Journal of Data Science, Statistics, and Visualisation

September 2024, Volume IV, Issue VI. *doi: 10.52933/jdssv.v4i6.95*

Measuring Emotions to Classify Songs: The Impact of the COVID-19 Pandemic on Music Streaming Data

Alessandro Albano
University of Palermo

Irene Carola Spera
University of Palermo

Mariangela Sciandra
University of Palermo

Antonella Plaia
University of Palermo

Abstract

Data from music streaming has gained increasing attention since it allows studying music preferences across diverse cultures and different periods of time. Indeed, the study of "music and emotion" is crucial for understanding the psychological relationship between human sentiments and music. The temporal study of musical emotions provides beneficial insights into the analysis of the mood of listeners during periods of particular relevance and stress (e.g., the COVID-19 pandemic). This study performs music streaming data analysis to retrieve the musical emotions of the top Italian streamed songs during the pandemic. To this end, we propose two new indices for measuring anger and joy in songs. We suggest a procedure for clustering music streaming data: the DISTATIS procedure and Partitioning Around Medoids (PAM) clustering algorithm are sequentially applied to identify intervals of time sharing similar sentiments. Finally, we employ the proposed procedure to investigate the relationship between the evolution of the pandemic spread and sentiments extracted from songs.

The results show that music streaming data analysis allow identifying five clusters of time intervals sharing similar sentiments, related to the evolution of the Italian restrictive quarantine measures.

Keywords: music streaming data, music emotion recognition, DISTATIS, PAM clustering, COVID-19.

1. Introduction

The increasing popularity of music streaming platforms has greatly contributed to the growth of music streaming data. Music is frequently referred to as a "language of emotion", and it is natural to categorise music in terms of its emotional associations (Kim et al. 2010). Indeed, music itself has the force to exorcise fear, anger and to instill emotions. In the Music Emotion Recognition (MER) literature, many studies have shown the impact of music on everyday well-being (Schwartz et al. 2017; Grey et al. 2020; Hu et al. 2021), but none have focused on how music streaming data can be exploited to describe people's well-being during the outbreak of Coronavirus disease (COVID-19). As a matter of fact, COVID-19 has deeply affected people's lives, with repercussions on the economy, work, health, social relations and everyday life. People are still facing stressful challenges that cause strong and overwhelming emotions in adults and children. Many countries have taken prompt and rigid public health measures, including lockdown and home quarantine, to prevent the uncontrolled spread of the virus. These public health actions were necessary to reduce the spread of COVID-19 but, on the other hand, may have had adverse effects, including making people feel isolated and lonely and increasing stress and anxiety. Indeed, relevant research projects show that social distancing and social isolation are distressing experiences for people and create negative emotions such as depression, fear, nervousness, and sadness (Wang et al. 2021). Zhao et al. (2020) reported that the emotional trend of people towards COVID-19 changes from negative emotions (during the initial stage) weakening to positive emotions (during the later stage) increasing.

Limaye et al. (2020) and Ni et al. (2020) studied how people heavily rely on digital social networks to maintain connections obstructed by physical distancing measures. Grey et al. (2020) focused on the relationship between perceived social support and mental health and sleep during the pandemic. They showed that perceived social support had significant inverse associations with anxiety, depression, loneliness, and irritability. Yao et al. (2021) investigated the influential mechanism of online social support on the public's belief to overcome the pandemic, showing that tangible support and esteem support can directly affect the public's beliefs. Doubtless, people primarily experienced loss of freedom, loneliness and were forced to face uncertainty over disease status (Brooks et al. 2020; Coelho et al. 2020; Lee et al. 2020; Porcelli 2020). In that challenging period, music represented a valuable aid against melancholy and acted as a companion providing relief.

In this work, we study the Italian most-streamed songs to investigate people's emotions induced by quarantine. The intuition behind our study is that retrieving the musical emotions induced by top streamed songs provides valuable insights into how Italians faced the pandemic from an emotional point of view. Specifically, we introduce a novel procedure for clustering musical data. The approach utilises two new indices that measure emotions in songs, combined with DISTATIS (Abdi et al. 2005; Abdi et al. 2007, 2009, 2012), a 3-way multidimensional scaling method. The procedure ends by applying the Partitioning Around Medoids (PAM) clustering algorithm (Kaufman and

Rousseeuw 2009) to identify groups of time-intervals, during the pandemic spread with similar sentiments and audio features. The results show that the proposed procedure allow identifying five clusters of time intervals sharing similar sentiments, related to the evolution of the Italian restrictive quarantine measures.

The paper is organised as follows: Section 2 presents a review of previous work on the topic, along with the development of the hypothesis and research questions. Section 3 outlines the proposed approach. Section 4 shows the application of the proposed methods to music streaming data. Finally, Sections 5 and 6 provide the discussion and conclusions, respectively.

2. Hypothesis Development and Related Works

The study of "music and emotion" is crucial for understanding the psychological relationship between human sentiments and music. The main beneficial effects of music are: relaxation, boredom reduction, concentration attention, higher well-being, emotion regulation, performance and general happiness improvement (Schwartz et al. 2017; Hu et al. 2021). Moreover, research has found that music to help alleviate sleeping problems in young adults (Papinczak et al. 2015) and improving the resilience of adolescents when facing stress (Gilligan 2000).

Music is also suitable for managing and regulating emotions and stress in everyday life since it has the capacity to both distract and engage listeners in a variety of cognitive and emotional ways (MacDonald 2013). All these factors make music highly relevant to the social relationship component of well-being. Indeed, it has been pointed out that understanding the reasons for listening to a particular type of music in a specific situation is necessary for the design and development of music information systems (Inskip et al. 2008; Demetriou et al. 2016).

It is generally recognized that there are two types of music emotions: *expressed* and *induced* (Li et al. 2018; Hu et al. 2021). The *expressed* refers to the emotions that composers want to convey through music. Artists use songs as a mean of communicating specific messages or crucial moments in their lives. The *induced* emotions represent the listener's emotional side. It is the emotion aroused in an auditor when they listen to a music piece. In this work, the musical *induced* emotion is addressed. Specifically, we aim at studying whether the mood swings, caused by the spread of the Coronavirus, influenced the types of music people listened to during home quarantine. The main contributions of this manuscript can be summarized as follows:

1. Introducing a novel procedure for clustering musical data based on sentiment-based indices.

2. Exploring a possible overlap between clusters of time-intervals (sharing similar sentiment-based indices) and particular pandemic phases determined by Italian government measures.

   To better understand the results and draw conclusions, a brief recap of the Italian pandemic's crucial phases is given below. The first restriction measures led to quarantine in 11 municipalities in northern Italy; this was followed by total closure of all schools, universities, and most economic activities on 4th March. These events were defined as **"Phase one"**. On 26th April 2020, Prime Minister Conte announced less restrictive measures in the so-called **"Phase two"**. **"Phase three"** began on 15th June 2020 with another DPCM, in which Italians were allowed to resume everyday life. Unfortunately, on **13th October 2020** another DPCM imposed new specific containment measures. The public restriction measures culminated on **4th November 2020** with another DPCM, which identified three areas in the Italian regions, corresponding to three increasingly critical levels (red, orange and yellow areas).

2.1. Music emotion recognition

Statistics in music has recently become a vast and rapidly growing field of research. The main reason for such widespread use is its broad applicability. Although research on automatic Music Emotion Recognition (MER) is not as mature as some other Music Information Retrieval (MIR) branches, it is clear that rapid progress is being made. One of the most frequent and promising uses of music search and retrieval is the building of emotion-based playlists, where automatic MER plays a significant role in providing usually unavailable emotion-related information.

Kim et al. (2010) provided a broad survey of state of the art in this field: they highlighted that many studies propose the use of a combination of audio features and lyrics to detect emotions conveyed by music. Yang and Lee (2004), using lyrics and a wide range of audio features, found strong correlations between particular lyrics and emotional categories (hostility, sadness, and guilt). In the work by Yang et al. (2008), three different methods of combining audio and text features were compared. They obtained the best results by first using audio and text separately to classify arousal and valence, using Support Vector Machine (SVMs) and then merging the results to determine a full Valence-Arousal classification. Laurier et al. (2008) explored the use of audio and multiple lyric features to classify emotions of songs in the four-quadrant Valence-Arousal space (Russell 1980). A classification in eighteen emotion classes was proposed by Hu et al. (2009) based on social tags from Last.fm[1]. Aljanaki et al. (2016) designed an experiment to study whether music can express and induce a complex fine-grained range of emotions; they also studied the extra-musical factors that influence induced musical emotion. Nalini and Palanivel (2016) combined Mel Frequency Cepstral Coefficients (MFCC) and Residual Phase (RP) features for emotion recognition in music (audio). A nonlinear SVMs learning algorithm was applied to obtain the optimal class boundary between the various emotions, namely anger, fear, happiness, and sadness, and obtained a recognition accuracy of 99%. Sciandra and Spera (2020) proposed a new class of models (Beta GLMM) for predicting the popularity index of songs, using the Spotify[2] audio features as covariates. Pergola et al. (2019), proposed a topic-dependent attention model for sentiment classification and topic extraction. Abdi et al. (2019), proposed a deep learning approach for sentiment classification employing a variant of Recurrent Neural Network (RNN) for sequential processing. Their approach to sentiment analysis involved extracting useful features to determine the polarity of

[1] http://www.last.fm
[2] https://www.spotify.com/it/

the sentences and using them to find subjective information, such as opinion, or finding desirable sentiment describing the mood from music lyrics.

