Music, mind and technology

MUSIC PLAGIARISM

CONTENTS

PROBLEM STATEMENT

In this project, we aim to explore automated ways of detecting and understanding plagiarism in music. We review existing work on plagiarism detection in music in terms of sampling, melody, rhythm, etc., and compare the results of these models. Ultimately we provide a thorough examination of the patterns of music plagiarism among Indian composers, drawing upon the insights obtained during the initial phase of the project.

INTRODUCTION

- Music plagiarism is when someone uses a significant portion of another person's music composition without permission or proper attribution. It can also refer to the act of claiming someone else's music as one's own. Plagiarism can occur in many forms, including using the melody, chord progressions, lyrics, or other significant aspects of a song without permission or credit. In many cases, music plagiarism is illegal and can result in legal action being taken against the plagiarizer.
- Each year, over 10,000 new albums of recorded music are released and over 100,000 new musical pieces are registered for copyright. However, there are no general rules that set a minimum number of similar notes or beats for music copyright infringement.

MILESTONES

Literature review to decide on a bipartite matching based approach for working with melody plagiarism

We have chosen to follow a Smith-Waterman based bipartite matching algorithm for melody inspection after an extensive and intensive literature review of different possible plagiarism analysis methods.

Generation of a dataset of 154 plagiarized Indian songs and their original versions

We refocus our project scope to analyze the patterns of plagiarism by popular Indian Composers typically from Hindi and Tamil languages. Hence we compile a dataset of snippets of their songs along with the original songs that were plagiarized.

Building on existing paper's implementation to give similarity scores for our dataset

The code obtained from the research paper that does a melodic similarity sequencing based on a function of pitch, downbeat, and tempo is expanded to give similarity scores for song pairs for every composer. These are recorded with the existing database.

Adding genre, county of origin, and musical feature information to expand the dataset

The database is expanded by adding genre and region information for all original songs. Additionally timbre, pitch, and tempo information is added for the purpose of understanding patterns of plagiarism during the analysis phase.

Identifying patterns of plagiarism through statistical analysis.

Conclusions were drawn by building multiple relevant representations of obtained correlations like a world heat map of plagiarism intensity based on cumulative similarity scores, correlation matrices between artist and genres copied, and language and musical feature based analytics.

LITERATURE REVIEW

1. Types of Music Plagiarism and respective inspection methods
    a. Sampling
    b. Rhythm
    c. Melody
2. Examples
3. Plagiarism or inspiration?
4. An adaptive meta-heuristic for music plagiarism detection based on text similarity and clustering
5. Identification and Detection of Plagiarism in Music using Machine Learning Algorithms
6. Music Plagiarism Detection via Bipartite Graph Matching

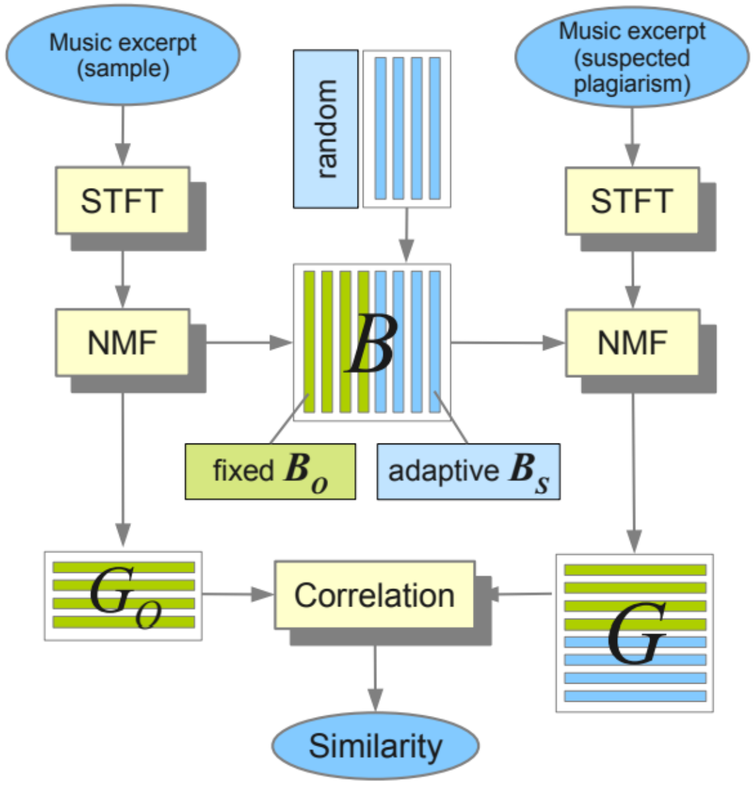
TYPES OF MUSIC PLAGIARISM

1. Sampling
2. Rhythm
3. Melody

SAMPLING

- **re-use of recorded sounds or music excerpts in another song**. Methods:
  a. The samples are often manipulated in pitch or tempo to fit the rhythm and tonality of the new song.
  b. mix additional instruments to the sample, such as additional vocals or drums.
  c. crop an excerpt of one or more bars and loop them.
  d. rearrangement and post-processing of the respective sample beyond recognition

INSPECTION METHODS



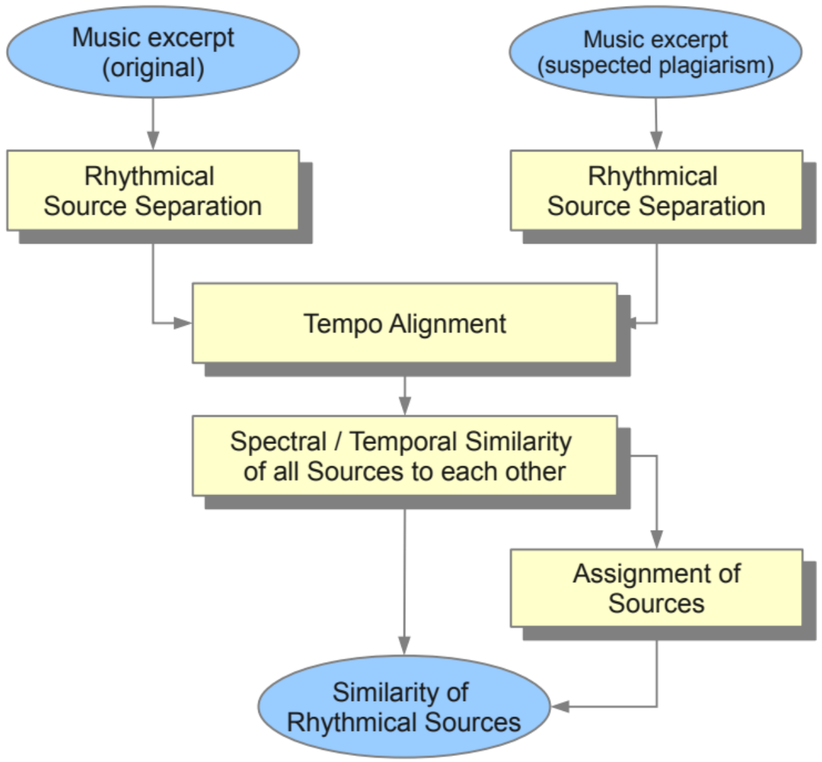
1. Due to the fact that sampling is basically the use of “a song in a song” it is related to the task of cover song detection. Cover song detection is commonly approached by **chroma features**, i.e., descriptor, which represents the tonal content of a musical audio signal in a condensed form that can be used for harmonicity similarity analysis.
2. Compare a time-frequency representation of both music excerpts by means of **Short-Term Fourier Transform (STFT)**.
   a. In order to retrieve the occurrences of **Xo** inside **Xs**, it is re-sampled both in time and frequency yielding **X_o**.
   b. Each **X_o** is shifted frame-wise along all frames of Xs and the accumulated, absolute difference **d** is computed between all corresponding time-frequency tiles.
   c. Assuming only re-sampling and looping were applied, **periodic minima will occur in d**. These correspond to the point, where an optimal matching can be found.

RHYTHM

Rhythm is the pattern of sound, silence, and emphasis in a song. In music theory, rhythm refers to the recurrence of notes and rests (silences) in time. When a series of notes and rests repeats, it forms a rhythmic pattern.

More formally, rhythm is formed by periodical pattern of accents in the amplitude envelopes of different frequency bands. Commonly the drums make up the beat or the guitar is playing the rhythm.

INSPECTION METHODS



1. Extract rhythmical features such as the beat spectrum or tempo in order to measure rhythmical similarity.
2. we assume, that the original rhythm may have undergone a number of manipulations, such as time stretching, pitch shifting, re-sampling or even shuffling of individual beats: Steps:
   a. **Rhythmic source separation** (using NMF – clustering of the components is necessary, since NMF often splits one instrument into several components. The assignment of components to each other is based on evaluating the correlation between the amplitude envelopes.)
   b. **Tempo alignment** to compensate for the difference in re-sampling factor.
   c. **Similarity of individual sources** using Pearson Coefficient.

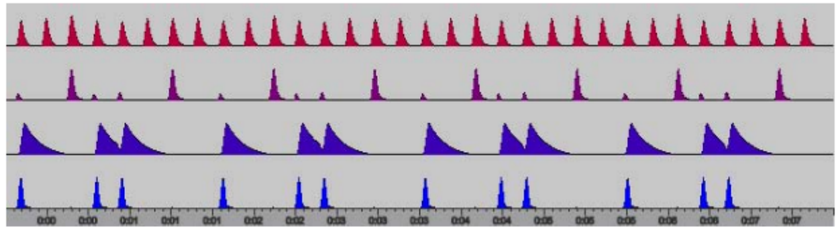
MELODY

Copied melodies are less obvious than the previously explained plagiarism types.

