

Temporal Analysis and Visualisation of Music

 Luan Misael Gomes de Moura^{1,5},  Carlos Henrique Quartucci Forster²,
 Emanuel P. Fontelles^{3,5},  Vinicius A. Sampaio^{4,5},  Mardônio J. C. França⁵

¹Itaú Unibanco

²Instituto Tecnológico de Aeronáutica

³Universidade Federal do Ceará

⁴Universidade Estadual do Ceará

⁵Boitatá Lab

luan.moura@itau-unibanco.com.br, forster@ita.br,

emanuelfontelles@fisica.ufc.br, vinicius.sampaio@aluno.uece.br,

mardfranca@gmail.com

Abstract. *This paper proposes a temporal analysis for music metadata using a generative probabilistic model for collections the discrete datasets such as text corpora. This method is also a topic model that is used for discovering abstract topics from a collection of documents. The method is then applied to audio metadata and song lyrics extracted with Echo Nest® engine, Spotify® Lyrics Genius® API. Song data time series are generated by grouping data items by release date, genre and dominant topics (from LDA analysis). Using a technique from Network Theory we visualise how these topics, in this case, genres, are related to each other through time.*

1. Introduction

Music is a core human experience, a fundamental part of our evolution. Probably, we sang before we spoke in syntactically guided sentences. In animal worlds we have different representations of song, dolphins, and birds produce sounds, though not always melodic, but still rich in semantically communicative functions. Also, songs have a single main purpose as calling attention to oneself, expanding oneself, selling oneself, deceiving others, reaching out to others, and calling on others [Schulkin and Raglan 2014]. As humans, we have a unique creative capability to express ourselves in different tones, emotions, instruments, genres.

Many of our social behaviour is expressed in music transformations through decades, for example, slavery with blues, '70s with psychedelic rock, reggae is deeply linked to Rastafari religion in Jamaica. This last one was spread into a commercialised jazz field, being known first as “rudie blues”, then “ska”, later “blue beat”, and “rock steady”.

As time went by our society changed, we stopped being nomads and became farmers, we changed our geographic space and built cities, we experienced wars, naval explorations, industrial revolutions and a huge advance in technology, these advances are well documented in our music and cultural movements that accompanied them.

This paper focuses on the period from the second half of the 20th century to the present day, seeking to understand how the rhythms intertwined over time, producing new rhythms, as well as the composition of their lyrics, are affected with the evolution of society. We seek to understand how people's behaviour over the decades has impacted the process of musical evolution along with technological advances.

The technique of topic modelling is often applied to group texts according to the co-occurrence of words, which may reveal some common meanings of discourse. The analysis of interactions between words in a collection of documents determines the probability of a given document to be grouped in a certain topic. This technique can be applied to lyrics to determine groups of songs with similar ideas.

There are several works involving musical lyrics and topic modeling, one in specific and very common used is the Latent Dirichlet Allocation (LDA) [Blei 2003]. LDA is a probabilistic model used to find latent semantic topics in a large corpus. We can cite the LyricsRadar [Sasaki 2014] which is a system based on latent topics of lyrics generated by LDA that enables a user to visualise the music topics interactively. LDA can also be applied to identify sentiments from songs through the lyrics. Emotion recognition can also be done with audio analysis [Sharma 2011, Dakshina 2013]. Attempts to use LDA directly in music audio was tested however it does not give a clear intuition about the formed topics [Hu 2009].

Temporal and regional evolution of topics were discussed with rap lyrics [Johnson-Roberson 2013]. They applied topic modelling techniques to extract the topic distribution, in this cases emotions, of rap songs. With these topics, they grouped similar emotions and analysed his evolution through the years.

Audio metadata can be used to facilitate musical analysis of a large amount of data for musicological purposes. The audio metadata has been used in tasks such as prediction of hit songs [Fan 2013], understanding how playlists were created by users [Pichl 2016] and also how these features influence musical choices [Barone 2017].