Russo et al. (2020) proposed a new approach for the recognition of emotional content in music based on a detailed cochlear model of auditory perception. To extract the appropriate affective music information, they employed a Convolutional Neural Network (CNN) for feature extraction on cochleogram images.

These studies underscore the significant advancements being made in the field of Music Emotion Recognition. The integration of diverse methodologies, such as combining audio and text features, employing advanced machine learning techniques, and leveraging deep learning models, is paving the way for more accurate emotion detection in music.

2.2. Clustering music data

Another critical use of statistics in musical research aims to cluster music data. There are several relevant articles in the literature covering this particular topic. Tsai et al. (2004) examined the feasibility of unsupervised clustering of acoustic song data based on the singer's voice characteristics extracted via vocal segment detection. A clustering algorithm that integrates features from both text and acoustic data sources was proposed by Li et al. (2007) to perform bimodal learning. Peng et al. (2007) used features from the signals to perform a K-Means clustering and used a number of labels as constraints to perform constraint-based clustering in order to group a set of songs by their artists. Wongso and Santika (2014) combined Dual Tree Complex Wavelet Transform (DTCWT) features and SVMs to propose a fast and accurate automatic method for music genre classification. They focused on classifying four genres of music, i.e., pop, classical, jazz, and rock, using strong features such as mean, standard deviation, variance, and entropy.

Sen (2014) explored a novel technique to cluster songs based on Echo Nest[3] Audio Attributes and the K-Means algorithm. Jondya and Iswanto (2017) clustered 101 songs from 18 provinces in Indonesia using audio data features selected with the Principal Component Analysis method. The selected features are extracted from the audio signal and clustered by the X-Means algorithm to find the correct number of clusters. Kartikay et al. (2016) employed four different classification algorithms, namely - SVMs, Naive Bayes, Linear Discriminant Analysis and Decision Trees, to classify a given track into a mood such as happy, sad, peaceful and angry. They used a combination of the valence and arousal values from a database of 1000 songs and several musical features such as tempo, energy and pitch.

Oramas et al. (2018) developed an approach using deep neural networks to learn and combine multimodal data representations for music genre classification. Their method allows a representation learning approach to classify music genres from different data modalities, i.e., audio, text, and images. Spera and Sciandra (2020) used the unsupervised classification algorithm PAM to analyse the Italian singer Luciano Ligabue's discography, studying whether the songs underwent a change in style over time.

The studies highlighted illustrate the diversity of approaches taken in clustering music data. From unsupervised techniques and constraint-based clustering to advanced ma-

[3] https://www.crunchbase.com/organization/the-echo-nest

chine learning algorithms and multimodal data integration, researchers are leveraging a wide array of methods to uncover patterns and relationships within music datasets.

3. Methods

This section details the proposed procedure used in our study to cluster time-intervals based on sentiment indices. The method includes the following steps:

1. Sentiment-based distances:
   (a) Extract sentiment-based indices from songs in a fixed time period.
   (b) Derive the distribution of each sentiment-based index in specific time intervals (e.g., weeks).
   (c) Compute a proper distance metric between weekly sentiment distributions.
   (d) Construct a list of distance matrices, where each matrix corresponds to a specific sentiment-based index.
2. Distance matrix aggregation:
   (a) Arrange the list of distance matrices into a three-way data structure.
   (b) Apply the DISTATIS method (Abdi et al. 2005; Abdi et al. 2007, 2009, 2012) to derive a compromise matrix that represents the consensus structure among the distance matrices. The compromise matrix is a unified measure of dissimilarity between weekly distributions across all sentiment indices.
3. Clustering:
   (a) Use the compromise matrix as input for the PAM clustering algorithm (Kaufman and Rousseeuw 1990).
   (b) Determine the optimal number of clusters using the Silhouette method.
   (c) Validate the identified clusters.

At the end of the procedure, we analyse the resulting validated clusters to identify patterns and trends in weekly sentiment indices.

3.1. Step 1: Sentiment-based distances

The several ways in which music communicates emotions are complex. For example, faster rates of speaking and faster tempo in music are associated with high-arousal emotions such as happiness and anger, whereas slow speech rate and tempo are markers of low-arousal emotions such as sadness and tenderness (Corrigall and Schellenberg 2013). However, most people are able to connect with the words of a song better than with its musical features. In most cases, emotions are strongly related to words, while the melodic aspects are generally adapted to fit the lyrical theme (Saluja et al. 2019). A way to deal with this complexity consists in combining information about both audio features and lyrics to define indices. To measure the emotions evoked by musical aspects we used the following audio features available on the Spotify Web API:

Energy: in a song, this measure varies from 0 to 1 and represents the percentage of intensity and activity.

Speechiness: it detects the presence of spoken words in a track. The more exclusively speech-like the recording, (e.g., talk show, audio-book, poetry), the closer to 1 the attribute value will be.

Valence: a measure from 0 to 1 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive, while tracks with low valence sound more negative .

Additionally, although not used for building the indices, the clustering method proposed in the following Sections also includes the following features:

Danceability: it describes how suitable a track is for dancing, based on a combination of musical elements, including tempo, rhythm stability, beat strength, and overall regularity. A value of 0 is the least danceable and 1 is the most danceable.

Loudness: the overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is primarily psychologically correlated with physical strength (amplitude). Values range between -60 and 0 db.

Tempo: the overall estimated tempo of a track, in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

To analyse the emotional aspect of musical texts, we chose to apply Sentiment Analysis through the use of an affective lexicon, which is a database of words in which each piece is classified according to its content in terms of subjectivity, polarity (positive or negative), ability to provoke particular emotions, and so on. In our study, we refer to the NRC Word-Emotion Association Lexicon (aka EmoLex) (Mohammad and Turney 2013), consisting of words and their associations with eight primary emotions, namely, anger, fear, anticipation, trust, surprise, sadness, joy and disgust, and two feelings that are negative and positive.

Song lyrics were extracted from Genius Web API[4], and the emotional aspect of musical texts was measured through the following indices:

Pct angry: percentage of angry words in the lyrics according to the NRC's lexicon of words.

Pct joy: percentage of joyful words in the lyrics according to the NRC's lexicon of words.

Specifically, the two indices proposed in this work are modified versions of the Lyrical and Sonic *Anger* indices proposed by Oppenheimer (2019).

$$\text{Anger} = w_1 \cdot \sqrt{\text{pct angry} \cdot \text{speechiness}} + w_2 \cdot \sqrt{(1 - \text{valence}) \cdot \text{energy}}, \tag{1}$$

$$\text{Joy} = w_1 \cdot \sqrt{\text{pct joy} \cdot \text{speechiness}} + w_2 \cdot \sqrt{\text{valence} \cdot \text{energy}}, \tag{2}$$

[4] https://genius.com/

where $w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$. In defining the two new indices, we decided to perform a convex combination of song lyrics and sound with weights $w_1 = 0.6$, $w_2 = 0.4$. As a matter of fact, around 85% of the pieces analyzed are written in the native language of the listeners (i.e., Italian). Therefore, it is straightforward for people to understand the semantic content of the lyrics. Moreover, we carried out a sensitivity analysis to investigate how the weighting structure influences the temporal proximity of weeks in clusters. We found out that the following weighting scheme ($w_1 = 0.6$, $w_2 = 0.4$) leads to highly cohesive clusters, which consist of adjacent weeks (see Section 4.3).

Both indices are built in such a way that they to vary from 0 to 1. *Speechiness* is used to follow the intuition that a song will be very angry (joyful) if it contains many angry (joyful) words in large lyrics. Contextually, the emotion evoked by audio features can be determined following Russell's Circumplex model of affection (Russell 1980). This model states that all affective states derive from two fundamental neurophysiological systems, one related to positivity (measured through *Valence*) and the other to arousal (measured through *Energy*). Thus, each emotion can be understood as a linear combination of these two dimensions, as shown in Figure 1.