A melodic motive is considered to be identical, even if it is transposed to another key, slowed down, sped up or interpreted with different rhythmic accentuation.

Thus, melody plagiarism is a gray area, where it is hard to discern copying from citation.

INSPECTION METHODS



1. In the MIR literature, a closely related task is Query-by-Humming (QbH). QbH can be used to retrieve songs from a database by letting the user hum or sing the respective melody. Melody plagiarism inspection can be done with basically the same approach, since means to identify and evaluate melodic similarity are required. The main difference is, that QbH searches across extensive databases while plagiarism detection concentrates on one single comparison, which has to be more precise.
2. Sequence Alignment - relies on the **Smith-Waterman algorithm** to find a local alignment between symbol-sequences. The algorithm tries to identify subsequences of symbols, which encode intervals between consecutive notes in the MIDI transcription. On execution, each of these melody fragments is compared to the entire suspect sequence. The resulting scores are ordered descending and presented via the graphical user interface.

EXAMPLES

POPULAR COURT CASES OF MUSIC PLAGIARISM

George Harrison Vs. The Chiffons

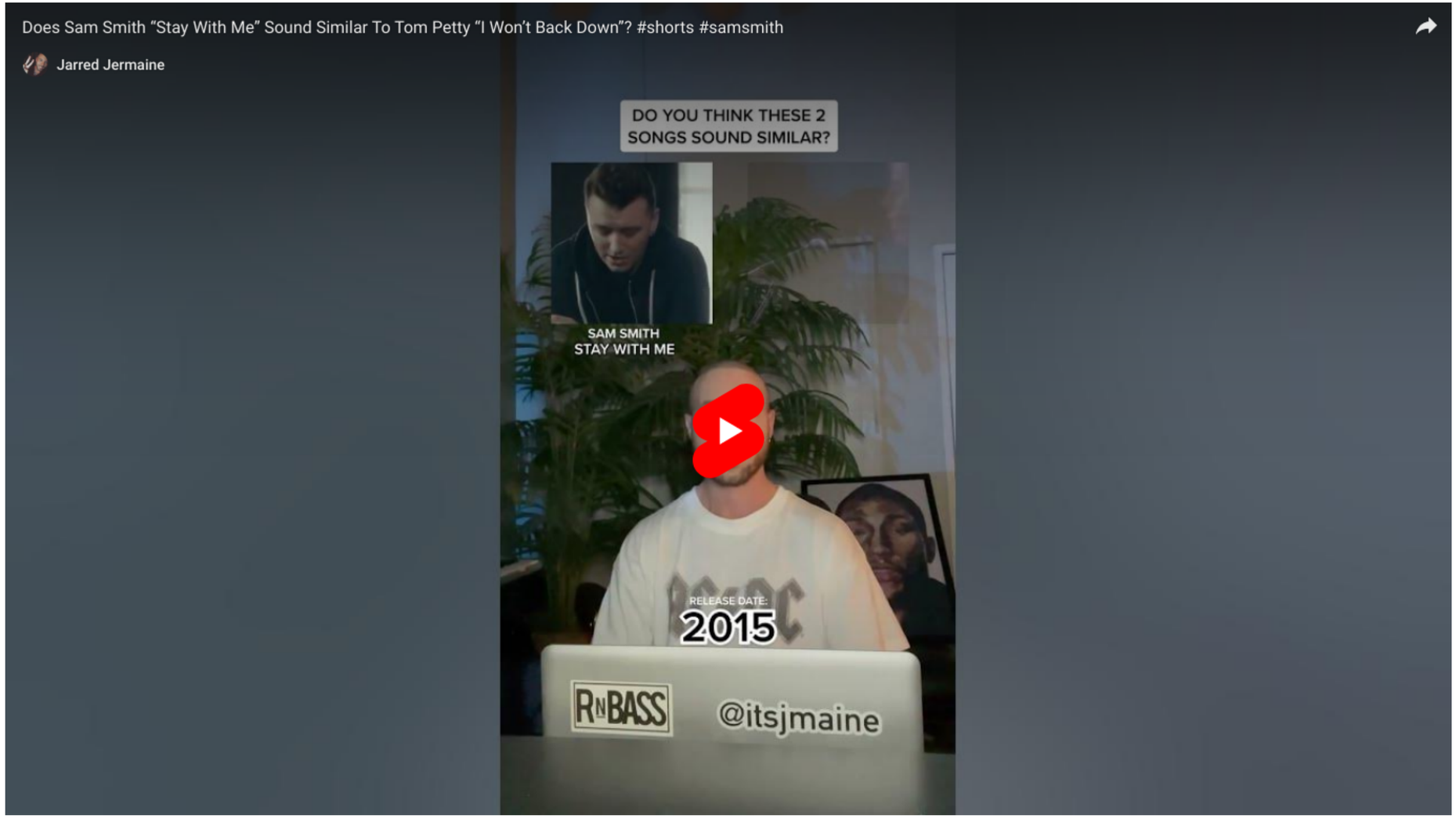
Harrison's "My sweet lord" sounded similar to Chiffon's "He's so fine". The judge declared the Beatles had "subconsciously" stolen the melody.

Sam Smith Vs. Tom Petty

"Stay with me" pitches are the same; the rhythms are the same; the chords are pretty much the same as "I won't back down". Petty gets 12.5% royalties off Sam's song.

Robin Thicke Vs. Marvin Gaye

"Blurred Lines" by Thicke and Pharell Williams claimed that their track was only inspired by the "feel" of Gaye's song, "Got to give it up" but to no avail. The jury awarded damages of nearly £5 million.



EXAMPLES

POPULAR COURT CASES OF MUSIC PLAGIARISM

We came across an article on the relevance of melody as a marker for plagiarism in Pop and Rock music. “Where lies the threshold of musical plagiarism?” Can we say that the similarities are coincidental which comes down to plagiarism or to something in between that may be labelled "inspiration? How the concepts of “idea” and “inspiration” translates to the language of music? Musical plagiarism is primarily a matter of a close similarity of a rather long melody. Inspiration concerns more subtle melodic similarities, chord progressions, special effects, style of arrangement, lyrical themes, and so on.

PLAGIARISM OR INSPIRATION?

Common chord progressions and generic rhythmical patterns cannot be copyright-protected. We don't know about any cases of plagiarism focusing primarily on rhythm. The rhythmical similarity is not even an easy-to-notice phenomenon unless you are focusing on it.
When comparing melodies, a perfect matching means both, identical pitch and identical rhythm (timing). Melody is usually the deciding element in cases of plagiarism. Similar but not identical fragments can instigate a sense of similarity.

PLAGIARISM OR INSPIRATION?

LITERATURE REVIEW

An adaptive meta-heuristic for music plagiarism detection based on text similarity and clustering

The paper tries to identify similarities between melodies of pop music. Their goal was to build an automated system able to take as input two melodies, as MusicXMLfiles, and provide an indication of their similarity (a percentage). The paper presents 2 methods- a text similarity-based method (in which they convert the music sheets into text strings using a technique called PINL representation. This representation uses symbols for pauses, intervals, and base note lengths to represent the melody as a text string.) and a clustering-based method (in which each melody in the knowledge base is first converted into text using the PINL representation and then into a vector of real numbers using the char2vec technique. The char2vec technique is particularly suitable for embedding music representations, and different vector sizes are used for the experiments) and further combine to get an improved hybrid method. To assess the effectiveness of the proposed methods the authors performed tests on a large dataset of ascertained plagiarism and non-plagiarism cases.

