_i * tag_i \tag{2}$$

where in our settings $T = 5$. Basing on (2), we define a distance measure between tracks, named *Track-Tag distance* (d_{TT}), as follow

$$d_{TT}(tr^{(1)}, tr^{(2)}) = \frac{1}{T} * \sum_{i=1}^{T} W_i^{(1,2)} * d_{tag}(tag_i^{(1)}, tag_i^{(2)}) \tag{3}$$

where the weight term is

$$W_i^{(1,2)} = \frac{w_i^{(1)} + w_i^{(2)}}{2 * max(w_i^{(1)}, w_i^{(2)})} \tag{4}$$

and the distance between two tags is defined as

$$d_{tag}(tag_1, tag_2) = \sqrt{2 * (1 - cos(tag_1, tag_2))} \tag{5}$$

$cos(tag_1, tag_2)$ represents the cosine similarity between tag-embeddings. The computation of the tag distances has been carried out with the Annoy Python library [8], which makes use of the approximate nearest neighbor technique for an efficient computation of the euclidean distance of normalized vectors. In the d_{TT} formula, $W_i = 1$ if $w_i^{(1)} = w_i^{(2)}$, so weights do not impact the distance. Otherwise, $W_i \in (0.5, 1)$, hence final distance decreases when multiplying the weight term with the tag distance. In detail, d_{TT} is near to 0 when two tracks have a high similarity, while it is around 1 when they are very different, according to their combination of user-generated tags.

For understanding a playlist diversity in terms of semantic annotations, we first compute the d_{TT} distance for every combination of two tracks in the playlist. Summing the distances and dividing by the number of total possible combinations, we obtain the *Playlist Track-Tag diversity index* (pDI):

$$pDI(playlist) = \frac{2}{M(M-1)} * \sum_{i,j} d_{TT}(tr_i, tr_j),$$
$$\forall tr_i, tr_j \in playlist, j > i$$

where M is the length of the playlist. pDI is near to 0 when there is a low diversity between tracks within the playlist, and almost 1 when tracks are extremely diverse, according to the d_{TT} distance. Table 3 provides two examples of playlists with different pDI values. The playlist with low pDI is formed by tracks mainly tagged as *"punk"*, *"rock"* or *"indie"*, while the playlist with high pDI has tracks with tags more diverse, passing from *"punk"* to *"downtempo"*, to *"hip hop"*.

4. RESULTS

4.1 Playlist Popularity Analysis

Results of the popularity analysis are shown in Table 4. We first observe the great differences between radio playlists from the CORN dataset and user-generated playlists from the other datasets. Tracks' popularity tPI in CORN varies significantly more than in the other cases, according to the *Shannon* and *Simpson* indexes, and the percentage of tracks with $tPI \in [0.0, 0.1)$ is smaller, indicating a large presence of popular tracks within the dataset. The mean of playlist popularity pPI is not extremely high, but it is more balanced than for DEEZ, SPOT and AOTM according to the *Gini* coefficient. Results can be interpreted as a consequence of the nature of radio playlists. Indeed, tracks rotation in radios is often constrained by commercial policies, because artists, or someone in their behalf, have to pay for broadcasting their tracks. This clearly makes difficult for artists with few resources to be on air in a radio. This phenomena is reflected in having few tracks from the long-tail, i.e. less popular, inserted in radio playlists. The balanced playlist popularity level also derives from the policy of alternating popular tracks with less know ones [9].

Regarding user-generated playlists, we observe how track popularity in DEEZ and SPOT are quite similarly distributed. On the contrary, in AOTM the presence of 99.99% of track with $tPI \in [0.0, 0.1)$ influences both the *Shannon* index, the *Simpson* index and the *Gini* coefficient, creating an unbalanced situation with no diversity in terms of track popularity. However, results of playlist popularity analysis give similar values among the three user-

[8] `https://github.com/spotify/annoy`

[9] `https://www.digitalmusicnews.com/2015/02/19/five-things-internet-radio-steal-broadcast-radio`

	AOTM	CORN	SPOT	DEEZ
Top tPI	4368	1746	1270	2523
Top tPI (normalized by dataset size)	0.045	0.112	0.008	0.034
Track with $tPI \in [0.0, 0.1)$ (%)	99.99	96.09	99.85	99.74
tPI Shannon index	$3.1{\cdot}E^{-5}$	0.212	0.013	0.020
tPI Simpson index	$4.3{\cdot}E^{-6}$	0.076	0.003	0.005
Avg pPI (normalized by Avg playlist length)	1.19	2.01	1.76	9.64
pPI Gini coefficient	0.66	0.39	0.67	0.55

Table 4. Summary of playlist popularity analysis results.

