University of Padova

Department of Information Engineering

Master Thesis in ICT For Internet and Multimedia

The Comparison of the Most Listened-to songs of 2000-2020 from within Network Science Perspective and the investigation and explanation of their effects on Human Psychology

Supervisor
Prof. Tomaso Erseghe
University of Padova

Master Candidate
Durmus Tulan

Academic Year

2023-2024

11.01.2024

Antalya-Turkey

Tulan: What is something you never want me to forget, Mom?
Miymet: My support and the sacrifices I make for you are unconditional.
You will always get what you want and I will always support you.

This success was possible with your support and sacrifices.
I'm glad you're here. This thesis is dedicated to you.
Thank you, mom.

Abstract

This thesis examines music from an engineering perspective, specifically through the analyzability of lyrics. It explores how machine learning and natural language processing techniques can be integrated alongside traditional music analysis methods to understand the evolution and emotional expression of music. The thesis aims to understand music from both an artistic and technical perspective by examining the history of music analysis, how it has changed with advancing technology, and what methods are used today. In particular, it examines the role of lyrics in musical structure and how these lyrics can be analyzed by forming a network. This network analysis allows us to gain a deeper understanding of the emotional expression and intrinsic connections of music. By evaluating the impact of machine learning and natural language processing techniques on music analysis, it explores the role of these technologies in enriching the meaning of music. Finally, this thesis examines the advantages and limitations of music analysis software. It sheds light on the future development of music analysis software, focusing on programs that are better able to analyze certain genres of music, as well as the inability to fully understand emotional expressions. It also offers a perspective on understanding the evolution of music and changes in listener preferences, using machine learning and natural language processing techniques to assess trends in the music world.

Contents

Listing of Figures

Listings

Listing of acronyms

SA . . . . Sentiment Analysis

NLP . . . . Natural Language Processing

US . . . . United States

API . . . . Application Programming Interface

NLTK . . . . Natural Language Toolkit

ASCII . . . . American Standard Code for Information Interchange

CSV . . . . Comma-Separated Values

UTF . . . . Unicode Transformation Format

ID . . . . Identification

POS . . . . Part of Speech

OS . . . . Operating System

CPM . . . . Component Process Model

t-SNE . . . . t-Distributed Stochastic Neighbor Embedding

1 Introduction

Music is marvelous. Not only because it is the common language for emotions and cultural incidents, but also due to its complex nature and ability to express its creators' inner world. Music has the power to overcome the restrictions of place and time, serving as an instrument that reaches people on an emotional level. Melody, rhythm, and lyrics work together to create a complex web. This web contains experiences, thoughts, and pieces of art of the people who created it. This web of music can carry a wide range of emotions, from anger to content, glad to bored [1]. Not only an aural experience, music captures the essence of many cultures. It carries the common emotions and experiences of people from all over the world since our culture is inherited by our ancestors, and their emotions and experiences are part of our culture.

Within this complex web, an interpretation for understanding the patterns we follow, and preferences we make is required. This is why, especially in the last 30 years, the concept of Sentiment Analysis was born. Sentiment analysis is used to analyze the sentiments expressed in written form data. It utilizes various techniques of natural language processing and machine learning algorithms. The expression 'newly' developed is used since it is fairly new; the term started to be used in the late '90s [2]. Sentiment analysis is based on categorizing the text's attitude as good, negative, or neutral [3]. This process involves analyzing written content to determine the sentiment expressed within it. In the field of music SA can be used to extract important information regarding the audience's emotional response to it is possible to use it for several purposes. With the application of sentiment analysis, the main aim is to acquire the dominant emotions and opinions linked to particular musical genres, performers, or separate songs.

Sentiment analysis is valuable in the music industry because it can identify and measure the feelings expressed in song lyrics, music reviews, and comments. The reason for that is it unveils a complex emotional spectrum that is merged into elements of music. The sentiment analysis is utilized by various people in the music industry nowadays. The most remarkable ones are producers, marketing teams, analysts, and artists. The composers use this spectrum to create their works, based on which emotions they want to include in their pieces of art. Marketing people are using these analyses to develop better marketing strategies [4]. Based on sentiment analyses it is possible to dive into deeper investigations based on the emotional influence of music with the scope of both individuals and communities. It is worth to be noted that, sentiment analysis in music has a wide range of applications. Psychology can use it to better understand the feelings that certain song genres or musical compositions arouse. From a marketing standpoint, observing consumer opinions and using them to develop more promoting music is crucial. Music historians can use it to observe the change in music over time since music is affected by culture pretty heavily as aforementioned. With community analysis, sociologists can analyze the patterns of music preferences and develop theories related to human interactions and behaviors.

The fact, that music has influence in psychology area, a new application and research areas developed for sentiment analysis. Psychologists use the assistance of sentiment analysis of the music to get a better insight into the emotional reactions and psychological effects triggered by musical content. Psychologists can systematically investigate and catch the new perspectives stated in music lyrics, reviews, or debates about particular songs or genres. They can identify patterns and trends between society and emotional reactions to music genres thanks to this method. Examples can be given, such as which genre leads to emotions like violation, aggressiveness, depression, and so on [5]. This correlation leads to the areas of research such as how music affects mood, thought processes, and general well-being. For example, determining the positive attitudes that are often associated with a certain genre of music may help guide therapeutic therapies aimed at reducing stress or improving mood [6]. On the other hand, identifying negative emotions associated with particular songs might help psychologists personalize therapies for those who are in emotional distress. Psychologists may use sentiment analysis as a useful tool to interpret their patients' negative emotions via not only music but also the words that their patients use. In this way, they can clarify the complex relationship between words and emotions by using the quantitative perspective of sentiment analysis. Via sentimental analysis, our knowledge of the psychological effects that listening to music extended. With this knowledge, better approaches to designing the cure for not mentally well people are achievable.

In order to grasp the patterns of lyrical elements in the songs, text-based sentiment analysis is used. The analysis material is selected as the lyrics of the songs that are on the charts of Hot 100 Billboard songs from 2000 to 2020 each year. The key objective is to use sentiment analysis to identify and measure the emotional density in the lyrics of the song collection [7]. This research is aimed to determine fundamental emotional patterns and changes by examining popular songs from this two-decade scope. Moreover, these analyzed patterns be will linked to the study of music with psychological discoveries, since it is assumed that selected lyrics are not random, but the density follows an order. The study objectives can be listed as to develop an accurate analysis and assess the accuracy as well as reliability of sentiment analysis instruments in interpreting the complex emotional characteristics of this dataset. By this analysis, the research aims to illustrate the percentage of popular emotions in most listened-to music. Moreover, it is also taken into account to establish a link between these findings and psychological facts by referring to psychological background, so offering a more profound comprehension of the emotional influence of music.

Three primary goals in the research, which seek to be achieved are listed as: First and foremost, to construct a strong sentiment analysis approach that is adapted to the selected dataset. It is important to adapt to the nuances of the changes in music over these two decades. The condition for the first goal is to adapt this dataset and methodology to have more accurate emotion detection and definition. The second objective is to examine and determine the emotional content of songs over the assigned period to spot followed patterns and yearly variances. Thirdly, to create links between psychological theories and the observed emotional patterns. By creating these links, the approach to results can be more real-life-oriented. It is crucial to create these links to demonstrate how music, more specifically lyrics affects emotions and overall well-being. Together, accomplishing these goals is the main objective of this analysis of the emotional background of popular music throughout the past two decades.

For research, defining steps are playing a vital role. It is crucial to achieve correct and wanted results. To achieve that, steps must be defined rigorously. The structure of the study and the steps taken were determined as follows: Chapter 2 goes background, mentioning the previously done research, and historical anecdotes. After that, in Chapter 3, the tools and methods that are used in this project are described. Also, challenges and data selection criteria are mentioned here. Chapter 4 is the application stage, some of the topics are: gathering data from sources, processing data, and cleaning. Next, Network Integrated Data Collection is presented. It is an intersection, a collaboration between Music Analysis and Network Science. This collaboration makes it easier to understand and handle data gathered from related sources. Later on, it dives into the use of several sentiment analysis techniques, compression between some of them, and the application of natural language processes to provide the annual sentiment analysis and data visualization. In Chapter 5, the focus is shifted to previously done psychological and sociological studies. This chapter is the link between the obtained results of music and emotional states and the associated conclusions from previous research. Chapter 6 examines patterns over 20 years, closely examining shifts in ongoing preferences and making social conclusions. The obtained results and comments are summarized in the following Chapter 7, which also highlights the thesis's limits and contributions.

2 Background

With the start of the new millennium, an important era in the history of the music industry has begun [8]. A rapid change in the music domain has started. The speed of change was dizzying and that changed how music is created, shared, and consumed. This transformation was mostly brought via advancements in technology. The adoption of digitization was a sweeping change in the music industry and replaced traditional methods of music creation and delivery techniques [9]. As digital forms progressively replaced physical ones, the industry's environment grew more and more integrated with the virtual world. This era's most notable feature was the exponential growth of streaming services. These services entirely altered how people could listen to and enjoy music. Spotify, Apple Music, and other platforms are used by millions of people nowadays, within just a fingertips, it is possible to access almost any song. People tend to use them for a couple of reasons like, they offer an extensive and user-friendly music library, ease of use, low prices compared to traditional methods, portable, etc. With the developments in technology, the music platforms developed in a way that can provide millions of songs. Social media also have given the chance to artists to interact with their listeners. Musicians could now reach a wider audience, while audiences had access to a greater variety [10]. As a result of this, the relationship between the artist and the audience has changed completely, in a more interactive way, the artist can take criticism from listeners and adapt it to be more popular, and successful. With this new bidirectional interaction, a blink from radios to the age of the Internet, the opinions of the audience gained importance. The audience's views, which were suddenly taken more seriously and accessible by communities, took on a significance never before seen. Thus, producers have to regulate their financial strategies based on the opinion of the crowd.

Other than serving as an envoy to carry emotions, thoughts, and artistic concerns, music is a kind of art and a social phenomenon. Music acts as a mirror reflecting the emotions, values, and tendencies of society [11]. The songs that were most popular during a period carry the stain of the cultural traditions, events, and emotions of the time of the production. During the digital revolution period, the music industry has also evolved, as artists, producers, and even us as individuals, to catch the pulse of society, attempting to navigate the complexities of a quickly changing digital world. Musical preferences are not just coincidences but also a complex result of emotions and cultural expressions. [12]. The tendency to prefer a genre of songs is knitted to the changing interaction between music and society.

Sentimental analysis is an awesome tool to understand what's underneath the lyrics but during the analysis, a classification system must be used. Since some words ignite positive emotions, some ignite negative ones. There are 3 main approaches to perform a classification. These are, machine-learning, lexicon-based, and hybrid [13]. As it is titled, hybrid is a mixture of both approaches. Both lexicon-based and machine-learning approaches for sentiment categorization have benefits and drawbacks[14]. The main advantage of machine learning approaches is their flexibility. It can optimized to be trained for specific datasets and parameters. Some of the most used algorithms are neural networks and support vector machines [15]. After being trained on large datasets, they are highly accurate in recognizing complex patterns. In contrast, its drawback is the possibility of overfitting and the necessity of labeled data. This occurs especially when confronted with noisy or distorted data, and the requirement for massive labeled datasets for efficient training. On the other hand, lexicon-based techniques offer speed and transparency compared to machine learning-based techniques. However, they are limited in handling contextual ambiguity. This is why, while using lexicon-applications, users of these techniques, must be attented. [16]. Beyond this, a hybrid strategy that combines the preset sentiment knowledge of lexicons with the flexibility of machine learning.

Sentiment analysis's introduction into the field of music has started a new point of view. It led to opportunities for recognizing the emotional agents of musical works. The researchers have developed new approaches to determine the emotional agents found in musical elements. For example, advanced natural language processing methods have begun to be used to analyze a wide range of emotions, and also the combination of emotions, as it is indicated in Ressel's emotion chart.[17]. With more analysis done, the success rate of analyses has increased, since there are more and more labeled data each day. The increment of ready-to-use data, in other words, 'the era of machines' led the newly done analyses to find themes and patterns that trigger particular feelings with a higher success rate. It is a feature of information, it is cumulative. It is possible to separate the elements and analyze them one by one. The separation of this made it possible to do the analysis of musical elements like speed, lyrics, melody, and harmony separately. Identifying the emotional tone in songs is nowadays possible with more accurate results, thanks to the advances in the machine learning field.

This study's main pursuit is to evaluate and discuss the emotions that are contained in the lyrics of the most popular songs from 2000 to 2020. The investigation of emotional language in the musical material of this particular era is guided by a set of clearly defined research questions and objectives that form the basis of the study.

The study's research questions aim to discover the emotional variations that popular song lyrics capture across the selected period. The study attempts to identify patterns, trends, and changes in emotional themes across time by examining the subtleties of the feelings expressed in these songs. It's crucial to examine the emotions expressed in the top songs from 2000 to 2020 for a few reasons. First of all, music is a servant of cultural symbols that reflects societal sentiments. Thus, we may learn about the feelings of people at this period by decoding the emotions expressed in the lyrics of well-known songs. Additionally, since digital platforms have made music widely available to a worldwide audience, we want to investigate if some feelings are culturally specific or universal. It is also worth mentioning that emotions are also carried, broadcasted, and spread out via music globally.

In summary, the aim is to identify the feelings conveyed in song lyrics written between 2000 and 2020. It is supposed to determine emotional tendencies over the last 20 years. Tendencies can lead to interesting findings, and the result of this study can be used for other studies later on, with various application areas.

3

Methods and Tools

3.1 Introduction of the tools

In the data analysis field, especially analyses containing psychological elements, constructing the appropriate dataset and selecting the right sources is crucial [18]. Selecting the right sources guarantees the accuracy, credibility, and applicability of any study project. Also, carefully examining the dataset throughout the selection process plays a vital role since the quality of the data directly affects the reliability and correctness of the conclusions [19]. The selection of songs from the years 2000 to 2020 is made by a few factors. First, in the context of this study, it can demonstrate the influences of how musical trends and emotional emotions are portrayed over a long period. Secondly, as aforementioned, the year 2000 can be entitled as a new area in the field of technology for various reasons. Within such a high velocity of change that we live through, 20 years is an appropriate scope of research. Furthermore, the study's comprehensiveness is greatly affected by the sources. The archives are well constructed after 2000 compared to before 2000[20]. The research guarantees a representative diverse sample that covers the wide range of popular music released during the specified period by precisely selecting data from reliable databases and music charts. The study's legitimacy is strengthened by the rigorous selection procedure, which also provides valuable insights into the emotional landscapes represented in the songs that have been played the most throughout the previous 20 years.

3.2 Target mass selection

For this research, the music preferences of the people from the United States are selected. Selection is done for several reasons. First off, the American population offers an adequate sample size for a thorough examination. The US is the most populous native English-speaking nation. [21]. Because English is so widely used as a first language, sentiment analysis tools—which are frequently tailored for the English language—can analyze and understand song lyrics with ease. [22]. In addition, the United States has a rich ethnic diversity. Percentages for the US as of 2022 are:[23].

- White: 60.1% (Non-Hispanic)
- Hispanic: 18.5%
- Black: 12.2%
- Asian: 5.6%
- Multiple Races: 2.8%
- American Indian/Alaska Native: 0.7%
- Native Hawaiian/Other Pacific Islander: 0.2%

Selecting a country with people from diverse ethnicities is also beneficial since music preferences are influenced by cultural factors, and to minimize this effect and for entitling the research globally, races have to be put into account equally, or as equally as possible [24]. The US is the best option in this case among native English-speaking countries, compared to the UK, Australia, New Zealand etc. The decision to focus on Americans also collides with the popularity of American music across the world, which makes it a useful case study for comprehending larger trends. The United States is a hub for monitoring and characterizing musical trends, having given rise to the well-known music chart app Billboard. [25] This makes the country a solid starting point for examining the emotional terrain contained in the songs that people will listen to the most between the years 2000 and 2020.

3.3 Constructing Dataset

Creating a dataset for sentiment analysis requires careful consideration. This consideration leads to how reliable the following analysis will be. The dataset's ability to be represented is the characteristic most attention-paid since it includes a wide variety of songs that reflect the wider musical preferences over the chosen years. Building a relevant dataset is important because it may provide information on the prevalent emotional themes in popular music, which can help us comprehend cultural and social attitudes more deeply. In the pursuit of balance and fairness, the selection of 100 songs for each year is motivated by the intention to provide equal opportunities for songs released throughout the year to be represented. This approach minimizes potential biases introduced by the popularity of songs released during specific seasons or events. By equally selecting songs across the year, the dataset aims to observe the trends for preferences over the whole year. ensuring This methodological choice aims to more accurate representation of the uncovered diverse emotions embedded in popular music.

3.4 Used platforms for data scarping

3.4.1 Python

Python is one of the most popular and adaptable programming languages used by developers from different fields nowadays [26]. The main characteristic that describes Python is its large library, user-friendly interface, and effective features. Python is selected as the main programming language in the processing, analysis, and extraction of significant insights from song lyrics in the context of this sentiment research project. Python is widely known for its ability to handle text data, adapted by developers for this purpose. These developments are making it a perfect tool for overcoming the complexities of sentiment analysis. For implementing Python's functionality, the Billboard and Genius Python packages are integrated [27],[28]. It can be said that providing quick access to Billboard's Hot 100 charts and making it easier to extract lyrics using the Genius API are the reasons why Python is tailored for this project.

In the foundation for the data collection in this study, Billboard and Genius are selected as essential sources, each providing distinct aspects for the examination. Billboard is an icon in the music industry with its extended history back to the 1900s[29]. Billboard picks the songs for the charts that represent the pulse of popular music. Quantitative statistics, such as song rankings, chart length, and audience reach, are provided by this reliable source. On the other hand, the dataset is qualitatively merged with the data provided by Genius, a collaborative platform for music lyrics and comments. It is entitled as a lyrical content archive, providing a trustworthy archive of song lyrics[30]. With the usage of these sources, a comprehensive analysis of music appeal and sentiment is aimed. Using Billboard's history charts, the patterns in song popularity throughout the designated years by using them as a temporal reference can be traced. Genius, with its user-generated annotations, uncovers the deep layers of emotions communicated via songs. By selecting and merging the right data sources, the study goals have been achieved successfully. The approaches used for data extraction will be covered in detail in the parts that follow, guaranteeing a careful approach to capturing the nuances of emotion in the most listened-to songs between 2000 and 2020.

3.4.2 Billboard

Billboard, a historical leader for criticism and ranking in the music industry, was founded in 1894. Back in the day, Billboard was publishing magazines of the music industry before developing into an entire resource for music lovers, artists, and industry professionals.[Billboard Wikipedia] The Hot 100 chart was first published by Billboard in 1958. It is a ranking system that ranks the top 100 songs in the US weekly. There are various categories of top 100 songs, such as all-time, yearly, weekly, and genre-based. Over time, this charting system has gained some important aspects. The characteristic features of Billboard are representing the success of music, influencing decisions made by the business, and illustrating the cultural influence of a song [31]. With the popularity and success of Billboard, taking a seat on the Hot 100 charts became seen as a huge achievement by song producers. Because Billboard's Hot 100 has well-developed parameters for the selection of song popularity and has a historical background, it is widely known and used all around the world to investigate the success of the song. This is why, the selection of the most popular 100 songs is made on these charts. A complete representation of music trends is ensured by the Hot 100, a dynamic reflection of public taste that gathers the most popular songs across genres [32]. Because of the chart's extensive history, a comprehensive dataset covering the two-decade period under investigation is available. The history of charts provides enough timeline to conduct a thorough analysis of sentiment patterns' dedicated scope. Furthermore, the study's goal is to detect the differences in song preferences over the years. This goal is perfectly aligned with the Hot 100's rigorous and consistent methodology since the appeal of the songs in these charts is based on respectable achievements.

3.4.3 Genius

Established in 2009, Genius is the leading site on the internet for group lyric annotation and analysis. Genius is founded on the concept that music is a global language and bonds every individual. The platform links everyone who produces and consumes music. The main purpose of creating such an environment is to analyze, debate, and enjoy the subtleties of lyrical content.

Genius provides a large repository of song lyrics tied with insightful comments. The platform works similarly to the system of popular social media nowadays, lyrics can be published and commented on. These comments let users see and comprehend different points of view in song lyrics with cultural, emotional, and historical information [33]. By these comments, producers can claim feedback for the producers for their future work. Also, be added by the producer with a verification system. These comments can cover the historical backgrounds, personal thoughts on the meaning of lyrics both by producers and listeners, and even disses. Genius was chosen as the main source to create the lyrics dataset for songs gathered from Billboard's Hot 100 charts. The main motivation for this selection is, that Genius provides the most accurate and correct lyrics among other options with the option of using the comments. The study's success in determining the underlying emotional content is improved by including the effect of these insightful comments. Genius is unique for this type of application. These two platforms are selected as base sources for constructing the dataset.

3.5 Dataset Cleaning

Data cleaning, often referred to as data cleansing or data scrubbing, is locating and fixing flaws or abnormalities in a dataset. Since the integrity of the data directly affects the quality of the study, the importance of data cleaning cannot be emphasized [34]. Abnormalities in data can occur for several reasons, such as entry problems, technical malfunctions, formatting discrepancies, and human errors [35]. These reasons emphasize why data cleansing is required. If ignored, these mistakes might produce biased or erroneous results. For the sake of the validity of the study's conclusions, the cleaning step is crucial.

Data cleansing is essential to ensuring that the emotional content in lyrics is correctly represented. Inaccuracies in the dataset, such as misspelled words, meaningless special characters, abbreviations, suffixes, inconsistent formatting, or incorrectly understood attitudes, might influence the analysis's conclusions if they are not properly cleaned up. One might draw false inferences about the dominant feelings within the designated time frame, for example, if a mislabeled sentiment distorts the overall emotional landscape shown in the songs.

NLTK, or the Natural Language Toolkit, serves as the initial step in the data processing pipeline. Leveraging NLTK, the lyrics undergo a comprehensive cleaning process, including lowercasing, removal of numerical characters tokenization, lemmetazing, and elimination of stopwords and punctuation. This step ensures laying the basis for accurate sentiment and emotion analysis. Implementation of data cleaning is designed step by step due to the special requirements of this project. Firstly, textual data from song lyrics needs to be thoroughly reviewed and standardized. This includes addressing abnormalities, such as misspellings, non-standard abbreviations, or slang words which are words used with unusual meanings [36]. Additionally, removing irrelevant information, such as metadata or special characters is vital for the dataset. To enhance accuracy, sentiment labels associated with each song should be carefully verified and standardized, ensuring a consistent dataset. This research attempts to improve the sentiment analysis findings' dependability by carefully using data-cleaning procedures. The well-done implementation of data cleaning strengthens the study's overall robustness and breeds a solid basis for further analysis and findings.

3.6 Sentiment Analysis

Sentiment analysis, or in other words, emotional assessment is the determination of the emotional tone and polarity of song lyrics. In the implemented methodology, two key analysis tools play pivotal roles; TextBlob for sentiment analysis, and NRCLex for emotion analysis.

TextBlob, a versatile Python library, is chosen for sentiment analysis due to its simplicity and effectiveness. Utilizing machine learning algorithms, TextBlob assigns a polarity score to each lyric, representing the degree of positive or negative sentiment. The decision to employ TextBlob is made by the fact that TextBlob is easy to implement, making it suitable for capturing overall sentiment trends within the song lyrics. [37] For a nuanced exploration of emotions, NRCLex is employed. NRCLex specializes in emotion analysis, providing detailed insights into various emotions associated with both individual words and chunks of words. This library relies on the NRC Emotion Lexicon to quantify fundamental emotions like fear, anger, trust, and joy. NRCLex is selected for its ability to offer a granular understanding of emotional nuances. Each tool brings various benefits to the methodology. Therefore, a blend of them is employed. NLTK ensures robust data cleaning, TextBlob facilitates straightforward sentiment polarity analysis, and NRCLex adds a layer of sophistication with detailed emotion analysis. The collective use of these tools aligns with the study's goals.

3.7 Data visualization Techniques

3.7.1 Matplotlib

Because of Matplotlib's adaptability and simplicity of integration with Python—the language of choice for data analysis—it was selected as the main visualization tool. Numerous chart kinds are available in Matplotlib, such as line, bar, scatter, and more. Because of its connection with Pandas, charting from DataFrame structures is made simple, which makes it appropriate for showcasing linkages, trends, and patterns within the collection. The ease of use of Matplotlib makes it possible to create static visuals. By using static visuals, the aim is to demonstrate the results side by side to make compression between each year and also decade.

3.7.2 Seaborn

Developed on top of Matplotlib, Seaborn is a robust data visualization library that improves both the ease of use and the visual appeal of producing statistic images. In this study bar charts, pie charts, and heatmaps are constructed by Seaborn. Compared to Matplotlib, Seaborn often requires fewer lines of code to create complex visuals. Its smooth integration with Pandas DataFrames makes it possible to visualize the data quickly. Color palettes and visibility options are among Seaborn's helped to increase readability. The library is a useful addition to Matplotlib's data visualization toolkit. This tool is especially tailored for applications involving statistical analysis and pattern discovery in data, which is the main aim of this study.

3.8 Scientific Computing and Machine Learning

3.8.1 Scipy

Scipy, a scientific computing library in Python, complements the data analysis workflow by providing a wide array of tools for mathematical calculations. Leveraging Scipy's functionalities, this code snippet employs the 'KMeans' clustering algorithm from the 'sklearn.cluster' module for clustering in high-dimensional spaces. The 'TSNE' module from 'sklearn.manifold' is used for dimensionality reduction, visualizing the word embeddings in a two-dimensional space. The integration with Matplotlib facilitates the creation of visually insightful plots. By utilizing Scipy's 'distance' module, the code assesses the proximity of words within each cluster. Beyond clustering, Scipy's rich feature set includes optimization, integration, interpolation, and signal processing. Scarped data analysis relies heavily on Scipy's seamless interface with other scientific libraries, like as NumPy and Matplotlib.

3.8.2 Scikit-Learn

Scikit-Learn is a sophisticated machine-learning toolkit in Python that proves invaluable for various aspects of clustering and analysis. To break it down, it taps into the power of the 'KMeans' algorithm from the 'sklearn.cluster' module, strategically dividing word embeddings into distinct clusters, forming a foundation for exploratory analysis. Moreover, Scikit-Learn's 'silhouette analysis' steps in to assess the quality of these clusters, shedding light on how tightly knit or separated they are. The library's versatile tools for distance calculations, sourced from the 'sklearn.metrics.pairwise' module, play a crucial role in quantitatively evaluating the relationships among words within and between clusters.