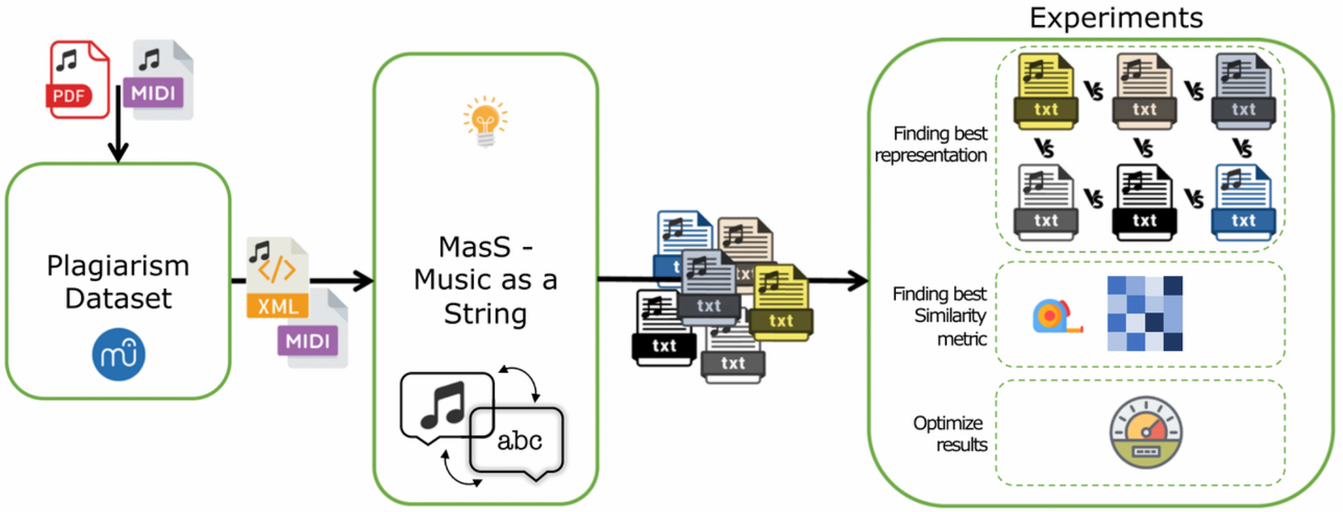
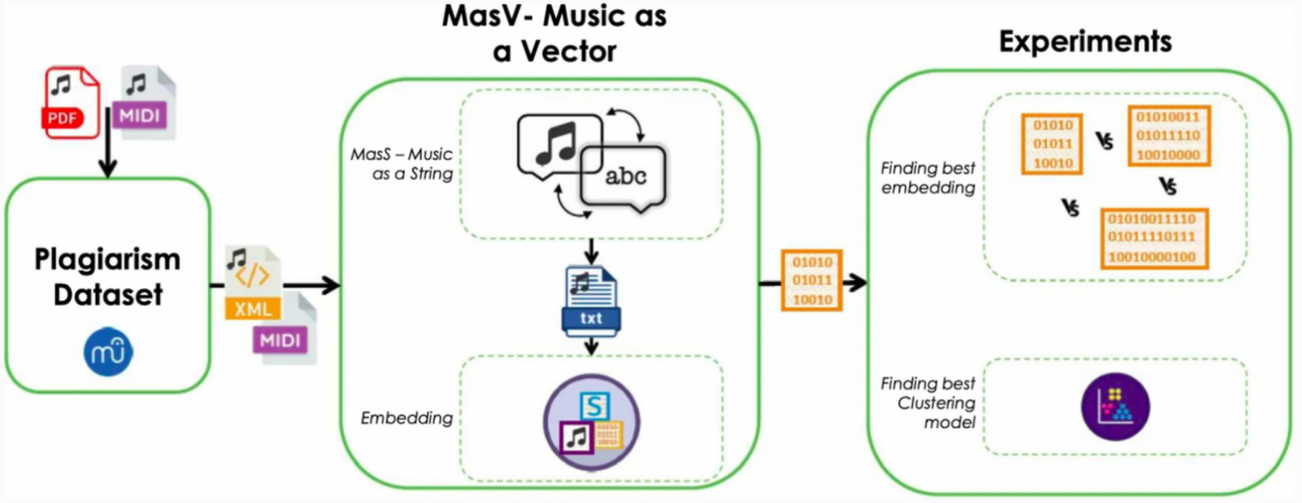


Fig. 1 The methodology followed to build the text similarity-based method for music plagiarism detection: (i) *Knowledge-base construction*, (ii) *MasS—Music as String*, (iii) *Experiments*



The methodology followed to build the clustering-based method for music plagiarism detection: (i) *Knowledge-base construction*, (ii) *MasV—Music as Vector*, (iii) *Experiments*

LITERATURE REVIEW

Identification and Detection of Plagiarism in Music using Machine Learning Algorithms

The authors first perform feature extraction to extract the note or the chord progression then, the harmonic reduction is performed to understand the structure of the music and then using Word2Vec model is applied to get the relationship between similar chords to perform chord substitution which will be the final data that is extracted for the classier models(KNN, Logistic Regression, Random Forest, DecisionTree, Gaussian Naïve Bayes) to predict plagiarism and the results were obtained.

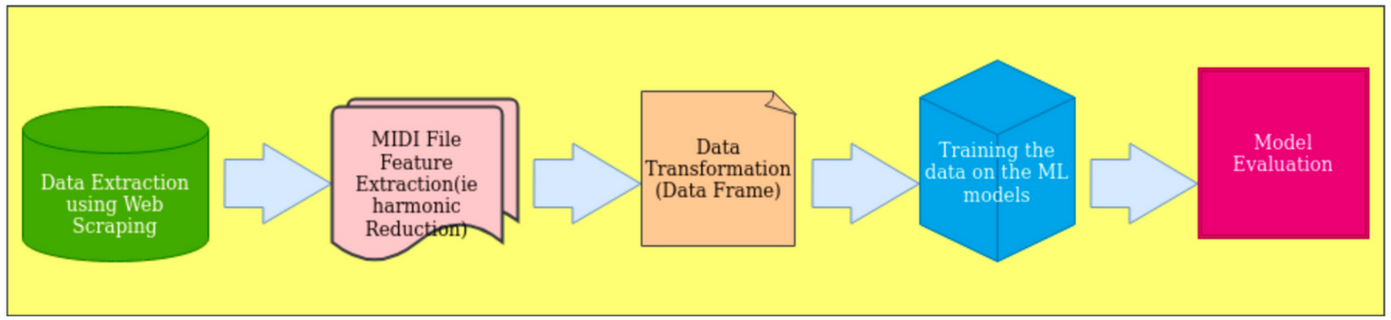


Figure 1: The flow indicating methods that is used to identify plagiarism in music

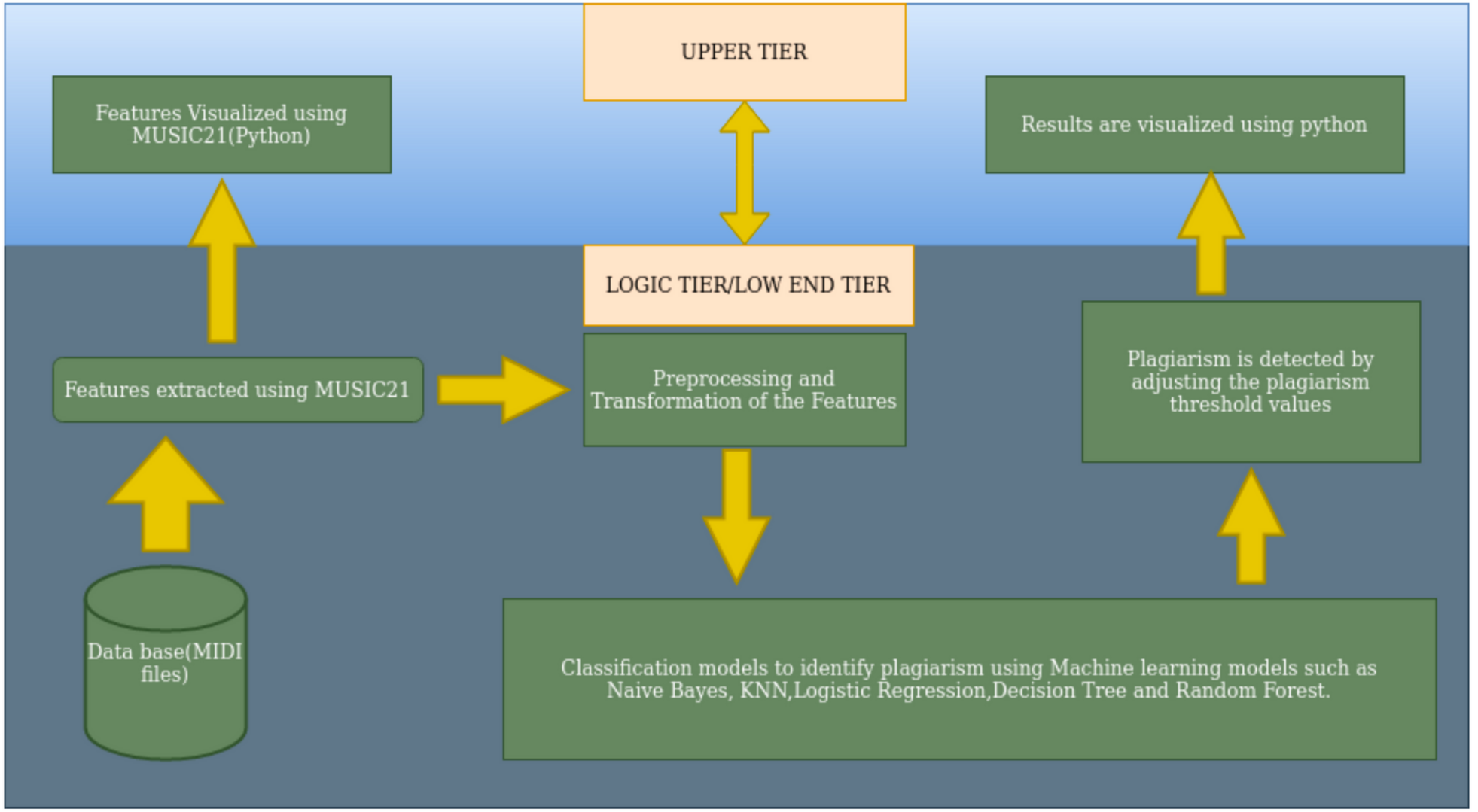


Figure 3: The flow indicating methods that is used to identify plagiarism in music

LITERATURE REVIEW

Music Plagiarism Detection via Bipartite Graph Matching

This paper proposes a new method (MESMF) which reduces the music plagiarism detection problem into a bipartite graph maximum matching task. The authors designed several kinds of melody representations and similarity computation methods.

Audio-based and sheet-based methods

Audio-based methods inspect music similarity using time-frequency representation or low-level features such as cepstrum coefficients. The whole song may be a copy of another one, but if the order of the notes is shifted then the audio-based algorithms cannot detect this situation.

Sheet-based methods measure **symbolic melodic similarity** which plays a crucial role in MIR. Symbolic melodic similarity evaluates the degree of similarity as human listeners can do.

MESMF (Maximum weight matching and edit distances model applied on the Sequences of Melodic Features)

Every note can be represented as a pair i.e. it's (pitch and duration).