	AOTM		CORN		SPOT		DEEZ	
	Original	Random	Original	Random	Original	Random	Original	Random
Mean	0.68	0.72	0.84	0.85	0.58	0.68	0.63	0.82
Std	0.12	0.10	0.09	0.03	0.19	0.11	0.19	0.11
Max	0.99	0.98	1.05	0.97	1.13	0.99	1.14	1.09
Min	$1.3{\cdot}E^{-5}$	0.21	0.33	0.69	$7.3{\cdot}E^{-7}$	0.22	$1.9{\cdot}E^{-5}$	0.27
Gini	0.10	0.08	0.06	0.02	0.19	0.09	0.17	0.08
QCD	0.11	0.10	0.07	0.03	0.21	0.12	0.22	0.09

Table 5. Playlist Track-Tag diversity index (pDI) descriptive statistics. QCD indicates the quartile coefficient of dispersion, computed as $QCD = (Q3 - Q1)/(Q3 + Q1)$, where Q_1 and Q_3 are the first and third quartiles.

generated datasets, where only the DEEZ stands out for having a high average value of pPI.

In general, the popularity of a playlist, intended as average of the frequency of its tracks within a dataset, can be influenced by several factors. As example, AOTM dataset has been created with playlists from 1998 to 2011, when music was consumed by means of different services than the ones which lead the market today. We suppose that playlists in AOTM do not often come from the interaction with a large music catalogues, or with algorithms for facilitating music search and discovery for playlist generation, and this can be related to a major presence of less popular tracks. Furthermore, current streaming services employ several tools to facilitate playlist sharing, to make this a collaborative process, and to incorporate tracks of a playlist into new playlists [14]. In terms of popularity, the possibility to share a playlist can have a positive impact, increasing the accessibility to much more content and then reducing the number of less popular tracks.

4.2 Playlist Semantic Diversity Analysis

We faced two limitations when retrieving tags from *Last.fm*: 1) we did not find tags for all queried tracks; 2) for part of the tracks, the associated tags were small, less than five. As a consequence, we follow a conservative approach when computing the Playlist Track-Tag diversity index: 1) we only consider playlists for which all tracks have associated tags (*complete information*); 2) tracks are only compared with other tracks with the same number of tags (*balanced information*).

After obtaining the semantic index for each playlist, we compute descriptive statistics for understanding how the computed descriptor characterizes these datasets. In addition, for every case we also extract the same statistics on playlists of average size, created with random tracks from the original playlists. In Table 5, we observe part of the differences between datasets.

As in the previous sections, the analysis on CORN radio playlists provides values of different order of magnitude than user-generated playlists. Indeed, the mean and the minimum value of the diversity index pDI are higher for this dataset. This can be related to the fact that radio playlists are rarely composed by tracks from the same artist, as it is the case for low diversity playlists according to our analysis. Moreover, we observe that CORN playlists are more balanced in terms of tag-similarity, as they are in terms of track popularity, as represented by a low *Gini* coefficient and low quartile coefficient of dispersion.

In order to better understand how the diversity index represents playlists' characteristics, we have carried out a qualitative analysis of the 10% of playlists with higher, and 10% with lower diversity. For every groups of playlists, we compute the following values, reported in Table 6: 1) average of Playlist Track-Tag index (Avg pDI); 2) average of unique tags for playlist (Avg tag count); 3) number of playlists with at least one tag in common between all the tracks (Common tags); 4) average of unique tags over tracks (Avg Tag/Tracks); 5) average of unique artists for playlist (Avg artist count); 6) playlist with tracks from the same artist (Single-Artist); 7) average of unique tracks for playlist (Avg track count) ; 8) average of unique tracks over artist (Avg Tracks/Artist).