This paper intends to contribute to temporal music analysis by proposing a method that combines the extraction of topics in song lyrics and a complex network representation to visualise the interaction of musical genres through the years. Our contributions are:

1. We propose to combine audio metadata with topics probabilities as features to group songs by musical genres and years;
2. We propose a complex network representation to compute the correlation between musical genres and years based on Minimum Spanning Tree algorithm;
3. We were able to visualise how musical genres are related over the past seven decades.
4. We have also contributed to creating a large dataset with song lyrics and audio metadata from the 50s until now.

2. Data

The audio data was scrapped using Echo Nest® API integrated engine with *spotipy* Python's package. The *spotipy* API permits the user to search for specific genres, artists, songs, release dates, etc. To obtain the lyrics we used the Lyrics Genius® API as base URL for requesting data based on the song title and artist name.

Table 1. Examples of manually selected songs of our dataset with a very high and low value for each acoustic feature.

feature	band/group	song	value
acousticness	Bob Dylan	I'll keep it with mine	1.0
	Nirvana	Scentsless Apprentice	0.0
danceability	Vanilla Ice	Ice ice baby	0.994
	Oasis	Slide Away	0.067
loudness	The Stooges	Death Trip	1.0
	Pink Floyd	Speak to me	0.059
instrument.	The Rolling Stones	Turd on the run	0.991
	Areta Franklin	It's so heartbreakin'	0.0
valence	The reggae boys	Mama look deh	1.0
	Jimi Hendrix	...and the gods made love	0.03
energy	Slayer	Altar of sacrifice	0.999
	King Crimson	Peace	0.004

Using *spotipy* we fetched data from songs of 7 musical genres (rock, reggae, jazz, blues, hip hop, country, pop) and release date ranging from 1950 to 2019. Our dataset consisted of 82452 songs distributed of 7 musical genres and release dates ranging from 1950 to 2019. The main information retained was the artist name, track name, release date, genre and track id. The track id is a unique id for each searched track. We used the 'track id' as input to the *spotipy*'s audio features tool and we kept only some of these features. The selected features were:

- **Acousticness:** Presence of acoustic instruments;
- **Danceability:** how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity;
- **Loudness:** The average loudness in decibels (dB) across the entire track;
- **Instrumentalness:** A high value describes whether a track contains fewer vocals;
- **Valence:** High (low) values means that the track is more happy, euphoric (sad, angry);
- **Energy:** Measures intensity and activity of music. Energetic tracks will be fast, loud and noisy.

More information can be found in [for Developers.]. These features were scaled between 0 and 1. For a better understanding of the features, we selected songs manually with a very high and low value of each feature and summarised in Table 1.

Using Genius API, we queried lyrics using the artist name and song name of the same songs in which we obtained the metadata. We started cleaning the texts by identifying the language using Google's library language-detection [Shuyo 2010] and removed all non-English texts. One of the patterns found in the retrieved lyrics was the presence of bracketed texts including the artist who sings that phrase and if the phrase is

either a verse, chorus, or intro. There were also texts in parentheses which in most cases were onomatopoeia or backing vocals. These texts between parenthesis/brackets were removed. We cleaned the remaining texts consisted of removing symbols, numbers, and stop words like common English words and proper nouns. The remaining words were lemmatized to its canonical form using WordNet Lemmatizer [Fellbaum 2005] provided by NLTK package[Bird 2009]. The data is available at [Moura 2020].

3. Topic Modelling of lyrics

Latent Dirichlet allocation (LDA) is a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities [Blei 2003]. This definition above may be a little difficult to understand but the main idea is that documents are represented as random mixtures over topics (*latent topics*). Here, these topics are characterized by a distribution over words. Finally, a document can be represented by the probability distributions it contains.

To understand the terms used in the LDA model, we first define some notation as following [Blei 2003]. A *word*, w , is the basic unit of discrete data, defined to be a *item* from a vocabulary indexed by $\{1, \dots, V\}$. Words are represented by a unit-basis vector, for example one-hot encoding representation. The i th word in the vocabulary is represented by a V -vector w such that $w_i = 1$ and $w_j = 0$ for $i \neq j$.