3.9 Natural Language Processing and Word Embeddings by Gensim

Gensim, a powerful natural language processing library in Python, takes center stage in this code snippet, driving the word embedding and processing tasks. Leveraging Gensim's 'Word2Vec' implementation, the code transforms words into dense vectors, capturing semantic relationships and contextual information. This embedding technique facilitates the exploration of semantic similarities and differences between words in a high-dimensional space. Gensim's intuitive interface and efficient algorithms make it an ideal choice for generating word embeddings from large textual datasets. The '`key_to_index`' functionality is harnessed to retrieve word vectors efficiently, enhancing the code's computational efficiency. Gensim's versatility extends beyond Word2Vec, offering additional models and functionalities for text summarization, topic modeling, and document similarity analysis. With its ease of use, Gensim stands as a key tool in natural language processing workflows

3.10 Complex Network Analysis and Community Detection

3.10.1 Igraph

Igraph, a versatile Python library tailored for complex network analysis, takes center stage in this code snippet. Its seamless integration allows for the creation, manipulation, and exploration of intricate graphs. Leveraging the igraph module, the code constructs an undirected graph from the provided edge and node data, offering a comprehensive platform for network representation. Igraph's extensive set of algorithms proves invaluable in the application of the Leiden community detection algorithm, enabling the identification and delineation of distinct communities within the graph. This library, with its intuitive design and compatibility with Matplotlib for visualization, stands as a fundamental tool for network science in Python, facilitating a deeper understanding of community structures and relationships embedded within the data.

3.10.2 Leiden Algorithm via IGraph

The Leiden algorithm within the igraph library is employed to unravel community structures within the constructed network of word embeddings. Integrated seamlessly into igraph, the Leiden algorithm excels in identifying hierarchical community formations and efficiently manages large-scale networks. By applying the Leiden algorithm, the code adeptly identifies and partitions the network into distinct communities or groups of nodes, shedding light on their interconnectedness. This subsection underscores the Leiden algorithm's effectiveness in revealing meaningful clusters and relationships within the graph, contributing to a nuanced understanding of semantic associations and contextual similarities. Despite the nomenclature difference, the application of the Leiden algorithm via igraph aligns with the broader goal of enhancing community detection and analysis within complex networks.

3.11 Psychological Analysis and Conclusion:

Sentiment and emotional analysis of lyrics reveal a range of psychological and sociological effects present in the musical domain. With the aid of the studies done so far, the range of emotions could be identified and enlightened on how people feel the emotions, how people react, and what were the effects temporary and permanent. These results are consistent with psychology ideas on emotional expression and social connection since music frequently acts as a mirror to society's sentiments [38]. A common emotional language found in music that cuts over linguistic and geographic barriers. Psychologists claim that universal emotions are expressed and recognized in music [39].

Furthermore, the presence of specific emotions within lyrics aligns with social psychological theories, such as social identity theory. [40]. The diverse emotional patterns identified in the lyrics may include the emotional experiences of different societal groups. Additionally, the emotional content within lyrics may serve as a means of emotional regulation for listeners, aligning with the theories of Gross, who highlights the role of music in influencing emotional states and facilitating emotional well-being.

A crucial area to investigate to correlate the findings of the research is the complex interaction that exists between the recognized emotions and known psychological components. The emotional aspects will be interpreted through the prism of Plutchik's wheel of emotions, a comprehensive framework that has received widespread recognition in psychological research. By placing the recognized emotions on Plutchik's wheel, the psychological foundations of the patterns that have been noticed are aimed [?]. As demonstrated by Plutchik's groundbreaking work in emotion theory, this method offers a nuanced viewpoint that enables a more sophisticated understanding of how the emotional content of songs interacts with basic human experiences and perceptions.

In conclusion, the investigation into emotional and sentiment analysis in song lyrics offers valuable insights into the psychological and societal dimensions within the musical domain. The study's progression, outlined in Chapter 3, involves meticulous steps such as target mass selection, the introduction of influential sources like Billboard and Genius, and the implementation of powerful tools including NLTK, NRCLex, and TextBlob, and analysis tools like NetworkX. In Chapter 5, the mentioned theories will be evaluated in depth. Subsequently, the next Chapter 4, Network Integrated Data Collection will encompass data collection, thorough cleaning, and in-depth analysis. This structured approach promises a comprehensive understanding of emotional landscapes across popular songs, bridging theoretical frameworks with practical insights.

4

Network Integrated Data Collection

4.1 Network Science-Music Analysis Collabration

The integration of network science and music analysis presents a unique approach to comprehending the complex domain of music. Network science offers a solid basis for modeling and analyzing the complex interactions found in the musical world, with its foundations in complex systems and graph theory. Music is commonly characterized as knitted produced of rhythms, harmonies, and melodies. Thus, music is an ideal fit for network science techniques.

At the core of this collaboration is the conceptualization of lyrical elements as nodes within a network, and the relationships between these components as edges. Each element blends to form a web of interconnected nodes. In this web, each node is influencing some other and being influenced by others. Network science offers tools to quantify these relationships, such as Node Degree Distribution, Community Detection, Centrality Measures, and Network Evaluation Analysis. With these tools provided by the network science area, unveiling patterns and trends that might be impossible to detect by traditional analytical approaches is possible.

This joint investigation aims to use network science's analytical capabilities to figure out the complexity of music. By doing this, the main objective is to push further the comprehension of how distinct components weave together to form the emotional coloration. This study provides fresh perspectives on how music studies can go beyond traditional studies with newly developed applications, interdisciplinary.

4.2 Extraction of Billboard and Genius Data:

4.2.1 Billboard Data:

The primary step involved in constructing the dataset is the retrieval of song data from Billboard. Leveraging the billboard library in Python, a script was devised to fetch this information for the chosen time frame systematically.

```
# Looping through dedicated period 2000-2020
for year in range(2020, 2021):
    get_hot100(year)
```

Listing 4.1: Code for Looping Through the Period 2000-2020

The Python script utilizes a nested loop structure to iterate through the range that is selected for mentioned reasons above. In this study, years between 2000 and 2020 are iterated.

```
# Function for scraping the charts from Billboard
def get_hot100(year):
    try:
        chart = billboard.ChartData("hot-100", date=f"{year}-12-31")
        print(f"Got Hot 100 chart for the year {year}")

        # Create the directory if it doesn't exist
        \texttt{directory = "./charts2/"}
        \texttt{os.makedirs(directory, exist\_ok=True)}

        with open(f"{directory}hot100_{year}.txt", "w") as f:
            f.write("title;artist;peakpos;lastpos;weeks;rank;isnew\n")
            for song in chart:
                f.write(f"{song.title};{song.artist};{song.peakPos};
                {song.lastPos};{song.weeks};{song.rank};{song.isNew}\n")
```

Listing 4.2: Code for scraping the charts from Billboard

The script is forming a date parameter for the Hot 100 chart. For each date, the script queries the Billboard ChartData API to obtain the list of top songs. The acquired data includes essential details such as song title, artist, peak position, last position, weeks on the chart, rank, and a flag indicating if the song is new to the chart. To organize the retrieved data systematically, the script generates a text file for each date, following a predefined format. Easy parsing and subsequent integration into the larger analytic process are guaranteed by the file structure. Exception handling is implemented in the script, to ensure the data extraction procedure is done without any room for errors.

```
except ValueError as ve:
    print(f"Date skipped since not valid: {year}-12-31. Error: {ve}")
```

Listing 4.3: Exception handling for invalid date

To avoid dates that might not provide acceptable chart data, the script generates an error message that says the chart for the specified date is not available.

The final dataset becomes a dynamic representation of the always-changing popular music stage, with recorded features like peak locations, chart lengths, and the progression of individual songs over time. Insights into the dynamics of musical tastes and chart performance can be obtained by the constructed dataset. It can be also used as a basis for further analyses since it is also possible to do genre, instrument, and cultural analyses, concerning insightful extracted features of the most popular songs. Categorized and constructed dataset is now ready to lyrics scarping by Genius API.

4.2.2 Genius Data:

The next step taken for data collection is extracting the songs' lyrics, based on the lists obtained from Billboard Charts. Genius, which is known for having the largest collection of song lyrics, is selected as the main source of information for this lyrical investigation.

Genius is an excellent option for lyrics collection since it provides a varied range of song lyrics, especially popular songs' lyrics are added on the platform immediately. It can be used for extracting textual data about music due to its publicly reachable database and user-friendly interface. Genius's structure aligns majorly with the study's goals since it offers textual samples that capture the creative expressions and nuanced emotions seen in popular song lyrics. To operationalize the extraction of lyrical data from Genius, a Python script leveraging the Genius API was employed. This script systematically navigates the Genius platform, searching for lyrics corresponding to each song identified in the Billboard Hot 100 chart for the specified years. The Genius API facilitates this process by allowing programmatic access to its extensive collection of song lyrics. The Genius API (Application Programming Interface) is a set of rules and protocols that allows developers to interact with and access data from the Genius.com platform programmatically. The Genius API provides developers with a standardized way to retrieve information related to music, lyrics, and artist details from the Genius database. It enables the integration of Genius features and data into third-party applications, websites, or scripts. Client ID and Client Secret, are used in the OAuth 2.0 authentication process. The client ID and client secret are unique identifiers associated with the application that is making requests to the API. They are used to authenticate and authorize the application with the Genius API. Client Token is a token provided by Genius after successful authentication. It serves as a temporary access credential that the application includes in each API request. The token grants the application the necessary permissions to access specific resources. The Genius API provides a standardized way for developers to access and integrate music-related data from the Genius.com platform into their own applications. API credentials are used to authenticate and authorize applications, and the API structure defines the endpoints and actions available for interacting with the Genius database.

```
with open("c:/Users/User/Desktop/Thesis Code/thesis_code/genius_credentials.txt", "r") as f:
    CLIENT_ID, CLIENT_SECRET, CLIENT_TOKEN = map(str.strip, f.read().split(","))
```

Listing 4.4: Reading Genius API credentials from a file

```
special_check = ["\\", "?", ".", ":", ",", ";", "/", "*", "'", "`", "#", "*", "+", "~", '"', "<", ">", "|"]
feat_check = ["Feat", "feat", "Feat.", "feat.", "Featuring", "featuring", "With", "with", "And", "and", "&", ",", ";", "X", "x"]
```

Listing 4.5: Python code for special character and featuring check

The script begins with manually constructed special characters, ‘`special_check`‘, and feature check, ‘`feat_check`‘ matrices. The elements that these matrices contained are searched for each song, handling potential complications such as special characters in titles or artist names, variations in artist names due to featured artists, and the detection of non-English songs. With these checks, the script is developed to guarantee the collection of relevant data.

```
for YEAR in range(2000, 2021, 1):
    Year = YEAR
```

```
    print(f"Getting top songs of {YEAR}...")
```

Listing 4.6: Python code for iterating over years

Lyrics scarping starts with a for loop, for each year in the loop, every other step is repeated. This loop iterates over a range of years, starting from the dedicated beginning year, 2000, and ending in year 2020. 2021 is determined as the end year since loops are ended 1 variable before.

```
file_path = f"c:/Users/User/Desktop/Thesis
    Code/thesis_code/charts2/hot100_{Year}.txt"
df = pd.read_csv(file_path, sep=";", encoding="cp1252")
```

Listing 4.7: Reading CSV file into a DataFrame

The next step is, reading chart data for relevant years. These chart data were saved into CSV files scarped by billboard API. CSV files that contain song data for the current year, specified by the 'YEAR' variable are read by using pandas.

```
if not os.path.exists(f"./songs/songs{YEAR}"):
    os.makedirs(f"./songs/songs{YEAR}")
    print(df.head())
```

Listing 4.8: Creating directory and printing DataFrame's head

This line creates a directory to store song lyrics for the current year if it doesn't already exist, specified by name 'songsYEAR'.

```
genius = Genius(CLIENT_TOKEN, timeout=50)
```

Listing 4.9: Creating Genius object

Initializes the Genius API client with the provided client token.

```
for title, artist in zip(df["title"], df["artist"]):
    print(f'Searching for "{title}" by {artist}')
    title = unidecode(title)
    artist = unidecode(artist)
```

Listing 4.10: Looping through DataFrame columns

Beginning now on, the lyrics data by title and artist name is started to scraping. This loop iterates through each row of the DataFrame '(df["title"], df["artist"])'. For each iteration, it extracts the "title" and "artist" values from the corresponding columns. The 'unidecode' function is applied to both the 'title' and 'artist' strings. The purpose of 'unidecode' is to convert Unicode characters to their closest ASCII equivalents. This can be useful for standardizing text and handling non-ASCII characters.

```
for flag in feat_check:
    if flag in artist.split(" "):
        % print("    ARTIST NAME PROBABLY CONTAINING A FEATURE! TRYING SIMPLER NAME...")
        artist = artist.split(" ")[:artist.split(" ").index(flag)]
        artist = " ".join(artist)
```

Listing 4.11: Checking for featuring keywords in artist name

'`feat_check`' is, as mentioned above, a predefined list of words that might indicate a featured artist in the song. If any of the words in '`feat_check`' are found in the artist's name (`artist`), it assumes that the artist's name contains a feature. It then attempts to simplify the artist's name by taking the part of the name before the detected feature (splitting and joining the words).

```
for symbol in special_check:
    if symbol in title:
        % print("    TITLE PROBABLY WEIRD! REMOVING SPECIAL CHARACTERS...")
        title = title.replace(symbol, "")
    if symbol in artist:
        % print("    ARTIST PROBABLY WEIRD! REMOVING SPECIAL CHARACTERS...")
        artist = artist.replace(symbol, "")
```

Listing 4.12: Removing special characters from title and artist

Similarly, `special_check` appears to be a predefined list of special characters. This `for` loop checks the '`symbols`' if any of these special characters are present in the song '`title`' and the name of the '`artist`'. If found, it removes these special characters from both the title and the artist's name.

Moreover, even though almost, all of the most listened songs are in English in US, there might be exceptions. For handling this case, this snippet is implemented :

```
try:
    if langdetect.detect(song.lyrics) != "en":
        % print("    NON_ENGLISH SONG DETECTED! CHECKING FOR FALSE TRANSLATION...")
        songs = genius.search_artist(artist, max_songs=5)
        d = []
        for s in songs.songs:
            try:
                d.append(langdetect.detect(s.lyrics))
            except langdetect.lang_detect_exception.LangDetectException:
                continue
        if d != []:
            lang = max(set(d), key=d.count)
            if langdetect.detect(song.lyrics) != lang:  % # TRANSLATION DETECTED
                % print("        TRANSLATION DETECTED! TRYING TO FIND ORIGINAL VERSION...")
                song_list = genius.search_songs(f"{title} {artist}", per_page=5)
                id_list = [song_list["hits"][i]["result"]["id"] for i in range(len(song_list))]
                lyrics_list = []
                song = 0
                for id in id_list:
                    s = genius.search_song(song_id=id)
                    lyrics_list.append(s.lyrics)
                    if langdetect.detect(lyrics_list[-1]) == lang:
                        song = s
                        break
except AttributeError:
    pass
```

Listing 4.13: Checking and handling non-English songs with translation detection

In the first three lines, the code uses the 'langdetect' library to detect the language of the song lyrics. If the detected language is not English, it enters this block. A message is printed indicating that a non-English song is detected and it will check for false translations. The Genius API is used to search for songs by the same artist, limiting the maximum number of songs to 5, in line 4. From lines 5 to 12, a loop iterates through the found songs by the same artist and attempts to detect the language of their lyrics. If there's an exception during language detection, it continues to the next song. If there are detected languages in the list 'd', it finds the most frequent one. Line 13 compares the detected language of the original song lyrics with the most frequent language detected among the other songs by the same artist. If they are different, it assumes that a translation is detected. Line 14 prints a message indicating that a translation is detected, and it will try to find the original version. Line 15-25 searches for songs with the same title and artist (presumably to find alternative versions or translations). They iterate through the results and selects the song whose lyrics are in the detected language.

The code continues with sections that handle writing the lyrics to a file, and in case of any errors or if the song is not found, it tries to search for the song again while considering translation detection and finding the original version.

```
try:
    with open(f"./songs/songs{YEAR}/{title} by {artist}.txt", "w") as f:
        f.write(unicodedata.normalize("NFKD",
            song.lyrics).encode("ascii", "ignore").decode("utf-8"))
except AttributeError:
    os.remove(f"./songs/songs{YEAR}/{title} by {artist}.txt")
    % print("    SONG NOT FOUND! TRYING AGAIN IGNORING THE ARTIST...")
    song = genius.search_song(title, get_full_info=False)
    try:
        if langdetect.detect(song.lyrics) != "en":
            songs = genius.search_artist(artist, max_songs=10)
            d = []
            for s in songs.songs:
                try:
                    d.append(langdetect.detect(s.lyrics))
                except
                    langdetect.lang_detect_exception.LangDetectException:
                    continue
```

Listing 4.14: Handling file operations for saving song lyrics

The first 3 lines attempt to open a file for writing the lyrics of the current song. It uses `unicodedata.normalize` to normalize Unicode characters, then encodes to ASCII ignoring errors, and finally decodes to UTF-8. This process is used to handle Unicode characters and potential encoding issues.

Lines 4-6 serve for Handling Attribute Errors: Catching `AttributeError` that might occur during the previous block, remove the file that was attempted to be created and prints a message indicating that the song was not found, and it will try again, this time ignoring the artist.

Line 7 searches for the song again, this time ignoring the artist. If the detected language of the lyrics is not English, it enters the following block, by lines 8-9.

Lines 10-16 search for more songs by the artist (up to a maximum of 10 songs). They attempt to detect the language of the lyrics for each song, by using `langdetect`. If there's an exception during language detection, it continues to the next song.

```
1  if d != []:
2      lang = max(set(d), key=d.count)
3      if langdetect.detect(song.lyrics) != lang:  % # TRANSLATION DETECTED
4          % print("        TRANSLATION DETECTED! TRYING TO FIND ORIGINAL
               VERSION...")
5          song_list = genius.search_songs(f"{title} {artist}", per_page=5)
6          id_list = [song_list["hits"][i]["result"]["id"] for i in
               range(len(song_list))]
7          lyrics_list = []
8          song = 0
9          for id in id_list:
10             s = genius.search_song(song_id=id)
11             lyrics_list.append(s.lyrics)
12             if langdetect.detect(lyrics_list[-1]) == lang:
13                 song = s
14                 break
```

Listing 4.15: Handling translation detection and finding the original version

Lines 1-2 find the most frequent language if there are detected ones in the list 'd' The rest of the code operates these steps in order: • Compares the detected language of the original song lyrics with the most frequent language detected among other songs by the same artist. • If they are different, it assumes that a translation is detected. • Prints a message indicating that a translation is detected, and it will try to find the original version.

In essence, with these snippets, data extraction from Billboard and Genius is performed. The compatibility between Billboard and Genius data extraction is worth noting since platform libraries are compact to each other and easy to implement. The synergistic use of these two dataset provider platforms made it possible to move to the next stage, the application of NLP for processing the used language tones to determine emotions.

4.3 Data Cleaning:

4.3.1 Introduction to Data Cleaning: A Prerequisite for Analytical Precision:

Data cleaning is a pivotal step in ensuring the quality and reliability of the textual corpus extracted from both Billboard and Genius datasets. The unstructured nature of song lyrics requires meticulous preprocessing to transform the raw text into a format suitable for analysis. In this section, the details for the implementation of data cleaning techniques, utilizing the NLTK library for natural language processing will be discussed.

The process commences by converting all text to lowercase, harmonizing the case of words to avoid discrepancies during subsequent analyses. Numerical characters are systematically removed from the lyrics, as they contribute minimal semantic value and may introduce noise into the dataset. Following, tokenization is applied, extracting parts-of-speech (POS). This step continues with a custom function to convert NLTK tags to Wordnet Tags, for categorization of the words to lemmantazion. With diving words to categories, lemmatization is applied to find the rootwords. Lastly, stopwords and significant words extraction are applied to obtain raw and sentiment analysis applicable datasets.

4.3.2 Data Cleaning with NLTK: A Symphony of Linguistic Precision:

NLTK offers a streamlined and efficient approach to these text-processing tasks. NLTK provides an arsenal of tools, ranging from tokenization to lemmatization and POS tagging, which are the main features of NLP operations. The cleaned and refined corpus serves as the basis for subsequent sentiment analysis. NTLK facilitates the cleaning process with a more accurate examination of emotional nuances within the lyrics of popular songs.

4.3.3 Lowercasing and Denumeralization:

In the initial step of Text Preprocessing, the objective is to refine the raw lyrics data to facilitate a more focused and meaningful linguistic analysis. This step is aimed to accomplish two key tasks to enhance the quality of the dataset. Firstly, all text is uniformly converted to lowercase, ensuring consistency in the representation of words throughout the corpus. This conversion eliminates the impact of case variations, providing a standardized basis for subsequent analysis. Secondly, numerical data is systematically removed from the lyrics, redirecting the focus exclusively to linguistic content.

```python
# Lowercasing
lyrics = lyrics.lower()

# Removing numbers
lyrics_nonum = re.sub(r'\d+', '', lyrics)
```

Listing 4.16: Lowercasing and removing numbers from lyrics

Achieved through the application of regular expressions, this process rids the dataset of numerical artifacts, allowing focus only on the linguistic nuances embedded within the lyrics. The output of this preprocessing stage, denoted as "`lyrics_nonum`", serves as the foundation for subsequent cleaning procedures, setting the stage for the extraction of meaningful linguistic features and patterns in the data.

An example song, the section of the lyrics of the song 'Applause' from Lady Gaga, and the lowercased & denumeralized version are demonstrated below:

I've overheard your theory, "Nostalgia's for geeks"
I guess sir, if you say so, some of us just like to read
One second I'm a Koons, then suddenly the Koons is me
Pop culture was in art, now art's in pop culture, in me

I live for the applause, applause, applause
I live for the applause-plause, live for the applause-plause

Live for the way that you cheer and scream for me
The applause, applause, applause

Give me that thing that I love (I'll turn the lights out)
Put your hands up, make 'em touch, touch (Make it real loud)
Give me that thing that I love (I'll turn the lights out)
Put your hands up, make 'em touch, touch (Make it real loud)

Denumeralized and Lowecased Lyrics :

i've overheard your theory, "nostalgia's for geeks"
i guess sir, if you say so, some of us just like to read
one second i'm a koons, then suddenly the koons is me
pop culture was in art, now art's in pop culture, in me

i live for the applause, applause, applause
i live for the applause-plause, live for the applause-plause
live for the way that you cheer and scream for me
the applause, applause, applause

give me that thing that i love (i'll turn the lights out)
put your hands up, make 'em touch, touch (make it real loud)
give me that thing that i love (i'll turn the lights out)
put your hands up, make 'em touch, touch (make it real loud)

4.3.4 Tokenization: Unveiling Linguistic Units:

In the pivotal stage of Tokenization, the objective is to unveil the linguistic units embedded within the lyrics data, breaking down the continuous stream of text into distinct and meaningful elements. Employing the Natural Language Toolkit (NLTK), the lyrics are initially tokenized into individual words, producing a tagged sequence through the integration of part-of-speech (POS) labels with each token. The resulting '`nltk_tagged`' sequence provides a granular representation of the grammatical roles of words within the lyrics.

```
nltk_tagged = nltk.pos_tag(nltk.word_tokenize(lyrics_nonum))
```

Listing 4.17: POS tagging with NLTK

- CC | Coordinating conjunction |
- CD | Cardinal number |
- DT | Determiner |
- EX | Existential *there* |
- FW | Foreign word |
- IN | Preposition or subordinating conjunction |
- JJ | Adjective |
- JJR | Adjective, comparative |
- JJS | Adjective, superlative |
- LS | List item marker |
- MD | Modal |
- NN | Noun, singular or mass |
- NNS | Noun, plural |
- NNP | Proper noun, singular |
- NNPS | Proper noun, plural |
- PDT | Predeterminer |
- POS | Possessive ending |
- PRP | Personal pronoun |
- PRP | Possessive pronoun |
- RB | Adverb |
- RBR | Adverb, comparative |
- RBS | Adverb, superlative |
- RP | Particle |
- SYM | Symbol |

- TO | To |
- UH | Interjection |
- VB | Verb, base form |
- VBD | Verb, past tense |
- VBG | Verb, gerund or present participle |
- VBN | Verb, past participle |
- VBP | Verb, non-3rd person singular present |
- VBZ | Verb, 3rd person singular present |
- WDT | Wh-determiner |
- WP | Wh-pronoun |
- WP | Possessive wh-pronoun |
- WRB | Wh-adverb |

A list for sample lyrics after POS tagging, which the original and lowercased&denumeralized version included above, is given below. Each punctuation mark, suffix, and word are separated and categorized by NLTK.

i	've	overheard	your	theory	,	``	nostalgia	's	for
NN	VBP	IN	PRP$	NN	,	``	NN	POS	IN
geeks	''	i	guess	sir	,	if	you	say	so
NN	''	NN	NN	NN	,	IN	PRP	VBP	RB
,	some	of	us	just	like	to	read	one	second
,	DT	IN	PRP	RB	IN	TO	VB	CD	JJ
i	'm	a	koons	,	then	suddenly	the	koons	is
NN	VBP	DT	NNS	,	RB	RB	DT	NNS	VBZ
me	pop	culture	was	in	art	,	now	art	
PRP	JJ	NN	VBD	IN	NN	,	RB	VBP	
's	in	pop	culture	,	in	me	i	live	for
VBZ	IN	JJ	NN	,	IN	PRP	VBP	VBP	IN
the	applause	,	applause	,	applause	i	live	for	the
DT	NN	,	NN	,	IN	JJ	VBP	IN	DT
applause-plause	,	live	for	the	applause-plause	live	for	the	way
NN	,	VBP	IN	DT	JJ	NN	IN	DT	NN
that	you	cheer	and	scream	for	me	the	applause	,
IN	PRP	VBP	CC	NN	IN	PRP	DT	NN	,
applause	,	applause	give	me	that	thing	that	i	love
NN	,	RB	VB	PRP	IN	NN	IN	JJ	VBP
(	i	'll	turn	the	lights	out	)	put	your
(	JJ	MD	VB	DT	NNS	RP	)	VB	PRP$
hands	up	,	make	'em	touch	,	touch	(	make
NNS	RB	,	VBP	PRP	JJ	,	JJ	(	VB
it	real	loud	)	give	me	that	thing	that	i
PRP	JJ	JJ	)	VB	PRP	IN	NN	IN	JJ
love	(	i	'll	turn	the	lights	out	)	put
VBP	(	JJ	MD	VB	DT	NNS	RP	)	VB
your	hands	up	,	make	it	real	loud	)	
PRP$	NNS	RB	,	VBP	PRP	JJ	JJ	)	

Figure 4.1: POS-Tagged Lyrics of An Sample Song

In the meticulous process of extracting Parts of Speech (POS), the focus is on discerning and categorizing the diverse grammatical roles played by words within the lyrics. Employing the Natural Language Toolkit (NLTK), each token is associated with its specific POS tag, providing insight into whether a word functions as a noun, verb, adjective, or adverb or else. To further streamline the dataset and concentrate on the linguistic elements that contribute most significantly to the meaning of the lyrics, a custom function, ‘`nltk_to_wordnet`’, is applied.

```
wordnet_tagged = map(lambda x: (x[0], nltk_to_wordnet(x[1])), nltk_tagged)
```

Listing 4.18: Mapping POS tags to WordNet tags

Part of Speech	Tag
Noun	n
Verb	v
Adjective	a
Adverb	r

Figure 4.2: WordNet Tags

```
def nltk_to_wordnet(nltk_tag):
    if nltk_tag.startswith('J'):
        return wordnet.ADJ
    elif nltk_tag.startswith('V'):
        return wordnet.VERB
    elif nltk_tag.startswith('N'):
        return wordnet.NOUN
    elif nltk_tag.startswith('R'):
        return wordnet.ADV
    else:
        return None
```

Listing 4.19: Function to convert NLTK tag to WordNet tag

This function selectively retains only adjectives, verbs, nouns, and adverbs, discarding other grammatical elements. These 4 categories are the content-carrying words, which means they are the fundamental elements of the determination of concept. The use of this subset of POS categories corresponds to the core of the lyrics, highlighting the key components. The whole list of POS tags is included below, and serves as a comprehensive reference guide, emphasizing the rigorous classification performed during this critical stage of the data-cleaning process

A list of related tags is given below :

Adjectives:

- JJ | Adjective |
- JJR | Adjective, comparative |
- JJS | Adjective, superlative |

All are compiled as adjectives, indicated as 'a'

Verbs:

- VB | Verb, base form |
- VBD | Verb, past tense |
- VBG | Verb, gerund or present participle |
- VBN | Verb, past participle |
- VBP | Verb, non-3rd person singular present |
- VBZ | Verb, 3rd person singular present |

All are compiled as verbs, indicated as 'v'

Nouns:

- NN | Noun, singular or mass |
- NNS | Noun, plural |
- NNP | Proper noun, singular |
- NNPS | Proper noun, plural |

All are compiled as nouns, indicated as 'n'

Adverbs:

- RB | Adverb |

- RBR | Adverb, comparative |
- RBS | Adverb, superlative |

All are compiled as adverbs, indicated as ‘r’

Others: Compiled as ‘None’

In the table below, for sample lyrics, results for Wordnet Tags are demonstrated. Lyrics are in Navy Blue, and lyrics are divided into 5 categories. Verbs are in pistachio green, indicated with ’v’, Adverbs are in Red, indicated with ‘r’, Nouns are in Yellow, indicated with ‘n’, Adjectives are in Forest Green, indicated with ‘a’, and the rest are in Orange, indicated with ‘Nthng’, which stands for ‘Nothing’.