From the analysis of these values, we confirm previous observations on the pronounced difference between CORN and the other datasets. Looking at the percentage difference of each parameter, we see how CORN playlists span

	AOTM			CORN			SPOT			DEEZ		
	Low	High	PD	Low	High	PD	Low	High	PD	Low	High	PD
Avg pDI	0.43	0.85	64	0.65	0.98	40	0.19	0.87	128	0.28	0.93	108
Avg Tag Count	20	44	75	25	25	0	15	30	67	11	22	67
Common Tags (%)	34	0	34	0	0	0	23	1	22	44	1	43
Avg Tag/Tracks	1	3	100	3	4	29	1	3	100	2	3	40
Avg Artist Count	6	15	86	9	7	25	2	7	111	2	5	86
Single-Artist (%)	55	2	53	0	0	0	83	22	61	69	7	62
Avg Track Count	17	16	6	10	7	35	12	11	9	8	7	13
Avg Tracks/Artist	11	1	167	1	1	0	11	4	93	6	2	100

Table 6. Qualitative analysis results of low/high 10% playlists, ranked by their pDI. Column "PD" reports the percentage difference (in %) between low and high cases values, calculated as $PD(a,b) = 100 * \frac{|a-b|}{(a+b)/2}$

a small range of diversity, in comparison to user-generated playlist datasets. SPOT and DEEZ present more variation in terms of diversity, reflecting the values obtained for the quartile coefficient of dispersion, presented in Table 5.

In general, these parameters are coherent with the analysis carried out before: playlists with a low pDI, hence with less diversity, present in average a smaller number of unique tags, more tags in common between tracks, few artists for playlist and a higher percentage of single artist playlists. Playlists with a high index present the inverse characteristics.

5. CONCLUSIONS

We have presented a statistical analysis of more than 400K playlists (2.3M songs) from four different datasets, three composed by user-generated playlists, while one, CORN, composed by radio playlists. Two of the user-generated datasets, SPOT and DEEZ, have playlists created between 2012 and 2018, while AOTM playlists between 1998 and 2011. We develop our analysis using descriptive statistics, and in addition we make use of indexes from the information retrieval literature for evaluating the distribution of the analyzed features within the sets. In particular, we focused on two aspects: popularity and diversity.

From the proposed metrics, we observe how differences between datasets emerge, reflecting the distinct context in which playlists have been created. On one side, radio playlists analysis shows clear different results from the ones obtained from user-generated playlist. On the user-generated side, we observe how the study of AOTM playlists reveals different characteristics than for SPOT and DEEZ datasets. Behind this fact, we identify as possible cause the change of music listening behaviours, shifting from the idea of personal music repositories in the beginning of the digital era, to the dominance of streaming services of today. We hypothesize that this paradigm change has also impacted the way users create playlists, and our results partially reflect this shift.

In our analysis, we have found a more balanced situation in SPOT and DEEZ datasets in terms of popularity, although they contain playlists with a high level of diversity in terms of semantic tags. Even if different explanations can be at the root of the different values, e.g. the larger song search space of streaming services, the lower cost to create and share a new playlist online, or the recommendation algorithms that support playlist creation, we cannot identify a specific cause with our analysis.

The proposed methodology can be applied to characterize playlists in terms of popularity and semantic diversity, allowing the comparative analysis of human-generated and algorithm-generated playlists in different contexts such as historical periods, platforms and musical genres. We find extremely valuable to compare different playlist datasets, as it allows to understand how changes in the listening experience are affecting playlist creation strategies.

We hypothesize that if we extend this analysis to a larger number of datasets, we would achieve a better understanding of these changes. For instance, one of the limitations of the considered datasets is that they provide a Western-centric view. Adding playlist creators country information could enrich our study. Similarly, a yearly-based temporal analysis would help to better understand temporal variations. Moreover, adding other content- and context-based features from playlists can help to explore factors that are hidden in the presented analysis. Finally, it has already been shown that considering the tracks ordering is helpful for extracting playlist characteristics [20], so we plan to include this information in future research.