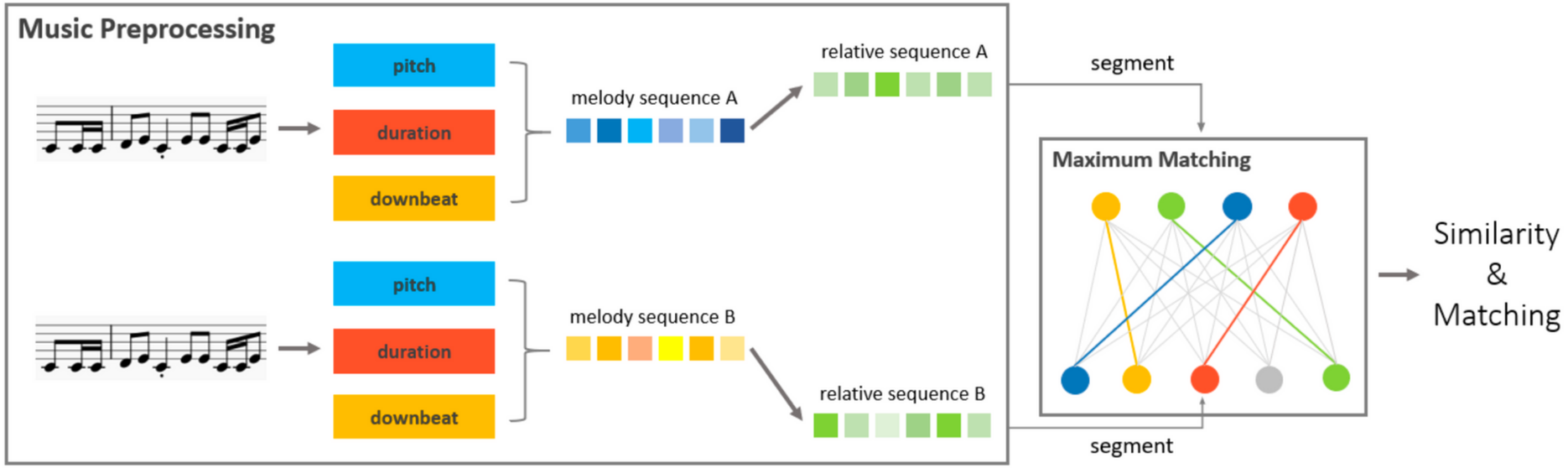
Given two songs A and B, the algorithm extracts their melodic features and transforms them into sequences.

A_s = {a_1, a_2, ... a_n), where the ith note of song A, a_i is represented as

ai = (pitch_ai, duration_ai, downbeat_ai), where downbeat_ai denotes whether this note is downbeat.

The sequences are transformed to relative form by subtraction of pitch and division of duration between neighbouring elements.

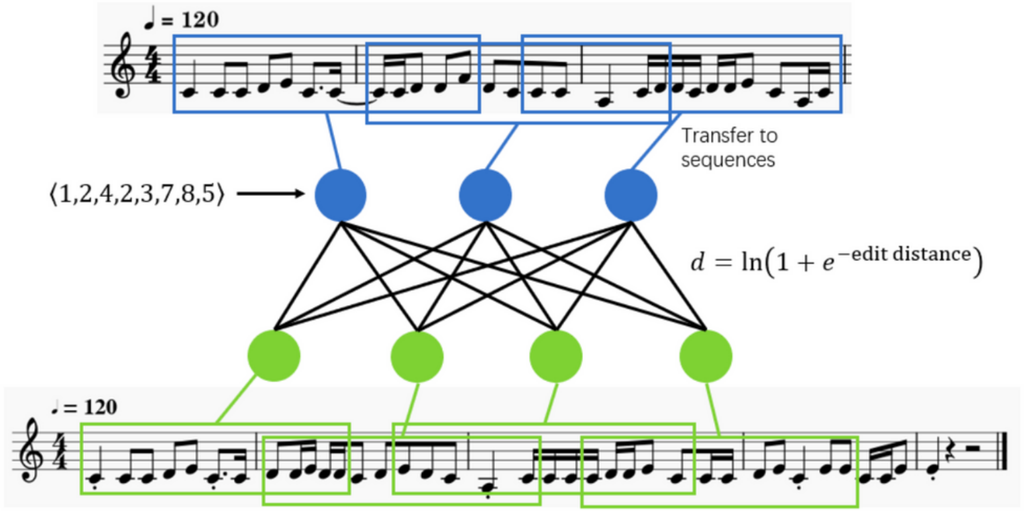
The two sequences A_s and B_s are divided into pieces with the same length l and overlapping rate r and are treated as nodes in the bipartite graph. We perform the maximum weight matching algorithm to get the final similarity.



The edge between two nodes has a weight equal to the function of the edit distance.
The dataset used in the paper consists of some well-known music plagiarism cases.

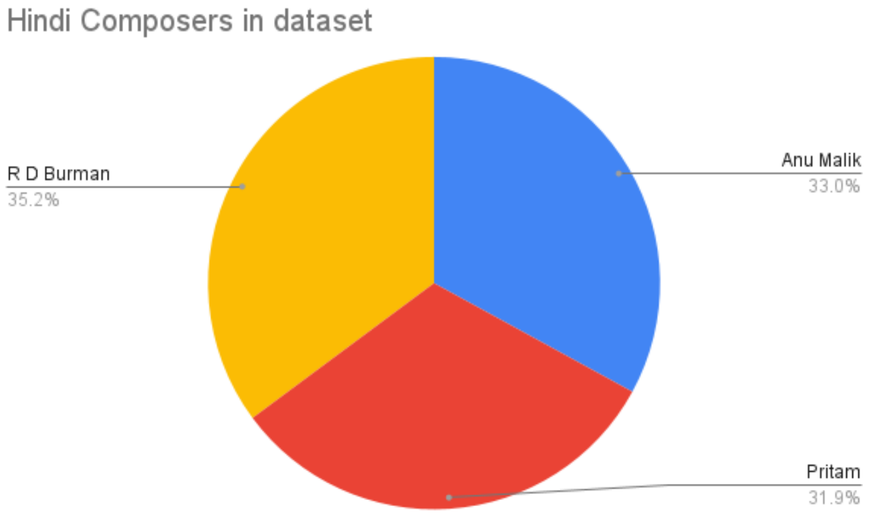
Pros

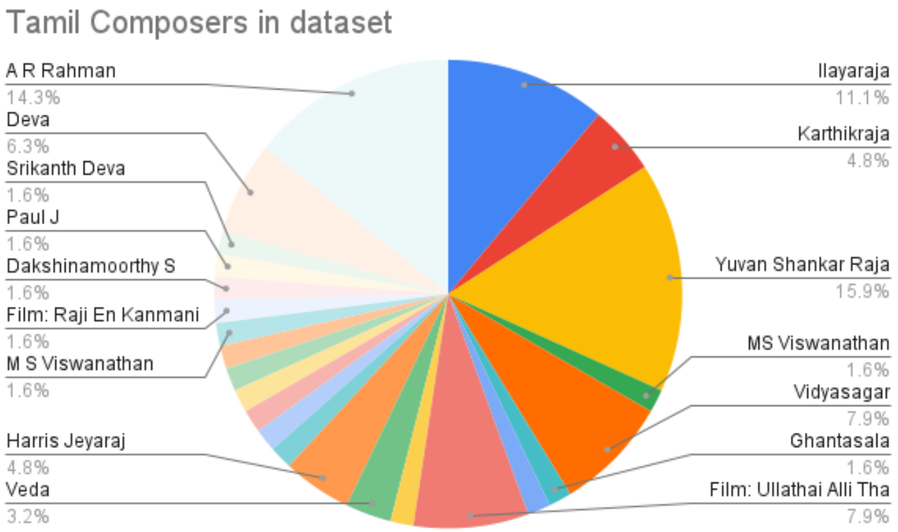
- The proposed methods deals with shift, swapping, transposition, and tempo variance problems in music plagiarism.
  - Transposition means increasing the pitches of all the notes to the same degree. (relative pitch)
  - Tempo invariance means that speeding up or slowing down the musical pieces will not influence the duration sequences we extract from them. (relative duration)
- It can also effectively pick out the local similar regions from two musical pieces with relatively low global similarity.



DATASET

- Our dataset consists of **91 Hindi** and **63 Tamil** song snippets along with the original song snippets that these Indian songs have most closely copied.
- We have hence made a dataset of a total of **300+** music files (original + plagiarized) available in **.rm** formats.
- This can be accessed on <drive link>





METADATA

- The dataset has been annotated to include the following information for each data point-
  a. Plagiarised song
  b. Original song
  c. Original Composer
  d. Plagiarising Composer
  e. Language
  f. Genre(s) of the original song
  g. Country of origin
  h. Tempo difference
  i. Timbral similarity
  j. Melodic similarity score (by Bipartite Graph Matching)

- Additionally, for each artist, we have cumulated similarity data country-wise and genre-wise for analysis purposes.
- <sheet link>

METHODS

Ground Truth

We consider the information on <website_link> as our ground truth.

