A Seminar Paper of <u>Selected issues in Digital humanities</u>
(Course code 833595401)

By

Ravid Motorin 201314374
Renana Sanders 311500805

Exploring Kanye West's Musical Journey (2002-2019): Topic Analysis using LDA and Text Summarization

Department of Information Science

Bar-Ilan University

September 2023

Abstract

This seminar paper delves into a comprehensive topic analysis of Kanye West's musical compositions spanning the years 2002 to 2019. A thorough literature review explores Kanye West's personal and professional life, as well as the transition from close to distant reading in connection with topic analysis, and topic analysis methodology in general.

We focused on the dimension of time to link Kanye West's personal and professional life. As a result, we decided early on to analyze our data by album (year) to create a clear timeline connecting Kanye's life to his music Data cleaning and exploration lead to notable findings, including Kanye's recurrent self-reference and the complex interplay between his lyrics and identity.

The study serves as a preliminary step, laying the foundation for an upcoming topic modeling phase that delves into the intricate relationship between Kanye West's life experiences and the themes prevalent in his lyrics. Beyond the scope of this artist's work, this research underscores the wider importance of analyzing rap music, offering a lens through which to glean valuable insights into contemporary cultural dynamics and urgent societal matters. Moreover, it highlights the adaptability of machine learning algorithms for text analysis, particularly when dealing with the distinctive feature of rap music—its propensity for producing extensive and verbose lyrical compositions.

Table of Content

1. Literature Review

1.1. Hip Hop and Rap

Hip hop and rap are closely related music genres that have origins in African American communities and have profoundly shaped contemporary culture and music. Hip hop refers to the broader cultural movement that originated in the 1970s in New York City and encompasses rapping, DJing, breakdancing, and graffiti art (Stapleton, 1998). As Chang (2005) notes, hip hop is recognized as both a musical style and a cultural force reflecting the experiences of marginalized urban youth.

Rap is a music genre that emerged as a component of hip hop culture, consisting of rhythmic speech delivered over beats and instrumentation (Edwards, 2009). Early pioneers of rap like Grandmaster Flash used linguistic techniques like metaphor, wordplay, and storytelling to give voice to inner-city struggles (Bradley & DuBois, 2010). Over decades, rap developed a range of stylistic variations and subgenres, often classified by region, technique or era (e.g. West Coast rap, trap, gangsta rap).

As a cultural movement and musical style, hip hop and rap provide a vital platform for self-expression, conveying the realities of life for disadvantaged groups, often focusing on themes like injustice, violence, poverty, identity and aspirations (Quinn, 2001). While often controversial for raw language and messaging, the storytelling power of hip hop gives it an unparalleled cultural resonance. As Rose (1994) emphasizes, the urgency and humanity of rap makes it "Black America's CNN", capturing experiences often excluded from mainstream media (p.187).

From modest roots in the Bronx, hip hop and rap music have spread globally, becoming among the most dominant cultural and musical forces worldwide (Mitchell, 2001). Far beyond just musical styles, they represent revolutionary

creative outlets allowing marginalized communities to shape culture on their own terms.

1.2 Kanye West

<u>Personal Life</u>

Kanye Omari West was born on June 8, 1977 in Atlanta, Georgia (Caramanica, 2014). His father, Ray West, was one of the first black photojournalists at The Atlanta Journal-Constitution and was also an activist for the Black Panther party. His mother, Donda West, worked as an English professor at Clark Atlanta University before retiring to manage her son's music career (Cole, 2016). West spent his early years in China with his mother as she taught as a visiting professor, before the family settled in Chicago when West was 10 years old (Bennett, 2018).
West initially pursued painting rather than music, gaining a scholarship to attend the American Academy of Art in 1997. However, he soon transferred to Chicago State University to study English, but ultimately dropped out to focus on his music career (Caramanica, 2014).

West's mother, Donda West, died unexpectedly in 2007 following complications from plastic surgery. Her death looms large in West's personal narrative (Cole, 2016).

In 2012, West began dating reality TV star Kim Kardashian. The pair had known each other for years but became romantic after Kardashian's divorce from NBA player Kris Humphries (Lee, 2022). Kardashian and West had their first child, daughter North West, in June 2013. They married the following year in a lavish ceremony in Florence, Italy in May 2014 (Mizoguchi, 2022). West and Kardashian have three additional children together: son Saint West born in December 2015, daughter Chicago West born via surrogate in January 2018, and son Psalm West born via surrogate in May 2019 (Lee, 2022).