I	've	overheard	your	theory	,	``	nostalgia	's	for
n	v	Nthng	Nthng	n	Nthng	Nthng	n	Nthng	Nthng
geeks	''	i	guess	sir	,	if	you	say	so
n	Nthng	n	n	n	Nthng	Nthng	Nthng	v	r
,	some	of	us	just	like	to	read	one	second
Nthng	Nthng	Nthng	Nthng	r	Nthng	Nthng	v	Nthng	a
i	'm	a	koons	,	then	suddenly	the	koons	is
n	v	Nthng	n	Nthng	r	r	Nthng	n	v
me	pop	culture	was	in	art	,	now	art	's
Nthng	a	n	v	Nthng	n	Nthng	r	v	v
in	pop	culture	,	in	me	i	live	for	the
Nthng	a	n	Nthng	Nthng	Nthng	v	v	Nthng	Nthng
applause	,	applause	,	applause	i	live	for	the	applause-p
n	Nthng	n	Nthng	Nthng	a	v	Nthng	Nthng	n
,	live	for	the	applause-p	live	for	the	way	that
Nthng	v	Nthng	Nthng	a	n	Nthng	Nthng	n	Nthng
you	cheer	and	scream	for	me	the	applause	,	applause
Nthng	v	Nthng	n	Nthng	Nthng	Nthng	n	Nthng	n
,	applause	give	me	that	thing	that	i	love	(
Nthng	r	v	Nthng	Nthng	n	Nthng	a	v	Nthng
i	'll	turn	the	lights	out	)	put	your	hands
a	Nthng	v	Nthng	n	r	Nthng	v	Nthng	n
up	,	make	'em	touch	,	touch	(	make	it
r	Nthng	v	Nthng	a	Nthng	a	Nthng	v	Nthng
real	loud	)	give	me	that	thing	that	i	love
a	a	Nthng	v	Nthng	Nthng	n	Nthng	a	v
(	i	'll	turn	the	lights	out	)	put	your
Nthng	a	Nthng	v	Nthng	n	r	Nthng	v	Nthng
hands	up	,	make	'em	touch	,	touch	(	make
n	r	Nthng	v	Nthng	a	Nthng	a	Nthng	v
it	real	loud	)						
Nthng	a	a	Nthng						

Figure 4.3: WordNet Tagged Lyrics for A Sample Song

The final `wordnet_tagged` output matches each token along with its corresponding WordNet POS tag, creating the usable dataset for the following lemmatization step.

4.3.5 Lemmatization: Harmonizing Linguistic Variations:

In the transformative stage of Lemmatization, the primary objective is to harmonize linguistic variations within the lyrics, bringing words to their base or root forms. The `'lemmatized_lyrics'` list serves as the vessel for this harmonization process. For each token in the previously tagged sequence, the corresponding WordNet POS tag is used to guide the lemmatization process. In instances where a valid tag is unavailable, the token is retained in its original form, allowing for the preservation of words without a clear grammatical tag. Conversely, when a WordNet tag is present, the *'lemmatizer'* is employed to lemmatize the token based on its grammatical role. The outcome is a harmonized representation of the lyrics, where words are unified into their base forms. The outcome mitigates variations arising from conjugations, inflections, or derivations. The ensuing list, '`unique_tokens`', captures these harmonized forms, offering a distilled collection of unique linguistic elements that collectively represent the essence of the lyrics. The use of a set ensures the exclusion of duplicates, yielding a refined set of lemmatized tokens that form the basis for subsequent analyses and interpretations of the linguistic content.

```python
# Tokenize the lyrics and find the POS tag for each token
nltk_tagged = nltk.pos_tag(nltk.word_tokenize(lyrics_nonum))
# Tuple of (token, wordnet_tag)
wordnet_tagged = map(lambda x: (x[0], nltk_to_wordnet(x[1])), nltk_tagged)

lemmatized_lyrics = []
for word, tag in wordnet_tagged:
    if tag is None:
        # If there is no available tag, append the token as is
        lemmatized_lyrics.append(word)
    else:
        # Else use the tag to lemmatize the token
        lemmatized_lyrics.append(lemmatizer.lemmatize(word, tag))

unique_tokens = list(set(lemmatized_lyrics))
```

Listing 4.20: Tokenizing and lemmatizing lyrics

A list of lemmatized lyrics of the sample song is given below:

['*hand*', ''*m*', '*now*', '*could*', '*geek*', '*scream*', '*...*', ''*s*', '*wait*', '*fame*', ''*em*', '*just*', '*light*', '*art*', ''*ll*', '*out*', '*suddenly*', '*touch*', ''*ve*', '*me*', '*your*', '*if*', '*applause*', '*read*', '*or*', '*find*', '*vein*', '*way*', *'a-p-p-l-a-u-s-e', 'baby', '(', 'sir', 'love', 'it', 'critic', 'lyric', 'have', 'loud', 'theory', 'up', 'iv', 'from', 'a-r-t-p-o-pembed', ',', 'right', 'and', 'make', 'second', 'to', 'also', 'stand', 'so', 'applause-plause', 'cheer', 'be', 'turn', 'thing', 'might', 'nostalgia', 'for', 'one', 'put', 'here', 'overheard', ')', 'an', 'koons', 'you', 'bang', 'guess', 'a', '?', 'that', 'crash', 'wrong', 'woo-oh-oh-oh', 'real', 'saying', 'then', 'woo', 'say', 'live', '"', 'bear', 'us', 'in', 'of', 'pop', '"', 'gong', 'only', 'some', 'culture', 'like', 'the', 'away', 'give', 'i'*]

4.3.6 Stopword Elimination: Distilling the Essence:

In the pivotal stage of Stopword Elimination, the aim is to distill the essence of the lyrics by systematically removing common and non-informative words that may dilute the meaningful content. To achieve this, a curated list of '`insignificant_words`' is created, including words like "*oh*', *ooh*' ", and also some commonly used abbreviations like "*gon*', *wan*'", which are abbreviations for "going to, want to" encompassing words deemed to carry minimal semantic weight or relevance in the context of the lyrics. This is added manually, since they are not categorized or detectable.

```
insignificant_words = ['embed', 'likeembed', 'might', 'also', 'like',
    'lyric', 'know','go', 'say', 'oh', 'ooh', 'get', 'well', 'come',
    'make', 'one','yeah', 'ay', 'ai', 'see', 'take', 'na', 'ca', 'let',
    'tell','gon', 'wan', "``", '...', "'s'", "n't", "'m'", "'cause', '-']
```

Listing 4.21: List of stopwords and non-meaningful words

Additionally, a broader set of stopwords, including '*punctuation*' marks and common English stopwords, is compiled in the '`stoplist`'. The union of these lists forms a comprehensive set of words to be excluded, ensuring that the resulting analysis focuses on the most content-rich linguistic elements.

```
stoplist = set(stopwords.words('english') + list(punctuation) +
    insignificant_words)
```

Listing 4.22: Creating a set of stopwords

The '`unique_nostop`' list is then generated, retaining only those tokens from the lemmatized set, '`unique_tokens`', that do not match any entries in the extended stoplist. Further refinement involves eliminating words containing apostrophes, contributing to a more streamlined and semantically relevant representation of the lyrics. This meticulous curation and removal process, based on predefined criteria, is instrumental in distilling the core essence of the lyrics, ultimately enhancing the precision and interpretability of subsequent analytical endeavors.

```
# Remove stopwords
unique_nostop = [word for word in unique_tokens if word not in stoplist]

# Remove words with apostrophes
unique_nostop = [word for word in unique_nostop if "'" not in word]

# Return the final list
return unique_nostop
```

Listing 4.23: Removing stopwords and words with apostrophes

And the final list of sample song, after stopwords removal, is given below:

[*'hand', 'could', 'geek', 'scream', 'wait', 'fame', 'light', 'art', 'suddenly', 'touch', 'applause', 'read', 'find', 'vein', 'way', 'a-p-p-l-a-u-s-e', 'baby', 'sir', 'love', 'critic', 'loud', 'theory', 'iv', 'right', 'second', 'stand', 'applause-plause', 'cheer', 'turn', 'thing', 'nostalgia', 'put', 'overheard', 'koons', 'bang', 'guess', 'crash', 'wrong', 'woo-oh-oh-oh', 'real', 'saying', 'woo', 'live', 'bear', 'us', 'pop', 'gong', 'culture', 'away', 'give'*]

4.3.7 The Refined Dataset: A Prelude to Network Integration and Conclusion

As the data emerges from this cleaning process, it leaves its rawness and becomes a polished corpus. The lyrical material of each song is distilled into a pure essence, free of linguistic imperfections. This revised dataset, which demonstrates the harmonic combination of computational accuracy and linguistic nuance, is now ready for incorporation into the complicated network of musical linkages. In summary, the data cleaning stage is not merely a technical necessity but a strategic process. With a well-developed data cleaning process, the quality of the dataset can be improved. This high-quality dataset enables a more nuanced exploration of emotional patterns in the lyrics of songs spanning the last two decades. The revised dataset is now ready for the next act: data integration into the networked tapestry of music analysis.

4.4 Sentiment Analysis:

4.4.1 Unveiling Emotions through Sentiment Analysis: An Introductory Prelude

Sentiment analysis, the interpretation through which the emotional nuances of lyrical content are unveiled, stands as a backbone chapter in this analytical narrative. Within this section, the implementation of TextBlob and NRCLex as the tools of choice lays the groundwork for deciphering the emotional tapestry woven into each song. The motivation behind the selection of these tools is as strategic as the analytical insights they are capable of revealing. Section 4.4: The Sentiment Analysis is constructed as follows: The first section 4.4.1 is the introduction. In section 4.4.2, the selection reasons and advantages of TextBlob and NRCLex is discussed. Section 4.4.3 is assigned as the implementation stage, which is divided into 2 subsections, for both algorithms. Lastly, in section 4.4.4, general comments about obtained results and the areas where will they be used is discussed.

4.4.2 Why TextBlob and NRCLex: A Symbiotic Selection

The choice of utilizing both TextBlob and NRCLex in this sentiment analysis approach is arbitrary, with each library contributing its unique strengths to the overall analysis. TextBlob, known for its simplicity and efficient polarity analysis, serves as the initial stage in the process. Its straightforward sentiment polarity measurement provides a quick overview of the sentiment expressed in the song's lyrics. By this quick application, a glance of the overall tone is caught, either positive or negative. Moreover, the subjectivity evaluation is also important to understand if the lyrics contain more subjective elements or objective ones. These two measurements are possible by TextBlob.

However, understanding sentiment in its entirety involves delving into the nuances of emotion, which is where NRCLex comes into play. NRCLex specializes in providing a detailed breakdown of emotional categories, offering a more granular perspective on the sentiments present in the text. By revealing the emotional intricacies embedded in the words, NRCLex complements TextBlob's high-level analysis and unveils the nuanced emotions that might be overlooked in a more simplistic approach.

Together, TextBlob and NRCLex work synergistically to illuminate the comprehensive emotional landscape portrayed in the lyrics of each song. This two-step approach ensures a more thorough understanding of the sentiment. With a richer interpretation they provide, varied emotions expressed are measured and categorized. The combination of these two libraries succeeds in the deep and accurate sentiment analysis.

4.4.3 Implementation Steps: A Dance of Algorithms

The implementation dance begins with the extraction of lyrics from the refined dataset. Each set of lyrics becomes a canvas awaiting the brushstrokes of sentiment analysis. TextBlob takes the lead, providing a binary verdict on the sentiment polarity—positive or negative, and subjectivity-subjective or objective. The basis for the ensuing emotional dissection is laid by this binary categorization. The lead is followed by NRCLEX, for deep examination, leading to deep analyses and accurate results.

TextBlob: A Linguistic Virtuoso

TextBlob adeptly wearing the hat of sentiment analysis. Its innate ability to identify sentiment polarity—which encompasses both positivity and negativity—fits very well with sentiment analysis's binary structure. TextBlob is the main component of this sentiment analysis.

To analyze the sentiment of a collection of songs, some Python code is issued. It utilizes the TextBlob library for natural language processing and sentiment analysis. The code processes each song in the specified songlist and computes both the polarity and subjectivity scores for the lyrics. The process involves several key steps:

```python
pol_score = []
sub_score = []
for song in songlist:
    source = open(os.path.join(path_songs, song), 'r', encoding='cp1252')
    lyrics = source.read()
    clean_tokens = clean_lyrics(lyrics)
    lyrics_string = ' '.join(clean_tokens)
    blob = TextBlob(lyrics_string)
    pol_score.append(blob.polarity)
    sub_score.append(blob.subjectivity)
```

Listing 4.24: Sentiment Analysis Scores for Each Song

Two lists, `pol_score` and `sub_score`, are initialized as empty lists. These lists will store the polarity and subjectivity scores for each song, respectively. After initializing matrices, the code enters a `for` loop that iterates through each song in the `songlist`. The code opens the corresponding song file using the `open` function with the file path constructed from the `path_songs` variable and the current song's name. The lyrics of the song are then read from the opened file using `source.read()`.

The `clean_lyrics` function is called to clean and tokenize the lyrics. The implementation and steps are not shown here since the `clean_lyrics` function is implemented and explained in the Data Cleaning stage. The cleaned tokens are stored in the `clean_tokens` variable. The cleaned tokens are joined into a single string, `lyrics_string`, using the `join` method.

A `TextBlob` object, `blob`, is created from the `lyrics_string`. `TextBlob` is a library that simplifies text processing, including sentiment analysis. The polarity and subjectivity scores of the lyrics are extracted from the `TextBlob` object using `blob.polarity` and`blob.subjectivity`, respectively. These scores are appended to the `pol_score` and `sub_score` lists, storing the sentiment analysis results for the current song.

The sentiment analysis results for the current song are printed to the console using `print(blob.sentiment)`. This provides information about the polarity and subjectivity of the lyrics.

```
Applause by Lady Gaga.txt
Sentiment(polarity=0.13172541743970315, subjectivity=0.576530612244898)
```

Figure 4.4: Polarity and Subjectivity Scores of A Sample Song

As an example, the output of the song, 'Applause by Lady Gaga is given above. This result offers a nuanced understanding of the emotional tone and subjectivity within the analyzed text. The polarity score, which ranges from -1 to 1, provides insight into the overall sentiment conveyed by the text. Negative values (less than 0) indicate a negative sentiment. Positive values (greater than 0) indicate a positive sentiment. A value of 0 suggests a neutral sentiment. In this particular case, a polarity score of approximately 0.13 suggests a slightly positive sentiment. This implies that the content carries a modest degree of positivity, contributing to an optimistic or favorable tone. On the other hand, the subjectivity score, measured on a scale from 0 to 1, gauges the level of subjectivity or objectivity present in the text. Subjectivity measures how subjective or objective the text is on a scale from 0 to 1. A score of 0 indicates very objective (factual) text. A score of 1 indicates very subjective (opinionated) text. With a subjectivity score of around 0.58, the text is characterized by a moderate degree of subjectivity. This indicates that the content contains a noteworthy element of personal opinion or perspective, contributing to a more subjective expression. Together, these sentiment analysis metrics provide valuable insights into both the emotional tenor and the subjective nature of the analyzed text, enhancing our comprehension of its overall affective qualities.

In summary, the code processes a list of songs, reads their lyrics from corresponding files, performs sentiment analysis using TextBlob, and stores the polarity and subjectivity scores in separate lists. The sentiment analysis results are also printed on the console for each song.

NRCLex: Navigating the Nuances of Emotion

While TextBlob provides a holistic sentiment polarity, NRCLex dives deeper into the ocean of emotions, dissecting them into eight fundamental categories—joy, anticipation, anger, trust, fear, surprise, sadness, and disgust—which are described in Plutchik's wheel of emotions, which this study relied on. NRCLex meticulously dissects the lyrical content, assigning weightage to each emotional category. This deep examination of eight emotions is the part where emotional analysis is done.

```
def song_emotions(song):
    source = open(os.path.join(path_songs, song), 'r', encoding='cp1252')
    lyrics = source.read()
    clean_tokens = clean_lyrics(lyrics)

    words = []

    for word in clean_tokens:
        words.append(NRCLex(word).top_emotions)

    avg_dict = {}
    counts = {}

    for word in words:
        for emo, score in word:
            if emo == 'anticip':
                emo = 'anticipation'
            avg_dict.setdefault(emo, 0)
```

```
            counts.setdefault(emo, 0)
            avg_dict[emo] += score
            if score > 0: counts[emo] += 1
```

Listing 4.25: Calculating Emotion Scores for Each Song

Here, a function called `song_emotions(song)` is described. This function is called for emotion analysis for each song extracted in previous steps.

The function starts by opening the song file specified by the `song` parameter using the `open` function. It reads the lyrics from the opened file and stores them in the `lyrics` variable.

The `clean_lyrics` function is called again to clean and tokenize the lyrics. The resulting clean tokens are stored in the `clean_tokens` variable.

An empty list named `words` is initialized to store emotion scores for each word in the cleaned lyrics.

A `for` loop iterates through each clean token.

For each token, the `NRCLex` class is used to obtain the top emotions and their scores. These emotions and scores are then appended to the `words` list.

Two dictionaries (`avg_dict` and `counts`) are initialized to store the average emotion scores and the count of words contributing to each emotion.

Nested `for` loops iterate through the list of words and their emotion scores.

The emotion '`anticip`' is mapped to '`anticipation`' for consistency.

The code initializes dictionaries if not present and accumulates scores and counts for each emotion.

```
for k in avg_dict.keys():
    if counts[k] > 0:
        avg_dict[k] /= counts[k]
        avg_dict[k] = round(avg_dict[k], 2)

return avg_dict
```

Listing 4.26: Calculating Emotion Scores for Each Song (Continued)

Finally, in the following snippet, average emotion scores are calculated for emotions with positive scores. The scores are rounded to two decimal places. The function returns the '`avg_dict`', which contains the average emotion scores for various emotions present in the lyrics.

In summary, the custom-defined `song_emotions` function processes the lyrics of a song, cleans and tokenizes the text, calculates emotion scores for each word using the `NRCLex` library, and then computes average emotion scores for different emotions present in the song. The resulting dictionary provides insights into the average emotional intensity associated with various emotions in the lyrics.

The obtained average emotions are saved in a specific type of Excel file, called comma-separated values (CSV). For saving, these snippets are implemented:

```
with open(csv_file_name, mode='w', newline='', encoding='utf8') as file:
    # Create a CSV writer object
    writer = csv.writer(file)

    # Get the header from the first song
    first_song = songlist[0]
    emo_dict_first_song = song_emotions(first_song)
    header = ['title_artist', *emo_dict_first_song.keys()]
    writer.writerow(header)

    # Iterate through songs and write data rows
    for song in songlist:
        print(song)
        emo_dict = song_emotions(song)
        print(song, emo_dict)

        # Write data row
        title_artist = song
        data_row = [title_artist, *emo_dict.values()]
        writer.writerow(data_row)
```

Listing 4.27: Writing Emotion Scores to CSV

Firstly, the code opens a CSV file '`csv_file_name`' in write mode, creating or truncating the file if it already exists. It creates a CSV writer object '`writer`' to facilitate writing to the CSV file.

The '`header`' for the CSV file is obtained from the first song in '`songlist`'. The emotional dictionary '`emo_dict_first_song`' is calculated using the `song_emotions` function, and the keys of this dictionary are used as column headers. The header is then written to the CSV file.

The code then iterates through each song in the songlist, with a for loop. For each song, it prints the song name to the console. It calculates the emotional dictionary '`emo_dict`' for the current song using the `song_emotions` function.

The song name and its emotional dictionary are printed to the console for debugging purposes. The data row, '`data_row`', is constructed, including the song name and emotional values, and is written to the CSV file.

4.4.4 General Comments: A Peek into the Emotional Spectrum

The findings provide a broad range of emotions present in popular music throughout the last 20 years. With each providing its insights, the strategic interaction between TextBlob and NRCLex provides a detailed examination of lyrical feeling. The results of TextBlob, positivity, and negativity scores provide a binary perspective. The subjectivity score allows us to comment on the tendencies more over the delusional songs or the realistic songs. With the binary results of Textblob, classification of the general mood can be done. On the other hand, NRCLex, enriches this binary perspective, revealing the predominant emotional hues within the lyrical landscape. The determined eight fundamental emotions, described by Plutnick are deep examinations. With various emotions examined, various mapping and observations are possible, such as combined emotions and their descriptions, trend analyses, tendency mappings et cetera Here you can add several analyses later on. The graphs and discussions with take place in section 6 deeply, after covering the most common and respected musical psychology theories in section 5: Psychological Theories.

In conclusion, sentiment analysis provides valuable insights that are essential components of the complex emotional structure. The emotional spectrum, now deciphered and quantified, intertwines with the relational web of musical connections. A comprehensive understanding, of where the emotional undercurrents of songs echo in the interconnecting channels of musical impact, will be provided by this merger of emotions and network science techniques.

5
Psychological Theories:

5.1 Introduction

The world of music is like a complicated woven fabric, made up of the delicate threads of how people feel, think, and understand things. As we start on this adventure, diving into the exploration of psychological ideas and their important impact on understanding why we like certain music emotionally, our goal is to untangle the complex ways that connect music to our feelings and thoughts. The chapters are aimed to explain and clarify the world of musical ideas, feelings, and how our minds work, preparing us for a deep exploration into the theories that help us understand the emotional sides of music. By looking closely at the colorful expressions in music, the interpretation of the psychological theories is aimed.

5.2 Theoretical Approaches on Lyrics Effect

Exploring the historical aspects of individuals' emotional music preferences reveals a complex narrative that goes beyond a simple list of songs that have gained popularity. This section aims to bring a different point of view to the lyrical analysis by including existing musical theories. The objective is to expose the everchanging history of the relationship between lyrics and psychology by mentioning the most famous research and theories A brief musical history of the analyzed scope, 2000-2020 is included since it is important to understand the dominant themes.

Moreover, by including theories, the results can be also commented on from a psychological perspective

Investigation of theories is necessary to explore the psychological processes behind the lyrical changes and the evolving emotional preferences. This investigation tries to understand the reasons for these changes rather than just demonstrating them. The approach aims to reveal the cognitive and emotional processes influencing lyrical expressions.

While investigating the lyrical effect of the music, Scherer's Component Process Model is one of the backbones in this field [41]. Scherer's CPM categorizes emotions into four dimensions. The dimensions are valence, power, arousal, and novelty. [42]The depth is increased by this examination of psychological investigations. For example, by Emmanuelle Hanet, Scherer's Component Process Model is developed to understand how cultural shifts affect music. This research explored how music affects each of the five elements in Scherer's CPM including Central representations and the subjective feeling component, the Emotional Component, the Physiological and motor expressive component, the Motivational component, and the Cognitive component. The study, employing a mixed-methods approach, validated the four dimensions in a larger dataset by integrating qualitative insights and quantitative evaluations. The obtained results show that valence and arousal were strongly impacted by music, confirming the theory that the Component Process Model framework may be used to explain the feelings that music evokes [43].