Data collection

We collected the rm files, genre, region and artist information for each song

METHODS

.rm file

Took each pair of .rm audio files (original and plagiarised)

.mp3 file

Converted the rm file to mp3 using pydub.AudioSegment. Clipped each mp3 file to 60s

.midi file

Converted mp3 file to midi using Spotify's basic_pitch library

melodic_similarity_score

Calculated the melodic similarity between the two midi files using the Music Plagiarism Detection via Bipartite Graph Matching algorithm

METHODS

.rm file

Took each pair of .rm audio files (original and plagiarised)

Tempo Difference

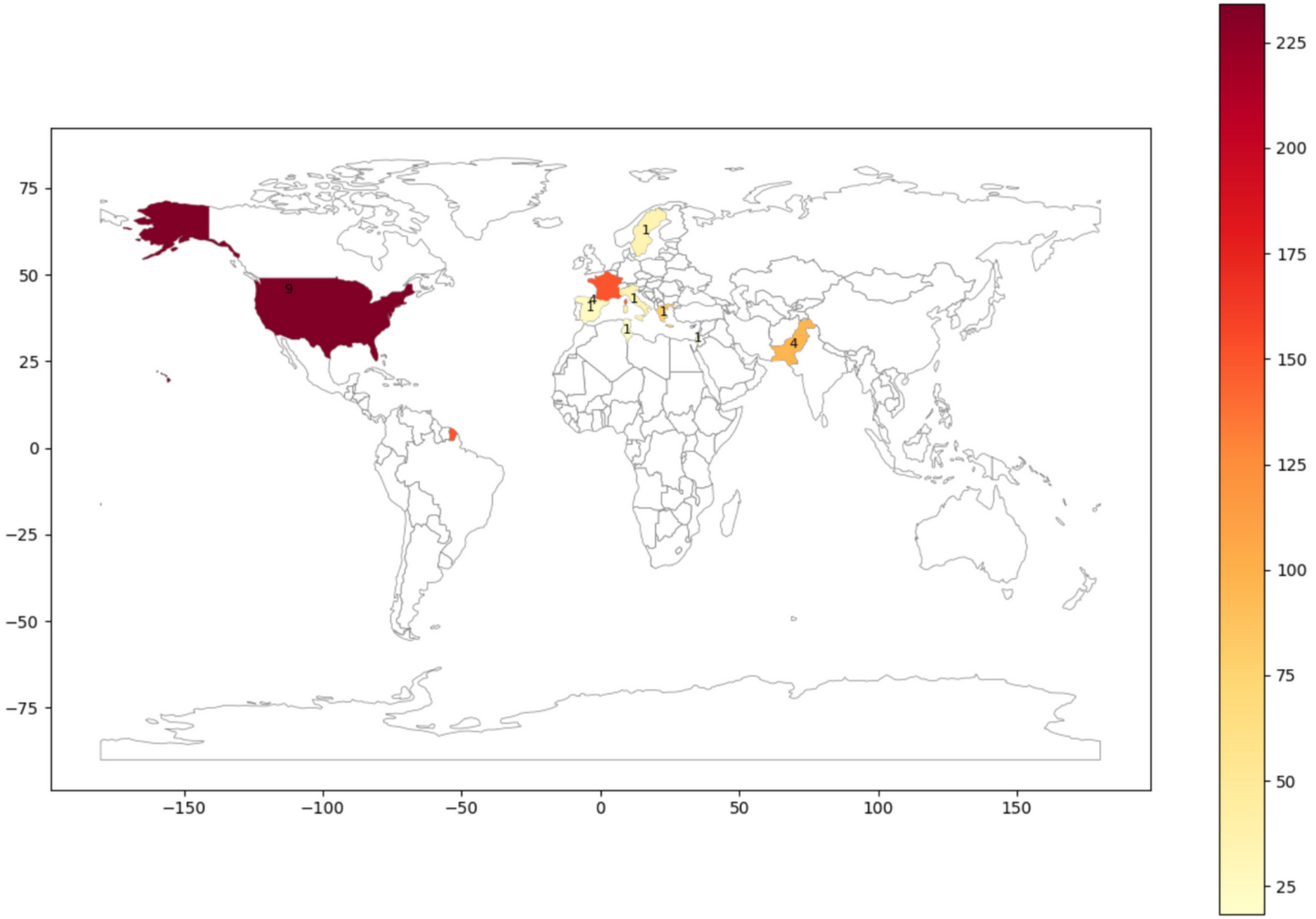
Found the tempo of each audio file using librosa.beat.beat_track. Calculated the absolute difference between the two.

Timbral Similarity

Computed the mel-spectrogram and chroma features for each file, flattened the feature matrices into 1D vectors, concatenated them into a single feature vector and computed the computed the cosine distance between the two.

Geographical Heatmap

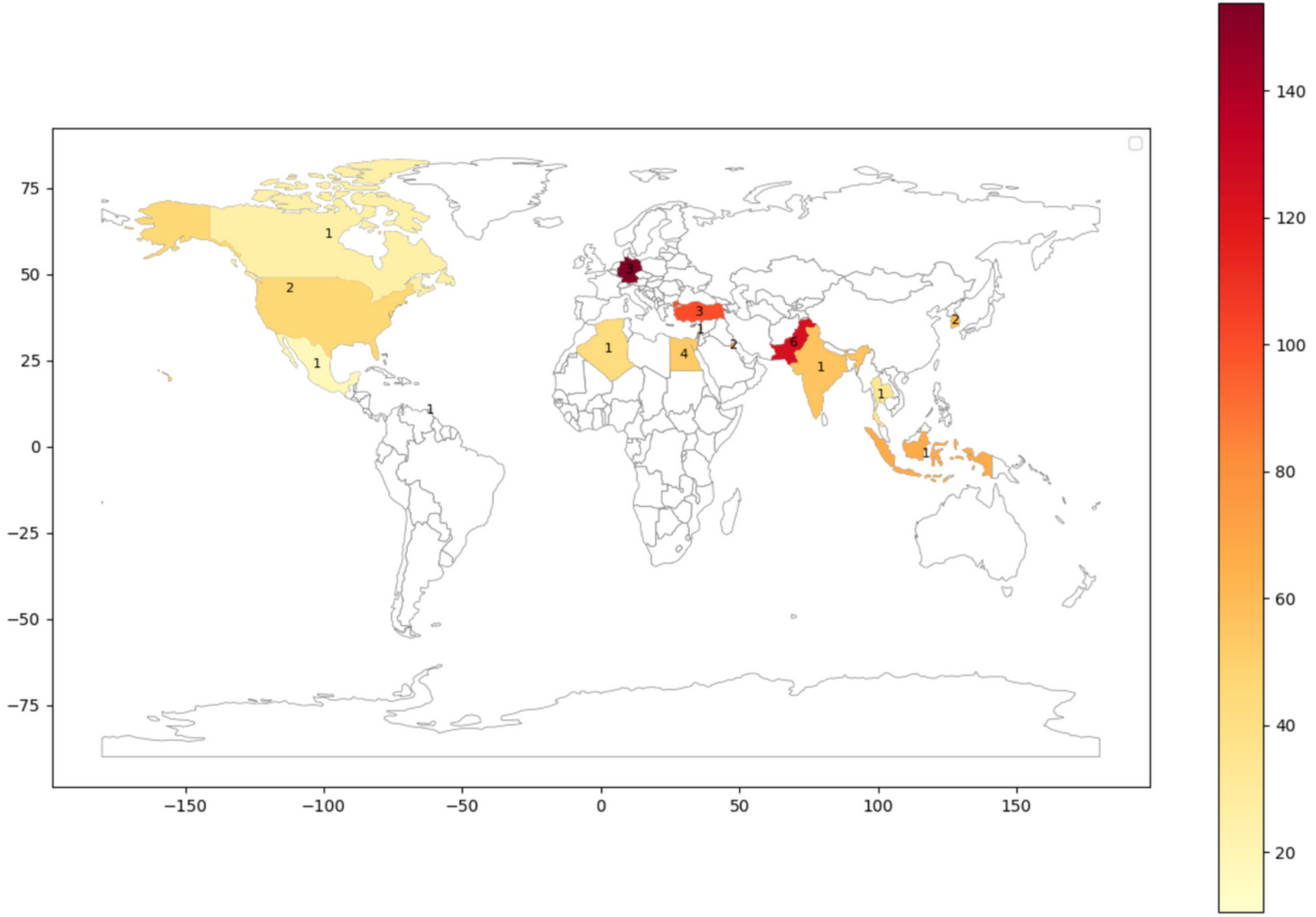
Visualised the region-wise cumulative similarity scores using geopandas



The above map marks the countries from which songs have been copied by Anu Malik. A majority of songs have been plagiarized from original **American music** with intense melodic similarities. Overall very high similarity scores were obtained for most songs indicating **minimum effort ripping off of original compositions** by Anu Malik.

