International Journal of Multimedia Data Engineering and Management
Volume 8 • Issue 4 • October-December 2017

Understanding User-Curated Playlists on Spotify:
A Machine Learning Approach

Martin Pichl, University of Innsbruck, Innsbruck, Austria

Eva Zangerle, University of Innsbruck, Innsbruck, Austria

Günther Specht, University of Innsbruck, Innsbruck, Austria

ABSTRACT

Music streaming platforms enable people to access millions of tracks using computers and mobile devices. However, users cannot browse manually millions of tracks to find music they like. Building recommender systems suggesting music fitting the current context of a user is a challenging task. A deeper understanding for the characteristics of user-curated playlists naturally contributes to more personalized recommendations. To get a deeper understanding of how users organize music nowadays, we analyze user-curated playlists from the music streaming platform Spotify. Based on the audio features of the tracks, we find an explanation of differences in the playlists using a PCA and are able to group playlists using spectral clustering. Our findings about playlist characteristics can be exploited in a SVD-based music recommender system and our proposed clustering approach for finding groups of similar playlists is easy to integrate into a recommender system using pre- or post-filtering techniques.

KEYWORDS

Clustering, Data Acquisition, Data Analysis, Information Retrieval, Machine Learning, Music Information Retrieval, Quantitative Study, User Modeling, User-Generated Content

INTRODUCTION

In the last decade, new technologies have paved way for new distribution channels for digital content (e.g., music streaming platforms like Spotify[1] or Apple Music[2]). At the same time, mobile devices as smartphones or tablets enable their users to access millions of tracks on those streaming platforms in various situations throughout the whole day. These developments make music organization a highly interesting topic: the challenge for the users is to find music they like in the overwhelming variety of music offered by music streaming platforms. In principle, users need to navigate through their music collection to find the music they aim to listen to during different activities or situations (Kamalzadeh, Baur, & Möller, 2012). In order to assist users in browsing these possibly extensive collections, streaming platforms heavily rely on recommender systems, but also on human editors. A deeper understanding for the characteristics of playlists, in particular how users curate their playlists can naturally contribute to more personalized and better recommendations.

In the field of music listening behavior analyses and recommender systems, social media platforms are exploited to gather relevant data for such analyses. Nowadays, a substantial number of people share what they are listening to at the moment using so-called #nowplayling tweets on Twitter. This

DOI: 10.4018/IJMDEM.2017100103

makes Twitter, which is the world's leading micro-blogging platform serving 320 million active users[3], a valuable data source. Twitter has already been exploited for various analyses of user listening behavior (Hauger, Schedl, Košir, & Tkalčič, 2013; Zangerle, Pichl, Gassler, & Specht, 2014) as well as for recommender systems (Pichl, Zangerle, & Specht, 2015; Schedl & Schnitzer, 2014; Zangerle, Gassler, & Specht, 2012). Earlier, automatic playlist generation, as a form of music recommendation, was studied intensively (Alghoniemy & Tewfik, 2001; Aucouturier & Pachet, 2002; Flexer, Schnitzer, Gasser, & Widmer, 2008; Logan, 2002; Pampalk, Pohle, & Widmer, 2005; Pauws & Eggen, 2002). In their analysis of user data derived from WebJay, a former web-based playlist service, Slaney and White (2006) found that people prefer different types of music and that users create playlists biased to these types of music. Furthermore, Cunningham et al. (2004) have shown that people categorize music after the intended use. Complementary to this, Kamalzadeh et al. (2012) found that people categorize music by activities and/or the mood in their music libraries. However, a problem for the general applicability of those qualitative studies is the small dataset in terms of users, playlists and listened tracks. In order to overcome this data problem, we exploit a recently published dataset of Spotify users (Pichl, Zangerle, & Specht, 2016). This dataset enables a profound quantitative analysis of the musical attributes of the tracks forming up different playlists.

In contrast to the well-researched field of automatic playlist generation, we aim to deepen our understanding for the characteristics of playlists created by human users and hence, shift our focus from automatic playlist generation to the analysis of playlists. To conduct this study, we require a data set containing information about users and their playlists. In a previous analysis we found that a substantial portion of so-called #nowplaying tweets refer to Spotify (Pichl, Zangerle, & Specht, 2014). In this work, we exploit a data set containing the subset of the Spotify users of the #nowplaying dataset and their playlists (Pichl et al., 2016). In total, we base our analyses on 1,137 users and their 18,296 playlists. We are particularly interested in studying the musical attributes of the tracks forming up different playlists. Therefore, we utilize the Echo Nest acoustical attributes contained in the dataset. Our analyses based on this data set are particularly driven by the following research questions (RQ):

- **RQ1:** How can we observe and explain acoustical differences between playlists using clustering techniques?
- **RQ2:** How do users utilize playlists of different types to organize their music?

The main contribution of this work is a quantitative analysis of the playlist generation behavior of Spotify users using machine learning techniques. We find that using a Principal Component Analysis (Pearson, 1901), we are able to explain differences using content-based music features. In a next step, we use spectral clustering to cluster playlists according to their musical features into five cluster. We determine the number of clusters by an analysis of the explained variance and an analysis of the eigenvalues. We observe that on average, each user creates playlists within three different clusters. Moreover, we observe, that 19.94% of all users create playlists in all five clusters, suggesting that users arrange different styles of music in different playlists. Complementary to that, we find that although nearly half of the users create playlists with classical and rap-style music, these playlists account only for 8 and 7% of all playlists. Moreover, we detect a cluster where 91% of all users create playlists in as it contains a form of "feel-good" popular music, serving as a common musical ground across all users. Our analyses also show that people do not necessarily group their music by genre. We consider the insights gained in this work to be useful for improved automatic playlist generation and music organization.

The remainder of this article is structured as follows: In the following section, we present works related to the presented analyses. Section 3 subsequently introduces the data set and in Section 4, we briefly present the methods used to analyze user-created playlists. Section 5 presents the results of the conducted analyses, which are further discussed in Section 6 where we also point out future work. Section 7 concludes this article.