According to Thayer's biopsychological model, mood control, music intervals, and physiological arousal are related [43]. According to Thayer's biopsychological model, mood control, music intervals, and physiological arousal are interconnected. This model posits that music interacts with biopsychological elements, such as autonomic nervous system activity, to influence emotional reactions. Music's tempo, rhythm, and harmony have the power to alter physiological states [44]. By these findings, Thayer's model reveals the dynamic interaction between biological functions and the feelings that music arouses in listeners.

Another theory is developed by Leonard Meyer, about the artistic decisions made by composers. He explains enlightening these artistic decisions. According to his theory, a variety of intricately intertwined elements, such as cultural background, psychological concerns, and devotion to long-standing musical traditions, influence these decisions [45]. Meyer questions the reasons for the repetition of some musical components while ignoring others. Through this examination, the complex selection process of the musical elements is demonstrated. The selection process between the range of composing options and the particular patterns preferred by composers was his investigation curiosity. Meyer's theory illuminates the multidimensional nature of musical expression by looking at the dynamics in between. This theory widely affected other studies by granting a fundamental theory. Provided insights are used later on for the detection of how style has changed throughout time.

In conclusion, by including current musical theories, the investigation of people's historical emotional music preferences goes beyond simple song popularity. The examination, which covers the years 2000–2020, provides insights into the dynamic link between psychology and lyrics. Through the integration of well-known theories, like Thayer's biopsychological model and Scherer's Component Process Model, the study clarifies the interaction between cultural, psychological, and physiological components that shape musical experiences. Leonard Meyer's approach emphasizes the complexity of musical expression and its progression throughout time by dissecting the complex choices made by composers. When taken as a whole, these theoretical frameworks improve our comprehension of the mental processes underlying lyrical modifications and changing emotional preferences in music.

5.3 Emotion and Cognitive Processing

In this chapter, brief information about the discoveries in the complex areas of emotion and cognitive processing is aimed. Understanding how music affects the brain's elicitation and processing of emotions allows for cognitive commentary on research findings. The research findings extend beyond mere statistical data and visual representation, delving into the intricate musical elements. The main focus is drawing attention to the significant impact of the lyrical material on the emotional and cognitive aspects of the listener's experience.

Knowledge about emotional experiences and cognitive is important to fully comprehend musical analyses. The lyrical elements—such as word choices, storytelling devices, and metaphors—intricately contribute to cognitive components. These cognitive components unveil the many layers that shape the listener's emotional and cognitive experience[46]. Lakoff and Johnson's Conceptual Metaphor Theory, for example, asserts that the use of metaphorical language in songs unveils how specific terms evoke distinct cognitive frames, influencing emotional reactions[47]. On the other hand, Schema theory explains how information is organized in the mind and shapes listeners' reactions to lyrical content [48]. Zajonc's Affective Primacy suggests that emotions may precede and influence cognitive processes, proposing that emotional reactions can occur more rapidly and instinctively than cognitive assessments. This hypothesis states that emotional reactions might happen more quickly and naturally than cognitive assessments [49]. Put another way, a picture or a piece of music might trigger an affective (or emotional) reaction that happens before conscious cognitive processing. Affective primacy underscores the swift and often reflexive nature of emotional responses in human cognition, asserting that emotions play a primary role in shaping initial reactions and judgments.

By this chapter, a journey on cognitive science and music relationship is made. At the end of this journey, revealing the harmonic interaction between thought, and feeling and how lyrical content shapes the whole musical experience is sought. The analysis of lyrical content in this context highlights the complex ways in which cognitive and emotional components interconnect inside the listener's awareness. With a focus on both experimental research and cognitive theories, a duality between theoretical and practical correlations was made possible. This duality interprets the mutually beneficial relationship between the listener's emotional resonances and the cognitive processing of lyrical content, emphasizing both experimental research and cognitive theories.

5.4 Emotion Regulation Through Music

One another area that applied lyrical analysis is emotional regulation. The usage of lyrical analysis in emotional regulation plays out as a real-life application area. This section discusses how music helps individuals regulate emotions during emotional difficulties.

How people intentionally use music to control their emotions has become an interesting research area, thanks to advanced mood regulation systems development. This area provides a sophisticated understanding of which songs become therapeutic, and preferred by people who deal with emotional difficulties. Intentional music use for emotion control gains clarity when rooted in psychological ideas, such as Gross's Process Model of Emotion Regulation. This study links the intentional application of music to describe the processes behind emotional regulation [50]. According to this model, individuals consume music to modify their emotional experiences, such as reappraisal or denial. This model brings a new perspective on how people intentionally utilize music to control their emotions.

One well-known psychological paradigm that emphasizes the critical role of cognitive evaluation and coping techniques in managing stressors is the Transactional Model of Stress and Coping by Lazarus and Folkman. People continually evaluate stresses, determine their importance, and employ mood regulation techniques based on cognitive assessments [51]. Evaluating one's perceived danger and coping skills is part of cognitive evaluation. After evaluation, the individuals use mood regulation techniques, such as emotional control and problem-solving skills. This approach is widely used by mood regulation system designers. It highlights the dynamic interplay between cognitive assessments and adaptive coping mechanisms, providing insights into how individuals regulate and control their emotional responses when faced with hardships.

In conclusion, the use of lyrical analysis in emotional regulation is shown to be an important field that we use in daily life intentionally or non-intentionally. In this chapter, this importance is discussed and the chapter includes brief information about the practical application area by mentioning prior research. Mentioning prior studies is purposeful since the developments of mood regulation systems of music are based on them. Studies such as Gross's Process Model of Emotion Control provide the intentional application of music and the underlying mechanisms behind emotional control. This research investigates how people consciously use music to navigate and manage their emotions, especially when faced with emotional challenges. Furthermore, Lazarus and Folkman's Transactional Model of Stress and Coping emphasizes the vital role that mental assessments and coping strategies play in handling stress sources. This model provides insightful information on the dynamic interaction between cognitive evaluations and adaptive coping mechanisms. Addressing the fundamental question of how people use music therapy to control their emotions, Lazarus and Folkman's work holds a central place in the field of mood regulation.

5.5 Cultural Influences on Emotional Perception

In the mosaic of emotional perception, the complex role of cultural variables on emotional reactions to music is undeniable. The theories of lyrical material as a prism to examine how cultural influences are intertwined with emotional themes in the songs are developed in other studies. There are some fundamental studies in the identification of cultural influences. Mentioning them is a must for crucial for understanding the ongoing effects of culture in music. The studies that connect cultural psychology theories to specific emotional experiences are mentioned in this section. These studies try to explain the bidirectional relationship dynamics between cultural and emotional nuances given through lyrical narratives. Markus and Kitayama's Cultural Theory of Self, for example, declares that individualism and collectivism are key factors determining cultural differences in emotional expression [52]. Their study is based on the analysis of cultural psychology and understanding the tendencies of self vs community benefit decisions. In the musical analysis area, their theory is used nowadays for evaluating cultural music whether it is made for society or personality[53]. In other words, independent, individualistic music versus interdependent collectivistic music.

Additionally, possible conflicts or synergies between personal and cultural emotional reactions can occur. Some researchers focus on this duality since it is crucial to observe how listeners react to conflicts that occur between their personal versus cultural morals. By examining the dynamic connection between individual emotions and bigger cultural representations, a subjective perspective is aimed to develop. Theories such as Triandis' theory of subjective culture, highlight the impact of common values, norms, and beliefs on emotional experiences within specific cultural contexts. In that theory, the possible alignments between individual emotional reactions and the cultural cues encoded in lyrics are searched. The study hopes to contribute to decoding the emotional nature of music to unveil cultural influences hidden in it. Additionally, possible conflicts or synergies between personal and cultural emotional reactions can occur. Some researchers focus on this duality since it is crucial to observe how listeners react to conflicts that occur between their personal versus cultural morals. By examining the dynamic connection between individual emotions and bigger cultural representations, a subjective perspective is aimed to develop. Theories such as Triandis' theory of subjective culture, highlight the impact of common values, norms, and beliefs on emotional experiences within specific cultural contexts[54]. In that theory, the possible alignments between individual emotional reactions and the cultural cues encoded in lyrics are searched. The study hopes to contribute to decoding the emotional nature of music to unveil cultural influences hidden in it [55]. Similarly, Hofstede's cultural dimensions theory is developed to measure differences across various cultures. Some examples of measured differences are power distance, individualism, masculinity, and uncertainty avoidance. Hofstede's paradigm investigates how power dynamics, individualistic or collectivist inclinations, gender roles, and attitudes toward uncertainty influence the emotional subtleties hidden in different pieces of music [56]. As a result, this paradigm offers a deep knowledge of cultural aspects. With this knowledge, the researchers obtain a new point of view of the emotional landscape of musical encounters across countries.

In conclusion, the investigation of the complex interaction between cultural characteristics and emotional responses to music creates a fascinating mosaic within the domain of emotional perception. Based on ideas such as Markus and Kitayama's Cultural Theory of Self, Triandis' Theory of Subjective Culture, and Hofstede's Cultural Dimensions Theory, this chapter aimed to explain the complex effect of cultural dynamics on the emotional themes encoded in music. These frameworks, which range from individualism-collectivism dynamics to the effect of shared values and cultural characteristics, serve as an interpretation through which we may discover the emotional nature of music in certain cultural contexts.

5.6 Psychological Theories and Music

The interactive relationship between psychology and music provides diverse theories. These theories cover the psychological aspects of musical elements such as melody, lyrics, harmony etc. Each element requires different assessment methods and each one proposes different prospects for further investigations. With the assistance of these theories, delving especially into lyrical content to uncover the delicate interplay between theory and practical examples was made possible in this research.

Applying psychological ideas to lyrical material results in a bridge that deepens musical analysis. The goal is not just to explain ideas, but also to demonstrate how certain features in lyrics elicit emotional reactions, so anchoring academic frameworks in the actual arena of lyrical expression. For example, Schachter and Singer's Two-Factor Theory of Emotion is used to show how physiological arousal and cognitive interpretation combine in poetic situations to shape emotional experiences[57]. This theory proposes that the emotional reaction to a stimulus consists of both physiological arousal and cognitive labeling of that arousal. It highlights the reaction between physiological and cognitive reactions towards a poetic expression.

Another fundamental in applied psychology in the music field is Cannon-Bard's theory of emotion. This theory is accepted as a milestone for the cognitive musical applications area. This theory is constructed on the simultaneous measurement of physiological and emotional responses to external stimulators[58],[59]. The hypothesis 'Lyrical elements may elicit both physiological and emotional reactions' is tested in this theory. Cannon-Bard proposes that emotional experiences are not solely dependent on physiological changes but can occur concurrently. The main question they worked on is 'How does music stimulate emotionally both the mind and body, through its lyrical components' which is very interesting and demanding. This insight extended the application area of lyrical analysis, being one most the most cited works in this field proves it.

The James-Lange theory of emotion, a fundamental concept in psychology, asserts a sequential link between physiological arousal and emotional experiences. In this model, an emotional stimulus initiates physiological responses, and individuals subsequently interpret these bodily changes as specific emotions [60]. Applied to music, this theory suggests that elements in music can trigger several physiological reactions. The James-Lange theory is one of the widely used theories to understand better how music influences emotions. Moreover, emphasizing the role of bodily responses in shaping emotional states is highlighted.[61]

We encounter several beautiful theories in the theoretical approach to psycho-musicology.

These theories bridge psychological concepts and musical elements. From delving into lyrical content with Schachter and Singer's Two-Factor Theory to exploring Cannon-Bard's simultaneous measurement of physiological and emotional responses, the research reveals the intricate dynamics of music and emotions. Each theory contributes unique insights into the emotional dimensions of music, uncovering how musical elements elicit both physiological and emotional reactions. This bridge between theoretical frameworks and practical examples enriches the understanding of the complex relationship between psychology and music, deepening musical analysis and paving the way for continued interdisciplinary exploration.

5.7 Impact of Lyrics on Emotional Experience

In this chapter, the studies done so far about the investigation of the influence of lyrics on emotional experiences are discussed. This discussion attempts to examine the theories of how lyrics build the emotional landscape. The focus is on the relationship between lyrics and sentiments, stimulators that ignite emotions.

The answer sought lies underneath is addressing the harmony between music and lyrics, by understanding a song's emotional characteristics. By presenting particular examples of developed theories of how lyrics enhance or modify emotional experiences, the explanation of the influence of the lyrics is aimed.

For instance, drawing upon the work of Huron and Biazzo on lyric-setting congruence, the exploration could illustrate how the alignment of lyrics with the melodic and harmonic elements enhances emotional congruence and intensifies the emotional impact. Moreover, the Congruence-Association Model posits that the emotional impact of a musical piece is heightened when the lyrics align with the melodic and harmonic elements, creating a congruent emotional experience [62]. This model proposed a general balance where lyrics misalign with the emotional tone set by the music encouraging reflection on the complexities of musical compositions. This model demonstrates the balance between verbal and nonverbal elements in shaping the overall emotional fabric of a musical work. Detection of the situations in which the lyrics contradict or stray from the emotional tone established by the melody established a new field, a new question: Are the emotions ignited by the songs affected more by lyrics or more by instrumentals?

Another insightful perspective is the Multimodal Integration Model, suggesting that the brain processes lyrics and music simultaneously, contributing to a unified emotional experience. This approach proposes the linguistic and musical aspects are processed together in specified brain areas, by explaining the synergy between lyrics and music. This research highlights the correlation of verbal and non-verbal elements of music triggering various emotional scales together, including both positive and negative emotions. [63]

In conclusion, the influence of lyrics on emotions is an area, with a lot of questions to comprehend and answer. Theories such as Huron and Biazzo's lyric-setting congruence, the Congruence-Association Model, and the Multimodal Integration Model are some fundamental theories in this field. They demonstrate the interaction of lyrics and melody, as well as how the brain interprets them at the same time. The aforementioned concepts suggested a balance of verbal and nonverbal aspects. By regarding the relative importance of lyrics and instrumentals in eliciting emotions, music is understood and produced efficiently.

5.8 Conclusion

In conclusion of this chapter, the journey through the interesting connection between psychology and music has been like an adventure. This chapter proved how much this connection really affects the way we like music. Thinking back on the theories mentioned, it's clear that music is more than just something heard – it's like a powerful way the listeners connect with our feelings and thoughts.

Through this chapter, how music and psychology work together is shown. Also, this chapter served to help us understand how culture, thinking, and managing our feelings all come together when we listen to music. Exploring how lyrics affect our emotions showed us that there's a kind of balance between the words and the music, making the emotional picture of music more complex.

Putting all these ideas together, the theories talked about and the real-life examples, it's not just about knowing more about why listeners love certain songs emotionally. It's also about seeing how music is a part of their emotional and thinking world. This chapter is like an invitation to keep exploring how psychology and music talk to each other and led to a strong knowledge about psychological agents since they are discussed in the next chapter, Chapter 6: Observed Trends.

6

Application Stage:A Multidisciplinary Approach:

6.1 Introduction

In the application stage, all aforementioned tools for the analysis are employed. The application stage is divided into 3 sections. Firstly, in the section 6.2: Yearly Sentiment Analysis, the gathered data is analyzed and average-median values of both emotions and sentiments are calculated. They are also demonstrated with graphs for visuality, in the ranges of 1 song, 100 songs for 1 year, and overall the research scope, 20 years, 2000 songs in total. Heatmaps are constructed for co-relationship. Also, emotion-triggering lyrics and percentages are demonstrated with one example song, for comprehension.

Apart from mathematical calculations, in section 6.3: Network Integrated Analysis, more complex analyses are demonstrated. Firstly, these analyses start with Network Density Analysis, followed by Assortivity values calculation. A Venn diagram for the overlapped nodes for specific nodes is added since it highlights an interesting finding. Moreover, community detection with two different algorithms, Leiden and t-SNE, various years communities are calculated and commented. T-SNE cluster Analysis is evaluated by the calculation of Silhouette Scores, for measuring the overall quality of the clustering algorithm. In the last section of this section, 6.4: Psychological Examinations, measured emotions and sentiments are commented on with psychological theories. Wheel visualization is implemented as a demonstration of Plutnick's emotion wheel theory, which this study is constructed on. Moreover, yearly emotions are visualized, with 5 major events occurring in the data scope, affecting the patterns. The effect of these events is discussed after visualizations.

6.2 Yearly Sentiment Analysis

In the enthralling journey of yearly sentiment analysis embark on a profound exploration into the ever-shifting landscape of emotions encapsulated within popular music lyrics. Each passing year initializes a new chapter, by refreshing a new chart. This section aims to breed a new perspective to understanding the lyrics-sentiment relationship. For deciphering lyrical sentiments annually, the study delved into the intricate lyrical symphonies in previous sections. These symphonies mirror the evolving emotional preferences of audiences. This annual voyage focuses on interpreting the emotional richness that defines the very essence of popular music.

6.2.1 Subjectivity Analysis

For initialization, the interplay between subjectivity and objectivity within song lyrics is employed. The demonstration is required to uncover the profound implications of this dual nature in musical expression. This exploration is crucial for understanding how artists weave their subjective experiences, emotions, and perspectives. A song's subjectivity provides listeners with a window into the artist's narrative, cultivating a profound emotional connection. However, an excessive embrace of subjectivity can sometimes lead to lyrics that may seem delusional, detached from reality, or overly introspective.

Within this context, consider the example of the year 2002, where the average subjectivity score stands at 0.53, positioning it squarely in the middle of the subjectivity spectrum. Notably, over 90 percent of the values fall within the range of 0.4 to 0.7, reflecting a nuanced balance between slight subjectivity and slight objectivity. However, outliers exist, with a few songs boasting exceptionally high subjectivity scores such as 0.82 and 0.91, suggesting extreme objectivity. Conversely, scores like 0.21 or 0.27 indicate extreme subjectivity. This diversity in scores underscores the possibilities of reaching the edge on both ends, but in general, the subjectivity score is weaving in around median values, reminds a Gaussian Distribution.

This analysis underscores the challenge of interpreting subjective scores and highlights the diverse range of lyrical expressions that artists employ. This analysis contributes to a richer understanding of the relationship between subjective introspection of the lyrics and the audience's interpretation

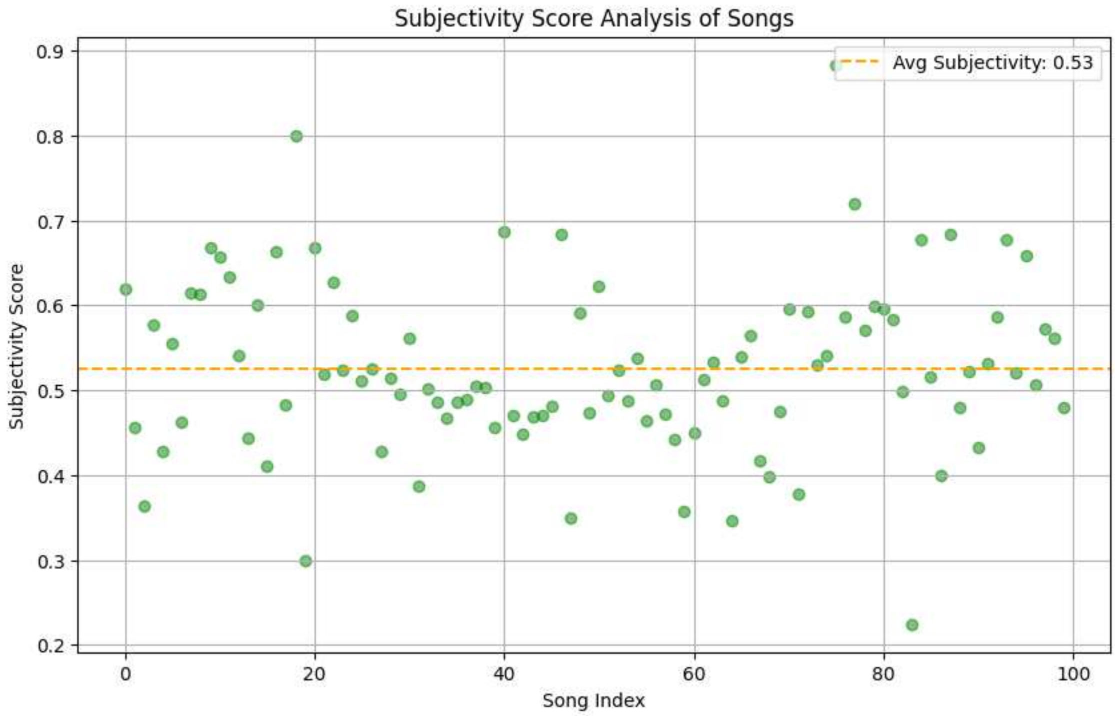

Figure 6.1: Subjectivity Scores for the Year 2002 Songs

The exploration immerses into the nuanced realm of average-median subjectivity scores, unraveling the intricate fabric of subjective expressions intricately woven within song lyrics. This comprehensive analysis provides a dual perspective on the central tendencies of subjectivity, utilizing both average and median subjectivity scores to discern patterns and variations in the emotional landscape of the dataset. The average subjectivity score, as the arithmetic mean, furnishes a broad overview of the collective subjectivity encapsulated in the analyzed lyrics, shedding light on the general tendency toward either objectivity or subjectivity. Meanwhile, the median subjectivity score, resilient to the influence of outliers, establishes a robust central point, ensuring a more representative measure of the dataset's overall subjectivity. This dual lens of average and median subjectivity scores enriches our understanding of the intricate interplay between lyrical subjectivity and objectivity, contributing to a holistic comprehension of the diverse emotional tapestry woven by musical artists across different compositions.

Furthermore, notable characteristic surfaces during this exploration—average subjectivity scores consistently fall within a narrow range for both average and median, oscillating around

0.47 and 0.53. The minimal deviation between median and average values underscores remarkable stability in the subjectivity trends across the dataset. This suggests that, unlike polarity scores influenced by general mood and song detections, the procured songs exhibit a balanced distribution in terms of subjectivity and objectivity. The subtle variations within this narrow range indicate a harmonious blend of subjective and objective elements in the lyrical content.

Moreover, the analysis of standard deviation values unveils an overall narrow range between 0.09 and 0.11, highlighting a general consistency in the dispersion of individual subjectivity scores. Exceptions in 2019 with a standard deviation of 0.128 and 2011 with 0.023 hint at years with distinctive characteristics that deviate from the usual trends. These insights contribute to a nuanced understanding of subjectivity dynamics within the dataset, emphasizing both stability and occasional variations that enrich the narrative of emotional expression in music.

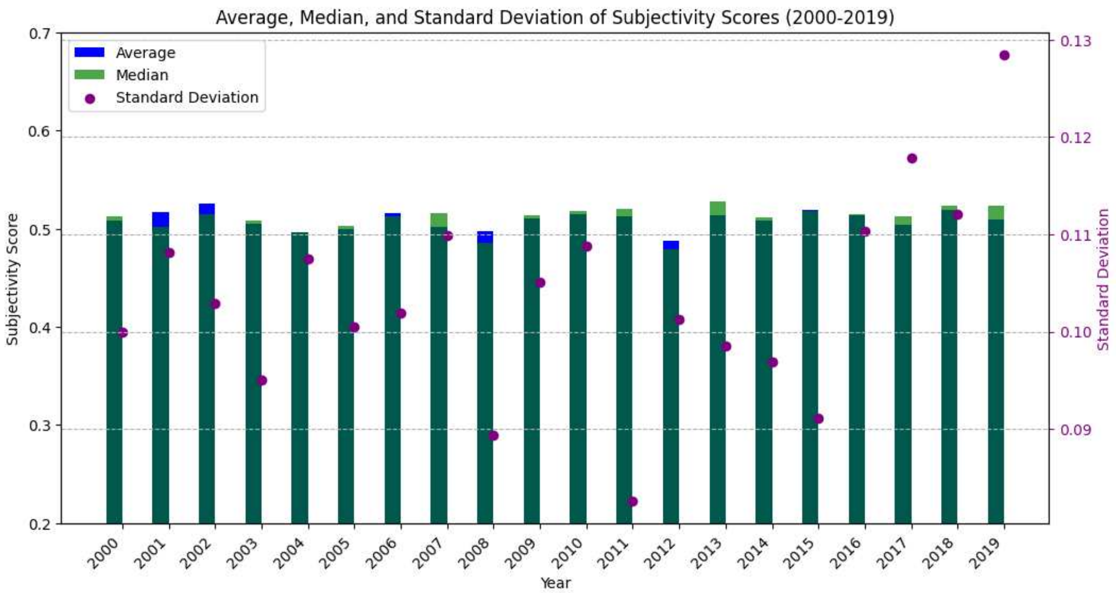

Figure 6.2: Subjectivity Scores for the years 2000-2019

6.2.2 The Spectrum of Positivity and Negativity: Polarity Sentiments Explored

Within in the section The Spectrum of Positivity and Negativity: Binary Sentiments Explored an illuminating journey is started to dissect the emotional dichotomy intricately woven into the fabric of song lyrics. This segment meticulously explores the binary realm of sentiments, distinguishing between the uplifting strains of positivity and the melancholic echoes of negativity within the rich tapestry of musical compositions. By delving into this emotional spectrum, the analysis seeks to unveil patterns, trends and shifts in lyrical tones over the years.

A notable aspect of this exploration lies in the revelation that songs exhibit a slightly positive inclination, with an average sentiment score of 0.08. Furthermore, mirroring the subjectivity score distribution, over 90 percent of the sentiment values fall within the range of -0.1 to 0.3, indicating a prevalent trend toward mildly positive language in song lyrics. This statistical pattern resembles a Gaussian distribution, underscoring the nuanced balance in the emotional expression of musical artists.