SISTAHA

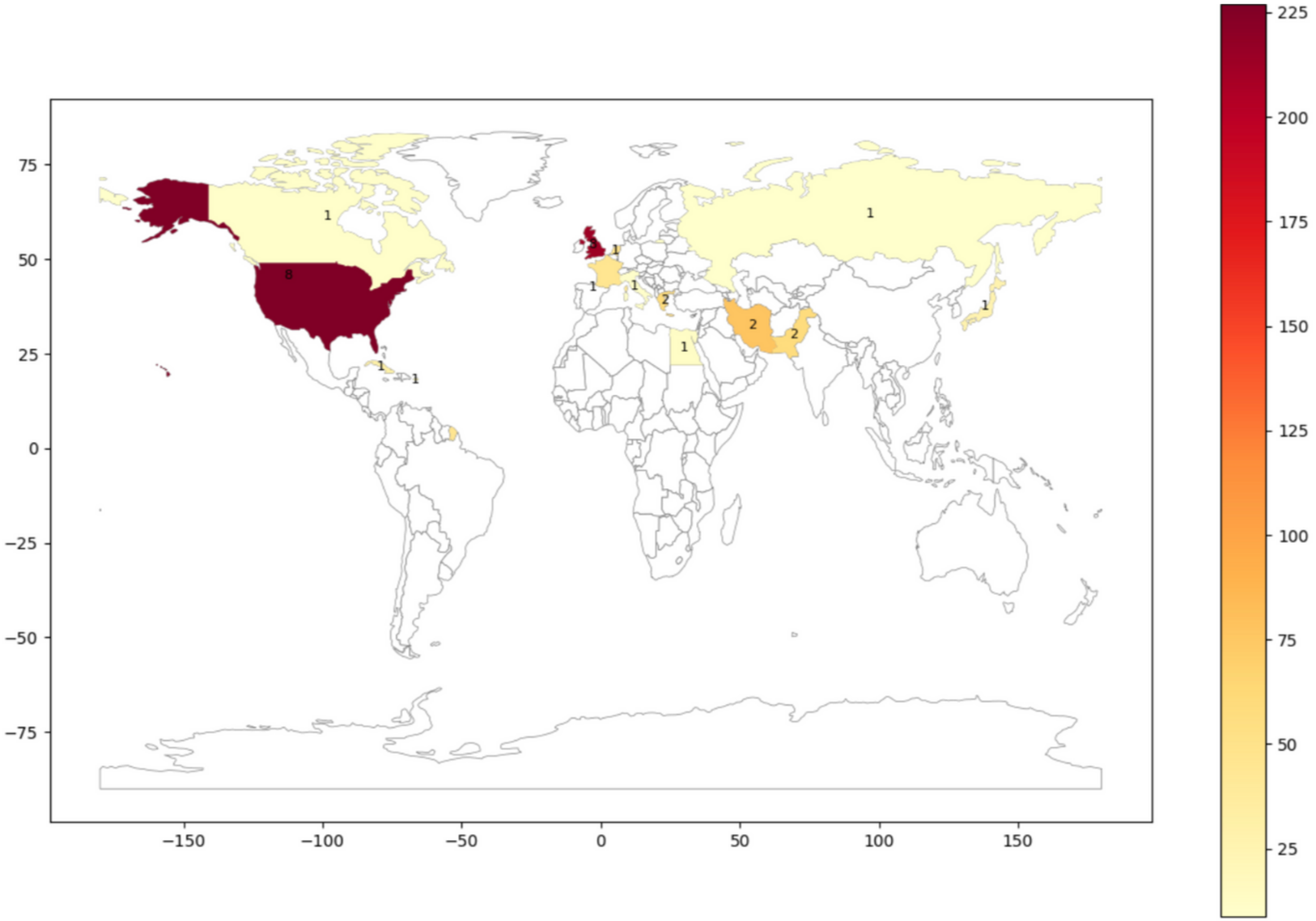
ANU MALIK



The above map marks the countries from which songs have been copied by Pritam. The results clearly indicate a high usage of German music as shown by intense melodic similarity between original and copied songs. Overall we observe a high influence of **German, Middle-eastern and Pakistani** music in Pritam's music composition.

SISLAYA

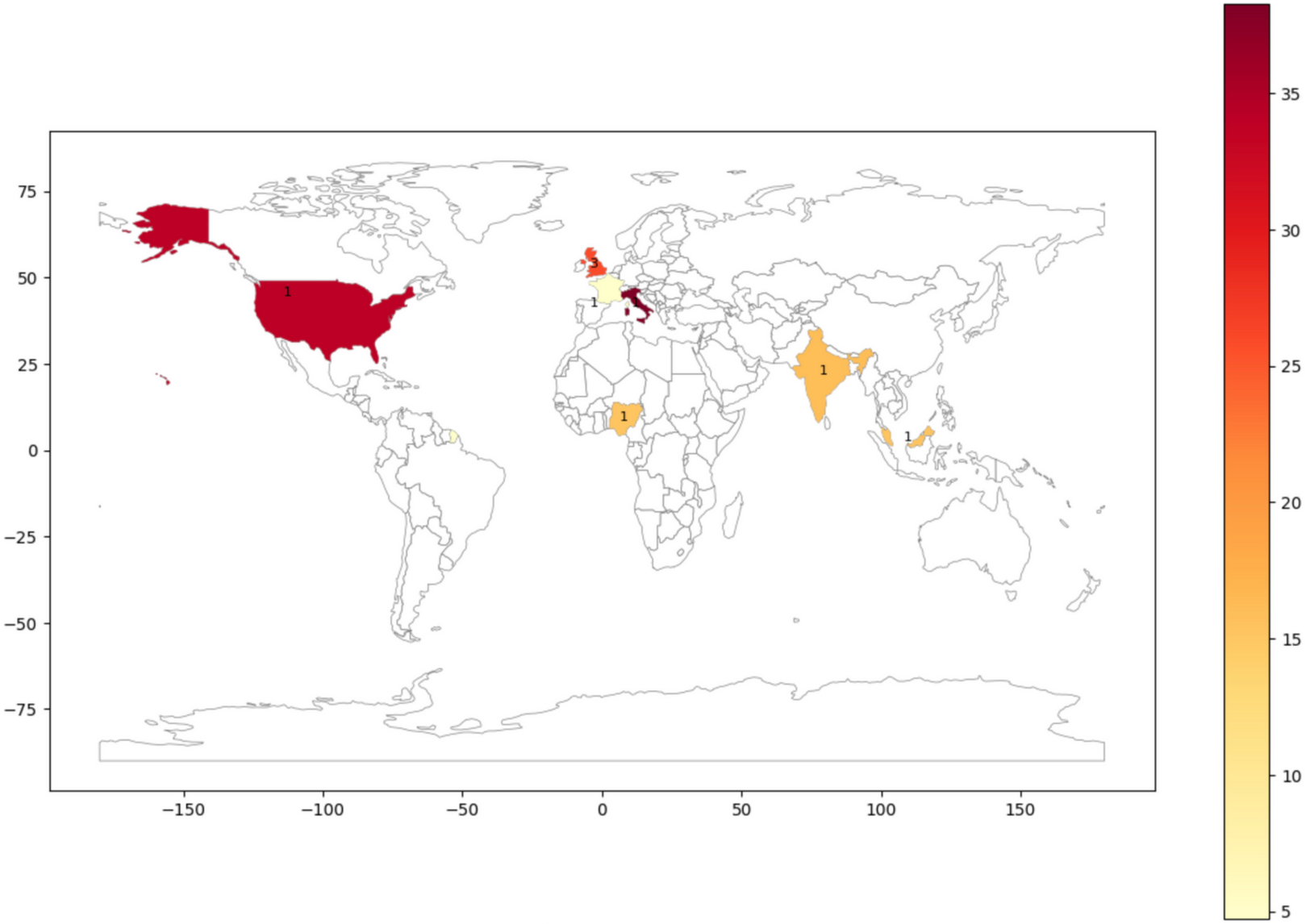
PRITAM



The above map marks the countries from which songs have been copied by RD Burman. A majority of songs have been plagiarized from **English** and **American music** with intense melodic similarities. Additionally music influence of **Middle Eastern** states like **Iran** and **Egypt** and **Eastern Europe** has been observed

SISTAHA

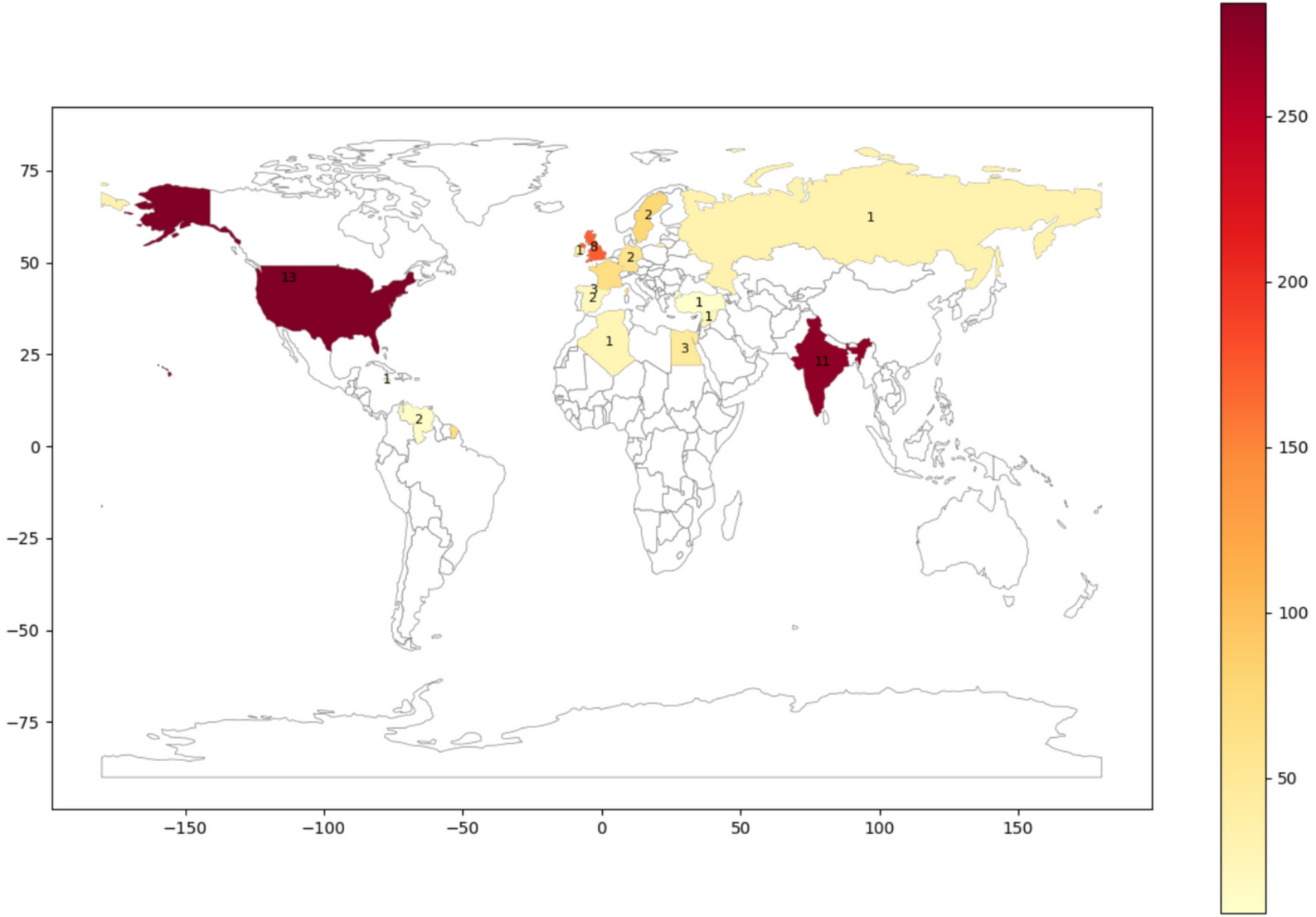
R.D. BURMAN



The above map marks the countries from which songs have been copied by AR Rahman. A majority of songs have been plagiarized from **English** and **American music** with intense melodic similarities.