In early 2021, Kardashian filed for divorce from West after 6 years of marriage. Their split played out publicly with West making pleas to reunite amid Kardashian's new relationship with comedian Pete Davidson (Mizoguchi, 2022). Their divorce was finalized in November 2022.

West remains close with Kardashian as they co-parent their four children.

Music career

West first made his mark in hip hop as a producer, utilizing detailed sampling methods to bridge soul and classical influences with hip hop beats (Caramanica, 2014).

West released his debut album The College Dropout in 2004 on Jay-Z's Roc-A-Fella Records. The album was both a critical and commercial success, selling over 3 million copies and effectively launching West's career into superstardom (Randall, 2009). He founded his own record label GOOD Music later that year. Over the next decade, West would release seven more solo studio albums that all debuted at #1 on the Billboard charts, along with the GOOD Music label compilation Cruel Summer in 2012 (Caramanica, 2014). His ability to story tell through sampling is cited as changing the soundscape of hip hop and paving the way for new waves of producers (Cole, 2016). As an MC, West portrayed a vulnerable persona on his early albums The College Dropout and Late Registration that examined faith, family, materialism, and self-consciousness through conscious rap lyrics like "We weren't supposed to make it past 25 / Joke's on you we still alive" (Kubrin, 2005; Brice-Finch, 2009). This introspective style connected West to both religious and secular audiences, positioning him as a broken hero in the public imagination (Zenger, 2012).

As West achieved mainstream success, his lyrics shifted to bold proclamations cementing his place among hip hop's greats with lines like "I've been working this grave shift and I ain't made shit / I wish I could buy me a spaceship and fly past the sky" (Randall, 2009). Musically, West continued

fearlessly innovating, incorporating darker themes through manipulations of form, pitch, and arrangement (McLeod, 2019).

Despite polarizing antics, West's relentlessly experimental approach earned him critical acclaim and cemented his status as a musical genius (Caramanica, 2014). Kanye West is one of the most critically acclaimed yet controversial hip hop artists, with over 100 million records sold and 21 Grammy Awards (RIAA, 2022).

West Influence

His net worth now exceeds $2 billion, stemming from music sales, touring, fashion ventures, and other business deals (Forbes, 2022). With over 30 million social media followers, West has cultivated a massive global fanbase through his controversial public persona (Twitter, 2022). West's Yeezy fashion brand earned billions in revenue through strategic partnerships (Forbes, 2022). As a cultural figure, his influence extends beyond music into design, business, politics, and faith (Kelley, 2022). Despite volatility, his cultural primacy persists.

Mental Health challenges

In 2016, West was hospitalized for psychiatric care after erratic behavior like ranting at concerts (CBC, 2016). He was diagnosed with bipolar disorder, a condition causing extreme mood swings between mania and depression (Johns, 2019).

West's 2018 album Ye candidly addressed his mental health, referencing suicidal thoughts and medication (Bernard, 2020). On the song "Yikes" he raps about his diagnosis: "That's my bipolar shit, nigga what? / That's my superpower, nigga ain't no disability" (Genius, 2018). West's social media activities indicate instances of manic episodes worsening his condition. A 2020 Twitter rant accusing his wife Kim Kardashian of trying to lock him up

exhibited apparent paranoia and delusions (Guardian, 2020). Kardashian subsequently posted about West's mental illness and asked for compassion.

Controversies

One of West's most infamous controversies occurred at the 2009 MTV Video Music Awards when he interrupted Taylor Swift's acceptance speech for Best Female Video to declare Beyonce should have won for "Single Ladies" (McIntyre, 2009). West apologized afterwards but the incident sparked widespread backlash against him in the media.

In a 2018 interview with TMZ, West stated "When you hear about slavery for 400 years ... for 400 years? That sounds like a choice" (TMZ, 2018). The comments provoked condemnation from many who found West dismissing the oppression faced by slaves. West later tried clarifying he meant mental enslavement, but the damage was done.

West has shown support for Donald Trump, wearing MAGA hats and visiting the White House in 2018 (Johnson, 2018). This led some fans to accuse West of endorsing Trump's controversial stances on issues like immigration. However, West also claimed he didn't agree with all Trump's policies and was supporting him based on prison reform efforts.