A *document* d_j is a sequence of n_j words represented by a vector $d_j = (w_{j,1}, \dots, w_{j,n_j})$ and a *corpus* is the collection of m documents: $D = (d_1, \dots, d_m)$. The vocabulary, then can be viewed as the collection of all unique words in the corpus.

The topics are distributions over words of the corpus. The generating process of a document d_j uses two variables generated by the Dirichlet distribution (*Dir*): $\varphi = (\varphi_1, \dots, \varphi_K)$ is $n \times K$ dimensional array where each φ_k is a n -dimensional variable and n is the number of unique words on the corpus, the vocabulary. The φ_K gives the proportions of words which describes the K th topic. The second variable, $\theta = (\theta_1, \dots, \theta_m)$ has a dimension $K \times m$ where m is the number of documents. Each vector θ_j is K -dimensional where K is the number of topics, θ_j represents the proportions of each topic in the document d_j . The variables θ and φ represents the probability distributions with $\sum_j^n \varphi_j = 1$, $\varphi_i > 0$ and $\sum_i^K \theta_i = 1$, $\theta_i > 0$.

The hyperparameters of the Dirichlet distribution for φ and θ are, respectively, β and α . More details about LDA mathematical description as also hyperparameters explanations can be found in [Blei 2003] and [Hoffman 2010].

3.1. Training and evaluating model

The input data of the LDA model was a term frequency matrix generated with the lyrics corpus. We have ignored terms that had a document frequency higher than 80% when we generated the vocabulary, to eliminate very common words that would lead to the creation of irrelevant topics [Salton and McGill 1986] and to find relevant words that can be distinguishable between topics. We also ignored terms with a document frequency lower than 1% to remove very rare and specific words.

We have used the batch learning method on the *scikit-learn* [Pedregosa 2011] implementation of the LDA. The batch method updates variational variables after each full pass through the data while the online method updates variational variables from mini-batch data points [Hoffman 2010]. We used the default parameters $\alpha = \beta = 1/K$ of scikit-learn’s implementation and varied the number of topics K .

To evaluate the best K , we have computed the marginal log likelihood based on the approximation algorithm chosen. To choose the best number of topics, we range K over the interval [5, 50] using a step size of 5. For a validation strategy, we have split the data into 30% for the test and evaluated the log likelihood metric in both sets for each K . The mean log likelihood for the stored K values in train and test data was very discrepant so we scaled the values dividing by the absolute value of the maximum K . We have found that the best K lied between [12, 23] so we adjusted the number of topics and calculated again the LDA model for each value K in that new range with unitary step. The best number of topics was $K = 19$. After the selection of the best number of topics, we have trained the LDA model with all the data.

3.2. Evaluating formed topics

We evaluated the cosine distance, Eqn. (1), of words per topic using a 100-dimension GloVe word embedding [Pennington 2014] vector representation. To visualize the best number of words to represent each topic, we select every word proportion per topic k represented by φ_k distribution by the most frequent term in that topic to the rarest term.

$$similarity = \cos(\Theta) = \frac{\varphi_i \cdot \varphi_j}{\varphi_i \varphi_j} \quad (1)$$

For each topic, we have chosen the first 10 words of φ_K and computed the cosine distance for every pair of their embedding vectors and stored the mean result in a variable $g_{k,10}$. We repeated the same methodology for the first 11 to 25 words of φ_k in each topic k calculating the mean cosine distance of the word embeddings. Based on those criteria, we select an appropriate number of words to represent each topic that maximizes the mean cosine similarity. Each topic received a label based on the main subject of the selected words. A stop word list was used to exclude common English words. The stop word list used is built-in Scikit-learn. Here we display some selected words of the principal topics and the number in parenthesis represents the mean cosine similarity of the topic words:

- **dating** (0.442): ‘baby’ ‘babe’ ‘loving’ ‘lady’ ‘tell’ ‘darling’ ‘crazy’ ‘cause’ ‘leave’ ‘say’ ‘right’ ‘girl’ ‘need’ ‘treat’ ‘bring’
- **violence** (0.492): ‘head’ ‘stand’ ‘black’ ‘come’ ‘dead’ ‘fight’ ‘kill’ ‘blood’ ‘hand’ ‘cold’ ‘face’
- **world/life** (0.631): ‘life’ ‘live’ ‘world’ ‘change’ ‘time’ ‘believe’ ‘days’ ‘remember’ ‘dream’ ‘things’ ‘come’
- **night/time** (0.576): ‘time’ ‘right’ ‘night’ ‘long’ ‘tonight’ ‘mind’ ‘wait’ ‘stay’ ‘ready’ ‘line’
- **shake the audience** (0.434): ‘yeah’ ‘need’ ‘everybody’ ‘whoa’ ‘right’ ‘know’ ‘cause’ ‘alright’ ‘gotta’
- **family/gospel** (0.369): ‘little’ ‘lord’ ‘mama’ ‘work’ ‘help’ ‘daddy’ ‘hard’ ‘brother’ ‘sister’ ‘woman’ ‘home’

Table 2. Examples of manually chosen songs with a high value of topic proportion for each of the six selected topics.

topic	artist	song	score
violence	Misfits	Scream	0.982
romantic	Doris Day	Here in my arms	0.932
obscene	Rihanna	Bitch better have my money	0.992
sound	Ten Years After	Hear me calling	0.902
sadness	Frank Zappa	Tears began to fall	0.925
feelings	Warren Hill	Do you fell what I'm feeling?	0.959

- **romantic** (0.407): ‘hold’ ‘love’ ‘sweet’ ‘kiss’ ‘woman’ ‘hand’ ‘goodbye’ ‘touch’ ‘arm’ ‘darling’ ‘heart’ ‘true’
- **communication/thinking** (0.727): ‘know’ ‘tell’ ‘think’ ‘say’ ‘cause’ ‘things’ ‘talk’ ‘thing’ ‘mean’ ‘maybe’
- **obscene** (0.363): ‘like’ ‘fuck’ ‘cause’ ‘money’ ‘better’ ‘shit’ ‘know’ ‘about’ ‘bitch’ ‘party’ ‘play’
- **sound** (0.484): ‘home’ ‘sing’ ‘hear’ ‘blue’ ‘song’ ‘play’ ‘music’ ‘come’ ‘bring’ ‘sound’ ‘call’
- **movement/places** (0.543): ‘say’ ‘go’ ‘drink’ ‘road’ ‘ride’ ‘round’ ‘come’ ‘home’ ‘walk’ ‘city’ ‘look’ ‘drive’ ‘take’
- **light/visual perceptions** (0.462): ‘light’ ‘dance’ ‘turn’ ‘eye’ ‘night’ ‘burn’ ‘high’ ‘moon’ ‘come’ ‘dream’ ‘like’ ‘happiness’
- **family/spiritual** (0.454): ‘people’ ‘free’ ‘shake’ ‘young’ ‘mother’ ‘land’ ‘trouble’ ‘children’ ‘send’ ‘child’ ‘father’
- **like/girls** (0.547): ‘like’ ‘girl’ ‘look’ ‘alright’ ‘make’ ‘pretty’ ‘beautiful’ ‘girls’ ‘know’ ‘feel’
- **sadness** (0.510): ‘away’ ‘heart’ ‘fall’ ‘break’ ‘leave’ ‘tear’ ‘walk’ ‘lonely’ ‘feel’ ‘inside’ ‘wish’ ‘hurt’ ‘stay’
- **feelings** (0.581) ‘feel’ ‘go’ ‘good’ ‘somebody’ ‘miss’ ‘real’ ‘morning’ ‘strong’ ‘feeling’ ‘luck’ ‘wrong’ ‘thing’

The mixture of very different words in φ_k makes some topics very hard to characterize. These topics were labeled with a tag “**undefined**” and were dropped. Songs with a very high proportion of a particular topic (~ 0.99) have a high degree of repeated words.