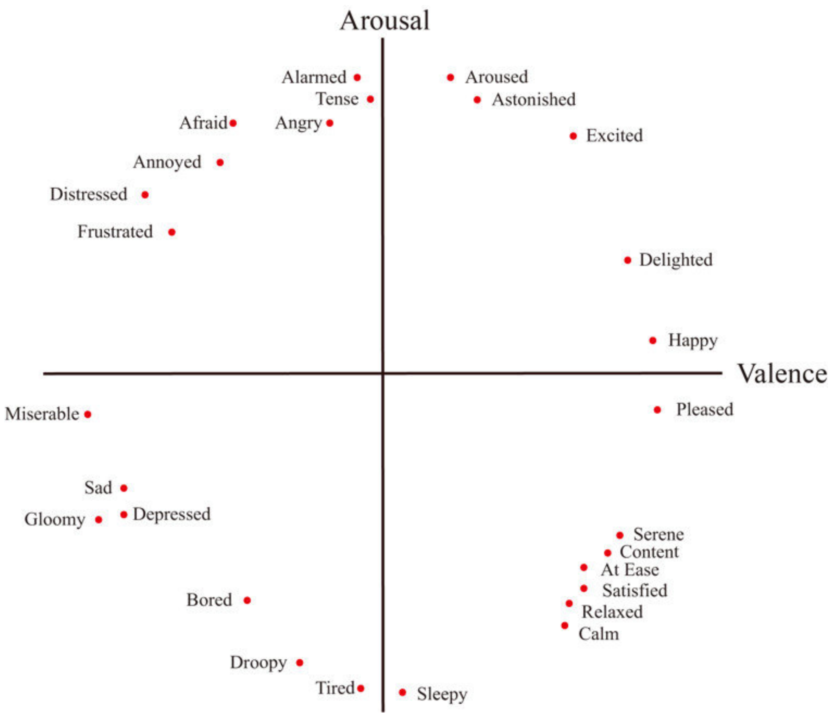

Figure 1: Russell's circumplex model.

In particular, the four quadrants can be interpreted as follows: songs evoking sadness will have low values of both valence and energy; songs expressing anger will have low valence and high energy; calm songs will have high valence and low energy; finally, cheerful songs will have high values of valence and energy.

Once the weekly distributions of sentiment-related indices have been derived (Steps 1-A and 1-B), we need to compute the distances among them (Step 1-C). We computed the distances by evaluating the distance between probability density functions (that is, by assessing the overlapping area between the two curves).

After an extensive review of the measures proposed in the literature, we chose a weighted version of the Manhattan distance, known as the *Canberra distance*. Although it is sensitive to small changes for values near zero, it is much more robust to outliers compared to other metrics. This metric belongs to the L_1 family. Its generalized equation is given by:

$$d_{can} = \sum_{i}^{n} \frac{|K_i - Q_i|}{K_i + Q_i}, \tag{3}$$

where K_i and Q_i are the ith components of vectors $\mathbf{K}$ and $\mathbf{Q}$, respectively, in an n-dimensional real vector space.

In our application, we consider 44 weekly distributions for each index. Therefore, $\mathbf{K}$ and $\mathbf{Q}$ represent the probability density functions of a specific index in the two weeks being compared. As a result, computing the Canberra distance between pairs of weeks yields five distance matrices (one for each index) with dimensions 44×44 (Step 1-D).

The five distance matrices were computed in R using the **philentropy** library (HG 2018).

3.2. Step 2: Distance matrix aggregation

To aggregate the different distance matrices (Steps 2-A and 2-B), we use the DISTATIS method, a generalisation of classical MultiDimensional Scaling (MDS) (working with a *single* distance matrix). DISTATIS (Abdi et al. 2005; Abdi et al. 2007, 2009, 2012) is a statistical technique, also knows as 3-way multidimensional scaling, that takes root in distance analysis. The name DISTATIS is derived from a method called STATIS in the French-speaking community, where it is attributed to Escoufier (1980, 1985) and L'Hermier des Plantes (1976).

The strength of DISTATIS lies in its able to project the *Compromise* where variables, observation and their scaled distance are all in the same space. The different steps involved in DISTATIS are:

1. Let $\boldsymbol{D_1}, \ldots, \boldsymbol{D_T}$ be the T squared distance matrices with I observations. In order to apply MDS each distance matrix is centred and normalized. The generic t-th normalized cross-product matrix is denoted $\boldsymbol{S_t}$.

2. Analyse the structure of the similarity between the cross-product matrices $\boldsymbol{S_1}, \ldots,$ $\boldsymbol{S_T}$. Abdi et al. (2005) suggested to evaluate the similarity between the $\boldsymbol{S_1}, \ldots, \boldsymbol{S_T}$ matrices, through the RV coefficient (Escoufier 1973):

$$RV(t, t') = \frac{\text{trace}\{\boldsymbol{S_t}^T \boldsymbol{S_{t'}}\}}{\sqrt{\text{trace}\{\boldsymbol{S_t}^T \boldsymbol{S_t}\} \times \text{trace}\{\boldsymbol{S_{t'}}^T \boldsymbol{S_{t'}}\}}}. \tag{4}$$

   The RV coefficient takes on values between 0 and +1 (the cross-product matrices are positive semi-definite matrices). $RV = +1$ means that the observations are grouped identically by the two matrices $\boldsymbol{S}_t$, $\boldsymbol{S}_{t'}$; conversely, a value of 0 implies a completely different grouping in the two matrices. The RV values for each combination of indices can be combined in a new *similarity matrix* denoted $\boldsymbol{C}$ with generic element $c_{t,t'}$, where $c_{t,t'} = RV(t, t')$ (Eq.(4)).

3. Derive an optimal set of weights for computing the compromise. From the eigendecomposition of $\boldsymbol{C}$:
$$\boldsymbol{C} = \boldsymbol{P\Theta P}^T, \tag{5}$$
we obtain the weights to be attributed to each $\boldsymbol{S}_t$ matrix, and to be used for computing the *compromise matrix* $\boldsymbol{S}_{[+]}$.

   In Eq.(5) $\boldsymbol{P}$ is the matrix of eigenvectors, such that $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}$, and $\boldsymbol{\Theta}$ is the diagonal matrix of the eigenvalues of $\boldsymbol{C}$.

   The elements of the first eigenvector (denoted with $\boldsymbol{p_1}$), after re-scaling ($\boldsymbol{\alpha} = (\mathbf{1}^T\boldsymbol{p_1})^{-1} \times \boldsymbol{p_1}$), return the optimal weights needed to compute the *compromise matrix* $\boldsymbol{S}_{[+]}$.

4. Compute the compromise as a weighted sum of the individual cross-product matrices
$$\boldsymbol{S}_{[+]} = \sum_t^T \alpha_t \boldsymbol{S}_t. \tag{6}$$

Defined in this way, the compromise matrix itself is a cross-product matrix, and therefore its eigendecomposition amounts to a PCA.

5. Compute the eigendecomposition of the compromise matrix.
The PCA of the compromise:
$$\boldsymbol{S}_{[+]} = \boldsymbol{V\Lambda V}^T, \tag{7}$$
allows to compute the compromise *factor scores* for the observations:
$$\boldsymbol{F} = \boldsymbol{V\Lambda}^{1/2}, \tag{8}$$
that are the new coordinates in the space defined by the components. Plotting the observations in the new component space, allows the distances on the map to reflect the similarities between them in the best way.

3.3. Step 3: Clustering

Once the compromise matrix, $\boldsymbol{S}_{[+]}$, has been derived, it becomes the input for the clustering procedure carried out using the PAM algorithm (Step 3-A).

PAM clustering can be considered an extension of the K-Means algorithm, with the advantage of robustness to outliers' presence. Moreover, PAM deals with general dissimilarity coefficients. The Silhouette coefficient is used to find the optimal number of clusters (Step 3-B). An advantage of the Silhouette approach is that it only depends on the actual partition of the objects, and not on the clustering algorithm. The optimal cluster number will be determined by maximizing the value of the Silhouette defined as:
$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{9}$$

where a_i is the average distance between i and the other points within the same cluster, while b_i is the average distance between i and the points of the nearest cluster. Therefore, the Silhouette is an internal validation metric ranging from -1 to 1; it measures to what extent an observation is similar to the own cluster compared to the nearest cluster. For each cluster, we define the average Silhouette width as the average of the s_i for all objects i belonging to that cluster. This allows us to distinguish "clear-cut" from "weal" clusters in the same plot; clusters with a larger average silhouette width will be more pronounced.

Cluster validation (Step 3-C) is performed following two different approaches. The first one consist in using the bootstrap elliptic tolerance regions for a global guarantee of the quality of clusters (all indices are considered simultaneously) (Hahn and Meeker 2011). Tolerance intervals express the variability in the allocation of the elements of the population in the several clusters that can be identified. Tolerance intervals indicate whether two categories are separable or "distinguishable". Suppose the tolerance intervals of two categories do not overlap. In that case, we can conclude that the accuracy of the assignment of observations to their respective categories is higher than the level chosen to compute the interval.

The second approach consists in comparing the quality of clusters for each index marginally by evaluating the average distance between and within clusters.

4. Data Analysis and Results

The objective of this section is to describe the emotions of Italians throughout the first months of the pandemic period using the proposed procedure, and highlight a possible relationship between the identified groups and the government's restrictions.