RELATED WORK

In literature, several studies about music organization can be found. Cunningham et al. (2004) conducted a study on how people organize CDs and MP3 files, based on interviews and on-site observations of focus groups. They found that facilities for creating playlists are a demanded feature. In a later study, based on an online survey, Kamalzadeh et al. (2012) found that people prefer a minimal amount of interaction. At the same time, users want the music to match their mood and want to be able to change the mood of the music played (Kamalzadeh et al., 2012). As for minimizing the required interaction with music systems, the automatic generation of playlists was studied intensively starting from the early 2000s. We categorize these approaches into (i) approaches mainly utilizing content or metadata and (ii) hybrid approaches incorporating user feedback in addition to the data sources mentioned before.

With respect to (i), there are approaches that facilitate a seed song along with the traditional k-nearest neighbors approach to find similar songs to the given start song (Logan, 2002). Further, approaches in which the user selects a start and an end song with a smooth transition in between (Flexer et al., 2008) and approaches based on user-defined constraints (Aucouturier & Pachet, 2002) have also been proposed. The used constraints may be content-based, i.e., the tempo of a song, or based on meta-information like the genre (Alghoniemy & Tewfik, 2001; Aucouturier & Pachet, 2002). With respect to (ii), we find approaches incorporating the contexts-of-use. In this case, metadata of tracks is used to cluster similar songs to playlists and users were asked to judge the suitability of this cluster for certain contexts-of-use (Pauws & Eggen, 2002). Besides this, also the skipping behavior combined with content-based features has been exploited. Skipping a song as an indicator for dislike is used in order to avoid adding songs with the same content-based features to the playlist as the skipped one (Pampalk et al., 2005).

Following up this prior research, in this work, we focus on how users facilitate their playlists on the music streaming platform Spotify. In contrast to Cunningham et al. (2004) and to Kamalzadeh et al. (2012), we approach this topic quantitatively using a broad user base gathered from the Spotify platform. This is done in order to lay a foundation for future music databases and libraries, recommender systems or new forms of playlist generation.

DATA SET

In this section, we provide details about our data set as well as the methods utilized for the performed analyses, before discussing and interpreting the results in the subsequent section.

For the analyses presented in this work, we use a novel, publicly available data set of Spotify users (Pichl et al., 2016). We use this dataset, as to the best of our knowledge it is the only publicly available dataset containing tracks along their audio characteristics organized in playlists. The dataset contains 1,133 Spotify users, organizing 796,024 distinct tracks in 18,296 playlist. On average, the data set features 18.25 playlists with a standard deviation (SD) of 19.07 and 1,084.07 (SD = 2,659.45) tracks per user. Furthermore, it contains the audio summary Echo Nest in terms of acoustical features. Hence, for each track it contains the *danceability*, *energy*, *loudness*, *speechiness*, *acousticness*, *liveness* and *tempo*. We give a statistical summary of the dataset in Table 1.

Table 1. Data set statistics

Measure	Mean	SD	Median
Average Tracks per User	1,084.07	2,659.45	478
Average Playlists per User	18.25	19.07	11.00
Average Tracks per Playlist	75.13	945.20	16.00

Data Cleaning and Aggregation

As we aim to get a deeper understanding for music playlists, we have to filter for musical tracks within our data set. Thus, we restrict the data set to tracks with a *speechiness* of 0.66 or below. The tracks with a *speechiness* that is higher than 0.66 are most likely audio books, according to the Spotify API documentation. To analyze the acoustic features of each playlist, we aggregate the acoustic features of the individual tracks for each playlist in the data set using the arithmetic mean. To show the dispersion of the tracks forming a playlist, we state the mean as well as the mean absolute deviation (MAD) of each acoustic attribute in Table 2. We make use of the MAD as it is a robust measure with respect to outliers (Leys, Ley, Klein, Bernard, & Licata, 2013). The table shows that except for loudness, the variance of each of the acoustic characteristics of the tracks inside a playlist is low and the MAD is rarely higher than the mean. Thus, we can conclude that aggregating the characteristics of the individual tracks to playlist characteristics using the mean is representative. Further, we argue that aggregating the loudness of the individual tracks to a playlist loudness is not reasonable: the variance among the loudness in the tracks of a playlist is too high. In 99.99% of all cases, the MAD is higher than the mean. Therefore, we drop the loudness characteristic for the conducted playlist analyses. Furthermore, as a principal component analysis (PCA) analysis works optimal with normal distributed variables, we performed a Shapiro-Wilk-Test (Royston, 1992). For all variables the test strongly indicates (p-value < 0.01) that the acoustic features are normally distributed (Royston, 1995).

METHODS

In a first step, we aim to identify variables that explain most of the variance in the data set and hence, differences in the user-generated playlists in regard to acoustic features, which reflects RQ1. In order to find these variables, we conduct a PCA (Pearson, 1901). The PCA is based on the standardized matrix to avoid problems with different scales. We compute the principal components (PCs) using

Table 2. Aggregated acoustical features - variance per playlist

Attribute	Number of tracks with MAD > Mean	Relative Portion
tempo	0	0.00%
energy	61	0.34%
speechiness	39	0.21%
acousticness	1,392	7.67%
danceability	2	0.01%
loudness	18,145	99.99%
valence	101	0.56%
instrumentalness	978	5.39%

the correlations matrix in contrast to the covariance matrix, a common method for conducting PCAs (Jolliffe, 1986). In a further analysis, we make use of k-Means clustering (MacQueen & others, 1967) to aggregate playlists into groups (or types). We estimate k using the PCA conducted in the first step as proposed by Ding and He (Ding & He, 2004).

PCA and k-Means are methods to explain linear relationships. To additionally find non-linear relationships and thus different (or more) groups of playlists, we apply normalized cut spectral clustering as proposed by (Shi & Malik, 2000). Spectral clustering is a graph-based approach, where the normalized cut criterion partitions the graph into groups with a high intra-group similarly and simultaneously a low inter-group similarity (Shi & Malik, 2000). The different clustering methods are applied aiming to answer RQ2 and hence, target at finding certain types of playlists. To find user types creating such playlists, we rely on several correlation and similarity measures as we aim to find correlations between users creating certain playlists in certain clusters.

In the next section, we presented the application of the introduced methods to the presented dataset.

RESULTS

In this section, we present the results of the analyses conducted using the methods described in Section 4. Firstly, we elaborate the results regarding RQ1, finding groups of playlists, before focusing on the users and thus, on RQ2.