Some of the topics formed were very generic, such **communication/thinking**, **light/visual perceptions**, **movement/places**, **world/life**, **night/time**. Given all these issues, we selected only six main topics to analyze: **feelings**, **sound**, **obscene**, **sadness**, **romantic**, **violence**. The table 2 summarises selected songs with a high score for each topic.

3.3. Time series analysis

At this point, we have generated time series for each topic grouping data per genre and release date. We computed the mean value of the topic probability for the songs in each genre and year. We have used a moving average with a window size of 5 years to smooth the data. Some of the results were summarised in Fig. 1 and we suppressed some of the genres to get a cleaner image.

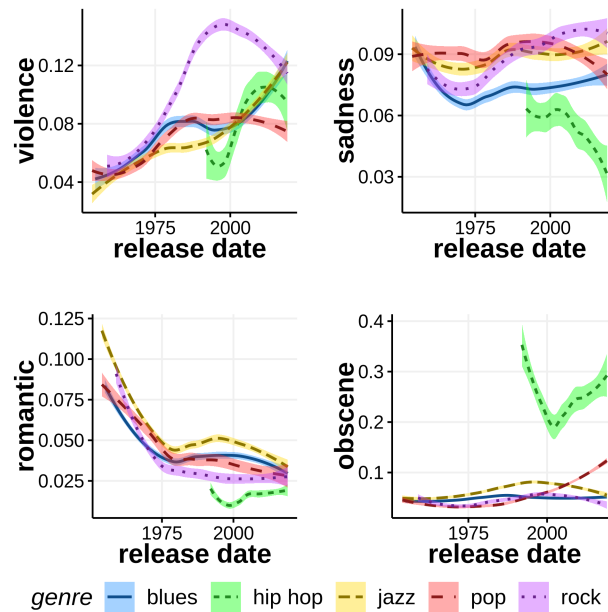


Figure 1. Songs were grouped by years and musical genres. Year by year, we have computed the mean proportion for each topic, thus generating a time series. A moving average with a window size of 5 years was used to smooth the data. The line plots have a confidence interval of 95%.

In the context of **violence**, our first topic, rock lyrics showed a very high peak near the 2000s. Also near the 2000s, we observed the emergence of the hip-hop genre. This genre presents a high probability of usage of words that were in the context of **violence** topic. However, it has a low score on **romantic** topic. Hip hop and reggae used to talk more about **feelings** but the trend slope decreased with time. For blues and jazz, we have observed a rising trend in the probability of words that were in the context of the **violence** topic also. Country and hip hop songs showed a decreasing trend of usage of words assigned to the **sad** topic. Most of the musical genres showed a decreasing trend on **romantic** topic.

A closer look at hip hop songs we had a very high probability of showing words present in the **obscene** topic and the model tended to emphasize that as the dominant topic in hip hop lyrics. Fine-tuning would be necessary to find better topics by running the LDA model just on hip hop lyrics. That approach was explored using rap songs [Johnson-Roberson 2013]. Using audio metadata we have applied the same procedure used to the topics probabilities, creating time series of the mean values of songs in the same genre and release date. The time series were illustrated in Fig. 2.

The genres with the highest **danceability** features were reggae and hip hop. We can see clear trends in that feature. The music became **louder** with time and the availability of plenty of synthetic instruments and modern recording devices can saw on the decreasing slope of the **acousticness** feature. Jazz songs had a tendency to have fewer vocals than the other genres. The overall **valence** showed a small decrease. Reggae and hip hop showed a clear decreasing trend in its valence meaning that the audio transmitted a sadder feeling.