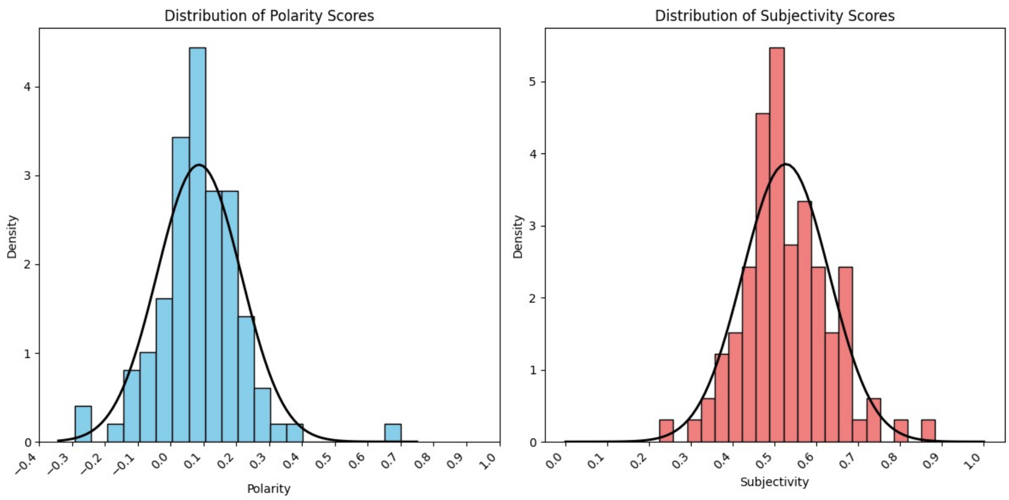

Figure 6.3: Polarity and Subjectivity Distribution of 2002 Songs

However, outliers exist within this emotional landscape. Notably, there are instances of songs that convey extremely positive sentiments, with scores reaching 0.7, or featuring two songs with a sentiment score of 0.35. Conversely, some songs delve into negativity, exemplified by two songs with a sentiment score of -0.3. Despite these extremes, the overall distribution maintains a similarity to Gaussian distributions, emphasizing the complexity and diversity of emotional expression within popular music.

This exploration not only identifies prevalent emotional themes but also offers profound insights into the evolving emotional landscape of popular music. Whether articulating tales of joyous highs or delving into the depths of despair, the nuanced examination of binary sentiments enriches our understanding of the emotional tapestry woven by musical artists, transcending the boundaries of conventional expression.

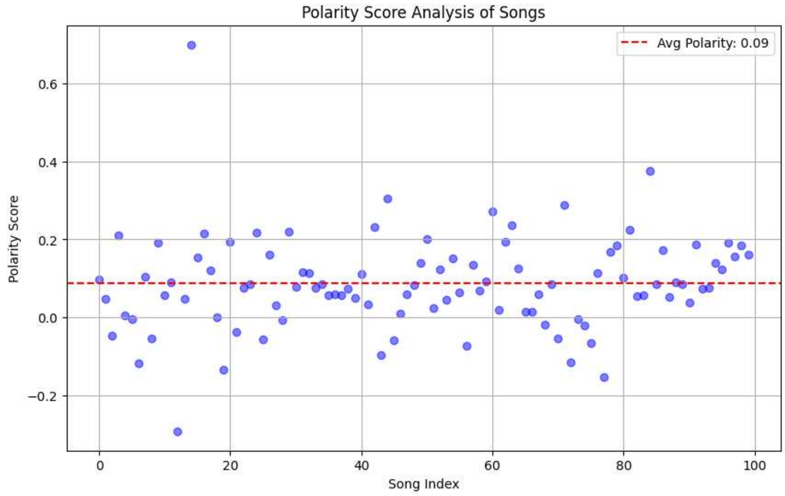

Figure 6.4: Polarity Scores for the Year 2002 Songs.

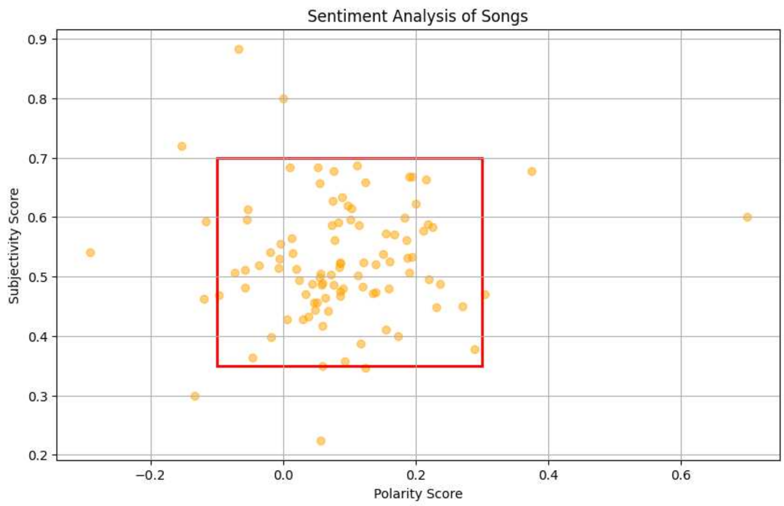

Figure 6.5: Polarity and Subjectivity Scores for the Year 2002

In the detailed analysis of binary sentiments, the exploration into average-median polarity scores unveils a nuanced comprehension of the central tendency in emotional expressions within song lyrics. Through a meticulous examination of both average and median polarity scores, we gain profound insights into the distribution's central tendencies, thereby illuminating the prevalent emotional tone characterizing the analyzed songs.

The average polarity scores exhibit a broad spectrum, ranging between 0.123 and 0.696, offering crucial points of consideration. This wide range prompts essential questions, particularly regarding the factors influencing these fluctuations over the years. Notably, a discernible decrease is observed from the years 2001 to 2003, hinting at a potential shift in the overall emotional tone of song lyrics during this period. Furthermore, despite the second-highest average in 2004 (surpassed only by 2011), a notable divergence between the average and median is evident. This divergence suggests that, while the average polarity score was relatively high, the presence of extremely positive songs contributed to a more negative general tendency.

The exploration of average-median polarity scores also reveals intriguing patterns in the relationship between average and median values. With exceptions in the years 2003, 2006, and 2008, where median values slightly surpass averages, the general trend indicates higher average scores. This phenomenon suggests a preference for extremely positive songs, even when the overall mood tends to be slightly negative. Additionally, the incorporation of standard deviation in the graphical representation enhances the depth of our analysis, providing a possibility to comment on the overall distributions.

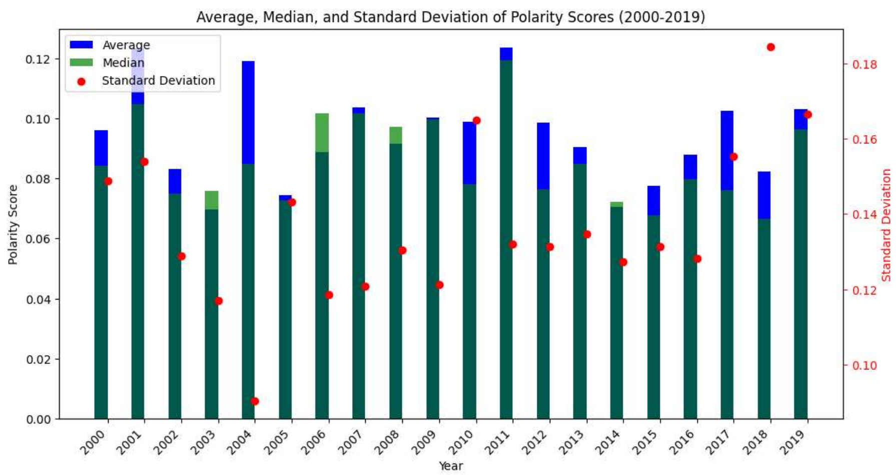

Figure 6.6: Average Polarity over Years

6.2.3 The Emotional Mosaic: NRCLex and the Eight Fundamental Emotions

The section The Emotional Mosaic: NRCLex and the Eight Fundamental Emotions is the stage where NRCLex is employed and the extensive framework of the eight fundamental emotions. In this part, sophisticated sentiment analysis methods break down lyrics to extract subtle emotional undertones that go beyond conventional sentiments. The study uncovers a complex tapestry of emotions that are delicately woven within the lyrical material by mapping these emotions onto a mosaic. When it comes to navigating through emotions such as joy, sadness, anger, anticipation, trust, fear, surprise, and disgust, NRCLex is our go-to guide. This incisive investigation reveals the richness and diversity of human expressions in the lyrics. As an illustrative example, the inclusion of the same song, "Applause" by Lady Gaga, after the cleaning step enhances our understanding. In the accompanying bar chart depicting the analyzed emotions, we observe a unique emotional profile. Notably, the negativity and positivity scores are measured equally at 0.52, indicating an overall neutral tone. The analyzed emotions for the song include 'fear' (0.28), 'anger' (0.29), 'anticipation' (0.35), 'trust' (0.39), 'surprise' (0.42), 'positive' (0.52), 'negative' (0.52), 'sadness' (0.21), 'disgust' (0.18), and 'joy' (0.33). This intricate emotional analysis suggests that the song's overall emotion, is based on lyrics. Observed results show that the selected example song leans towards trust and surprise, with lower values for sadness and disgust. In interpretation, this signifies that the song carries an exciting and promising tone, rather than of hate or melancholy. The incorporation of individual emotions enriches our comprehension, allowing us to discern the nuanced emotional landscape within the lyrical composition.

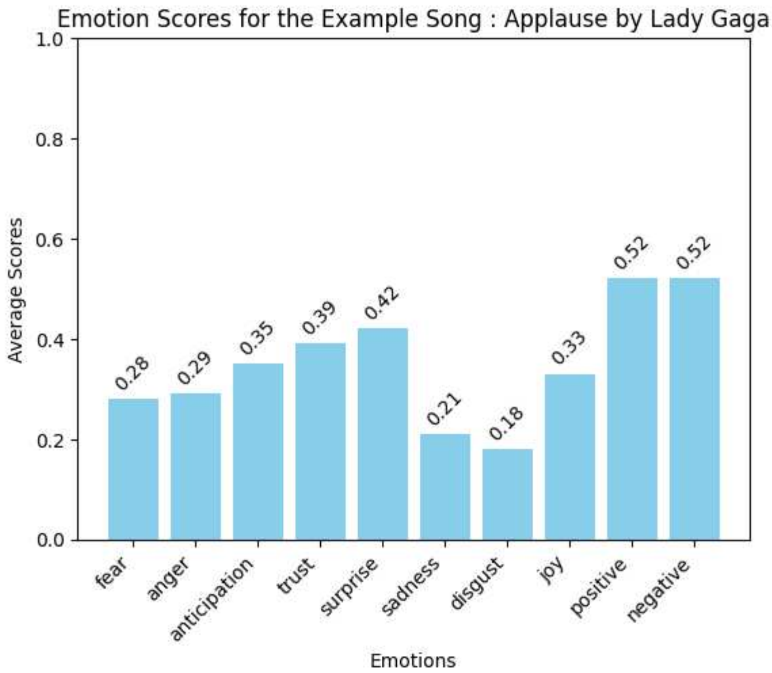

Figure 6.7: Emotion Analysis of A Sample Song

Unveiling Emotional Peaks and Valleys: Peaks of Joy, Valleys of Sadness

Unveiling Emotional Peaks and Valleys: Peaks of Joy, Valleys of Sadness intricately charts the emotional contours of music, leveraging the analytical prowess of NRCLex and its alignment with Russell's emotion chart. This section not only discerns emotional peaks and valleys but also distinguishes between contrasting emotions such as joy and sadness. Through the sophisticated linguistic analysis provided by NRCLex, the analysis comprehensively explores the positive-negative spectrum, unraveling the intricate emotional tapestry woven into song lyrics. In the presented chart, the average emotions for 100 analyzed songs for each year are demonstrated, revealing noteworthy insights. A salient observation is the consistent dominance of negative tones over positive ones, suggesting a prevailing inclination towards a more optimistic tone in the analyzed songs. The prominence of the emotion 'anticipation' across all years captures attention, and its prevalence may be attributed to the averaging nature of the analysis, where words can carry multiple emotions. Further exploration into the data reveals a decade-long dominance of the emotion 'anticipation' when comparing the first three years (2000-2002) with the last three years (2017-2019). This observation raises intriguing questions about societal trends, such as shifts in anticipation levels, expectations from the future, or even a potential increase in depressive sentiments. The emotion 'trust' maintains a stable profile across all years, which could be influenced by the lexicon's encoding of fewer stimulators for this emotion. Notably, the year 2002 exhibits a significant spike in the emotion 'disgust,' correlating with increases in 'anger,' 'sadness,' and 'fear.' This spike can be linked to the impactful events of 9/11, illustrating the tool's capability to capture and reflect societal responses to significant events. This graph serves as a powerful tool for understanding the evolving trends of emotions over the years, offering valuable insights into the emotional craftsmanship of musical compositions and their resonance with societal dynamics. The exploration of these emotional peaks and valleys is also mentioned in Subsection 6.4.2 :Key Events Influences with the reasons and commentary.

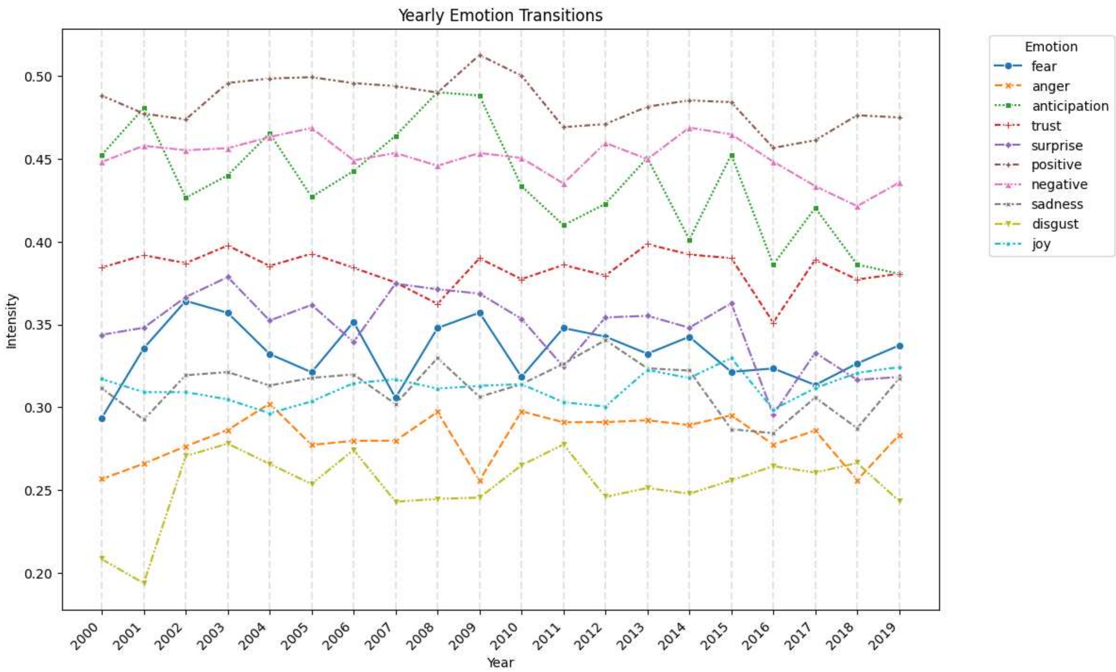

Figure 6.8: Emotion Averages of the Songs Between the Years 2000-2019

Heatmaps of Emotion Triggering Lyrics

Embarking on the exploration of Heatmaps for Emotion Triggering Lyrics, we immerse ourselves in the intricate choreography of emotions within song lyrics. This visual journey unfolds through heatmaps, where each token contributes to a color-coded display, unveiling the intensity and frequency of different emotions. To delve deeper, we undertake a meticulous analysis of a chosen example: ”Lady Gaga - Applause.” Through comprehensive data cleaning, all tokens from this iconic song are gathered, and their sentiment scores are individually extracted. Each token undergoes sentiment analysis, with a focus on retaining non-zero values for at least one emotion, ensuring a concentrated display of emotions expressed in the lyrics.

In the resulting heatmap, vibrant colors illuminate the emotional landscape, vividly portraying the sentiment dynamics inherent in Lady Gaga's lyrical masterpiece. As an illustrative example, the obtained and cleaned words from this song's lyrics have undergone separate sentiment analysis. Words that do not trigger any sentiments are excluded from this heatmap, creating a focused representation. The x-axis showcases the layout of emotions, while the y-axis displays sentiment-triggering words. For instance, the word 'critic' is coded as a negative sentiment trigger by itself, assigned a value of 1. Similarly, words like 'fame' and 'culture' are positive sentiment triggers without any other associated emotions.

Notably, certain words serve as multiple triggers, exemplified by 'Theory,' which triggers the emotions of anticipation and trust with values of 0.5 each. Another example is 'bang,' triggering the emotions fear, anger, surprise, negative, sadness, and disgust, each with a value of 0.2. This insightful heatmap provides a valuable understanding of how emotions are detected and measured by NRCLex, shedding light on how specific words can trigger and contribute to various emotions. The exploration enriches our comprehension of the intricate relationship between lyrics and emotions, offering a nuanced perspective on the emotional nuances encapsulated in the lyrical content.

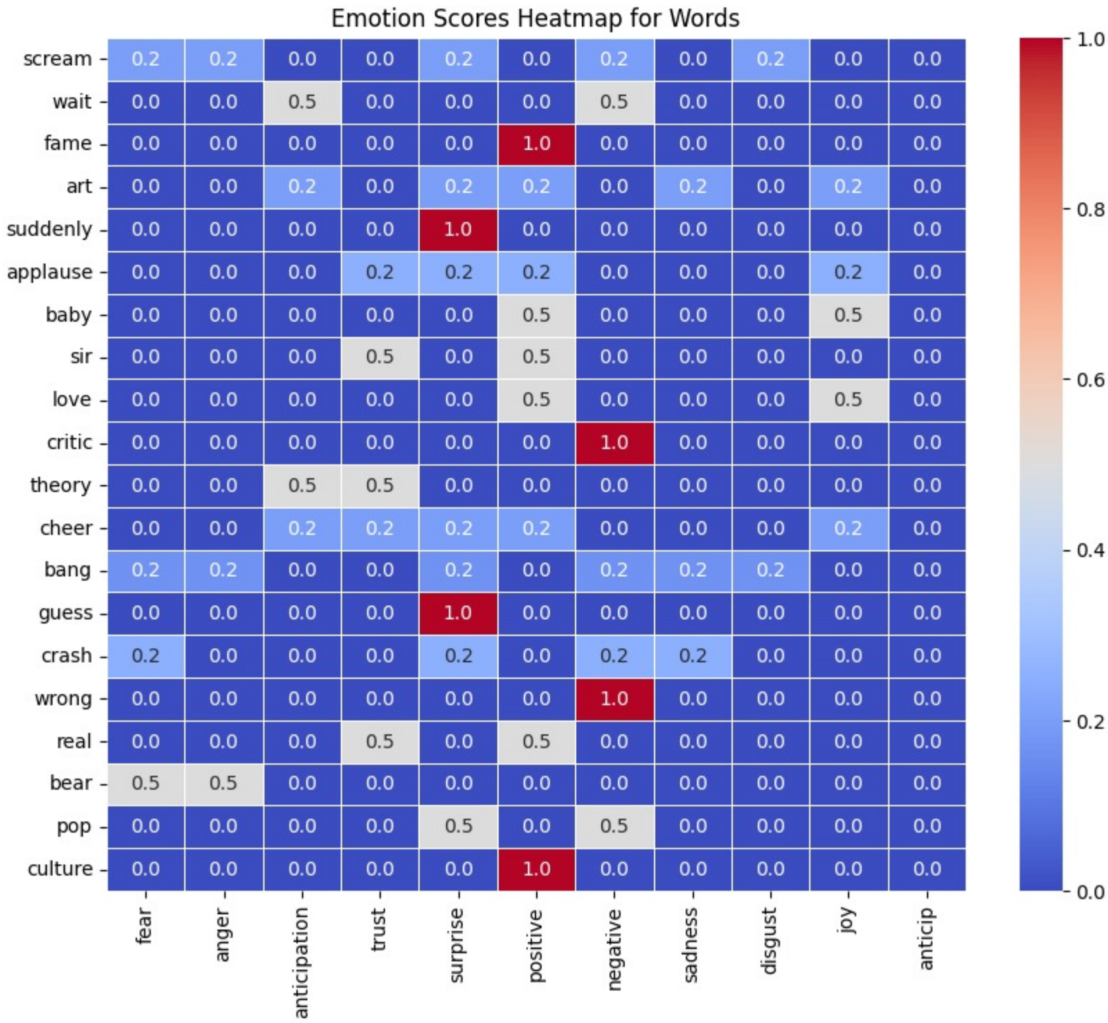

Figure 6.9: Heatmap of A Sample Song's Lyrics

A Symphony of Themes: Themes Woven into Emotional Patterns

The section A Symphony of Themes: Themes Woven into Emotional Patterns embarks on a captivating journey to unveil the thematic undercurrents that contribute to the emotional tapestry of music. Beyond individual emotions, this section delves into how recurring themes in lyrics intertwine to create intricate emotional patterns. Recognizing that specific topics or motifs may evoke distinct emotional responses, the exploration enriches the lyrical landscape by unraveling these thematic threads through NRCLex analysis. This reveals a symphony of interconnected themes that collectively shape the emotional character of songs.

As the analysis continues, a quadruple emotion matrix is introduced, comprising four components: 1) Upper Left - Absolute Differences, 2) Upper Right - Absolute Sums, 3) Lower Left - Minimum, and 4) Lower Right - Maximum. These matrices serve as invaluable tools, offering insights into the relationships between different emotions. Absolute differences provide an easy way to track the disparity between two emotions, occurring in a similar percentage. Absolute sums help trace pairs of emotions that co-occur frequently, depicted by a darker shade. The minimum and maximum matrices contribute by highlighting the compression between the lower and higher values of corresponding tiles, showcasing the spectrum of emotional relationships.

Moreover, the exploration extends to the calculation of correlation matrices and sentiments vs valence sections. These analyses further strengthen and demonstrate the correlation and relationships between emotions, offering a comprehensive understanding of the intricate emotional patterns embedded in the thematic content of song lyrics.

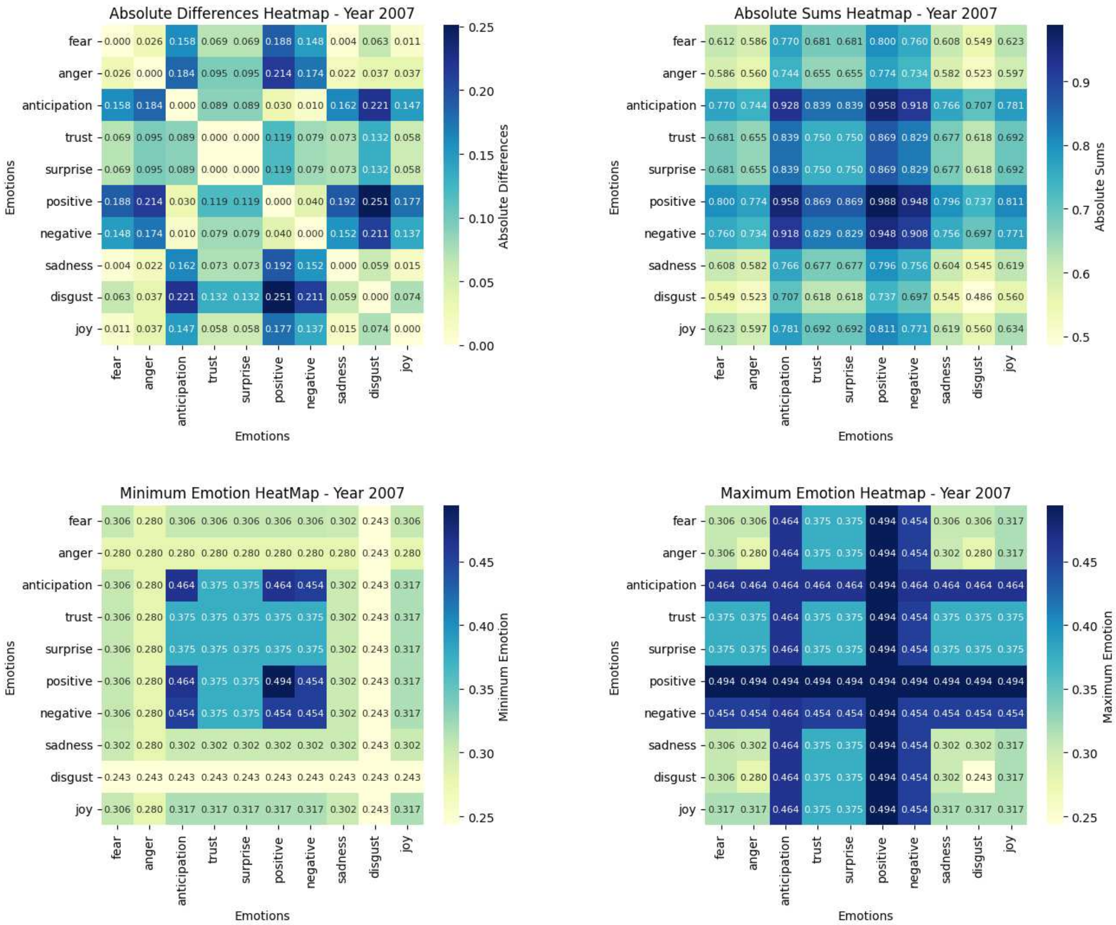

Figure 6.10: 4 Heatmaps of the Average Emotions

6.2.4 Emotion Co-occurrences

Correlation Matrix: Exploring Emotional Relationships

The section Correlation Matrix: Exploring Emotional Relationships, the provided matrix embarks on a compelling journey to unravel the intricate connections between sentiment and individual emotions in popular music lyrics. By calculating the correlation matrix for each year, the code presents a visual representation that vividly illustrates how emotions coalesce or diverge about one another. The resulting heatmap serves as a graphical servant, offering an illustration of the complex emotional interplay within song lyrics. This exploration forms a cornerstone for understanding the emotional landscape of music through a quantitative lens.

With the calculation of the correlation matrix, the relationships between emotions become easily visualized and comprehensible. Taking the example of the correlation matrix for 2002, the main diagonal consistently shows a value of 1.00, as expected since these tiles represent the relationship of an emotion with itself. Examining the rest of the matrix, the highest correlation is observed between the emotion 'surprise' and the sentiment 'positive,' with a value of 0.75. This correlation aligns with the intuitive understanding that surprises are generally associated with positive emotions.

The second-highest correlation is between anticipation and fear, measuring 0.71. This raises an intriguing psychological question: "Are we afraid of what we expect?" This correlation suggests a significant connection between the emotions of anticipation and fear, pointing toward the complex interplay between expectation and anxiety.