Groups of Playlists

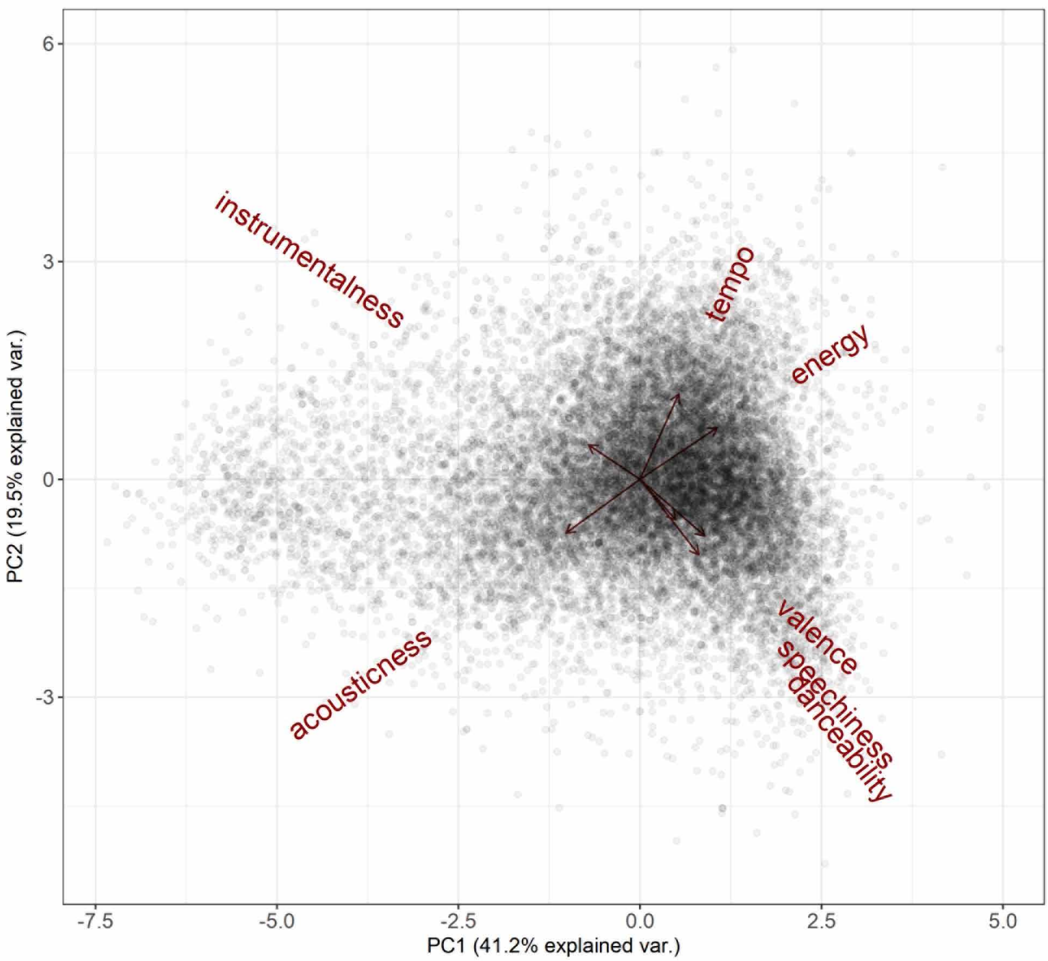
Based on the aggregated data set described in the preceding section, we conduct a PCA. Figure 1 depicts a biplot of the first two Principal Components (PCs), where each playlist is represented as a dot. This allows analyzing half of the variation within the playlists data set. The first PC on the x-axis distinguishes *acoustic* and *instrumental* playlists from playlists focusing on t*empo* and *energy* as well as playlists focusing on *valence* and *danceability*. This is, as the loading vector of PC1 only has negative signs for *acousticness* and *instrumentalness* and thus contrasts those two attributes from the other attributes. By only using the first PC, we are able to explain 27% of the variation.

Analogously, we observe that the second PC on the y-axis divides more *instrumental* playlists and playlists with high *tempo* and energy from playlists which are more *acoustic* as well as playlists with high *danceability*, *valence* and *speechines*s values. Again, this is as the loading vector of PC2 has negative signs for latter attributes, whereas the former three attributes are positively signed. By using the second PC, we are able to explain another 19% of the variation. By using our web-based analysis tool[4], we allow multimedia researchers to investigate arbitrary combinations of PCs. In this work, we complement our analysis by looking at PC3: PC3 separates tracks with high *speechiness* values from the rest. Using the first three PCs, we are able to explain 61% of the variance. Each further added PC adds 10% or less explained variance.

Based on the findings of the conducted PCA, we aim to partition our set of playlists into clusters of playlists: *instrumental* and *acoustic* playlists, playlists focusing on *valence* and *danceability* along with *speechiness* and playlist focusing on tempo and *energy*. Hence, we apply k-Means clustering with k = 3 to k = 7. Clustering into 3 clusters leads to clusters that are based on the first two PCs (as described above), whereas clustering into 7 clusters leads to clusters based on each of to the 7 acoustical features. This is shown in Figure 2, where different k-Means solutions are plotted for different k. Each point represents a playlist, plotted against PC1 and PC2. The color and shape of the points represent the cluster membership.

In order to formally determine the optimal number of clusters for our next analyses, we rely on the wide spread method utilizing the gap statistic (Tibshirani, Walther, & Hastie, 2001). This method is based on the Elbow Curve (Bishop, 2006) or rather on the idea that it is important how much the within-cluster sum of squares (WCSS) decreases with an increasing number of clusters, as the WCSS naturally decreases with the number of clusters. In our approach, the gap statistic indicates that five