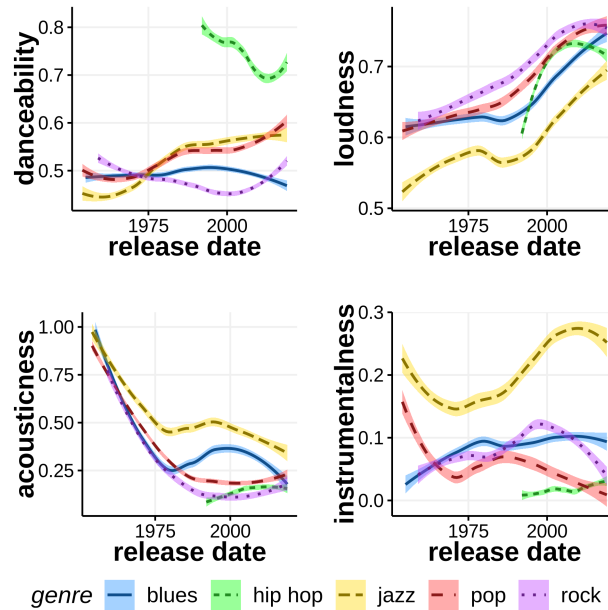


Figure 2. Songs were grouped by years and musical genres. Year by year, we have computed the mean proportion for each acoustic features, thus generating a time series. A moving average with window size of 5 years were used to smooth the data. The line plots have a confidence interval of 95%.

A different approach was also tried by grouping the audio metadata by its release date and the dominant topic. We have searched for the songs that had a topic proportion bigger than 30% for one of the six selected topics for analysis. These songs received a topic label for the dominant topic and corresponded to $\sim 34.4\%$ of the data. From that point, we worked only with the data with dominant topics. The generated series can be visualized in Fig. 3.

Most of the selected topics showed an increasing trend in its danceability. Lyrics with obscene language showed to be more danceable and a high slope close to 80s can be seen, this possibly happened due to the emergence of hip hop records. The series generated for the loudness and acoustics features does not seem to differ much for each genre. The mean slope of these features matches the information of Fig. 4. The topic **feelings** showed to have a very distinct behavior compared to the other genres for the **instrumentality** feature. Music that talks about feelings had less instrumental parts in their songs but increased over the years while obscene songs became less instrumental. We can see that all topics showed a rising trend in its valence feature and a decreasing slope can be seen after the 80s.

4. Complex Networks Analysis

To understand how the musical genres were correlated through time we grouped data by its release date and musical genre extracting the mean value. We computed the correlation distance $d_{ij} = \sqrt{2(1 - r_{ij})}$ where r_{ij} is the correlation coefficient between the points i and j . We used d_{ij} as edge weights for edges E on an undirected graph $G(V, E)$ with vertex V the set of all grouped mean points, and computed the minimum spanning tree $MST(S, F)$ with a subset of vertex S and edges F [Prim 1957]. The drawing of the formed tree was

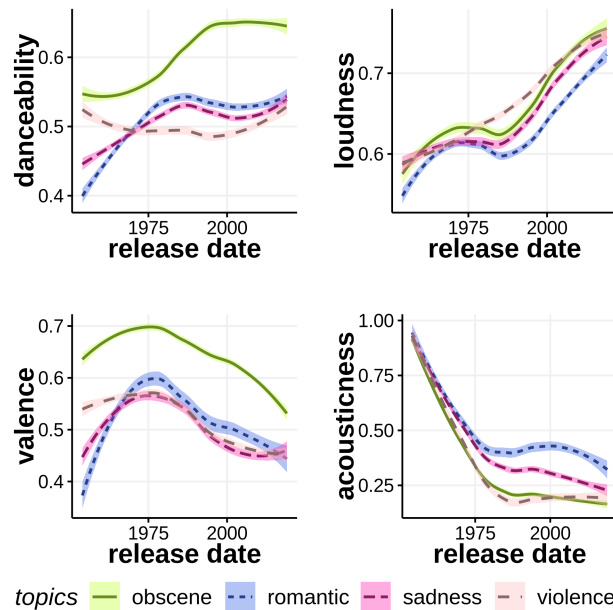


Figure 3. Grouped audio per topic and release date. Samples were grouped by years and the dominant topic taking the mean value of the acoustic features generating a time series for each feature. The dominant topic consists of a topic with a high proportion of a sample. A moving average with a window size of 5 was used to smooth the data. The line plots have a confidence interval of 95%.

made using the Kamada-Kawai algorithm which was designed for general undirected graphs [Kamada 1989]. We generated a tree coloring vertex by its genre and increased node size based on its release year. Bigger nodes correspond to older mean values of grouped data.