Conversely, the two lowest correlations are observed between fear-joy and disgust-joy, with values of -0.44 and -0.39, respectively. These findings indicate a notable contrast between the theme 'joy' and the concepts of 'fear' and 'disgust,' supported by the negative correlations observed in the study. This nuanced exploration of emotional relationships through correlation matrices enhances our understanding of how sentiments and individual emotions intersect, offering valuable insights into the intricate emotional fabric of music lyrics over the years.

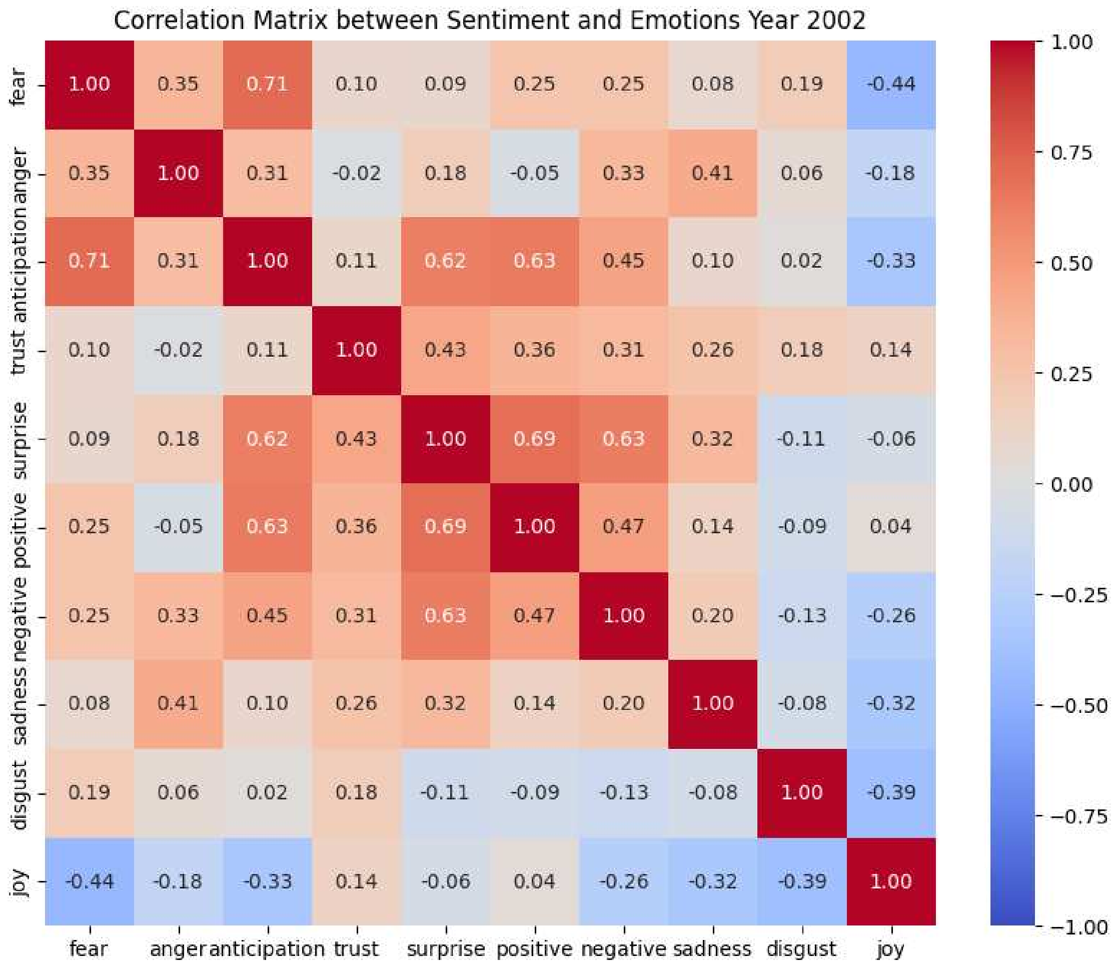

Figure 6.11: Correlation Matrix of the Average Emotions in the year 2002

Sentiment versus Valence

In the exploration of Sentiment vs. Valence, we delve into the intricate relationship between sentiment and valence in popular music lyrics. The graph demonstrates the correlations between positive and negative sentiment scores and specific emotions. This analysis is important because it clarifies how sentiments relate to the valence dimension. By this analysis, which emotions express the general positive or negative is demonstrated.

A bar chart is presented to measure the correlation between two sentiments and emotions for the year 2007. Similar to the correlation matrix, the highest value in the bar chart is the correlation of sentiment itself, with a value of 1. Additionally, the positive sentiment exhibits a strong correlation with the emotion 'surprise,' having a value of 0.76, followed by 'anticipation' with a value of 0.71. The least correlated emotions with positive sentiments are 'disgust' with -0.37 and 'anger' with -0.15.

Conversely, in the negative sentiment category, 'surprise' and 'anticipation' also have the highest correlation values, with 0.61 and 0.43, respectively. 'Disgust' remains the least correlated emotion with a value of -0.38. Two noteworthy observations can be made: Firstly, the strength of the correlation is based on the co-occurrence of words; thus, emotions that occur more frequently together with sentiments demonstrate stronger relationships. Secondly, although 'surprise' and 'anticipation' are the highest correlated emotions with both positive and negative sentiments, their correlation power is significantly stronger with positive sentiments. This is exemplified by the values of 0.76 versus 0.61 for 'surprise' and 0.71 versus 0.43 for 'anticipation.' The slight difference in the correlation with 'disgust,' -0.38 for positive and -0.37 for negative, further supports this observation. This analysis provides insights into the correlation between sentiments and emotions. By this analysis, the understanding of the cooccurrence between positive and negative sentiments and specific emotions enhanced

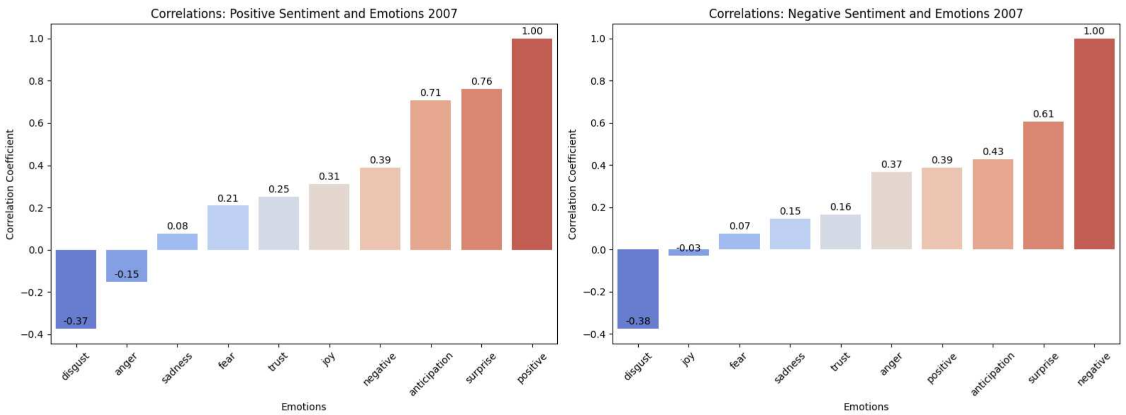

Figure 6.12: Correlation of the Binary Sentiments

6.3 Network Integrated Analyses

6.3.1 Network Density Analysis

Network density is crucial in graph theory and is calculated based on the connections of a point. It is calculated by measuring the ratio of actual edges to the total possible edges. Represented as a decimal or percentage, higher network density signifies a more interconnected and robust structure. This metric is invaluable for understanding the efficiency and resilience of a network, as denser networks tend to facilitate quicker information flow and are more resistant to disruptions. By analyzing network density, researchers gain insights into the cohesion and complexity of social, biological, or technological systems, aiding in the identification of key nodes and potential vulnerabilities crucial for strategic decision-making and system optimization.

The examination of network metrics reveals notable trends. Post-2010, network density experienced a remarkable upswing, excluding a dip in 2016. Conversely, the number of nodes witnesses a sharp decline after 2010. Plausible explanations for this phenomenon emerge. The analysis underscores a pronounced disparity in unique word usage between the first and second decades of this century. The dynamic evolution of the music industry, potentially marked by genre dominance, could contribute to the observed decrease in unique word usage. Additionally, to enhance precision, edges with values below 10 were excluded, focusing on highly occurring pairs. This, however, may entail the omission of a substantial dataset. In summation, the substantial surge in network density during the second decade signifies the ever-changing landscape of the music industry.

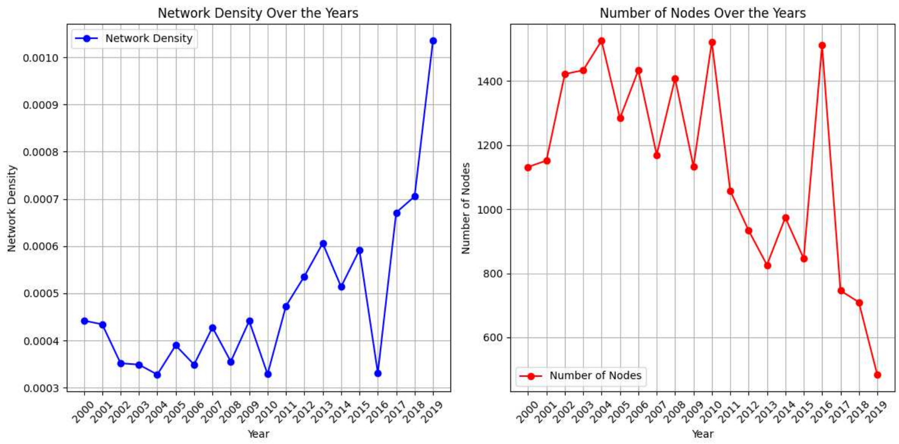

Figure 6.13: Network Density and Node Numbers Graphs

6.3.2 Network Assortativity and Metrics Exploration

Within the fascinating field of music network analysis, the complex interactions between emotional patterns and network structure are highlighted. The section "Emotional Trends in Network Structure: Exploring Degree Distributions" begins a more in-depth investigation into how degrees have changed within music networks over several decades. Log-binning is used by the algorithm to interpret the degree distribution and extract the frequency of emotional connections in the network. A year is represented by each subplot, which displays the degree frequencies' logarithmic transformation as well as the ensuing linear regression lines. These visualizations explain how emotional expressions become entangled in the complex network of musical connections throughout time, illuminating not just the emotional dynamics inherent in song interactions but also the broader patterns.

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In the context of the provided linear regression formulas for each decade, the dependent variable is represented by the slope of the regression line, and time (in decades) serves as the independent variable.

$$y = mx + b$$

The linear regression formula takes the form: where y is the dependent variable, x is the independent variable, m is the slope of the line, and b is the y-intercept. In this case, the slope, m, represents the rate of change of the dependent variable over time. A negative slope indicates a decrease in the dependent variable as time progresses, while a positive slope signifies an increase. The regression formulas provided for each decade include the slope, m, and a constant term, b.

The slopes are calculated using statistical methods to minimize the difference between the observed values and the values predicted by the regression line. The negative value of the slope indicates a negative trend, suggesting a decreasing pattern over the years In the analysis of the linear regression formulas representing the trends over the first 5 decades (2000-2005) compared to the last 5 decades (2015-2019), a notable evolution is observed. The average slope of the regression lines, signifying the rate of change in the dependent variable concerning time, exhibited a shift from approximately -0.86 in the earlier period to around -1.02 in the more recent years. This transition suggests a heightened pace of decline in the latter half of the dataset, indicative of a potentially accelerated negative trend. The steeper negative slope implies an increased rate of change, pointing towards factors contributing to a more pronounced decrease in the dependent variable. The evolving slope values underscore the dynamic nature of the underlying phenomena, emphasizing the importance of temporal considerations and potential shifts in influencing factors over the examined decades. This observation contributes valuable insights into the changing dynamics of the dataset and prompts further exploration into the specific contextual factors driving these.

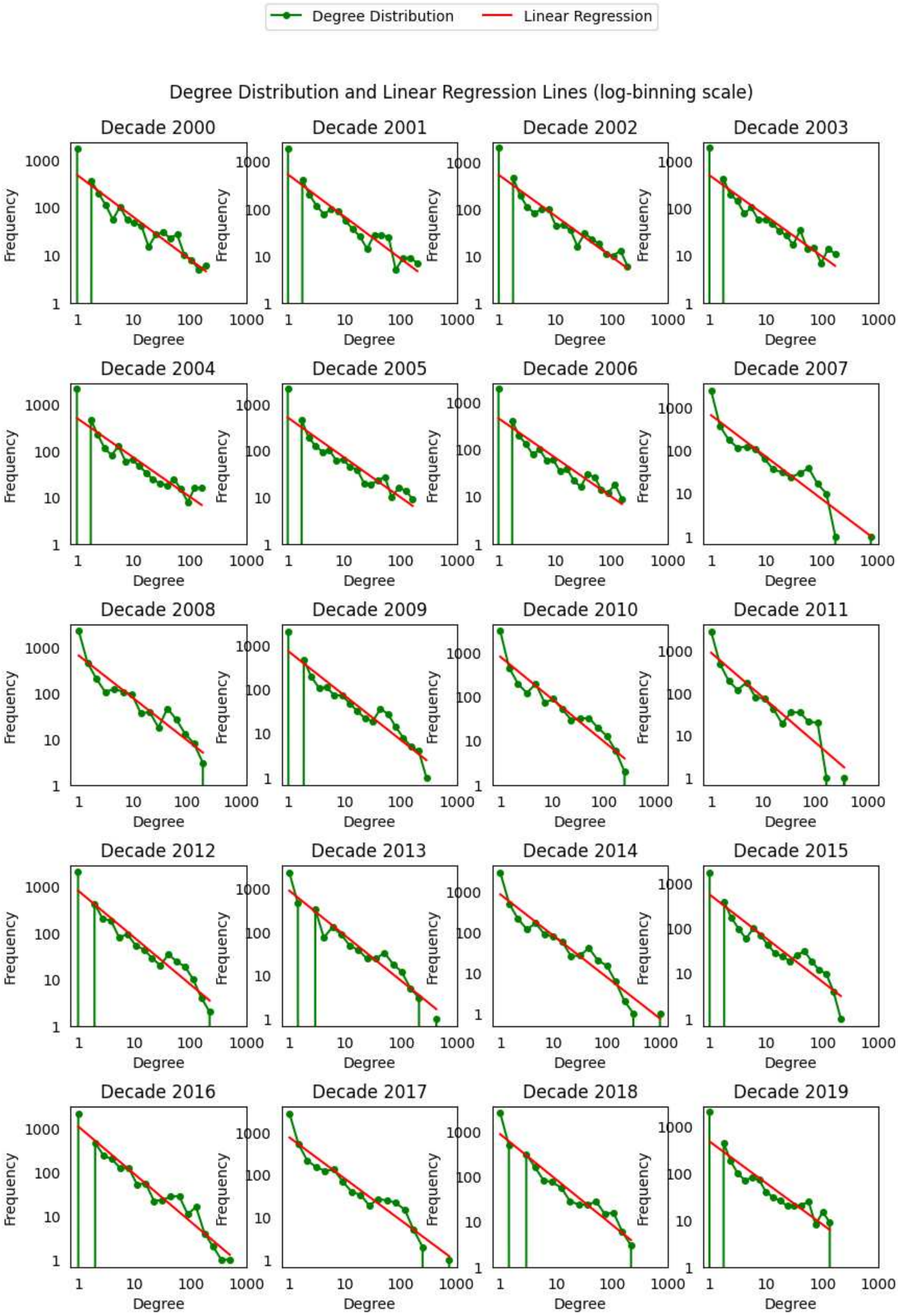

Figure 6.14: Degree Distributions over Years

Figure 6.15: Slope Values of the Power Law over Years

6.3.3 Temporal Evolution of Network Overlaps: A Venn Diagram Analysis

In the pursuit of understanding the dynamic evolution of network structures over time, a comprehensive analysis was conducted utilizing a Venn diagram approach. The code, situated within the context of the title Temporal Evolution of Network Overlaps: A Venn Diagram Analysis focuses on the interplay of nodes among distinct years (2000, 2009, and 2018) in the provided network data. Through the implementation of the `'matplotlib_venn'` library and leveraging three individual network graphs corresponding to each year, the code discerns overlapping nodes and reveals intricate relationships. The resulting Venn diagram distinctly delineates the nodes exclusive to each year, as well as the shared nodes between pairs of years and the triple overlap across all three temporal snapshots. This visual representation provides a nuanced understanding of the changing network composition, shedding light on nodes that persist across multiple years and those that emerge or recede over time. Such insights contribute to unraveling the temporal dynamics inherent in the network structures under investigation. A triple Venn diagram illustrating detected nodes is presented below, focusing on the years 2000, 2009, and 2018. Specifically, 18 nodes were unique to 2000, 23 to 2009, and 9 to 2018. A noteworthy observation emerges concerning the exclusive usage of explicit language; swear words commencing with 'b' 's' 'n' or 'f' appear solely in 2018, absent in 2000 or 2009. This raises an intriguing question: 'Are censorship laws loosening over the years, or is society becoming coarser?' The findings prompt an exploration into evolving societal norms and legislative frameworks.

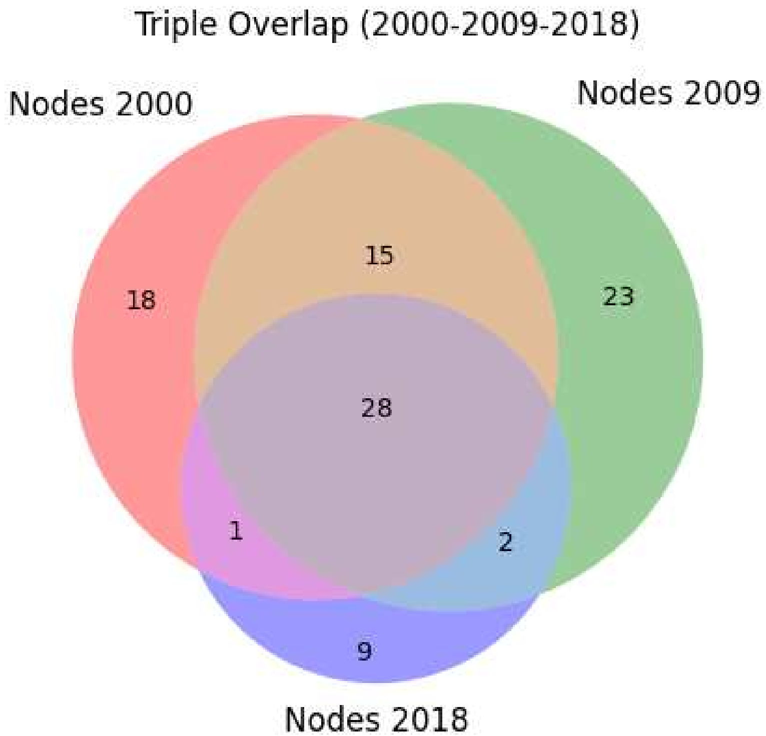

Figure 6.16: Venn Diagram of Overlapping Nodes

```
Year 2000 Nodes: {'long', 'want', 'need', 'around', 'time', 'could', 'try', 'away', 'feel', 'thing', 'back', 'think',
Year 2009 Nodes {'move', 'ta', 'fall', 'want', 'need', 'around', 'hey', 'time', 'could', 'away', 'stop', 'try', 'feel'
Year 2018 Nodes {'still', 'fall', 'want', 'need', 'big', 'around', 'time', 'could',       ,       , 'find', 'feel',
All Years Overlapping Nodes: {'want', 'need', 'around', 'could', 'time', 'feel', 'thing', 'back', 'think', 'leave',
2000 & 2009, not 2018: {'mind', 'turn', 'hear', 'lose', 'hold', 'world', 'away', 'try', 'even'}
2000 & 2018, not 2009: {'friend'}
2009 & 2018, not 2009: {'head', 'fall'}
Only 2000: {'long', 'little', 'hand', 'wo', 'show', 'always', 'would', 'eye'}
Only 2009: {'move', 'ta', 'live', 'ever', 'everything', 'nothing', 'hey', 'boy', 'stop', 'start', 'light', 'tonight'}
Only 2018: {'still', 'sh t', 'god', 'b tch', 'run', 'bad', 'hit', 'big', 'f ck', 'money', 'n gga', 'find'}
```

Figure 6.17: Overlapping and Sole Nodes

6.3.4 Community Detection for Song Clusters

Community detection is a widely used analysis method, based on the grouping of the related elements in a network. In the song clusters, this methodology proves to be a potent tool for revealing inherent patterns, more specifically, related words that are used frequently in the songs. Leveraging community detection algorithms enables the identification of word communities that share meaningful associations. This section, titled as Community Detection for Song Clusters dives into the application of two different community detection methods. Via employment of these two different algorithms, a richer understanding of the intricate relationships within obtained and cleaned raw lyrics.

Leiden Community Detection: Unveiling Structural Patterns

The section Community Detection for Song Clusters employs advanced network analysis techniques to uncover the underlying emotional connections between songs. In this exploration, songs are conceptualized as interconnected nodes, and their emotional bonds are revealed through community detection algorithms. The Leiden algorithm, featured in this analysis, dynamically groups songs into distinct communities based on their emotional similarities. The resulting network visualization represents songs as nodes and emotional connections as edges. As an example, the community detection graph for the year 2010 is implemented below, where 4 communities are detected by the Leiden algorithm. The graph exhibits a distinct visual organization, with node values arranged in decreasing order from left to right, creating a clear hierarchy within the structure. Additionally, the varying sizes of nodes, determined by their respective values, contribute to the overall emphasis on key elements within the graph. Notably, the communities, identified from 1 to 4, are systematically grouped, adding an intuitive layer to the visualization. This deliberate arrangement aids in conveying a comprehensible representation of the network, highlighting both the importance of individual nodes and the cohesive relationships within each community.

Further analysis reveals the notable nodes within each community:

- **Community 1:** Never, Time, Love, Back
- **Community 2:** Think, Could, Still, Day
- **Community 3:** Want, Girl, Put, Man
- **Community 4:** Look, Life, Turn, Little

The identified nodes within each community offer distinct patterns, painting a vivid picture of the thematic content within each group. In Community 1, encompassing nodes such as "Never," "Time," "Love," and "Back," a prevalent theme emerges, featuring melancholic love songs that delve into the complexities of relationships. Moving to Community 2, nodes like "Think," "Could," "Still," and "Day" collectively form a cluster of songs reflecting contemplation and a sense of lingering emotions. In Community 3, where nodes like "Want," "Girl," "Put," and "Man" are prominent, a collection of regretful tunes unfolds, exploring themes of longing and missed opportunities. Lastly, Community 4, with nodes like "Look," "Life," "Turn," and "Little," embodies a cluster of songs with a diverse range, potentially capturing narratives of introspection, transformation, and the intricacies of life's journey. This detailed breakdown unveils the emotional tapestry woven within each community, providing valuable insights into the nuanced relationships and diverse themes present in this musical landscape.

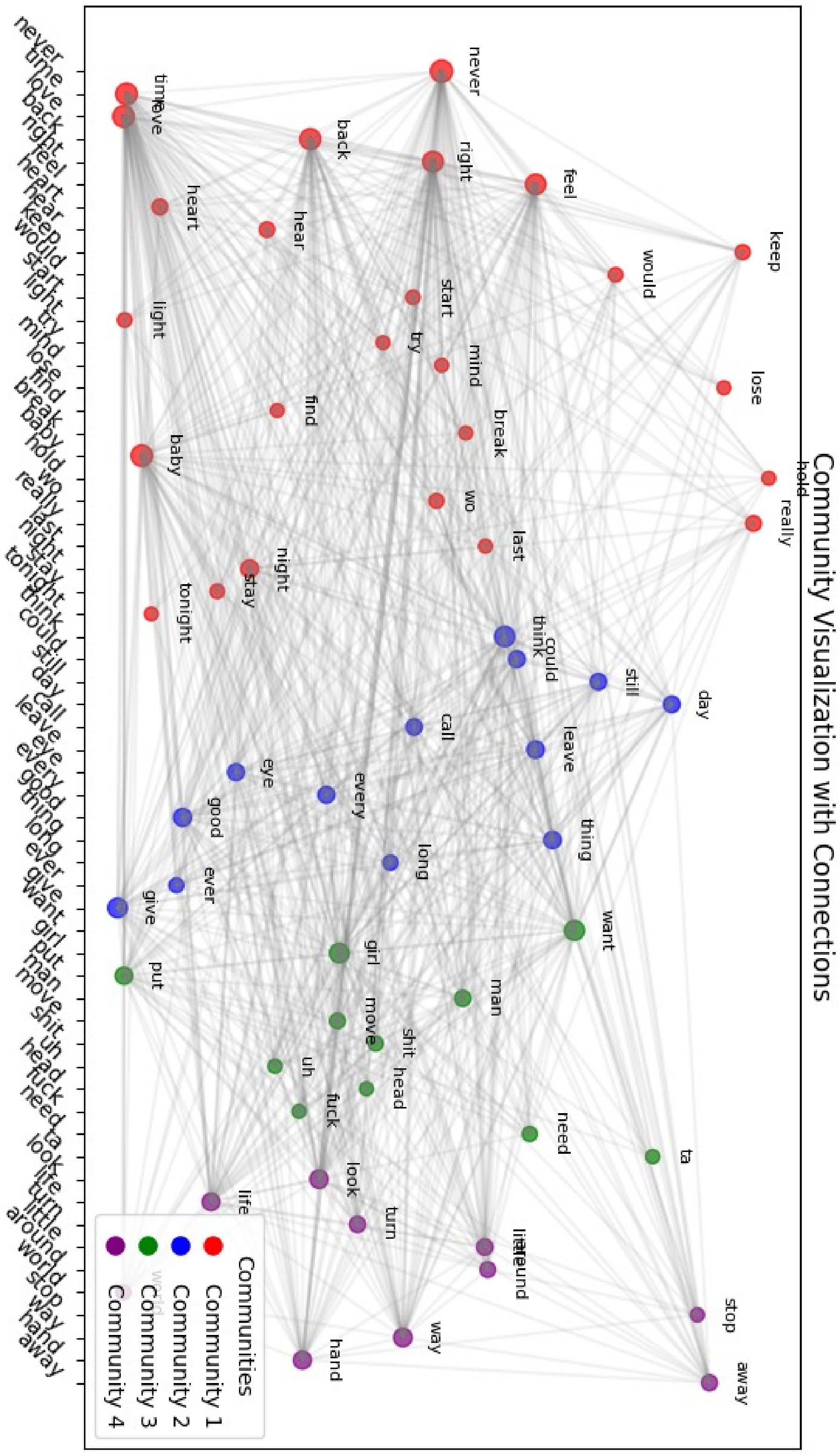

Figure 6.18: Community Detection Graph

Emotional Clustering of Song Lyrics: Visualizing Song Similarity

In this comprehensive exploration of song lyrics, sophisticated techniques are employed to reveal underlying patterns and emotional similarities within the intricate landscape of lyrical content. With the help of Word2Vec embeddings, words are transformed into numerical vectors. The application of t-SNE (t-Distributed Stochastic Neighbor Embedding) plays a pivotal role in this analysis. This algorithm facilitates the reduction of high-dimensional embeddings, in this case, 100 dimensions, into a two-dimensional space. These two dimensions, denoted as (x, y) coordinates, provide a visually interpretable representation of emotional clusters within the song lyrics.