Figure 1. Biplot using PC1 and PC2

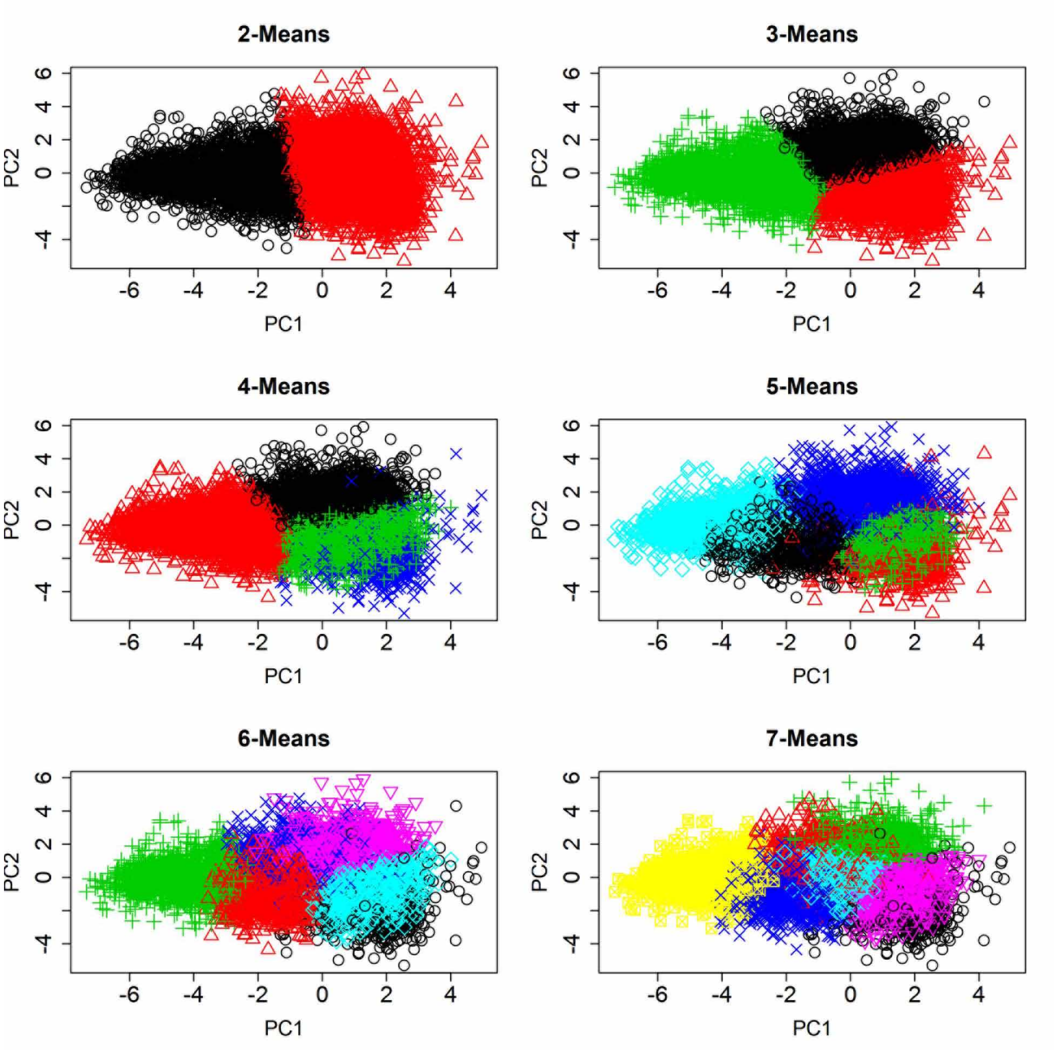


clusters are an appropriate solution. We confirm the result by plotting the "Elbow Curve", which is in our case plotting the number of clusters vs. the WCSS.

In a further analysis, we apply spectral clustering as introduced in Section 4. We apply this method in order to find non-linear separable groups of playlists. By considering the explained variance to determine the number of clusters (as for k-Means), we observe that analogously to k-Means, also for spectral clustering the optimal number of clusters is five. However, as we can observe in Figure 3, we get partly different solutions, although the partitioning based on the acoustical features (explained by the PCA) is similar. Hence, the groups of playlists share the same characteristics, but the assignment of playlists to the groups differs. This is, as the thresholds for the acoustical features determining the cluster memberships are different.

To compare our two clustering strategies, we rely on R^2 computed as $R^2 = \frac{BSS}{TSS}$, where BSS is the between cluster sum of squares and TSS is referred to as the total sum of squares. We find