To facilitate the reproducibility and transparency of our study, the data and the software used are made publicly available [10] .

6. ACKNOWLEDGMENTS

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

7. REFERENCES

[1] J.-J. Aucouturier, and F. Pachet. "Scaling up music playlist generation", *IEEE International Conference on Multimedia and Expo* , pp. 105–108, 2002.

[10] `https://github.com/MTG/playlists-stat-analysis`

[2] A. Berenzweig, B. Logan, D.P.W. Ellis, and B. Whitman. "A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures", *Computer Music Journal*, Vol. 28, No. 2, pp. 63–76, 2004.

[3] O. Celma. *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*, Springer, 2010.

[4] C.-W. Chen, P. Lamere, M. Schedl, and H. Zamani. "Recsys challenge 2018: automatic music playlist continuation", *Proc. of the 12th ACM Conference on Recommender Systems*, pp. 527–528, 2018.

[5] S. Chen, J.L. Moore, D. Turnbull, and T. Joachims. "Playlist prediction via metric embedding", *Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, 2012.

[6] K. Choi, G. Fazekas, and M. Sandler . "Understanding Music Playlists", *Machine Learning for Music Discovery Workshop at the 32nd International Conference on Machine Learning*, 2015.

[7] S. Cunningham, D. Bainbridge and, A. Falconer. "'More of an art than a science': Supporting the creation of playlists and mixes", *Proc. of the 7th International Conference on Music Information Retrieval*, pp. 240–245, 2006.

[8] R. Dias, D. Gonçalves, and M.J.Fonseca. "From manual to assisted playlist creation: a survey", *Multimedia Tools and Applications*, Vol. 76, No. 12, pp. 14375–14403, 2017.

[9] B. Fields, P. Lamere, and N. Hornby. "Finding a path through the jukebox: the playlist tutorial", *Proc. of the 11th International Society for Music Information Retrieval Conference*, 2010.

[10] C. Gini. " Concentration and dependency ratios (in Italian)", *English translation in Rivista di Politica Economica*, Vol. 87, No. 1997, pp. 769–789, 1909.

[11] I. Kamehkhosh, Di. Jannach, and G. Bonnin. "How automated recommendations affect the playlist creation behavior of users", *CEUR Workshop Proc.*, 2018.

[12] B. McFee, and G. Lanckriet. "Hypergraph models of playlist dialects", *Proc. of 13th International Society for Music Information Retrieval Conference*, pp. 343–348, 2012.

[13] F. Pachet, P. Roy, and D. Cazaly. "Combinatorial approach to content-based music selection", *IEEE Multimedia*, Vol. 7, No. 1, pp. 44–51, 2000.

[14] Y.S. Park, and B. Kaneshiro. " An Analysis of User Behavior in Co-Curation of Music Through Collaborative Playlists", *Extended Abstracts for the Late-Breaking Demo Session of the 18th International Society for Music Information Retrieval Conference*, 2017.

[15] J. Pennington, R. Socher, and C. Manning. "Glove: Global Vectors for Word Representation", *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543, 2014.

[16] M. Pichl, E. Zangerle, and G.Specht. "Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name?", *Proc. of the 15th IEEE International Conference on Data Mining Workshop*, pp. 1360–1365, 2016.

[17] J. C. Platt, C. J. C. Burges, S. Swenson, C. Weare, and A. Zheng. "Learning a Gaussian process prior for automatically generating music playlists", *Advances in Neural Information Processing Systems*, Vol. 2, No. 14, pp. 1425–1432, 2002.

[18] C. Shannon. "A Mathematical Theory of Communication", *The Bell System Technical Journal*, Vol. 27, No. 1948, pp. 379–423, 1948.

[19] E. H. Simpson. "Measurement of diversity", *Nature*, Vol. 163, No. 688, 1949.

[20] A. Vall, M. Schedl, G. Widmer, M. Quadrana, and P. Cremonesi. "The importance of song context in music playlists: Enabling recommendations in the long tail", *CEUR Workshop Proc.*, 2017.