In 2022 West and conservative pundit Candace Owens wore shirts that said "White Lives Matter" as a dig against the Black Lives Matter movement (Elich, 2022). Many saw the act as West dismissing issues impacting the Black community. The move furthered accusations of West allying himself more with rightwing figures than marginalized groups.

In October 2022, West threatened on Twitter he would go "death con 3 on JEWISH PEOPLE" (ESPN, 2022). The tweets led to his accounts being restricted for hate speech. Several brands also cut ties with West following the antisemitic comments.

Kanye West's albums have pushed hip hop into stadium-sized spectacle while retaining soulful confessionalism (Coleman, 2007). Yet public breakdowns and inflammatory opinions threaten to undermine goodwill toward his artistic contributions. West remains one of music's most fascinating figures for reflecting the uneasy negotiation between raw talent, mental health, fame, and morality still unraveling in American society.

Kanye West's place in hip hop culture

Kanye West occupies an eminent position in hip hop culture and rap music due to his groundbreaking artistry, influence, and commercial success. West's innovative sampling techniques bridged soul music inspirations with rap beats, helping pioneer a new wave of hip hop production (Bradley & DuBois, 2010). As Caramanica (2014) notes, West's debut album The College Dropout (2004) established him as both a skilled MC and visionary producer, earning critical and popular acclaim with its blend of wit, vulnerability, and sonic intricacy.

West's subsequent albums contain bold experimentation with form and style, weaving together genres while retaining hip hop at the core (McLeod, 2019). He frequently subverts hip hop conventions, seen in the stark electronics of 808s and Heartbreak (2008) and abrasive industrial textures on Yeezus (2013) (Shepherd, 2013). As a lyricist, West utilizes stream-of-consciousness storytelling and pulls no punches addressing controversies around fame, mental health struggles, and politics (Kelley, 2022).

Through a relentlessly innovative approach, West has been instrumental in pushing hip hop into unprecedented territories, pioneering new trends spanning production, vocals, lyricism, and sound (Caramanica, 2021). Described as a genius, he has won 21 Grammys among countless other accolades (RIAA, 2022). His influence permeates contemporary music, seen in the vulnerable lyrical styles of rappers like Drake and Kid Cudi, and the fusion of rap with other genres by artists like Lil Nas X (Zisook, 2022).

While controversial for inflammatory statements, West's cultural imprint is undeniable, cementing him as "the most consistently groundbreaking artist of his generation" within hip hop and rap (Kelley, 2022, p. 34). His boundary-shattering artistry has profoundly expanded the horizons of what rap music can embody and accomplish.

1.3. The Transition from Close to Distant Reading

The shift from close reading to distant reading represents a significant transformation in the field of literary studies. Franco Moretti, a prominent figure in the field of digital humanities, has played a pivotal role in advocating and popularizing distant reading techniques. As Moretti famously quipped, transitioning to distant reading often involves making "a little pact with the devil" (Moretti, 2013), implying a willingness to sacrifice the depth of individual textual analysis for the broader insights that distant reading can provide.

In this chapter, we will explore the main reasons for this transition, emphasizing the advantages and disadvantages of distant reading while keeping Moretti's provocative statement in mind.

Defining Close Reading

Close reading, a traditional approach to literary analysis, involves a meticulous examination of individual texts (Moretti, 2013). Scholars engaged in close reading scrutinize a text's language, style, and structure, often focusing on minute details to uncover deeper meanings and themes. This method has historically dominated literary studies, enabling scholars to provide insightful interpretations of specific texts.

Emergence of Distant Reading

Distant reading, a term popularized by Franco Moretti in his book "Distant Reading" (2013), represents a paradigm shift in literary analysis. Distant reading involves the analysis of a vast corpora of texts using computational tools and techniques, allowing scholars to identify broader trends, patterns, and connections that transcend individual texts.

Main Reasons for Transitioning

The Big Data Revolution

The advent of digital technologies and the proliferation of digitized texts have enabled scholars to access and analyze vast amounts of literary data (Jockers, 2013). Distant reading leverages this abundance of data to uncover patterns that were previously impossible to discern through close reading alone.

Addressing Limitations of Close Reading

Close reading has inherent limitations, such as its inability to analyze large-scale trends across multiple texts (Ramsay, 2011). Distant reading complements close reading by offering a macroscopic perspective on literature, allowing researchers to investigate questions related to genre, authorship, and cultural shifts.