Figure 2. k-Means for k between 2 and 7

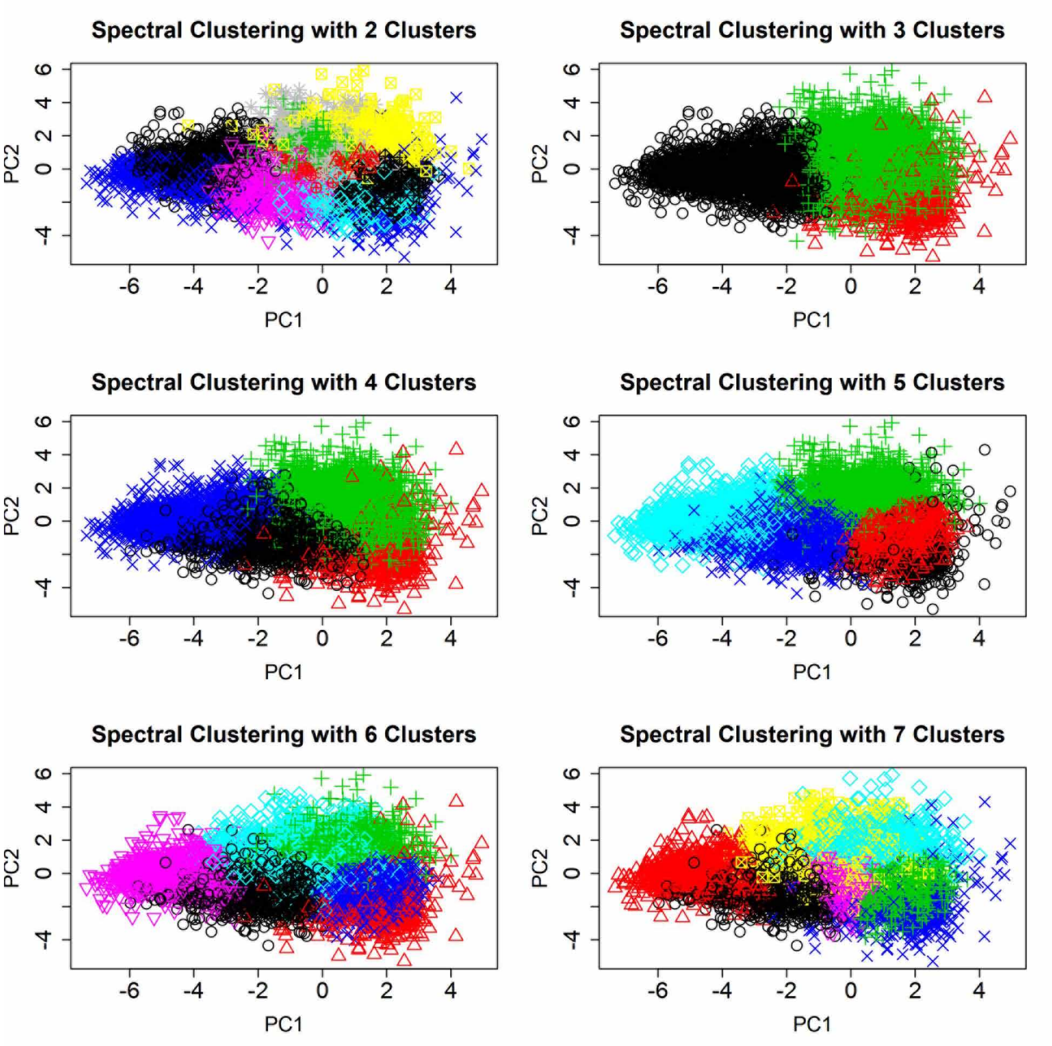


that spectral clustering delivers a 53.7% better result ($R^2 = 75.8\%$) compared to k-Means clustering ($R^2 = 49.3\%$). Hence, we argue that k-Means and spectral clustering find the same cluster archetypes, however the cluster assignment is more precise if spectral clustering is used.

In a next step, we aim to get an overview of the acoustical attribute characteristics of the five clusters. Therefore, we visualize these as a star diagram for each cluster as shown in Figure 4. This diagram shows the different features and their manifestation in the five clusters computed using spectral clustering.

Cluster 1 contains tracks focusing on *energy* and *tempo*, whereas cluster 2 contains tracks with high *speechiness*, *energy*, *valence* and *danceability* values. Cluster 3 is rather similar to cluster 2, besides the high *speechiness* values. This is, as the former one contains mostly rap music, in contradiction to the latter, which contains different forms of pop music. This observation is underpinned by the genre distribution as discussed in Section 5.2. Furthermore, we witness that high *danceability* values correlate with high *valence* values (Clusters 2 and 3). Cluster 5 contains tracks focusing on *acousticness*

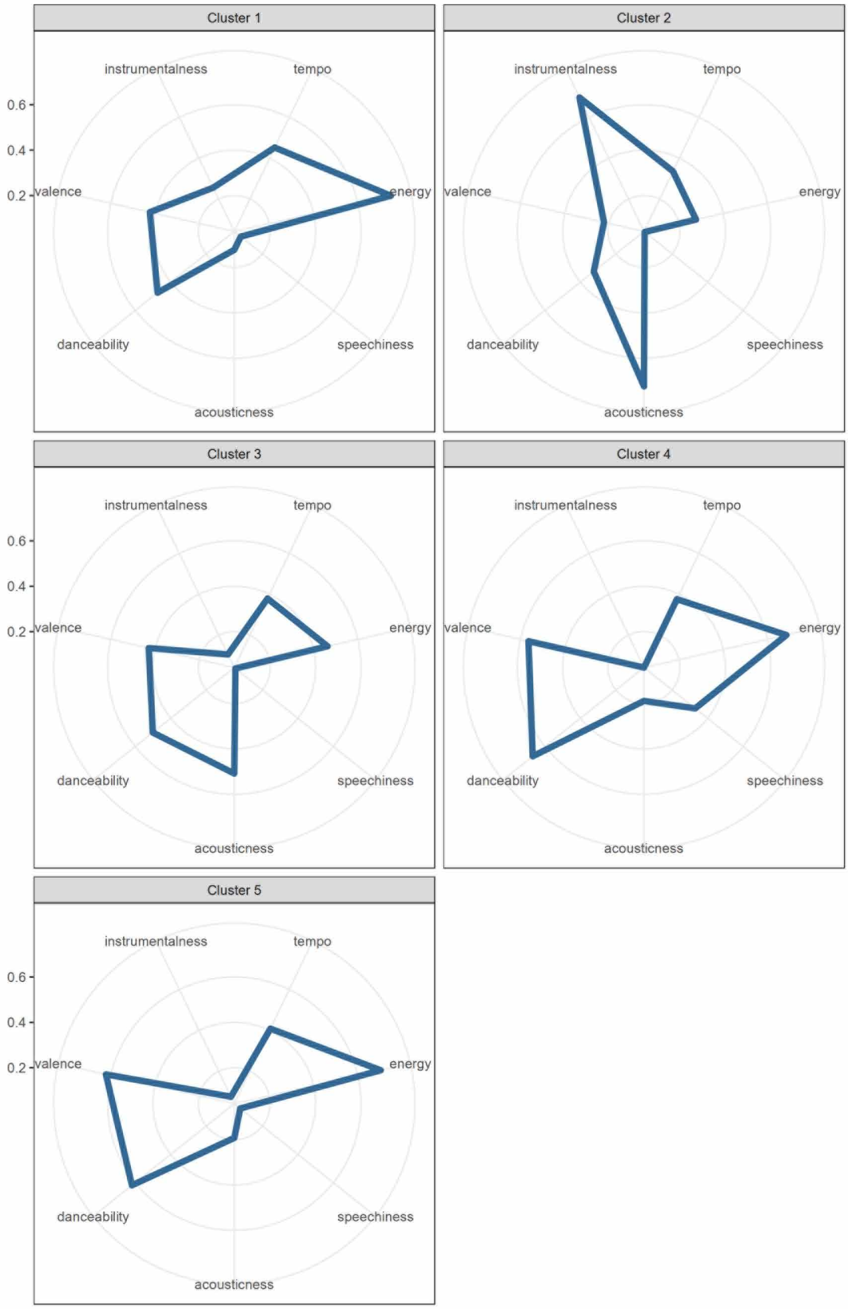
Figure 3. Spectral clustering from 2 to 7 clusters



and *instrumentalness* as this cluster mostly contains classical music. Again, this is reflected in the genre distribution.