SISTAHA

A.R. RAHMAN



The above map represents the influence of various countries on Tamil music. A majority of songs plagiarized from foreign music copy **English** language songs, predominantly from the UK and US. Additionally a high percentage of Tamil music has in fact been copied from **ghazals**, Indian music originally composed for **Hindi films**, or **folk/classical** music from other languages.

SISALAYA

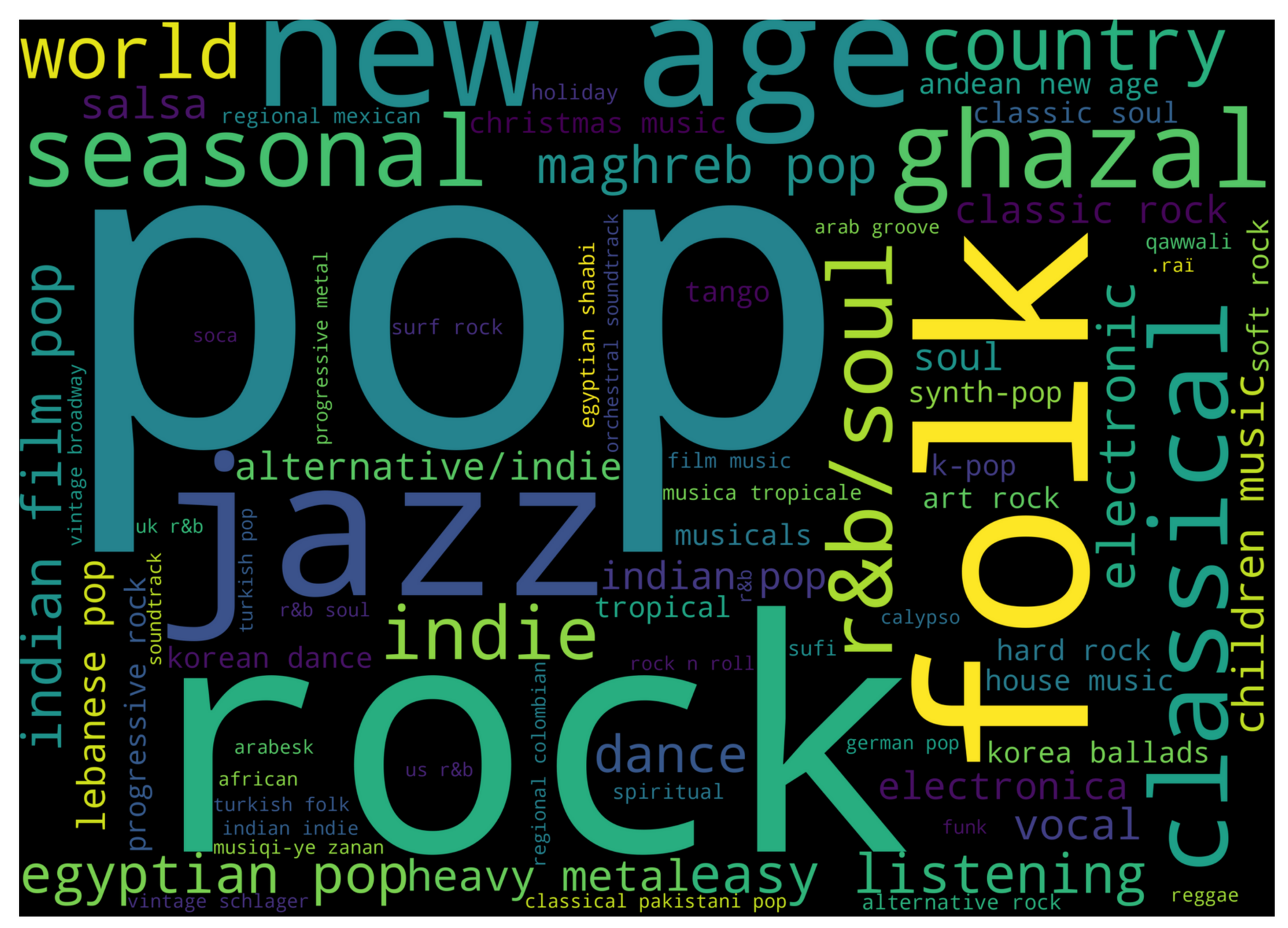
TAMIL

	Pop	Electronic	Soundtrack	Classic	Country	Folk	R&B	Rock	New Age	Jazz	Ghazal	Indie	Heavy Metal
Anu Malik	35.70	6.81	0.00	3.00	2.06	4.31	7.73	21.88	2.10	3.43	2.24	4.56	6.17
Pritam	38.33	0.00	0.00	12.43	2.00	9.99	0.00	15.67	9.40	0.00	3.03	6.87	2.28
RD Burman	35.30	2.44	0.71	9.58	5.38	6.71	4.28	19.89	2.53	7.49	4.04	1.67	0.00
AR Rehman	30.48	34.68	8.24	0.00	0.00	0.00	5.45	0.00	0.00	18.37	0.00	2.78	0.00
Tamil	48.50	8.23	8.67	3.87	0.00	14.10	1.08	11.57	0.00	0.00	3.98	0.00	0.00

1. Pop and Electronic are the most common genres for plagiarism across all artists, except for AR Rahman who seems to borrow more from Jazz.
2. Pritam has a higher percentage of plagiarized songs in Classic, Country, Folk, and New Age genres compared to other artists.
3. Anu Malik has a high percentage of plagiarized songs in the Rock, R&B, and Indie genres.
4. RD Burman seems to borrow more from Classic, Country, Jazz, and Ghazal genres compared to other artists.
5. Tamil seems to borrow more from Pop, Electronic, and Folk genres.