4.1. Spotify music streaming dataset

In order to run our procedure, we downloaded the top 200 streamed songs per week from the Spotify website[5]. The period considered spans from 7th February to 11th December 2020 (44 weeks). We considered only the top 20 songs from each ranking, assuming that these represent the weekly musical preferences of Italian users. The total number of records is 880. For each song, the number of weekly streams is counted (streams are considered valid in Spotify if a piece is played for over 30 seconds). Since each song could appear in many different rankings over the weeks, the total number of different songs is 181. The lyrics corpus is pre-processed through tokenization and stopword deletion (Vijayarani et al. 2015); and then the *Anger* and *Joy* indices are computed as defined in Eqq.(1)–(2).

The distributions of the two indices over the weeks (adequately weighted for the number of streams of each song) are plotted in Figure 2.

The graph shows how the highest levels of *Anger* are associated with the lowest levels of *Joy*. In particular, from week 4 (28th Feb.-5th March) to week 6 (13th March-19th March), the *Anger* index takes the highest values, with a peak in week 5 (6th March–12th March). In the same week, the *Joy* index reaches a negative peak. This is

[5] `https://spotifycharts.com/regional/it/daily/latest`

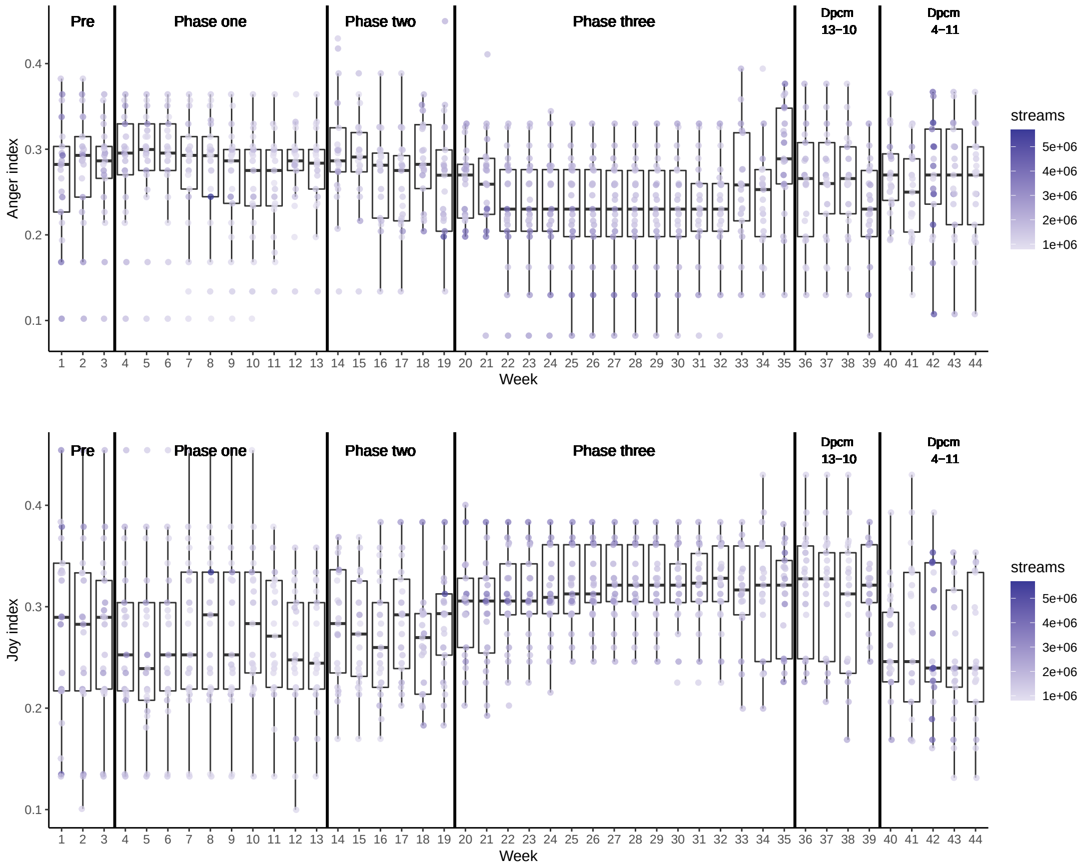

Figure 2: Distributions of the *Anger* (top-frame) and *Joy* (bottom-frame) indices over the weeks.

reasonable, considering that these are the first three weeks of lock-down. Conversely, it is essential to underline that during Phase three, where the new DPCM allowed Italians to return to everyday life, we spot the highest levels of *joy* and the lowest levels of *Anger*. This is particularly evident during the period from week 22 (3rd July - 9th July) to week 32 (12th Sep.-18th Sep.). Finally, once summer was over, the epidemiological curve resumed an upward trend. This led to new restrictions and, as expected, to a sharp change in the two indices. However, to be more precise, Figure 3 shows the median values over weeks for the whole set of examined indices (i.e., *Anger*, *Joy*, *Danceability*, *Loudness*, *Tempo*). In this graphical representation *Loudness* and *Tempo* were rescaled according to the $g(x) = \{x - \min(x)\}/\{\max(x) - \min(x)\}$ transformation, in order to set their range to [0,1], making the distribution comparable among variables

The musical characteristics show a fluctuating trend. Furthermore, the plot reveals a relationship between *Danceability* and *Loudness*. Specifically, we note that weeks with high levels of *Danceability* have, at the same time, low levels of *Loudness*. As regards *Tempo*, its highest value is reached in the first week, while we observe a wide interval of time, spanning from week 16 to week 32, i.e., from 16th May to 18th September, where *Tempo* is steady and low. As already stated, we intend to detect groups of weeks that share similar emotion and audio features. The heatmaps of each distance matrix are reported in Figure 4.

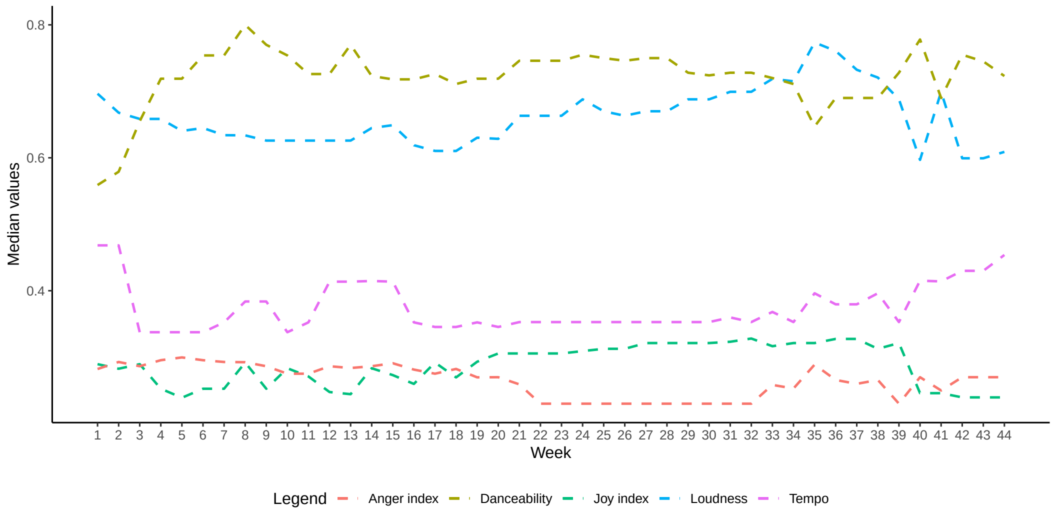

Figure 3: Median values over weeks of the five indices analyzed.

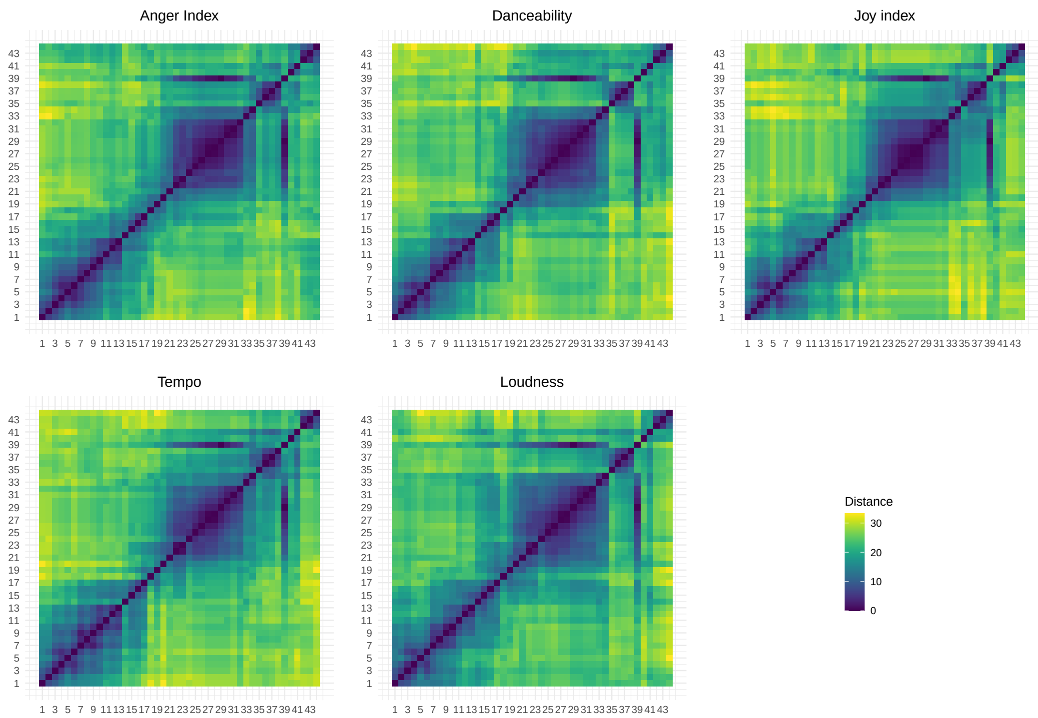

Figure 4: Heatmaps of the five Canberra distance matrices.