To answer RQ1, there exist differences based on the audio characteristics of playlists. By conducting a PCA, we are able to explain 60% of the variance using the first three PCs: we observe that the first PC separates *acoustic* and *instrumental* playlist from the rest. The second PC separates playlists with high *valence* and *danceability* from the rest. The third PC separates tracks with high *speechiness* values. Based on these characteristics, we are able to cluster playlists into five different groups using spectral clustering. This is already a valuable insight. However, aiming to get a better understanding of the different clusters, we explore the genre distribution among each of the clusters in the next section.

Figure 4. Star diagrams



Genre Distribution and Playlist Names

We obtain genre information for each track using the genre tags provided by Spotify. To derive a genre distribution for each cluster we count the number of appearances of each genre in each cluster. Complimentary to the genre distributions, we furthermore analyze at the playlist names facilitated by

Table 3. Top-5 Genres of each cluster

Cluster	Top-5 Genres
Cluster 1	folk-pop, indiefolk, singer-songwriter, chamberpop, folkrock
Cluster 2	hiphop, poprap rap, alternativehiphop, gangsterrap
Cluster 3	pop, dancepop, popchristmas, permanentwave, synthpop
Cluster 4	permanentwave, alternativerock, indietronica, indierock, indiepop
Cluster 5	classicalchristmas, classical, soundtrack, romantic, bowpop

Table 4. Top-5 Terms appearing in playlist names of each cluster

Cluster	Top-5 Playlist Terms
Cluster 1	christmas, love, james, motion, jazz
Cluster 2	rap, hip, hop, hiphop, roots
Cluster 3	love, summer, party, dance, soul
Cluster 4	black, love, day, dance, punk
Cluster 5	motion, classical, piano, orchestra, bach

users. In Table 3, we list the top-5 genres for each cluster. Analogously, we list the top terms (after having removed English stopwords using WordNet (Miller, 1995)) appearing in the playlist names in each cluster in Table 4. These findings show that for clusters 2 and 5 the playlist names and the genre names are very consistent and use the same vocabulary. In contrast to that, for the remainder of the clusters, the used vocabulary differs. The playlist names contain temporal- as well as activity-related terms. This is congruent with the findings of Pichl et al. (Pichl et al., 2015) and is an interesting topic for future work, as a profound analysis of playlist names is out of scope of this work.

In a next step, we look into whether there is a difference in the genre distribution among the clusters. Therefore, we rely on the Pearson Similarity to compute similarities between the different genres appearing in the individual clusters. Thus, in a first step, we count how many times each of the distinct genres occurs in each cluster. In a second step, we apply the two similarity measures on all pairs of clusters.

In Equation 1, we show the computation of the Pearson correlation coefficient. In this Equation, X and Y represent vectors containing the counts of the different genres in Cluster x and respectively the counts of the genres Cluster y.

Equation 1: Pearson Correlation Coefficient

Table 5. Genre Similarities between the clusters using Pearson Correlation

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	1	0.22	0.77	0.62	0.35
Cluster 2	0.22	1	0.50	0.24	0.04
Cluster 3	0.77	0.50	1	0.77	0.24
Cluster 4	0.62	0.24	0.77	1	0.29
Cluster 5	0.35	0.04	0.24	0.29	1

$$Á_{x,y} = \frac{\mathrm{COV}\left(\mathrm{X},\mathrm{Y}\right)}{\mathrm{SD}\left(\mathrm{X}\right)*\mathrm{SD}\left(\mathrm{Y}\right)}$$

Refer to Table 5. We observe one strong correlation between cluster 1 and 3 $\left(\rho = 0.77\right)$ as well as a moderate correlation between cluster 1 and 4 $\left(\rho = 0.62\right)$, which we lead back to the different forms of pop-music genres in those three clusters. Hence, our method based on acoustical features provides different results than categorizing playlists using the track genres. Categorizing playlist solely by the genre would lead to a category containing the playlists of cluster 1 and 3 as well as parts of cluster 4.

Moreover, we observe weak correlations of several clusters. This implies that the same genres, mainly different forms of pop music, appear among several clusters. E.g., we can find the "popchristmas"-genre in almost all clusters. Hence, we argue that users do not necessarily group tracks by same ways as genres group tracks. In other words, users use the same genres in different playlists. In addition, we observe that the correlation coefficient is nearly 0 between Cluster 2 (the "rap Cluster") and Cluster 5 (the "classical music Cluster"), confirming that rap-style music is rather different from classical music. These results are consistent for Pearson and Jaccard Similarity.

Besides analyzing the genre distribution of the playlist-clusters, we also study the user distribution among the clusters in the next section.

Users Among Clusters

In this section, we analyze the user distribution among the clusters representing playlists with similar acoustic features. We investigate how many users create playlists only in a single cluster (i.e., they only listen to a single type of music with respect to acoustic features) and how many users create playlists in different clusters. In Table 6, we state the number of users and the number of clusters in which they created playlists. We observe that 65.75% of the users organize their music in playlists belonging to three or more clusters. 19.94% of the users create playlists among all five clusters, the maximum. On average, a user is represented in 3.08 clusters with a median of 3 (SD = 1.36). From the median and mean, we can see that the number users with respect to the number of clusters is equally distributed. The average number of users per cluster is 631.60 with a median of 183 (SD = 232.39).

We are also interested in whether we can find clusters, which are populated by the same users. I.e., whether if users that create a playlist in cluster A, are the same user that create a playlist in cluster B. Therefore, we look at the correlation between the clusters in terms of users having created playlists in those clusters. As the data is ordinal or at least discrete between 1 and 54, which is the maximum number of playlists a user created within a cluster, we apply Spearman's rank correlation coefficient as shown in Table 7.