SISTAWA

GENRES



SISALTAH

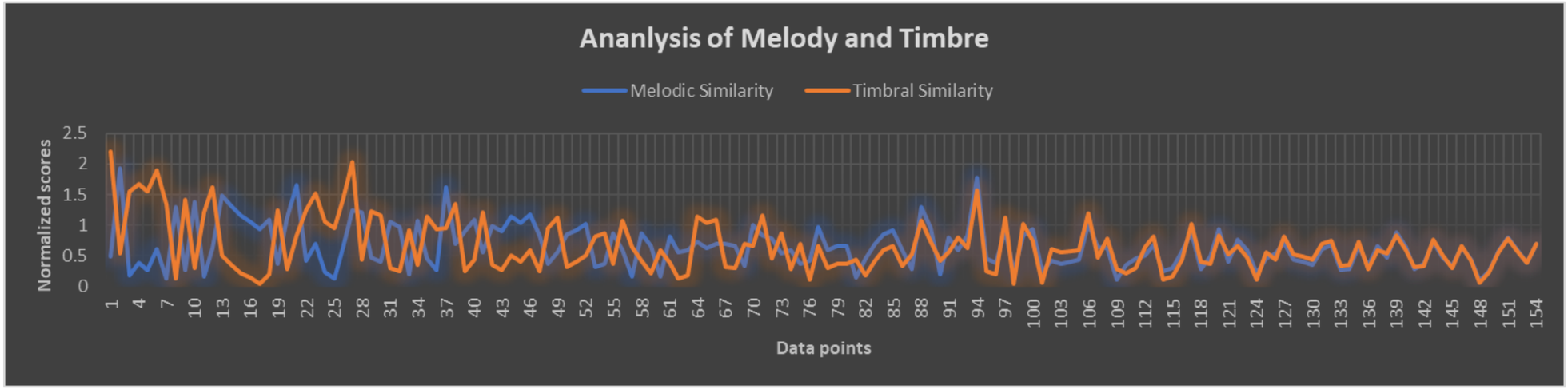
GENRES

SISATAHA

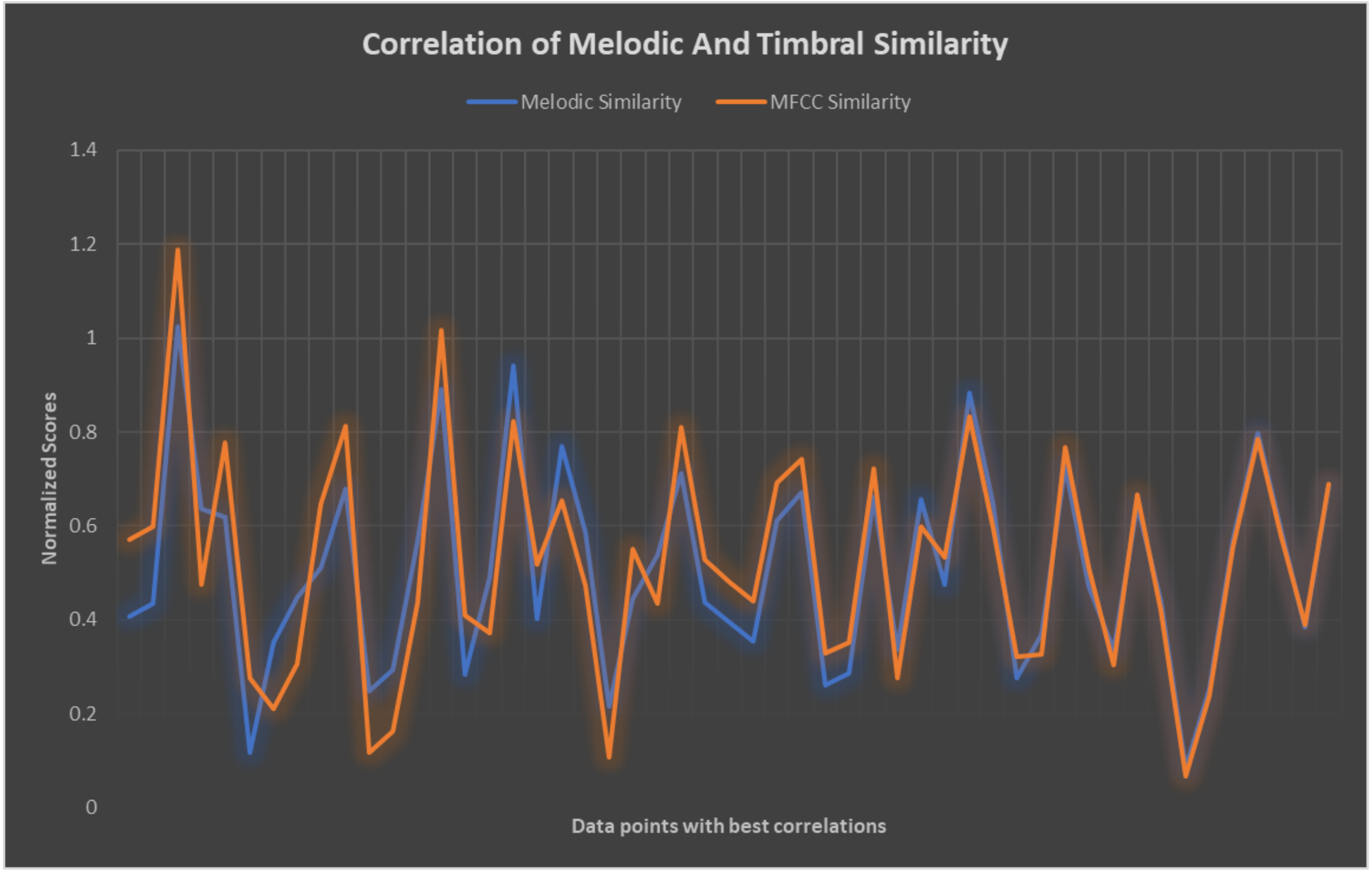
GENRES

INFERENCES

1. Timbral similarity refers to the degree of similarity or resemblance between the sound qualities (characteristic tone colour or texture) of two audio signals.
2. The timbral similarity is assessed by analyzing the spectral content, temporal characteristics, and perceptual features of the audio signals.
3. We observe a positive correlation between the melodic similarity scores derived from the bipartite graph matching of midi files (acoustic features) and the timbral similarity extracted from the .rm files (perceptual features).
4. From this, we can reason that there is no significant loss of information when we convert rm to midi files and the Bipartite Graph Matching algorithm to detect music plagiarism is in line with how people perceive audio similarity.

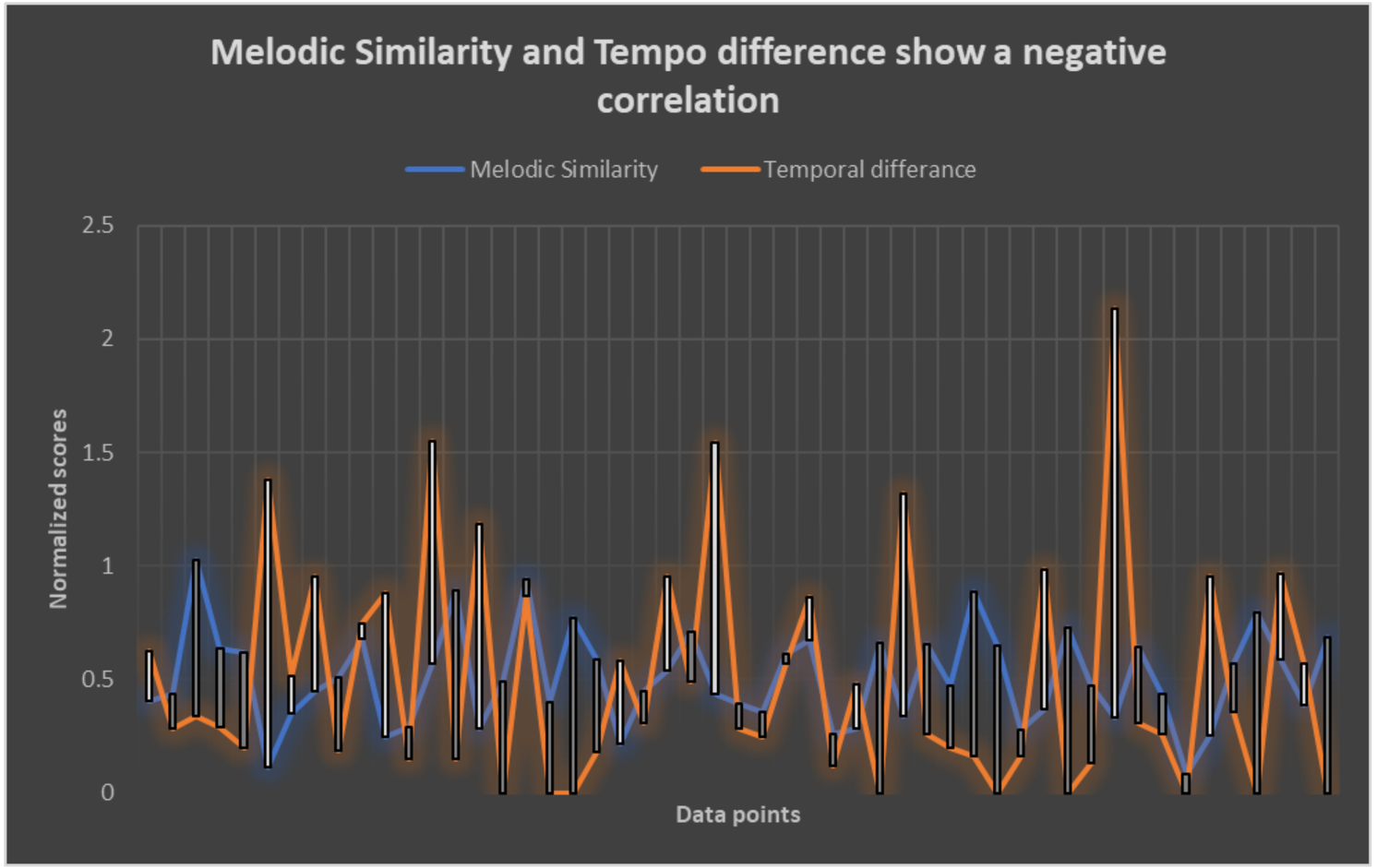


REFERENCES


Correlation of Melodic And Timbral Similarity
Melodic Similarity
MFCC Similarity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Normalized Scores
Data points with best correlations

INFERENCES

1. Tempo is the speed or pace of a musical composition.
2. We observe a negative correlation between the melodic similarity scores and tempo difference between the two audio files.
3. From this, we can infer that the way in which we perceive and process tempo similarity is an important aspect of how we relate to music.



CONCLUSION

1. Prevalence: our analysis revealed instances of plagiarism where the songs share significant similarities in pitch, tempo and timbre.
2. Pop music is a genre that is known for its catchy melodies and memorable hooks, and it is often targeted by artists looking to replicate its success. Rock, electronic, and jazz are other common genres for plagiarism.
3. Plagiarism can occur across different genres. Our analysis has revealed instances of plagiarism occurring across different genres, such as Bollywood songs copying from Western pop or Indian classical music.
4. Finally, the Smith Waterman algorithm using tempo, pitch and downbeat for calculation of melodic similarity is an efficient tool not only for the purpose of plagiarism inspection but also mapping timbral similarity.

FUTURE WORK

1. Expand the dataset to add more artists and languages. This could potentially reveal more instances of plagiarism and provide a more comprehensive picture of the prevalence of plagiarism in the music industry.
2. Explore and evaluate other automated ways of music plagiarism detection.
3. Explore other audio features such as rhythm and harmony.
4. Explore the legal implications of plagiarism in different jurisdictions and investigate how plagiarism cases have been handled in court.
5. Explore how plagiarism is perceived and dealt with in different musical traditions and cultures around the world.

THANKS

Pratham Gupta (2020101080)
Harshita Gupta (2020101078)
Nukit Tailor (2020114012)