4.2. Building the compromise space through DISTATIS

After collecting musical data, we arranged our data-set in a 3-dimensional $44 \times 44 \times 5$ array, in which each stratum consists of a squared distance matrix (whose generic element is the distance (3) between a pair of weeks).

In order to create a *compromise matrix* $\boldsymbol{S}_{[+]}$ (Eq.(6)), we evaluated the similarity between indices by computing the RV coefficients (Eq.(4)). The RV coefficients between each pair of indices are reported in the $\boldsymbol{C}$ matrix (Table 1).

Table 1: Similarity matrix ($\boldsymbol{C}$ matrix) based on RV coefficients.

	Anger	Tempo	Danceability	Loudness	Joy
Anger	1.00	0.95	0.92	0.94	0.95
Tempo	0.95	1.00	0.95	0.94	0.95
Danceability	0.92	0.95	1.00	0.94	0.92
Loudness	0.94	0.94	0.94	1.00	0.94
Joy	0.95	0.95	0.92	0.94	1.00

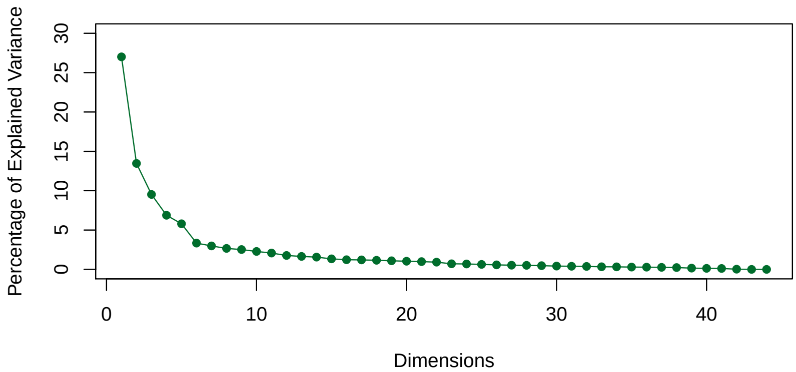

Figure 5: Compromise: Explained Variance per Dimensions.

The $\boldsymbol{C}$ matrix shows that distance matrices are highly similar. Thus, we that expect each index will contribute equally to the determination of the consensus matrix. The eigendecomposition of the $\boldsymbol{C}$ (Eq.(5)) matrix returns the weights given by the first eigenvector, that are $\boldsymbol{\alpha} = \{0.200, 0.201, 0.199, 0.200, 0.200\}$. The quality of the consensus can be evaluated by computing the percentage of total variance explained by this first eigenvector of the $\boldsymbol{C}$ matrix eigendecomposition, in our case equal to 95%. This considerable value indicates that the five indices substantially agree on the similarity structure between the weeks under analysis.

Once the 44×44 consensus matrix $\boldsymbol{S}_{[+]}$ is obtained, the PCA (Eq.(7)) of this matrix reveals the similarity structure of the weeks jointly driven by the five indices.

In Figure 5, the scree plot shows the percentage of variance explained by each dimension.

The plot suggests reducing the space to the first three dimensions, which explain 50% of the total variance of the compromise matrix. The factor scores and the projections of the weeks on the first two dimensions are shown in Figure 6 (the interpretation of colours is given in the following Subsection).

4.3. Clustering results

The PAM method was performed in R using the `pam` function of the **cluster** library (Maechler et al. 2019). The Silhouette plot (Figure 7) suggests that the optimal number of clusters is five.

Table 2 and Figure 8 jointly exhibit the cluster features.

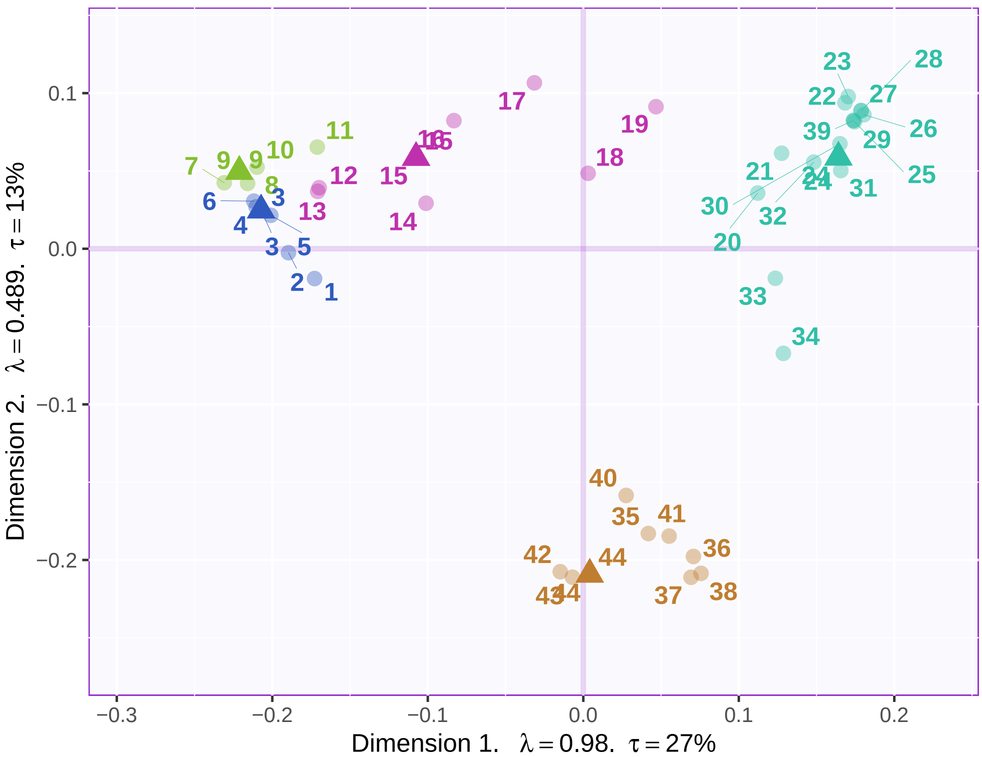

Figure 6: Analysis of the compromise: plot of the PAM 5 groups of weeks in the plane defined by the first two dimensions the of compromise PCA. The eigenvalue is denoted by λ, the proportion of explained variance is τ.

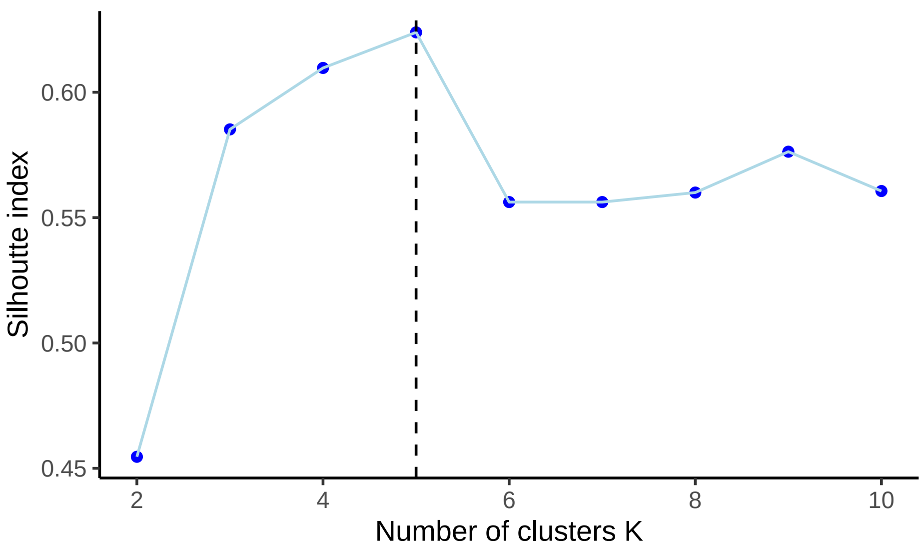

Figure 7: Silhouette plot.