Table 6. User distribution in number of clusters

Clusters	Users	Relative Portion
> 1	1027	100.00%
> 2	936	82.61%
> 3	745	65.75%
> 4	503	44.40%
> 5	226	19.94%

Table 7. User-cluster correlations

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	1.00	0.36	0.61	0.54	0.54
Cluster 2	0.36	1.00	0.42	0.34	0.25
Cluster 3	0.61	0.42	1.00	0.54	0.35
Cluster 4	0.54	0.34	0.54	1.00	0.46
Cluster 5	0.54	0.25	0.35	0.46	1.00

We do not observe any strong correlation between the individual clusters $\left(\rho > 0.7\right)$, nevertheless there are several moderate correlations $\left(\rho > 0.5\right)$ between the clusters. It is worth to mention that cluster 2, the "rap cluster", does not have any moderate correlations with other clusters. Cluster number 5, the "classical music cluster" only shows low moderate correlation with cluster number 1. However, every cluster except cluster 2 (rap) does have these moderate correlations to cluster number 1, the "folk cluster", a cluster containing different forms of folk music according to the genre distribution. Further, clusters 3 and 4 also show a moderate correlation. With respect to the acoustic attributes of these two clusters, they are rather similar, except for the fact that cluster number 3, containing pop music, shows higher values for *valence* and *danceability*. We interpret this as "feel good music".

Complementary to this, to estimate the overall popularity of the clusters, we compute the number of users and playlists in each cluster as shown in Table 8.

We find that 91% of all users created playlists in cluster number 3, the "feel good music"-cluster. In addition, 38% of all playlists are located in this cluster. Interestingly, about 40% of all users created playlists in the "rap" or "classical music" clusters, however playlists in those clusters only account for 7 and respectively 8% of all playlists. This means that a high number of persons create playlists with rap or classical music, while at the same time, the number of playlists with respect to the total number of playlist is low. This means, that classical music or rap music can be considered as niche music with respect to the number of playlists

DISCUSSION AND FUTURE WORK

Summing up our results, we find and explain differences in terms of acoustic features across the playlists using a linear PCA. Based on this, we cluster playlists into five clusters (or groups) of playlists using spectral clustering. Using k-Means yields the same results considering the number of clusters and cluster characteristics however R^2 is much smaller. Hence, we argue for using spectral clustering.

Table 8. Users and playlist per cluster

Cluster	Users	%	Playlists	%	Pls./Users
1	712	69%	4,385	25%	6.16
2	388	38%	1,241	7%	3.20
3	932	91%	6,664	38%	7.15
4	722	70%	3,978	23%	5.51
5	404	39%	1,331	8%	3.29

On average, a user is represented in 3 clusters (SD = 1.36), which indicates that one user prefers different styles of music. This supports qualitative studies that people prefer different styles of music dependent on the intended use or the mood (Cunningham et al., 2004; Kamalzadeh et al., 2012) quantitatively. Moreover, we show that those studies are also valid nowadays for music steaming platforms. Along with that, we found, that the genre classifies music different to our classification based on acoustical attributes. As the same genres are present in several clusters and playlists, we argue that classifying music for certain playlists based on the genre is not reasonable. Furthermore, we see that different types of music (in terms of acoustical attributes) are tagged with the same genre. Based on these findings we argue, that novel approaches for classifying tracks in music databases and libraries, as presented in the next paragraph, could be valuable to users. Connected to this we find that a high number of users create playlists with rap or classical music, while at the same time, the number of playlists with respect to the total number of playlist is low. This means, that classical music or rap music can be considered as niche music with respect to the number of playlists, however not as niche music when considering the number of users. Additionally, the number of users creating playlists in *both* clusters is low. Getting a closer look, we find that those users mostly create playlists in all clusters.

The insights presented above (i.e. utilizing the rap music vs. classical music clusters) can be a valuable input for a recommender system to assess the (dis-) similarity of users. Therefore, we plan to integrate the insights gained in this work into a recommender system in a future work. Our data set contains roughly 700,000 tracks. The challenge of a recommender system is to select a small number of tracks a user is likely to enjoy (i.e., only 16, which is the median number of tracks in a playlist). By applying our presented clustering technique, the number of potential recommendation candidates can be reduced: If a user is currently listening to tracks belonging to a specific cluster, a recommender system should only consider tracks as relevant belonging to this cluster. This technique is known as pre-filtering (Adomavicius & Tuzhilin, 2010) and naturally requires the recommender system to be able to infer the current cluster from the current song selection. The recommendation algorithm of choice we apply pre-filter to would be singular value decomposition (SVD), as both algorithms yield to the same results using different computation strategies. Another possible application using our findings and data is to create user classifications for recommender systems, i.e. based on mining for association rules in our data set. Possible rules would be (i) that users who create playlists in rap and classical music clusters are users creating playlists in all cluster or (ii) users solely creating playlists in the rap cluster would not create a playlist in the classical music cluster.

Besides this, even more interesting than pre-filtering or creating association rules, as it is a trending topic in current research, are context-aware recommender systems. Hence, we plan to tag each of the clusters with a certain mood or the intended use. As already mentioned in Section 2, people want to have very little interaction with their music databases and libraries, but still want to get music matching their mood or their activities. Thus, a possible application could provide search facilities capable of finding music fitting a given situation. This is why tagging our clusters with this information would enable presenting music to users based on clusters matching their activities and moods. One approach will be to exploit the playlist names as proposed by Pichl et al. (2015).

CONCLUSION

In this article, we present an analysis of user-generated playlists on the music streaming platform Spotify. This study facilitates a recent dataset containing Spotify playlist data for a profound quantitative analysis. Our main contribution is a novel study explaining the differences and commonalities among user created playlist. We show that “feel-good” popular music is serving as a common musical ground across all users. 91% of all users create at least one playlist in the “feel good music”-cluster. Additionally, we observed, that classical music and rap music can be considered as niche music with respect to the number of playlists, however not as niche music when considering the number of users. Furthermore, users creating playlists in both, the rap and the classical music cluster, are rare.

Further, we found that users listen to different styles of music (or at least organize different styles of music in their libraries). We consider these findings as important foundations for understanding and creating user-centric music recommendation systems in future works: Our novel findings about playlist characteristics can be easily modeled in a SVD-based recommender system as a PCA can be computed using SVD. Furthermore, it is possible to leverage the presented computation of groups of playlists in pre- or post-filtering recommendation approaches.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2010). Context-Aware Recommender Systems. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (1st ed., pp. 217–253). New York, NY, USA: Springer-Verlag New York, Inc.

Alghoniemy, M., & Tewfik, A. H. (2001). A Network Flow Model for Playlist Generation. *Proc. of the Intl. Conf. on Multimedia and Expo (ICME)* (pp. 329–332).