Some genres have a high similarity with each other, principally reggae, jazz, rock, and pop. There is also a big mixture of older rock, blues, jazz, and country forming a branch with great colorful nodes. We find in the year-granularity tree some groups of small nodes from the same genre indicating the specialization of the genres differing from each other. We also grouped data by its decade instead of years, this allows us to see a big picture of the total evolution. The resulting trees are in Fig. 5. The Year-Granularity Tree (YGT) here proposed, showed a correlation between reggae songs and hip hop that the decade's tree could not find. The country genre was very mixed with many genres like jazz, pop, blues, rock. The YGT also shows that the rock and pop groups join all the tree branches while in the decades tree the pop and blues nodes are in the middle of the tree connecting all branches.

5. Conclusions

We have shown that the LDA technique is capable of finding good topics in a collection of music lyrics. The topics found seem to talk about specific feelings with a positive mean cosine distance between the words (transformed into word embedding) of a given topic. On that side, the mean probability of topics in a group of lyrics presented in the same year and genre gave us a way to investigate the evolution of topics in different musical

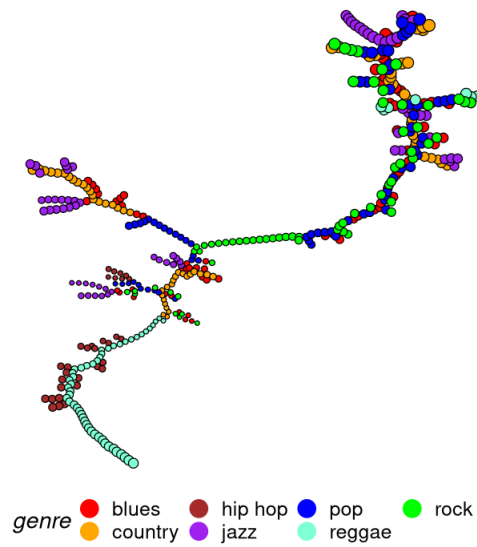


Figure 4. Minimum Spanning Tree generated using the correlation distance of songs represented by acoustic and lyric topic data grouped by years and taking the mean value. The colors represent different musical genres while the diameter represents the year, old music group nodes have a bigger diameter.

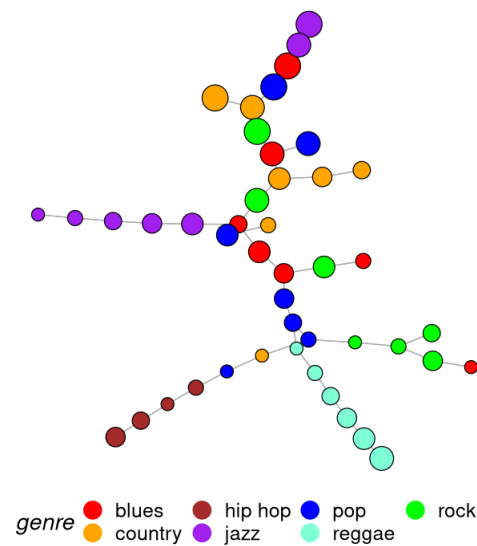


Figure 5. Minimum Spanning Tree generated using the correlation distance of songs represented by acoustic and lyric topic data grouped by decades and calculating the mean value. The colors represent different musical genres while the diameter represents the decade, music from ancient decades is represented by group nodes with a bigger diameter.

genres through time. That idea was already explored in [Johnson-Roberson 2013] with rap music. We have generalized to different kinds of topics and genres showing the music evolution of the most common music genres. We also investigated the evolution of audio metadata through the years/genres and years/topics.

The main contribution of this paper is the characterization of music by its topic probabilities and audio metadata which condenses both audio and lyric information of the songs. In our second main contribution, we proposed here a way of visualizing music time series evolution, we called Year-Granularity Tree, which is a tree where each node is a (genre, year) structure interacting with other nodes by a correlation distance d . The correlation distance between these nodes is the input of the MST algorithm which is capable of calculating the closest genres of our graph and generate a tree structure that connects genres and years in a way that we can see the relationship of different epochs in a nonlinear time direction.