The first three clusters consider the lockdown period corresponding to the first and second phases. In particular, Cluster 1 covers the first three weeks before and after the first DPCM (the beginning of Phase one). Cluster 2 embraces the central part of Phase two, and Cluster 3 includes the final part of Phase one and the entire Phase two.

Table 2: Clusters.

Cluster	Weeek	Medoid
Cluster 1	1-6	3
Cluster 2	7-11	9
Cluster 3	12-19	15
Cluster 4	20-34, 39	24
Cluster 5	35-38, 40-44	44

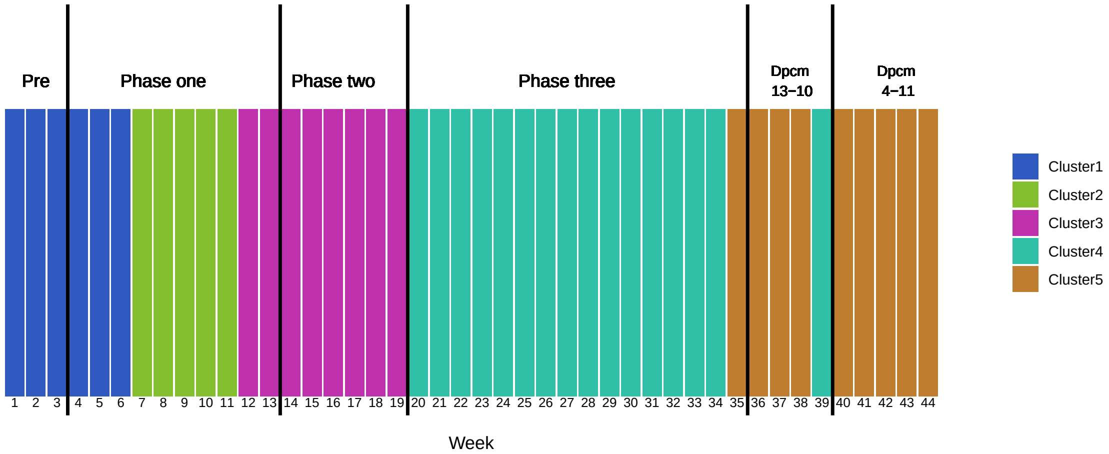

Figure 8: Relationship between clusters and Government's measures.

A confirmation of the relationship between identified clusters and the Government's restrictions is given by Cluster 4, which strongly identifies Phase three, namely the period with fewer restrictions, including summertime. Finally, Cluster 5 covers the weeks characterized by new constraints due to increased infections. Figure 6 facilitates the interpretation of the relationship between clusters and weeks, reporting the results of the two procedures applied sequentially (DISTATIS and PAM algorithm). Figure 9 graphically compares the average values of the five indices examined in the five identified groups.

The graph shows a specular trend of *Anger* and *Joy*. Specifically, the first cluster (i.e., the three weeks before and after the first DPCM) is characterised by the maximum value of the *Anger* index. On the contrary, *Anger* has a prominent negative peak in Cluster 4. In contrast, the *Joy* index has an inverse trend with a flattering rise in Cluster 4 (the happiest and the least angry group). The latter also includes weeks with high *Loudness* and *Danceability* and low *Loudness* levels, which perfectly characterize the summer period. Finally, Cluster 5 denotes a considerable increase of *Anger* and *Loudness* compared to Cluster 4. Moreover, it shows a moderate decrease in both *Loudness* and *Danceability* and a substantial decrease in *Joy*.

In order to display intervals in a 2-dimensional space, we drew an ellipse that includes the proportion of points corresponding to the chosen significance level.

Figure 10 shows that elliptical regions of confidence of elements from the same cluster tend to overlap. This is particularly true for Cluster 5 (in brown), showing very cohesive

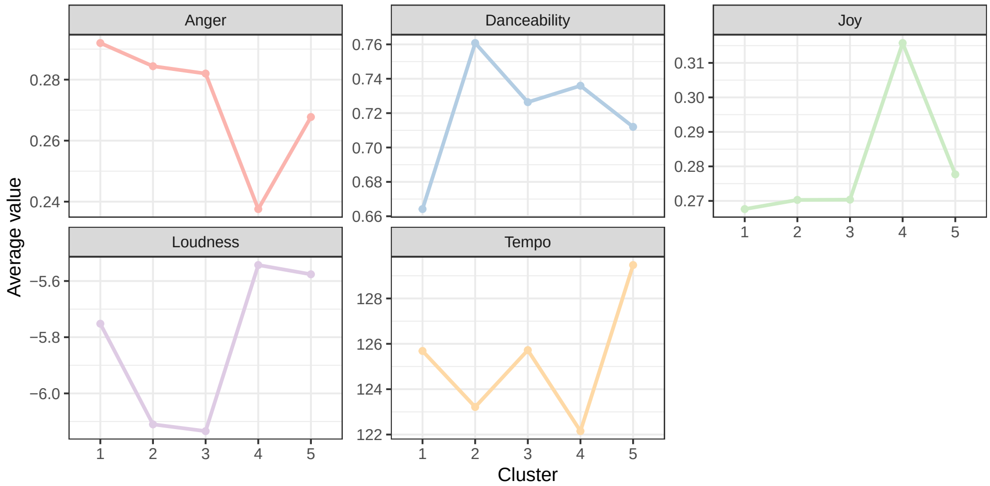

Figure 9: Attributes for each cluster.

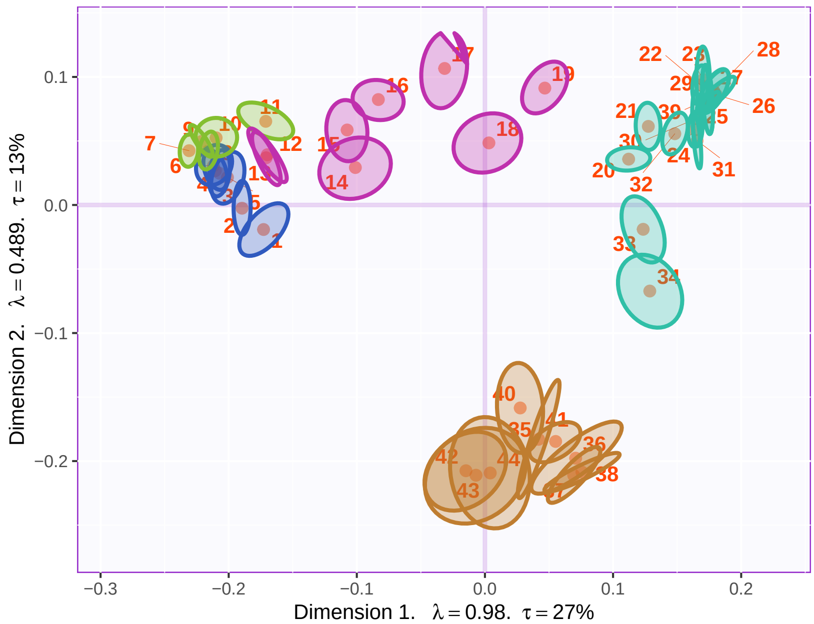

Figure 10: Global factor scores and their bootstrapped tolerance intervals. Cluster colours: 1 in blue, 2 in green, 3 in purple, 4 in petrol green watercolour, 5 in brown. The eigenvalue is denoted with λ, the proportion of explained variance is τ.

intervals. On the contrary, Cluster 3 (in purple) represents a more sparse case. On the other hand, weeks belonging to different clusters tend to be reliably separated, except for a few weeks belonging to Clusters 1 (in blue) and Cluster 2 (in green). The second approach to validation is presented in Figure 11; the horizontal axis gives the average distance between medoids. The vertical axis gives the average distances within clusters. The purpose of this plot is to investigate whether the dissimilarity caused by one index overwhelms the others. In Figure 11, the optimal area is on the bottom right and indicates high homogeneity between weeks within clusters and high heterogeneity between different clusters. We note that none of the five indices lies in the optimal

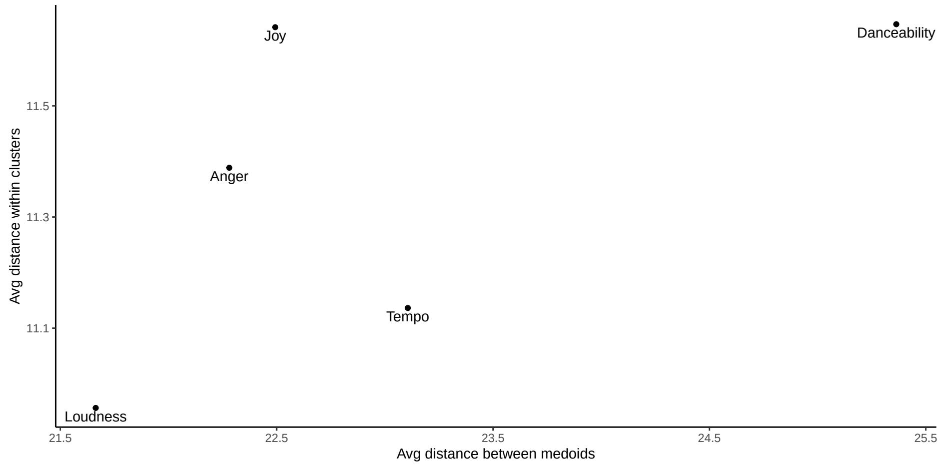

Figure 11: Cluster validation plot.

region. For *Danceability*, the distance between clusters is maximized, but there is high heterogeneity inside. The opposite is true for *Loudness*. In comparison, the other three indices lie in intermediate situations. This means that no index has prevailed over the others, thus confirming the similarity of the role played by each distance matrix identified by DISTATIS.