Aucouturier, J.-J., & Pachet, F. (2002). Scaling up Music Playlist Generation. *Proc. of the Intl. Conf. on Multimedia and Expo (ICME)* (pp. 105–108). doi:10.1109/ICME.2002.1035729

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Cunningham, S. J., Jones, M., & Jones, S. (2004). Organizing digital music for use: an examination of personal music collections. *Proc. of the 5th Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 447–454).

Ding, C., & He, X. (2004). K-means Clustering via Principal Component Analysis. *Proc. of the Twenty-first Intl. Conf. on Machine Learning (ICML)* (pp. 29–36). doi:10.1145/1015330.1015408

Flexer, A., Schnitzer, D., Gasser, M., & Widmer, G. (2008). Playlist Generation using Start and End Songs. *Proc. of the 9th Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 173–178).

Hauger, D., Schedl, M., Košir, A., & Tkalčič, M. (2013). The Million Musical Tweets Dataset: What Can We Learn From Microblogs. *Proc. of the 14th Intl. Society for Music Information Retrieval Conf. (ISMIR)* (pp. 184–194).

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer Verlag. doi:10.1007/978-1-4757-1904-8

Kamalzadeh, M., Baur, D., & Möller, T. (2012). A Survey on Music Listening and Management Behaviours. *Proc. of the 13th Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 373–378).

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. doi:10.1016/j.jesp.2013.03.013

Logan, B. (2002). Content-Based Playlist Generation: Exploratory Experiments. *Proc. of the 3rd Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 259–296).

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297).

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, *38*(11), 39–41. doi:10.1145/219717.219748

Pampalk, E., Pohle, T., & Widmer, G. (2005). Dynamic Playlist Generation Based on Skipping Behavior. *Proc. of the 6th Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 634–637).

Pauws, S., & Eggen, B. (2002). PATS: Realization and user evaluation of an automatic playlist generator. *Proc. of the 3rd Intl. Symposium on Music Information Retrieval (ISMIR)* (pp. 222–230).

Pearson, K. (n. d.). Notes on regression and inheritance in the case of two parents. Proc. of the Royal Society of London (pp. 240–242).

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, *2*(6), 559–572. doi:10.1080/14786440109462720

Pichl, M., Zangerle, E., & Specht, G. (2014). Combining Spotify and Twitter Data for Generating a Recent and Public Dataset for Music Recommendation. *Proc. of the 26nd Workshop Grundlagen von Datenbanken*, Ritten, Italy (pp. 35–40).

Pichl, M., Zangerle, E., & Specht, G. (2015). Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name? *Proc. of the 15th IEEE Intl. Conf. on Data Mining Workshops (ICDM)* (pp. 1360–1365). doi:10.1109/ICDMW.2015.145

Pichl, M., Zangerle, E., & Specht, G. (2016). Understanding Playlist Creation on Music Streaming Platforms. *Proc. of the 18th IEEE Symposium on Multimedia (ISM)* (pp. 475–480). IEEE. doi:10.1109/ISM.2016.0107

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*, *2*(3), 117–119. doi:10.1007/BF01891203

Royston, P. (1995). Remark AS R94: A remark on Algorithm AS 181: The W test for normality. *Applied Statistics*, *44*(4), 547–551. doi:10.2307/2986146

Schedl, M., & Schnitzer, D. (2014). Location-Aware Music Artist Recommendation. *Proc. of the 20th Intl. Conf. on MultiMedia Modeling (MMM)* (pp. 205–213). doi:10.1007/978-3-319-04117-9_19

Shi, J., & Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905. doi:10.1109/34.868688

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *63*(2), 411–423. doi:10.1111/1467-9868.00293

Zangerle, E., Gassler, W., & Specht, G. (2012). Exploiting Twitter's Collective Knowledge for Music Recommendations. *Proc. of the 2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages* (pp. 14–17).

Zangerle, E., Pichl, M., Gassler, W., & Specht, G. (2014). #nowplaying Music Dataset: Extracting Listening Behavior from Twitter. *Proc. of the 1st ACM Intl. Workshop on Internet-Scale Multimedia Management* (pp. 21–26). doi:10.1145/2661714.2661719

ENDNOTES

1 http://www.spotify.com
2 http://www.apple.com/music/
3 http://about.twitter.com/de/company
4 http://dbis-pla.uibk.ac.at

Martin Pichl is a research assistant and data scientist at the University of Innsbruck since 2014. He is working in the research group Databases and Information Systems (DBIS), a subdivision of the Department of Computer Science. He earned his master's degree in Information Systems at the University of Innsbruck in 2014 and is currently pursuing his PhD at the University of Innsbruck. His main fields of expertise are information retrieval and recommender systems. In his PhD project, he is concerned with user centric music recommender systems. For this, he is focusing on incorporating a user's context during music consumption into recommender systems to provide better music recommendations.

Eva Zangerle is a postdoctoral researcher at the University of Innsbruck at the research group for Databases and Information Systems (Department of Computer Science). She earned her master's degree in Computer Science at the University of Innsbruck and subsequently pursued her PhD from the University of Innsbruck in the field of recommender systems for collaborative social media platforms. Her main research interests are within the fields of social media analysis, recommender systems and information retrieval. Over the last years, she has combined these three fields of research and investigated music recommender systems based on data retrieved from social media platforms aiming to exploit new sources of information for recommender systems. She was awarded a Postdoctoral Fellowship for Overseas Researchers from the Japan Society for the Promotion of Science allowing her to make a short-term research stay at the Ritsumeikan University in Kyoto. Eva Zangerle teaches at the University of Innsbruck, Austria and at the University of Applied Sciences in Salzburg, Austria.

Günther Specht is full professor and chair of the research group Databases and Information Systems. He received his doctorate and his habilitation at the Technical University of Munich (TUM), where he also held his first professorship. In 2000, he became chair of the Database and Information Systems research group in Ilmenau/ Thüringen (Germany). In 2001 he changed to a professorship for Database and Information Systems at the University of Ulm (Germany). In 2007, Günther Specht founded the DBIS group at the University of Innsbruck, where his research interests focus on big data analyses, NoSQL-databases, information retrieval, recommender systems, plagiarism detection and genome databases.