For a clearer understanding, we exemplify this process using the song "03 Bonnie and Clyde" by Jay-Z, leveraging only one song for simplicity. The t-SNE algorithm, influenced by parameters such as perplexity and the number of iterations, transforms the Word2Vec embeddings into a plot where each word is represented by a distinct (x, y) coordinate pair. The resulting graph visually illustrates the spatial relationships between words, with neighboring points indicating lyrically akin content.

In the broader context of cluster detection, our analysis encompasses a subset of 10 songs, yielding the identification of 5 distinct emotional clusters. The main thematic focus of each cluster offers a glimpse into the diverse emotional landscapes captured within the song lyrics. Cluster 1, characterized by themes of "soldier," "Mercedes," and "dudes," reflects a narrative centered around resilience and a vibrant lifestyle. Cluster 2 delves into themes of "girlfriend," "lack," and "whatever," embodying a spectrum of emotions related to relationships and personal struggles. Meanwhile, Cluster 3, with themes like "girl," "sex," and "focus," centers on themes of romance and desire. Cluster 4, incorporating terms such as "working," "necessary," and "breathe," reveals a thematic emphasis on perseverance and life's essentials. Lastly, Cluster 5, with themes like "work," "shoulder," and "treat," explores a range of emotions related to professional life and interpersonal dynamics.

It is crucial to note that the outcomes of cluster detection are inherently dependent on the parameters selected during the analysis. In this case, the choice of parameters, including the number of clusters and the clustering algorithm, significantly influences the thematic interpretation. To delve deeper into the nuanced meanings, dictionary-based applications are applied. Using several community detection algorithms allows for a richer understanding of the emotional relationships and thematic nuances inherent in the song lyrics. As a result, 5 different clusters are obtained.

In addition to cluster detection, an outlier detection approach is employed. Identifying the words that do not align with specific emotional clusters is crucial since outliers indicate many insights. Noteworthy outliers include words such as ”wheel,” ”today,” ”city,” ”soon,” and ”mami.” These outliers pose challenges in the context of machine learning-based approaches, as they either serve as slang terms embedded in the song’s language or carry metaphorical meanings that transcend traditional machine learning interpretations. For instance, while ”wheel” conventionally refers to a tool for vehicles, its metaphorical usage, especially in conjunction with ”Mercedes Benz,” introduces layers of meaning beyond the literal interpretation. This highlights a limitation in cluster determination, as the ever-evolving nature of lyrical language poses challenges for detection algorithms to swiftly adapt to rapid linguistic changes. This acknowledgment emphasizes the importance of considering evolving language trends and potential metaphorical nuances in song lyrics during the cluster analysis process.

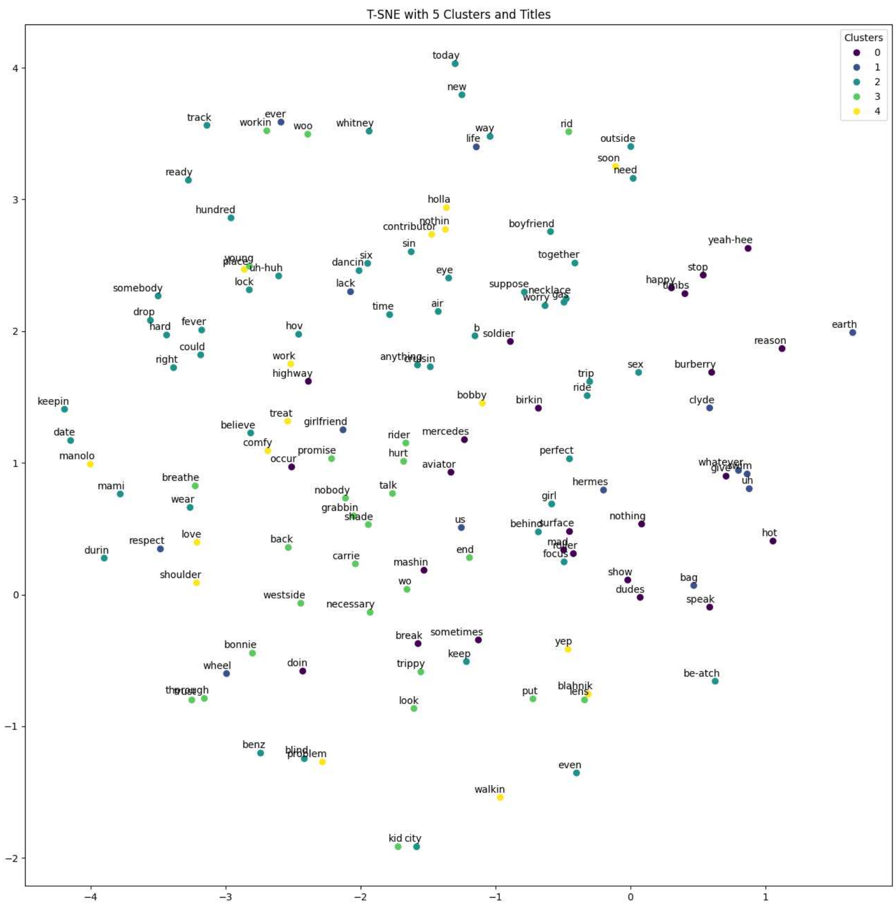

Figure 6.19: Community Detection with T-SNE

6.3.5 Assessing Clustering Quality: Silhouette Analysis

Silhouette Analysis is a technique for assessing the quality and cohesion of clusters. This analysis is used to measure overall clustering quality during the clustering process. The ranking is made by measuring the compactness and separation of clusters. The silhouette score ranges from -1 to 1, where a high positive score indicates that the data point is well-matched to its cluster and poorly matched to neighboring clusters, implying a robust and well-defined cluster. Conversely, a low or negative silhouette score suggests that the data point may be assigned to the wrong cluster or that overlapping and poorly separated clusters exist. Silhouette Analysis offers a thorough metric to assess the overall appropriateness and quality of the clustering technique by computing silhouette scores for every data point and averaging their overall points. This method helps determine the ideal number of clusters to create and guarantees that the final clusters accurately represent the underlying patterns and relationships in the dataset.

The obtained average silhouette score for the clustering process is -0.0628, suggesting scope for improvement in achieving more well-defined and distinct clusters. It is essential to note that the clustering task is inherently challenging, and the negative silhouette scores across clusters highlight areas where enhancements can be made. Cluster 1 and Cluster 2 exhibit the lowest silhouette scores at -0.0972 and -0.0991, respectively, indicating opportunities for refinement in achieving better internal cohesion within these clusters. Moreover, Cluster 0, Cluster 3, and Cluster 4 also resulted in negative silhouette scores. The adjustments of parameters and algorithms might lead to better silhouette scores by enhancing the overall quality of the clusters. This initial analysis lays a basis for further refinement. Acknowledging the complexities involved is beneficial for optimizing the clustering algorithm, both from a developer and researcher perspective. The potential for optimization might lead to more meaningful insights from the song lyrics dataset.

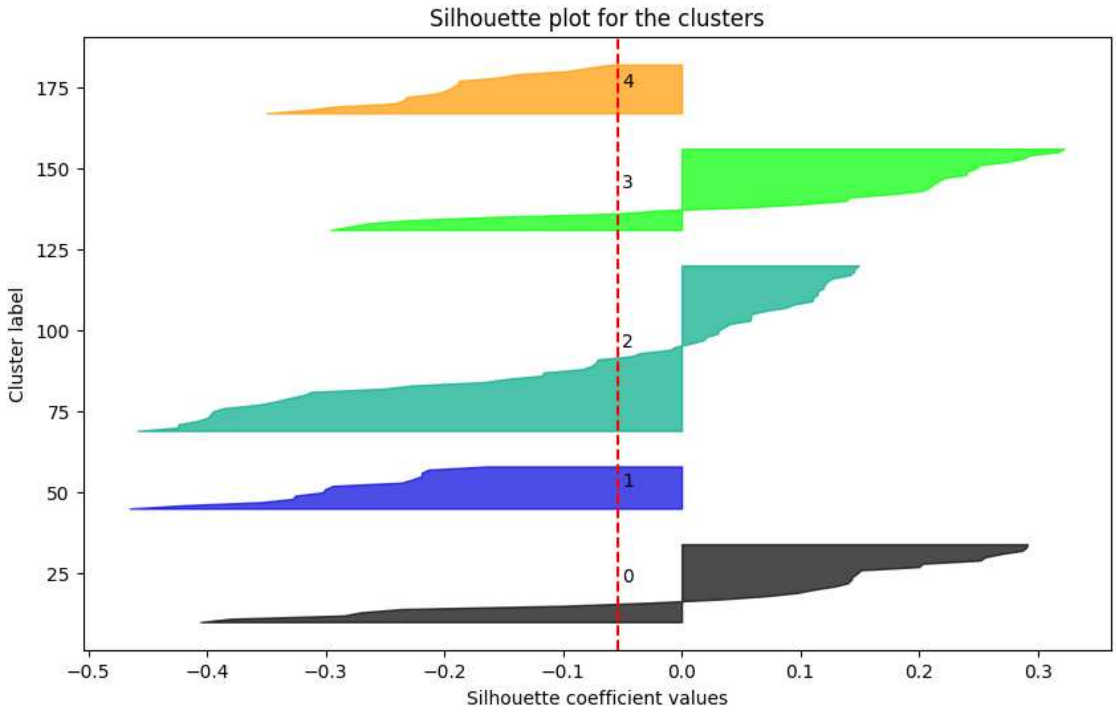

Figure 6.20: Silhouette Analysis Graph

6.3.6 WordClouds: An Colorful Visualization

Word clouds provide a visual representation of the most frequently occurring words. In this section, word cloud representations are constructed to demonstrate the most prominent words. The font size is based on the number of nodes, which is terminated by iGraph as it did in previous sections. Word clouds transform textual data into visually striking displays by assigning greater font size to words that appear more frequently. This approach eases the observation of recurring motifs, and distinctive vocabulary within the song lyrics. The resulting word cloud serves as a mesmerizing snapshot.

The word cloud is meticulously crafted by prioritizing the most frequently occurring words, allowing us to witness the evolving linguistic trends across different years. To illustrate, let's consider the word clouds for the years 2008 and 2010. Only 2 years' word cloud is illustrated below for simplicity. In 2008, the largest nodes included 'right,' 'time,' and 'love,' closely trailed by 'baby,' 'way,' and 'never.' These predominant words paint a vivid picture of the prevalent themes within the songs of that year. Fast forward to 2010, and we observe a notable shift with 'never' taking the lead, followed by 'time' and 'love.' The rest of the nodes, although slightly smaller, bear witness to the dominance of these three central themes. Via this transformation, it is proven that word clouds are one of the strongest visual materials for the determination of the predominant themes. For further investigations, the rest of the word clouds can be constructed easily with the same approach, by changing the dataset.

Figure 6.21: WordCloud of the Nodes 2008

Figure 6.22: WordCloud of the Nodes 2010

6.3.7 Network Visualization

Network graphs are powerful tools for visualizing relationships and structures within complex datasets. In the context of song lyrics, nodes represent individual words, and edges denote connections or associations between words. This visualization technique is widely used for the demonstration of complex webs into connection trackable graphs.

In our case, the construction of the network graph involved loading an edge list and node list from the song lyrics dataset. Each edge in the graph is associated with a weight, reflecting the strength of the connection between two words. for community detection, the Leiden method is employed

In terms of visual representation, the node sizes in the graph were determined by the frequency of each word in the dataset, emphasizing words with higher occurrence. Node colors were assigned based on their values, with darker colors indicating greater values. This color scheme highlights the significance of nodes with larger sizes, providing a visual cue for nodes with higher frequency. The observed results, exemplified using data from the year 2000, reveal prominent nodes within the network graph. Keywords such as ”love,” ”never,” ”time,” ”baby,” and ”feel” stand out with notable frequency, emphasizing their central role in song lyrics and underscoring their significance within the dataset. For instance, the word ”love” has a substantial value of 70, followed by ”never” with 53 occurrences and ”time” with 52 occurrences. Additional noteworthy terms contributing to the lyrical landscape include ”baby” (47), ”feel” (44), ”think” (42), ”look” (42), ”way” (41), ”right” (41), ”could” (40), ”thing” (39), and ”want” (38). It's crucial to acknowledge that these findings are based on the year 2000 data for simplicity. The potential to extend this analysis to subsequent years (2001-2019) offers an opportunity to explore evolving patterns and trends in song lyrics over the specified time frame.

Excluding nodes with values below 20, we focus on illustrating predominant nodes, ensuring a clearer representation of significant words within the network graph. Via this criteria, the significance of the highly occurred words is enhanced. Noteworthy examples of less occurring words include ”lose,” ”hear,” ”hand,” ”even,” ”put,” ”turn,” and ”always.”, ranging from 20 to 22. This filtering contributes to a more insightful analysis of the lyrical content.

Via illustrating an example network graph, the importance of visually understandable graphs is underscored. The interconnection of the lyrical content, represented by words as words and lines as connections, is sky-clear with this visualization approach.

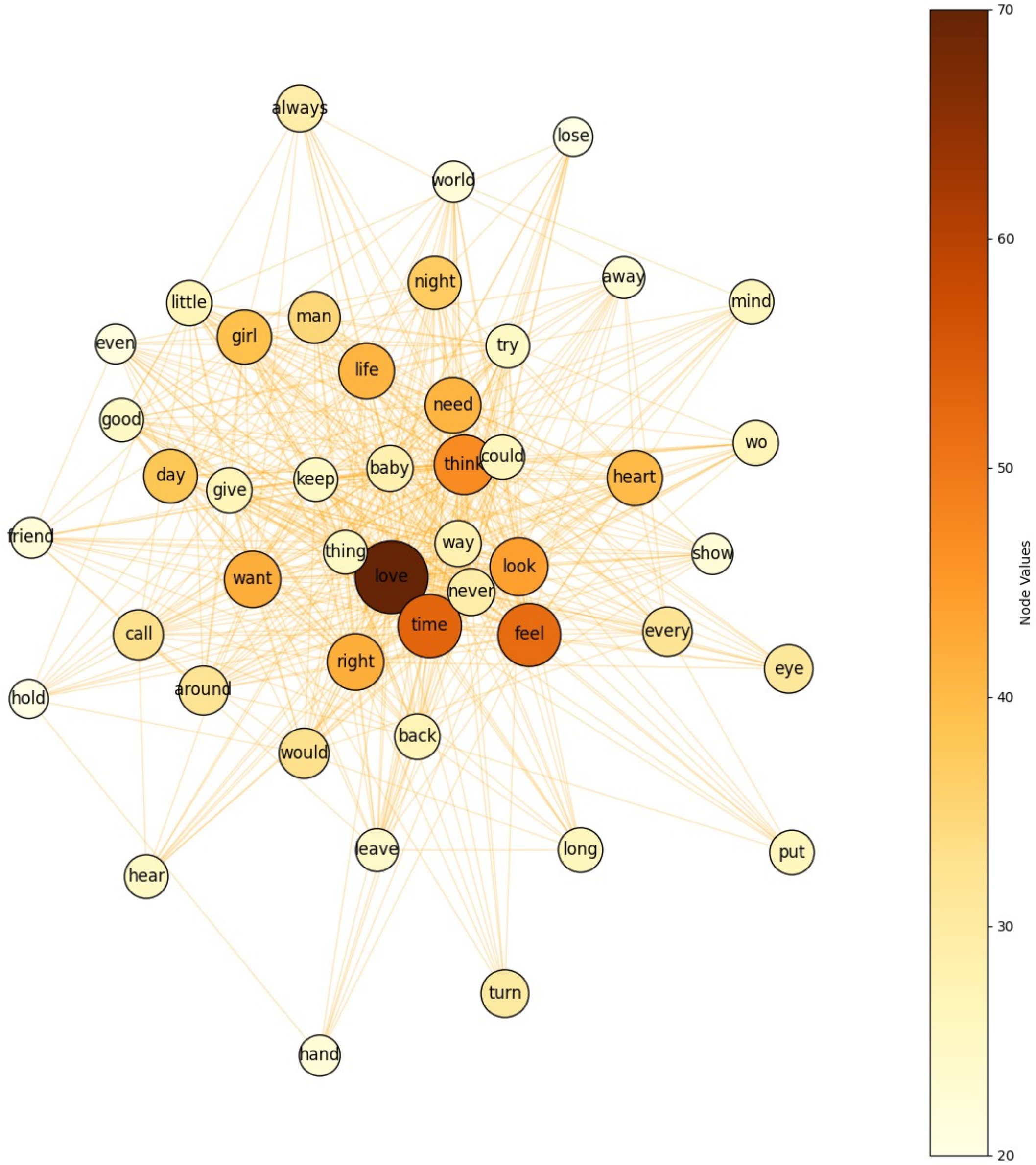

Figure 6.23: Network Graph of the Nodes of the Year 2000

6.4 Psychological Examinations

6.4.1 Wheeel Visualization for Clarity

In examining the emotional evolution of song lyrics between the years 2007 and 2012, significant patterns and shifts across Plutchik's emotion wheel categories have been identified. The analysis reveals subtle but noteworthy changes in the emotional landscape of the lyrical content. Notably, there is a slight reduction in overall positive emotions, with joy scores experiencing a modest decline. Trust, another positive emotion, exhibits a decrease, indicating potential shifts in perceived reliability or trustworthiness within the lyrics. Conversely, there is a marginal increase in fear, suggesting a subtle trend towards more emotionally intense or apprehensive themes. The theme of surprise sees a reduction, indicating a potential decrease in unexpected or surprising lyrical elements. The emotion of sadness experiences a noteworthy decrease, hinting at a shift towards more positive or uplifting lyrical themes. Disgust, related to emotionally repulsive elements, also shows a decline. The emotion of anger has remained stable over the years. Anticipation witnesses a subtle increase, indicating a rise in themes related to anticipation or expectation

In scrutinizing the emotional trajectory of song lyrics from 2007 to 2012, discernible alterations in Plutchik's emotion wheel categories are evident. A nuanced analysis reveals intriguing nuances in the emotional tapestry of lyrical content during this timeframe. Notably, the overall emotional tone appears to undergo a subtle shift, characterized by a decrease in joy and trust. Joy experiences a decline, indicating a potential departure from exuberant or jubilant lyrical expressions. Similarly, trust exhibits a modest reduction, hinting at a potential evolution in the perceived reliability or trustworthiness portrayed in the lyrics. Fear, on the other hand, sees an increase, suggesting a potential trend towards more emotionally intense or apprehensive themes. The emotions of surprise and sadness both witness marginal decreases, pointing towards a potential shift away from unexpected or melancholic lyrical elements. Disgust, associated with emotionally repulsive elements, experiences a decline. Meanwhile, anger remains relatively stable across the examined years. The emotion of anticipation undergoes a subtle increase, suggesting a rise in themes related to anticipation or expectation. This comprehensive examination offers valuable insights into the nuanced interplay of emotions within song lyrics, shedding light on the evolving emotional landscape in the realm of music between the years 2007 and 2012. The year selection for the wheel chart in 2007 and 2012 is random, for the sake of simplicity. Only values from these two years are demonstrated below, and they are commented on for this section. The year 2007 is colored blue, and 2012 is colored red. Overlapped values are in burgundy color.

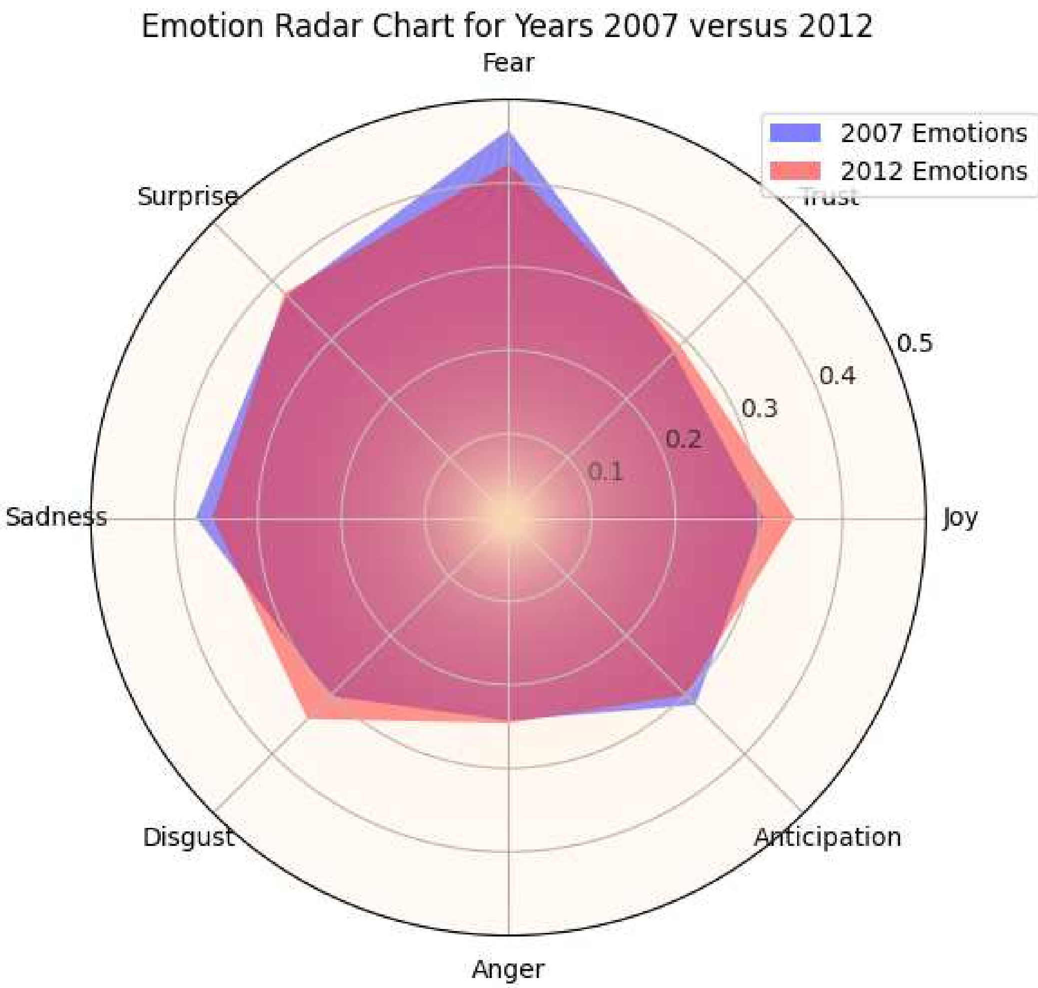

Figure 6.24: Radar Graph for the years 2007 and 2012

6.4.2 Key Events Influences

In this section, major events occurred in the data range and their effects are discussed. The four emotions are selected for this discussion: anticipation, fear, anger, and disgust. These emotions are selected for several reasons. Firstly, as the result has investigated, these emotions are the emotions which have sharp peaks and valleys. Secondly, these emotions are affected by selected events. Selected events have more triggering effects for these emotions rather than others, for instance, protests trigger hatred and elections trigger both anticipation and also fear of change. Analyzing anticipation unveils expectations and hope amid uncertainties, while fear, anger, and disgust shed light on the emotional responses triggered by major global events. Based on data range, 5 major events are selected which are: 9/11, the Iraq War, the 2008 economic crisis, the Black Lives Matter movement, and the Trump election.

Anticipation

Anticipation can be defined as the emotion that is expectation towards the future, very simply. In the aftermath of the 9/11 attacks in 2001, anticipation scores exhibit a decline, reflecting a period marked by heightened fear and uncertainty. The subsequent Iraq War (2003-2005) sees a gradual increase in anticipation values, potentially signaling a shift toward expectations or hope amid ongoing geopolitical turmoil.

The economic crisis in 2008 introduced a notable decrease in anticipation, aligning with the overall apprehension and lack of optimism during the global financial downturn. However, anticipation values rebounded during the Black Lives Matter movement in 2013, suggesting a renewed sense of hope or anticipation for social change. The 2015-2016 Trump election period witnessed a mixed response, with anticipation values showing fluctuations, possibly influenced by the uncertainty surrounding political shifts.

These variations in anticipation scores highlight the connection between major societal events and the anticipation and sociological expectation trends, which can be further investigated as are we getting more depressive or the expectation trends of communities.

Fear:

Examining fear values in song lyrics across major historical events reveals intriguing patterns reflective of societal dynamics. In the aftermath of the 9/11 attacks in 2001, fear scores experienced a noticeable surge, illustrating the widespread anxiety and uncertainty following the unprecedented events. The subsequent Iraq War (2003-2005) maintains elevated fear values, indicating the sustained impact of geopolitical tensions on emotional expression.

The economic crisis in 2008 introduced a complex emotional response, with fear values exhibiting fluctuations. This suggests the multifaceted influence of economic downturns on societal sentiments, encompassing both anxiety and resilience. The fear scores during the Black Lives Matter movement in 2013 show a dip, reflecting a potential shift towards themes of social justice and activism.

However, the 2015-2016 Trump election witnesses an upturn in fear values, suggesting heightened apprehension during this politically charged period. These variations in anger scores underscore the dynamic interplay between the perception of fear and stimulus. Further research based on the 'fear' concept can be referred to based on this interplay, especially on nation-based research.