5. Discussion

In this work, we assumed that the emotional state of Italians during the pandemic could be analysed by retrieving the emotional content of the most streamed songs in Spotify. To better justify such a link, we check whether the musical patterns found in 2020 were already present in the previous year (Figure 12). The figure shows that, except for the valence index, the indices trends are unrelated in the two years, so it is reasonable to assume that they are not determined by some annual seasonal effect.

Summarising, in this work, we propose two indices for measuring the levels of *Anger* and *Joy* of songs that can be used, together with other indices characterising music and songs, to identify periods of time sharing similar instilled emotions. Both indices have been computed based on the top streamed songs listened to by Italians on Spotify during the first ten months of the COVID-19 pandemic. A multivariate-multidimensional analysis, considering the two proposed indices (*Anger* and *Joy*) and three indices provided by Spotify WEB API (*Loudness*, *Danceability* and *Tempo*) was run through the DISTATIS procedure. Five clusters strongly related to pandemic phases were identified.

6. Conclusions

In this work, we propose two indices for measuring the levels of *Anger* and *Joy* of songs that can be used, together with other indices characterising music and songs, to identify periods of time sharing similar instilled emotions. Both indices have been computed based on the top streamed songs listened by Italians on Spotify during the first ten months of the COVID-19 pandemic. A multivariate-multidimensional analysis,

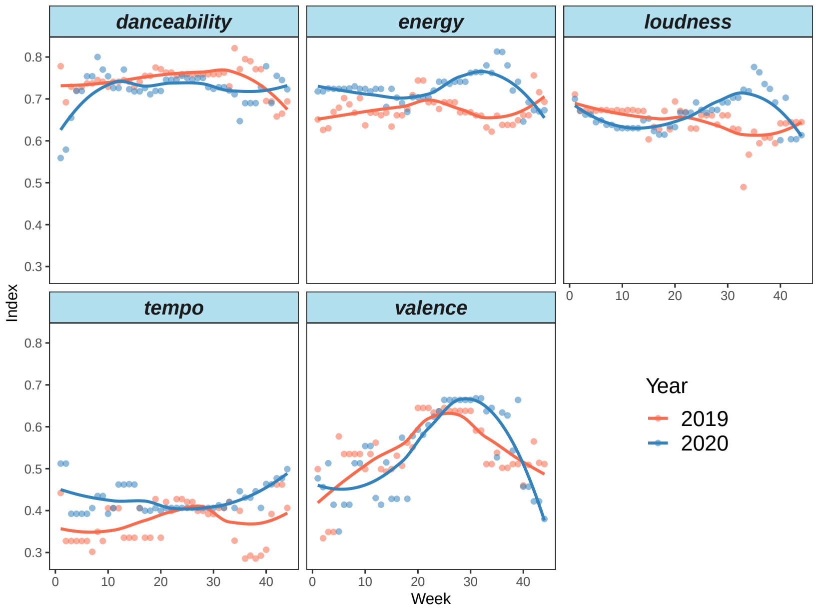

Figure 12: Indices time series.

considering the two proposed indices (*Anger* and *Joy*) and three indices provided by Spotify WEB API (*Loudness*, *Danceability* and *Tempo*), was run through the DISTATIS procedure. The results of DISTATIS were used as input for applying the PAM clustering algorithm to identify optimal clusters of weeks sharing similar emotions and features. Finally, we applied two validation measures. The proposed procedure allows not only to find the clusters of weeks and verify their robustness but also to quantify the weight of each index in affecting the results. Five clusters were identified; the first three cover the lockdown period, i.e., the first and second phases. Cluster 4 entirely covers summertime (as expected, summertime weeks fall into a unique cluster), and Cluster 5 includes the weeks characterised by new constraints due to increased infections. We have shown that all the variables considered play an important role in describing the relationship between Spotify Italian users' mood and government restriction. Specifically, high levels of *Anger* and low levels of *Joy* and *Danceability* are related to states of frustration due to quarantine and fear of the looming threat. Opposite values characterised the summertime-Cluster. Finally, the top streamed songs of Cluster 5 reflect that people's mood were worsening than summer. However, *Anger* levels are not as high, and *Joy* levels are not as low as in the first phase. Instead, we observe high levels of *Loudness* and *Tempo* and medium levels of *Danceability*, which perhaps shows that people were resigned to the bad situation.

Acknowkedgments

The authors want to sincerely thank Prof. Giorgio Fazio for his ideas and thoughtful discussions, which were key in developing this research.

References

Abdi, A., Shamsuddin, S. M., Hasan, S., and Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. *Information Processing & Management*, 56(4):1245–1259, https://doi.org/10.1016/j.ipm.2019.02.018.

Abdi, H., Dunlop, J. P., and Williams, L. J. (2009). How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage*, 45(1):89–95, https://doi.org/10.1016/j.neuroimage.2008.11.008.

Abdi, H., O'Toole, A. J., Valentin, D., and Edelman, B. (2005). DISTATIS: The analysis of multiple distance matrices. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pages 42–42. DOI: 10.1109/CVPR.2005.445.

Abdi, H., Valentin, D., Chollet, S., and Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, 18(4):627–640, https://doi.org/10.1016/j.foodqual.2006.09.003.

Abdi, H., Williams, L. J., Valentin, D., and Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):124–167, https://doi.org/10.1002/wics.198.

Aljanaki, A., Wiering, F., and Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management*, 52(1):115–128, https://doi.org/10.1016/j.ipm.2015.03.004.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., and Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, 395(10227):912–920, https://doi.org/10.1016/S0140-6736(20)30460-8.

Coelho, C. M., Suttiwan, P., Arato, N., and Zsido, A. N. (2020). On the nature of fear and anxiety triggered by COVID-19. *Frontiers in Psychology*, 11:3109, https://doi.org/10.3389/fpsyg.2020.581314.

Corrigall, K. A. and Schellenberg, E. G. (2013). Music: The language of emotion. *Handbook of Psychology of Emotions: Recent Theoretical Perspectives and Novel Empirical Findings*, 2:299–325, https://www.researchgate.net/publication/257431972_Music_The_language_of_emotion.

Demetriou, A., Larson, M. A., and Liem, C. (2016). Go with the flow: When listeners use music as technology. https://repository.ubn.ru.nl/bitstream/handle/2066/161892/161892.pdf

Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, pages 751–760, https://doi.org/10.2307/2529140.

Escoufier, Y. (1980). L'analyse conjointe de plusieurs matrices de données. *Biométrie et Temps*, 58:59–76.

Escoufier, Y. (1985). Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données. *Statistique et Analyse Des Donnees*, 10(1):1–10, http://www.numdam.org/item/SAD_1985__10_1_1_0/.

Gilligan, R. (2000). Adversity, resilience and young people: The protective value of positive school and spare time experiences. *Children & Society*, 14(1):37–47, https://doi.org/10.1111/j.1099-0860.2000.tb00149.x.

Grey, I., Arora, T., Thomas, J., Saneh, A., Tohme, P., and Abi-Habib, R. (2020). The role of perceived social support on depression and sleep during the COVID-19 pandemic. *Psychiatry Research*, 293:113452, https://doi.org/10.1016/j.psychres.2020.113452.

Hahn, G. J. and Meeker, W. Q. (2011). *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons, DOI: 10.1002/9780470316771.

HG, D. (2018). **Philentropy**: Information theory and distance quantification with R. *Journal of Open Source Software*, 3(26):765, http://joss.theoj.org/papers/10.21105/joss.00765.

Hu, X., Chen, J., and Wang, Y. (2021). University students' use of music for learning and well-being: A qualitative study and design implications. *Information Processing & Management*, 58(1):102409, https://doi.org/10.1016/j.ipm.2020.102409.

Hu, X., Downie, J., and Ehmann, A. (2009). Lyric text mining in music mood classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pages 411–416, https://archives.ismir.net/ismir2009/paper/000095.pdf.

Inskip, C., Butterworth, R., and MacFarlane, A. (2008). A study of the information needs of the users of a folk music library and the implications for the design of a digital library system. *Information Processing & Management*, 44(2):647–662, https://doi.org/10.1016/j.ipm.2007.05.004.