As future works, we propose a different vectored representation of lyrics and audio of different languages. One can use the M-BERT [Devlin 2019] as a vectored representation of the multilingual lyrics in a document embedding approach. The audio metadata can be complemented by the embedding representation of the audio files using the Speech2vec technique [Chung 2018]. With our new features, we can apply clustering techniques and analyze the characteristics of the formed clusters and propose recommendation systems based purely on the music.

6. Acknowledgment

This research was supported by Itaú Unibanco. Any opinions, findings, and conclusions expressed in this manuscript are those of the authors and do not necessarily reflect the views, official policy or position of Itaú Unibanco. We thank our colleagues from Itaú, ITA, UFC, and Boitatá Lab who provided insights and expertise that greatly assisted this research.

I thank Carlos Henrique Quartucci Forster for assistance with methodology review, Mardônio França for the proposition, and aid with Latent Dirichlet Allocation and Ludwing Ferney Marenco Camacho for the proposition of Minimum Spanning Tree usage and for comments that greatly improved the manuscript.

References

- Barone, M. D., B. J. . W. M. H. (2017). Acoustic features influence musical choices across multiple genres. *Frontiers in psychology*, 8:931.
- Bird, S., K. E. . L. E. (2009). *Natural language processing with python*. O'Reilly Media Inc.
- Blei, M., N. A. . J. M. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, pages 993–1022.
- Chung, Y.; Glass, J. (2018). Speech2vec: A sequence-to-sequence framework for learning word embeddings from speech. *arXiv preprint arXiv*.
- Dakshina, K., . S. R. (2013). Lda based emotion recognition from lyrics. *Advanced Computing, Networking and Informatics*, 1:187–194.

- Devlin, J., C. M. L. K. T. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL*.
- Fan, J., . C. M. (2013). Study of chinese and uk hit songs prediction. *Proceedings of International Symposium on Computer Music Multidisciplinary Research*, pages 640–652.
- Fellbaum, C. (2005). Wordnet and wordnets. *Encyclopedia of Language and Linguistics, 2nd ed*, pages 665–670.
- for Developers., S. Get audio features for a track. Available at <https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/> (2019/03/11).
- Hoffman, M., B. F. R. . B. D. M. (2010). Online learning for latent dirichlet allocation. *Advances in neural information processing systems*.
- Hu, J. (2009). Latent dirichlet allocation for text, images, and music. *University of California, San Diego*.
- Johnson-Roberson, C., . J.-R. M. (2013). Temporal and regional variation in rap lyrics. *NIPS Workshop on Topic Models: Computation, Application and Evaluation*.
- Kamada, T. K. S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15.
- Moura, L., F. E. S.-V. F. M. (2020). Music topics and metadata. <https://data.mendeley.com/datasets/3t9vbwxgr5/1>.
- Pedregosa, F. (2011). Scikit-learn: Machine learning in python. *Journal of machine Learning research*, 12:2825–2830.
- Pennington, J., S. R. . M. C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Pichl, M., Z. E. . S. G. (2016). Understanding playlist creation on music streaming platforms. *IEEE International Symposium on Multimedia (ISM)*, pages 475–480.
- Prim, R. C. (1957). Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401.
- Salton, G. and McGill, M. J. (1986). Introduction to modern information retrieval. *McGraw-Hill, Inc., USA*.
- Sasaki, S., Y. K. N. T. G. M. . M. S. (2014). Lyricsradar: A lyrics retrieval system based on latent topics of lyrics. *Ismir*.
- Schulkin, J. and Raglan, G. B. (2014). The evolution of music and human social capability. *Frontiers in Neuroscience*.
- Sharma, Govind, . N. M. (2011). Mining sentiments from songs using latent dirichlet allocation. *International Symposium on Intelligent Data Analysis.*, pages 328–339.
- Shuyo, N. (2010). Language detection library for java.