Anger:

Analyzing anger values in song lyrics across major historical events provides insights into the emotional landscape. In the aftermath of the 9/11 attacks in 2001, there was a notable increase in anger scores, reflecting the collective outrage and frustration during that period. The subsequent Iraq War (2003-2005) maintains elevated anger values, echoing the prolonged and contentious nature of the conflict.

The economic crisis in 2008 introduced a nuanced emotional response, with anger values showing a fluctuating pattern. This likely mirrors the complex and multifaceted impact of the financial downturn on societal sentiments. In 2013, coinciding with the Black Lives Matter movement, there is a dip in anger scores, suggesting a potential shift toward themes of social justice and activism.

However, the 2015-2016 Trump election marks a significant upturn in anger values, indicating heightened emotional intensity during this politically charged period. Within general patterns of anger, also a social media investigation research or polls can be done, not just only on the music realm, but also on different variations of other interactive multimedia.

Disgust:

Analyzing the disgust values in song lyrics about significant events provides insights into the emotional response of lyrical content to key moments in recent history. The aftermath of the 9/11 attacks in 2001 is reflected in a notable increase in disgust values, indicating a potential emotional reaction to the collective trauma and upheaval caused by the events. The Iraq War (2003-2005), characterized by the horrors of war and the emotional toll it takes on society, corresponds to a period of heightened disgust and values.

The Black Lives Matter movement's emergence in 2013, advocating for racial justice and equality, coincides with fluctuations in disgust values. This may reflect the societal tension and emotional response to issues of systemic racism and social injustice. The 2015-2016 Trump election, a period marked by divisive political discourse, also shows nuanced changes in disgust scores, possibly linked to the polarizing nature of the election and its impact on public sentiment.

In summary, the analysis of disgust values in song lyrics aligns with major historical events, indicating the emotional resonance of music with societal experiences, trauma, and socio-political developments. The fluctuations in disgust scores suggest a dynamic interplay between global events and our reaction to them.

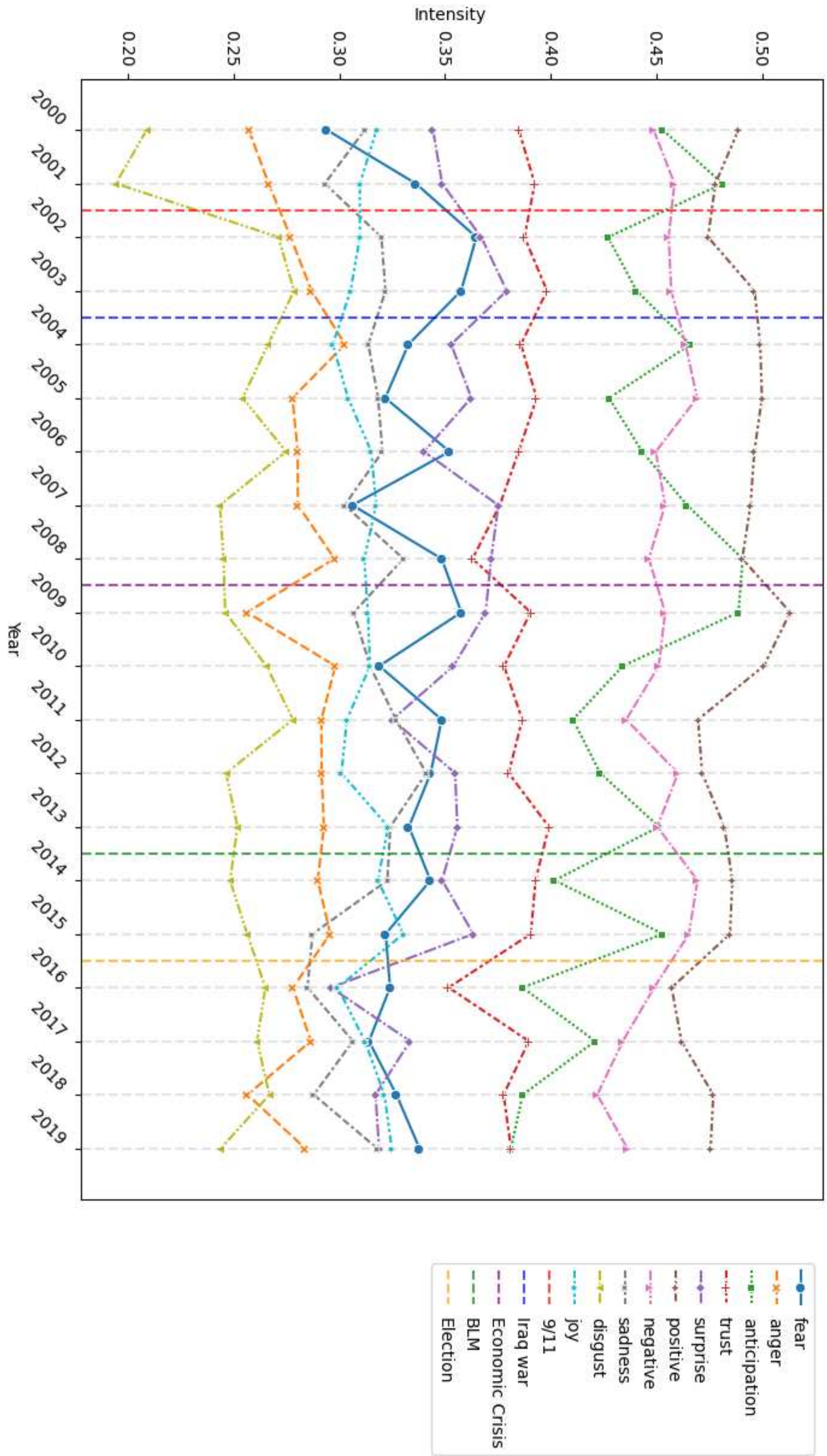

Figure 6.25: Events Marked Emotion Averages over the Years 2000-2019

6.4.3 Commenting Result with Psychology Theories

As it is explored, the historical events have massive effects on the emotion fabric of the lyrics, especially on the emotions anticipation, fear, anger, and disgust. This multidisciplinary exploration is guided by Scherer's Component Process Model, Thayer's Biopsychological Model, and Leonard Meyer's Theory. These three theories are selected as each one of them are providing nuanced perspectives on the cognitive, physiological, and musical dimensions the emotions.

Scherer's Component Process Model:

Serving as our compass, Scherer's Component Process Model unveils the cognitive evaluations underpinning emotional expressions in song lyrics. The post-9/11 decline in anticipation scores underscores the profound impact of the rises in the fear and uncertainty. This rises are aligning with Scherer's emphasis on the interplay between cognitive appraisals and emotional experiences. The explorations made on binary sentiments, as mentioned in the Spectrum of Positivity and Negativity, correlates with Scherer's Component Model.

Thayer's Biopsychological Model:

Thayer's Biopsychological Model sheds light on the physiological underpinnings influencing emotional expressions amid socio-political events. The surge in fear scores post-9/11 and during the Iraq War aligns with Thayer's notion of arousal influencing emotional responses. Nuanced patterns during the Black Lives Matter movement, marked by a dip in fear scores, hint at a potential shift towards themes of social justice. This underscores the intricate relationship between physiological states and emotional expressions, providing a comprehensive understanding of the emotional landscape.

Leonard Meyer's Theory :

Breathing life into the emotional mosaic through NRCLex and the eight fundamental emotions, Leonard Meyer's Theory unveils the ebb and flow of emotions during the 2015-2016 Trump election, mirroring Meyer's concept of tension and resolution. Our intricate emotional analysis, spanning joy to sadness, trust to disgust, reveals a dynamic interplay between evolving societal events and the expressive nature of song lyrics. Exploring binary sentiments through the Spectrum of Positivity and Negativity aligns with Meyer's emphasis on tension and resolution, resonating with Thayer's Biopsychological Model.

Navigating the Peaks of Joy and Valleys of Sadness through Meyer's lens, we recognize the subtle interplay between cognitive evaluations, physiological responses, and musical elements. The prevalence of negative tones aligns with Meyer's theory, mirroring societal trends. Moreover, the prevalence of anticipation raises questions about evolving emotional expectations, like are humans becoming less delusional, creative or more depressive? Via these theories, the layers to the narrative of how musical compositions are added. The question if the musical pieces are carrying any traces from the cognitive appraisals or physiological states from both producers and customers is answered. Answer is, yes, they are.

In summary, this multidisciplinary approach, with the help of Scherer's Component Process Model, Thayer's Biopsychological Model, and Leonard Meyer's Theory, enriches the psychological understanding about song lyrics. It unveils a profound connection between the evolving societal landscape and the expressive tapestry woven by musical artists.

7

Results and Discussion:

7.1 Summary of Analysis Findings:

This research delves into the intersection of data engineering, network science, and psychology to examine the emotional intricacies in popular music over two decades. With three primary objectives, the research aimed to construct a robust sentiment analysis approach tailored to the nuances of evolving musical trends, examine the emotional content of songs within this timeframe and establish links between observed emotional patterns and psychological theories.

Methodologically, our study integrates cutting-edge engineering techniques to comprehensively explore emotional patterns in music. The application of network science brings forth intricate analyses, with notable results in network density and structure. The findings show a substantial surge in network density during the second decade. Network-integrated analyses go deeper, exploring emotional trends in network structure through metrics like network assortativity. This exploration uncovers the complex interactions between emotional patterns and the evolving network, shedding light on the broader dynamics of song interactions.

In retrospect, examining the linear regression formulas delineating trends across the first five decades (2000-2005) in contrast to the last five decades (2015-2019) furnishes a nuanced comprehension of the evolving emotional patterns in music. The computed negative slopes within these formulas symbolize a consistent declining trend over the years, marking a discernible shift. This shift implies a heightened pace of decline in emotional patterns, particularly accentuated in the latter half of the dataset, indicating a potentially accelerated negative trend. The steeper negative slope further suggests an increased rate of change, pointing towards contributing factors that amplify the pronounced decrease in emotional patterns. The evolving slope values underscore the dynamic nature of the underlying phenomena, accentuating the significance of temporal considerations and potential shifts in influencing factors over the examined decades.

Further engineering insights arise from the correlation analyses of sentiments and emotions, enhancing our understanding of the intricate emotional fabric of music lyrics. Correlation matrices vividly illustrate relationships between emotions, offering valuable insights into the complex interplay between sentiments and individual emotions. Noteworthy correlations, such as between 'surprise' and 'positive' sentiment, highlight the nuanced emotional dynamics embedded in music.

The analysis incorporates both cluster and outlier detection methods to gain insights into evolving language trends and potential societal shifts reflected in song lyrics. An intriguing observation emerges regarding the exclusive appearance of explicit language in 2018, prompting questions about evolving censorship laws and societal norms. In addition, outlier detection highlights words that defy emotional clustering, emphasizing the challenges posed by slang terms and metaphorical language in machine learning-based analyses. This underscores the need to consider the dynamic nature of lyrical language and its potential metaphorical nuances when interpreting cluster analysis results, offering valuable insights into evolving societal frameworks and linguistic trends.

The examination of song lyrics across major historical events reveals a significant connection between societal dynamics and musical expression. Patterns in lyrical content demonstrate nuanced shifts in response to key moments, reflecting the complex interplay between external events and emotional expression. Notably, post-9/11, there is a discernible decline followed by subsequent increases during the Iraq War and the 2008 economic crisis, illustrating the impact of these events on lyrical sentiment. The emergence of the Black Lives Matter movement and the 2015-2016 Trump election elicits varied responses, highlighting the dynamic nature of the relationship between global occurrences and the emotional resonance found in music lyrics. This exploration provides valuable insights into the evolving socio-political landscape and collective emotional responses to pivotal historical moments, underscoring the unique role of music as a cultural and emotional barometer.

The findings present a refined sentiment analysis approach adept at capturing the dynamic changes in music. The research identifies emotional patterns and variances in popular songs from 2000 to 2020, unveiling connections with psychological theories. Noteworthy is the incorporation of Scherer's Component Process Model, Thayer's Biopsychological Model, and Leonard Meyer's Theory, among others, to contextualize emotional experiences.

The significance of this research lies in providing insights into the emotional landscape of popular music and its profound impact on individual and collective well-being. By contributing to the understanding of emotional patterns in music over two decades, the study bridges the gap between psychology and music, offering valuable implications for the broader fields of data engineering, network science, and psychology.

7.2 Contribution and Limitations of the Thesis:

This thesis makes notable contributions to the interdisciplinary exploration of emotional patterns in music, laying a robust foundation for future research in data engineering, network science, and psychology. The integration of sentiment analysis, complex network analysis, and psychological theories offers a novel approach, providing a nuanced perspective on the emotional landscape in the contemporary music industry. Additionally, the study delves into linguistic trends, offering insights into evolving societal norms and legislative frameworks.

The comprehensive investigation spans two decades, contributing significantly to our understanding of how music influences and reflects societal emotions. By adopting a multifaceted methodological approach, including literature review, data scraping, sentiment analysis, and complex network analysis, the thesis advances methodologies in emotional analysis in music. The inclusion of outlier detection highlights the need for adaptive techniques in the face of rapidly evolving language trends, pushing the boundaries of current machine learning applications in the field.

However, the study acknowledges certain limitations. The dataset construction, relying on the last day of the year, introduces a potential bias due to the influence of the Christmas theme on song preferences, with approximately 8 percent of the dataset comprising Christmas songs. This may impact the generalizability of results, and while these songs are not excluded to maintain dataset integrity, understanding this bias is crucial for interpreting the findings. Moreover, the dynamic nature of lyrical language poses challenges in cluster determination and outlier detection, emphasizing the need for continuous adaptation of algorithms to stay attuned to evolving language trends. Additionally, the scope of psychological theories is limited, and a broader exploration of additional frameworks could further enrich the analysis. Recognizing these contributions and limitations enhances the study's transparency, and reliability, and inspires future research in emotional analysis within music.

References

[1] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, Netherlands, Aug. 2010.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[3] B. Liu, *Sentiment Analysis and Opinion Mining*, ser. Synthesis Lectures on Human Language Technologies, 2012, vol. 5, no. 1.

[4] M. Rambocas and J. Gama. Marketing research: The role of sentiment analysis. [Online]. Available: https://ideas.repec.org/p/por/fepwps/489.html

[5] B. Krahé and S. Bieneck, "The effect of music-induced mood on aggressive affect, cognition, and behavior," *Journal of Applied Social Psychology*.

[6] T. Hillecke, A. Nickel, and H. V. Bolay, "Scientific perspectives on music therapy," *Annals of the New York Academy of Sciences*, vol. 1060, pp. 271–282, 2005.

[7] R. Malheiro, R. Panda, P. Gomes, and R. P. Paiva, "Emotionally-relevant features for classification and regression of music lyrics," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 240–254, April-June 2018.

[8] D. Throsby, "The music industry in the new millennium: Global and local perspectives," 01 2002.

[9] V. Langley and A. Ahmed, "How global record companies adapt to digitisation in the music industry," 2019. [Online]. Available: https://core.ac.uk/download/287780163.pdf

[10] T. M. Fountain. (2021, September 13) The evolution of the music industry — and what it means for marketing yourself as a musician. Forbes Business Council. [Online]. Available: https://www.forbes.com/sites/forbesbusinesscouncil/2021/09/13/the-evolution-of-the-music-industry---and-what-it-means-for-marketing-yourself-as-a-musician/?sh=7d433a60297a

[11] L. Peralta. (2021, November 3) How does music affect society? Save the Music Blog. [Online]. Available: https://www.savethemusic.org/blog/how-does-music-affect-society/

[12] A. North and D. J. Hargreaves, *Musical preference and taste*. Oxford University Press, 2008, p. 0. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780198567424.003.0003

[13] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Information Processing & Management*, vol. 53, no. 4, pp. 764–779, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457316302205

[14] C. Dhaoui, C. M. Webster, and L.-P. Tan, "Social media sentiment analysis: lexicon versus machine learning," *Journal of Consumer Marketing*, vol. 34, no. 6, pp. 480–488, 2017.

[15] M. Wiering, M. van der Ree, M. Embrechts *et al.*, "The neural support vector machine," in *The 25th Benelux Artificial Intelligence Conference (BNAIC)*, 2013.

[16] N. Mukhtar, M. A. Khan, and N. Chiragh, "Lexicon-based approach outperforms supervised machine learning approach for urdu sentiment analysis in multiple domains," *Telematics and Informatics*, vol. 35, 08 2018.

[17] A. F. N. Ressel and A. L. Name], "Machine learning approaches to recognize human emotions," *Frontiers in Psychology*, vol. 14, no. Sec. Emotion Science, September 20 2023, part of the Research Topic "Machine Learning Approaches to Recognize Human Emotions.".

[18] H. Taherdoost, "Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects," *In-*

ternational Journal of Academic Research in Management (IJARM), vol. 10, no. 1, pp. 10–38, 2021, ffhal-03741847f.

[19] M. Suer. (2023, September 22) What is data quality and why is it important? Alation Blog. [Online]. Available: https://www.alation.com/blog/what-is-data-quality-why-is-it-important/

[20] T. Nuggets. (2017, December 25) 4 ways storage technologies have evolved. CBT Nuggets Blog. [Online]. Available: https://www.cbtnuggets.com/blog/technology/data/4-ways-storage-technologies-have-evolved

[21] Wikipedia contributors. (2024) List of countries by english-speaking population. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population

[22] K. Dashtipour, S. Poria, A. Hussain *et al.*, "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognitive Computation*, vol. 8, pp. 757–771, 2016.

[23] I. Ghosh. (2020, December 28) Visualizing the u.s. population by race. Visual Capitalist. [Online]. Available: https://www.visualcapitalist.com/visualizing-u-s-population-by-race/

[24] Steve. (2023, March 2) How culture impacts musical taste by exploring the endless diversity of music. Eat This Music. [Online]. Available: https://eatthismusic.com/2023/03/02/how-culture-impacts-musical-taste-by-exploring-the-endless-diversity-of-music/

[25] Billboard hot 100. Wikipedia contributors. [Online]. Available: https://en.wikipedia.org/wiki/Billboard_Hot_100

[26] DataScientest. (2024, February 8) Python: The most popular language. DataScientest. [Online]. Available: https://datascientest.com/en/python-the-most-popular-language

[27] GitHub contributors. guoguo12/billboard-charts. [Online]. Available: https://github.com/guoguo12/billboard-charts

[28] Genius api documentation. Genius. [Online]. Available: https://docs.genius.com/#/getting-started-h1

[29] Noah Passion. (2019, October 25) The evolution of billboard and birth of the "hot 100". Penn State University Blog. [Online]. Available: https://sites.psu.edu/noahpassion/2019/10/25/the-evolution-of-billboard-and-birth-of-the-hot-100/

[30] Genius (company). Wikipedia contributors. [Online]. Available: https://en.wikipedia.org/wiki/Genius_(company)

[31] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova, "Musical trends and predictability of success in contemporary songs in and out of the top charts," *Royal Society Open Science*, vol. 5, p. 171274, 2018.

[32] C. A. McDermott, "An interdisciplinary analysis of popular music genres in the recording studio and billboard charts," 2021. [Online]. Available: https://opus.uleth.ca/server/api/core/bitstreams/97f22868-c0da-4aa2-a9ee-7cd70c65ba20/content

[33] How genius works. Genius. [Online]. Available: https://genius.com/Genius-how-genius-works-annotated

[34] Unimrkt. (2023, August 5) The importance of data cleaning. [Online]. Available: https://www.unimrkt.com/blog/the-importance-of-data-cleaning.php#:~:text=Improved%20data%20accuracy%3A%20Data%20cleaning,insights%20drawn%20from%20the%20data.

[35] (2023, August 11) Data anomaly & quality monitoring: Importance, impact & roadmap. Atlan. [Online]. Available: https://atlan.com/data-anomaly-and-quality-monitoring/

[36] W. Wong. (2020, Nov 13) How to preprocess social media data and text messages. Towards Data Science. [Online]. Available: https://towardsdatascience.com/how-to-preprocess-social-media-data-and-text-messages-b011efacf74

[37] P. Shah. (2020, Jun 27) Sentiment analysis using textblob. Towards Data Science. [Online]. Available: https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524

[38] T. Schäfer, P. Sedlmeier, C. Städtler, and D. Huron, "The psychological functions of music listening," *Frontiers in Psychology*, vol. 4, p. 511, Aug 13 2013.

[39] S. E. Trehub, J. Becker, and I. Morley, "Cross-cultural perspectives on music and musicality," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1664, p. 20140096, Mar 19 2015.

[40] K. I. Batcho, M. L. DaRin, A. M. Nave, and R. R. Yaworsky, "Nostalgia and identity in song lyrics," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 2, no. 4, p. 236–244, 2008.

[41] K. R. Scherer, "The dynamic architecture of emotion: Evidence for the component process model," *Cognition and Emotion*, vol. 23, no. 7, pp. 1307–1351, Nov 1 2009. [Online]. Available: https://doi.org/10.1080/02699930902928969

[42] P. C. Ellsworth and K. R. Scherer, "Appraisal processes in emotion," in *Handbook of Affective Sciences*, R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, Eds. Oxford University Press, 2003, p. 572–595.

[43] R. E. Thayer, *The Biopsychology of Mood and Arousal*. United Kingdom: Oxford University Press, 1990.

[44] I. N. Clark and J. Tamplin, "How music can influence the body: Perspectives from current research," *Voices: A World Forum for Music Therapy*, vol. 16, no. 2, Apr. 2016.

[45] L. B. Meyer, *Style and Music: Theory, History, and Ideology*. United Kingdom: University of Chicago Press, 1996.

[46] C. Toader, C. P. Tataru, I. A. Florian, R. A. Covache-Busuioc, B. G. Bratu, L. A. Glavan, A. Bordeianu, D. I. Dumitrascu, and A. V. Ciurea, "Cognitive crescendo: How music shapes the brain's structure and function," *Brain Sci.*, vol. 13, no. 10, p. 1390, Sep 2023.

[47] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago: The University of Chicago Press, 2003, original work published 1980.

[48] S. A. Widmayer, "Schema theory: An introduction," 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:30998082

[49] V. T. Lai, P. Hagoort, and D. Casasanto, "Affective primacy vs. cognitive primacy: Dissolving the debate," *Frontiers in Psychology*, vol. 3, 2012. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00243

[50] K. Dahlgren. (2014, 10) Don't go wasting your emotion: The process model of emotion regulation. [Online]. Available: https://sites.tufts.edu/emotiononthebrain/2014/10/10/dont-go-wasting-your-emotion-the-process-model-of-emotion-regulation/

[51] S. Folkman and R. S. Lazarus, "If it changes it must be a process: Study of emotion and coping during three stages of a college examination," *Journal of Personality and Social Psychology*, vol. 48, no. 1, pp. 150–170, 1985.

[52] G. J. Walker, J. Deng, and R. B. Dieser, "Culture, self-construal, and leisure theory and practice," *Journal of Leisure Research*, vol. 37, no. 1, pp. 77–99, 2005.

[53] H. R. Markus and S. Kitayama, "Culture and the self: Implications for cognition, emotion, and motivation," *Psychological Review*, vol. 98, no. 2, pp. 224–253, 1991.

[54] H. C. Triandis and V. Vassiliou, "A comparative analysis of subjective culture," University of Illinois at Urbana, Group Effectiveness Research Lab, Tech. Rep., 10 1967. [Online]. Available: https://apps.dtic.mil/sti/pdfs/AD0663889.pdf

[55] H. C. Triandis, "The self and social behavior in differing cultural contexts," *Psychological Review*, vol. 96, no. 3, pp. 506–520, 1989.

[56] H. Wale. Hofstede's cultural dimensions theory. [Online]. Available: https://corporatefinanceinstitute.com/resources/management/hofstedes-cultural-dimensions-theory/

[57] S. Schachter and J. Singer, "Cognitive, social, and physiological determinants of emotional state," *Psychological Review*, vol. 69, no. 5, pp. 379–399, 1962.

[58] W. B. Cannon, "The james-lange theory of emotions: A critical examination and an alternative theory," *The American Journal of Psychology*, vol. 39, no. 1/4, pp. 106–124, 1927. [Online]. Available: https://doi.org/10.2307/1415404

[59] ——, *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Researches into the Function of Emotional Excitement*. D Appleton & Company, 1915. [Online]. Available: https://doi.org/10.1037/10013-000

[60] C. L. Golightly, "The james-lange theory: A logical post-mortem," *Philosophy of Science*, vol. 20, no. 4, pp. 286–299, 1953.

[61] E. B. Titchener, “An historical note on the james-lange theory of emotion,” *The American Journal of Psychology*, vol. 25, no. 3, pp. 427–447, 1914. [Online]. Available: https://doi.org/10.2307/1412861

[62] S.-L. Tan, A. J. Cohen, S. D. Lipscomb, and R. A. Kendall, *The Psychology of Music in Multimedia*. United Kingdom: OUP Oxford, 2013.

[63] C. Liao, D. Liao, and J. Guessford, “Multimodal lyrics-rhythm matching,” 2023.

Acknowledgments

I would like to express my deepest gratitude to Prof. Tomaso Erseghe, my thesis supervisor, for his continuous support, and mentorship, and for being more than just a professor but also a source of inspiration. I consider myself truly fortunate to have had the opportunity to work under his mentorship.
I am also indebted to Prof. Leonardo Badia, whose belief in my potential and encouragement allowed me to start on this master's journey. His belief made it possible for me to live through these amazing years in Italy.
My special thanks are extended to Prof. Federica Battisti for her inspiring courses that were wonderful for me to navigate my remaining academic journey. With her amazing courses, I was able to polish my abilities and academic interests.
I am grateful to Prof. Marco Cagnazzo for his exceptional guidance in an interactive, joyful, and intellectually stimulating course. His approach on the topics he covered was mesmerizing and made my learning experience both enriching and enjoyable.
Collectively, the support and inspiration from Prof. Leonardo Badia, Prof. Federica Battisti, Prof. Marco Cagnazzo, and Prof. Tomaso Erseghe are milestones in my academic journey. I am sincerely thankful for their contributions and the influences they have had on my education. Con I miei ringraziamenti, Professori.
Son olarak, teşekkürler Mürüvvet Tulan, teşekkürler Uğur Tulan, teşekkürler Derya.
İyi ki varsınız.