Jondya, A. G. and Iswanto, B. H. (2017). Indonesian's traditional music clustering based on audio features. *Procedia Computer Science*, 116:174–181, https://doi.org/10.1016/j.procs.2017.10.019.

Kartikay, A., Ganesan, H., and Ladwani, V. M. (2016). Classification of music into moods using musical features. In *2016 International Conference on Inventive Computation Yechnologies (ICICT)*, volume 3, pages 1–5. IEEE, DOI: 10.1109/INVENTIVE.2016.7830197.

Kaufman, L. and Rousseeuw, P. J. (1990). *Partitioning Around Medoids (Program PAM)*. John Wiley & Sons, Inc., Hoboken, NJ, USA, ISBN: 978-0-470-31680-1, DOI: 10.1002/9780470316801.ch2.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, DOI: 10.1002/9780470316801.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., and Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, volume 86, pages 937–952, https://archives.ismir.net/ismir2010/paper/000045.pdf

Laurier, C., Grivolla, J., and Herrera, P. (2008). Multimodal music mood classification using audio and lyrics. In *2008 Seventh International Conference on Machine Learning and Applications*, pages 688–693. IEEE, https://doi.org/10.1007/978-3-319-14442-9_3.

Lee, S. A., Mathis, A. A., Jobe, M. C., and Pappalardo, E. A. (2020). Clinically significant fear and anxiety of COVID-19: A psychometric examination of the coronavirus anxiety scale. *Psychiatry Research*, page 113112, https://doi.org/10.1016/j.psychres.2020.113112.

Li, F., Ng, T., and Hu, X. (2018). Emotion-aware music information retrieval based on physiological signals and user profile. In *International Society for Music Information Retrieval Conference (ISMIR)*. Télécom ParisTech and IRCAM., http://hdl.handle.net/10722/259797.

Li, T., Zhu, S., and Ogihara, M. (2007). Hierarchical document classification using automatically generated hierarchy. *Journal of Intelligent Information Systems*, 29(2):211–230, https://doi.org/10.1007/s10844-006-0019-7.

Limaye, R. J., Sauer, M., Ali, J., Bernstein, J., Wahl, B., Barnhill, A., and Labrique, A. (2020). Building trust while influencing online COVID-19 content in the social media world. *The Lancet Digital Health*, 2(6):e277–e278, https://doi.org/10.1016/S2589-7500(20)30084-4.

L'Hermier des Plantes, H. (1976). Structuration des tableaux à trois indices de la statistique. *Université de Montpellier II, Thesis.*

MacDonald, R. A. (2013). Music, health, and well-being: A review. *International Journal of Qualitative Studies on Health and Well-being*, 8(1):20635, https://doi.org/10.3402/qhw.v8i0.20635.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). ***cluster****: Cluster Analysis Basics and Extensions*. R package version 2.1.0 — For new features, see the 'Changelog' file (in the package source).

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465, https://arxiv.org/abs/1308.6297.

Nalini, N. and Palanivel, S. (2016). Music emotion recognition: The combined evidence of MFCC and residual phase. *Egyptian Informatics Journal*, 17(1):1–10, https://doi.org/10.1016/j.eij.2015.05.004.

Ni, M. Y., Yang, L., Leung, C. M., Li, N., Yao, X. I., Wang, Y., Leung, G. M., Cowling, B. J., and Liao, Q. (2020). Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: Cross-sectional survey. *JMIR Mental Health*, 7(5):e19009, DOI: 10.2196/19009.

Oppenheimer, E. (2019). Using data to find the angriest death grips song. https://towardsdatascience.com/angriest-death-grips-data-anger-502168c1c2f0.

Oramas, S., Barbieri, F., Nieto Caballero, O., and Serra, X. (2018). Multimodal deep learning for music genre classification. *Transactions of the International Society for Music Information Retrieval. 2018*, 1(1): 4–21, http://doi.org/10.5334/tismir.10.

Papinczak, Z. E., Dingle, G. A., Stoyanov, S. R., Hides, L., and Zelenko, O. (2015). Young people's uses of music for well-being. *Journal of Youth Studies*, 18(9):1119–1134, https://doi.org/10.1080/13676261.2015.1020935.

Peng, W., Li, T., and Ogihara, M. (2007). Music clustering with constraints. In *ISMIR*, pages 27–32, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3ff3057452675d7c3b86b777416739f240353e6c.

Pergola, G., Gui, L., and He, Y. (2019). TDAM: A topic-dependent attention model for sentiment analysis. *Information Processing & Management*, 56(6):102084, https://doi.org/10.1016/j.ipm.2019.102084.

Porcelli, P. (2020). Fear, anxiety and health-related consequences after the COVID-19 epidemic. *Clinical Neuropsychiatry*, 17(2), DOI: doi.org/10.36131/CN20200215.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161, https://doi.org/10.1037/h0077714.

Russo, M., Kraljević, L., Stella, M., and Sikora, M. (2020). Cochleogram-based approach for detecting perceived emotions in music. *Information Processing & Management*, 57(5):102270, https://doi.org/10.1016/j.ipm.2020.102270.

Saluja, V., Jain, M., and Yadav, P. (2019). L, M&A: An algorithm for music lyrics mining and sentiment analysis. In *CATA*, pages 475–483. https://doi.org/10.29007/wkj6.

Schwartz, R. W., Ayres, K. M., and Douglas, K. H. (2017). Effects of music on task performance, engagement, and behavior: A literature review. *Psychology of Music*, 45(5):611–627, https://doi.org/10.1177/0305735617691118.

Sciandra, M. and Spera, I. C. (2020). A model-based approach to spotify data analysis: A Beta GLMM. *Journal of Applied Statistics*, 49(1):214–229 , DOI: 10.2139/ssrn.3557124.

Sen, A. (2014). Automatic music clustering using audio attributes. *International Journal of Computer Science Engineering*, 3(6):307–312. https://www.ijcse.net/docs/IJCSE14-03-06-096.pdf.

Spera, I. C. and Sciandra, M. (2020). Ho perso le parole: Come ritrovarle con la sentiment analysis: Metodi statistici per l'analisi della produzione discografica di Luciano Ligabue. *TAB*, pages 1–145.

Tsai, W.-H., Rodgers, D., and Wang, H.-M. (2004). Blind clustering of popular music recordings based on singer voice characteristics. *Computer Music Journal*, 28(3):68–78, https://www.jstor.org/stable/3681510.

Vijayarani, S., Ilamathi, M. J., Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1):7–16.

Wang, J., Wang, Z., Liu, X., Yang, X., Zheng, M., and Bai, X. (2021). The impacts of a COVID-19 epidemic focus and general belief in a just world on individual emotions. *Personality and Individual Differences*, 168:110349, https://doi.org/10.1016/j.paid.2020.110349.

Wongso, R. and Santika, D. D. (2014). Automatic music genre classification using dual tree complex wavelet transform and support vector machine. *Journal of Theoretical & Applied Information Technology*, 63(1).

Yang, D. and Lee, W.-S. (2004). Disambiguating music emotion using software agents. In *ISMIR*, volume 4, pages 218–223, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.103.666.

Yang, L., Cheng, L., and Chen (2008). *Advances in Multimedia Information Processing-PCM 2008*, volume ch. 8, pp. 70–79. Springer.

Yao, Z., Tang, P., Fan, J., and Luan, J. (2021). Influence of online social support on the public's belief in overcoming COVID-19. *Information Processing & Management*, page 102583, https://doi.org/10.1016/j.ipm.2021.102583.

Zhao, Y., Cheng, S., Yu, X., and Xu, H. (2020). Chinese public's attention to the COVID-19 epidemic on social media: Observational descriptive study. *Journal of Medical Internet Research*, 22(5):e18825, DOI: 10.2196/18825.

Affiliation:

Alessandro Albano
Department of Economics, Business and Statistics
University of Palermo
Viale delle Scienze, 8, 90128 Palermo PA, Italy
E-mail: alessandro.albano@unipa.it

Irene Carola Spera
Department of Economics, Business and Statistics
University of Palermo
Viale delle Scienze, 8, 90128 Palermo PA, Italy
E-mail: irenecarola@libero.it

Mariangela Sciandra
Department of Economics, Business and Statistics
University of Palermo
Viale delle Scienze, 8, 90128 Palermo PA, Italy
E-mail: mariangela.sciandra@unipa.it

Antonella Plaia
Department of Economics, Business and Statistics
University of Palermo
Viale delle Scienze, 8, 90128 Palermo PA, Italy
E-mail: antonella.plaia@unipa.it

Journal of Data Science, Statistics, and Visualisation https://jdssv.org/
published by the International Association for Statistical Computing
http://iasc-isi.org/

ISSN 2773-0689
September 2024, Volume IV, Issue VI *Submitted:* 2023-07-30
doi:10.52933/jdssv.v4i6.95 *Accepted:* 2024-08-01