Broadening the Scope of Literary Studies

Distant reading extends the boundaries of literary studies by encompassing non-canonical texts, popular culture, and lesser-known authors (Moretti, 2013). This expansion of scope encourages a more inclusive approach to literature and culture in general.

Advantages of Distant Reading

- Macroscopic Analysis

Distant reading allows scholars to uncover macroscopic patterns and trends in literature (Jockers, 2013). By analyzing extensive corpora, researchers gain insights into the evolution of literary genres, the dissemination of ideas, and cultural influences.

- Efficiency and Speed

Distant reading techniques expedite the analysis process, enabling scholars to process and analyze large volumes of text efficiently (Moretti, 2013). This

increased speed facilitates the exploration of research questions that require extensive data.

- Interdisciplinary Collaboration

Distant reading encourages collaboration between literary scholars and experts in fields such as computer science and data analysis (Jänicke, S et al. 2015). This interdisciplinary approach fosters innovation and widens the methodological toolkit available to scholars.

Disadvantages of Distant Reading

- Loss of Nuance

Distant reading may overlook the nuances and subtleties present in individual texts (Moretti, 2013). The emphasis on quantitative analysis can sometimes disregard the qualitative richness of literature.

- Data Selection and Bias

The selection of corpora and data preprocessing can introduce biases into distant reading analyses (Jockers, 2013). Researchers must be cautious in their data selection and interpretation to avoid reinforcing existing biases.

- Technological Barriers

Not all scholars have access to the necessary computational tools or possess the technical skills required for distant reading. This technological barrier can create disparities in the field.

Exploring the Symbiosis: Distant Reading and Topic Analysis

Topic analysis is a fundamental component of distant reading, facilitating the identification and extraction of latent themes, subjects, or topics within a given corpus (Blei, Ng, & Jordan, 2003). This connection is firmly rooted in the computational underpinnings of both approaches, wherein algorithms and statistical models are employed to reveal the semantic content of texts and map their thematic landscapes. By employing topic modeling techniques such as Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF) and text summarization, distant reading scholars are able to unearth prevalent subjects and their evolution across time and space, thereby advancing our understanding of literary, historical, or cultural trends (Lee & Seung, 1999).

Consequently, the integration of distant reading and topic analysis empowers researchers to explore extensive textual datasets, illuminating previously inaccessible dimensions of human expression and intellectual exploration.

As scholars navigate the transition from close to distant reading, they must consider the trade-offs inherent in Moretti's metaphorical "pact with the devil" (Moretti, 2013). Moretti's insightful words remind us that while distant reading offers a broader perspective on literature, it necessitates sacrifices in terms of the depth of analysis applied to individual texts. Nevertheless, this transition holds great promise for the future of literary studies, fostering innovation and inclusivity while challenging scholars to strike a balance between the devil and the details.

1.4. Text Analysis Models

Topic Modeling

Topic modeling is a powerful computational technique used in natural language processing and text analysis to uncover latent themes or topics within a large collection of documents or texts. It is especially valuable in distant reading, where researchers aim to analyze and understand patterns and trends across extensive textual datasets.

Topic modeling algorithms

1. Latent Dirichlet Allocation (LDA):

Description: LDA, introduced by Blei, Ng, and Jordan (2003), is a foundational topic modeling algorithm. It operates on the premise that documents are mixtures of topics, and topics are mixtures of words. LDA assumes a generative process where each document is created by selecting a set of topics and then generating words based on those topics. LDA involves probabilistic modeling, which helps uncover hidden thematic structures within a corpus. We will use in our work LDA model to uncover Kanye west evolving themes in his album's lyrics.

2. Non-Negative Matrix Factorization (NMF):

Lee and Seung (1999) elucidate NMF as a factorization method applied to topic modeling. NMF factorizes a document-term matrix into two lower-dimensional matrices, one representing topic and the other representing the distribution of terms in topics.

3. Latent Semantic Analysis (LSA):

Deerwester et al. (1990) originally proposed LSA as a method for dimensionality reduction and semantic analysis. LSA aims to capture latent semantic relationships among terms by applying singular value decomposition (SVD) to a term-document matrix. LSA reduces the dimensionality of the term-document matrix, enabling the discovery of semantic connections between words and documents.

4. Probabilistic Latent Semantic Analysis (pLSA):

Hofmann (1999) presents pLSA as a probabilistic extension of LSA. It introduces a probabilistic framework to model the generation of terms in documents based on a set of latent topics.

Text summarization

Text summarization, on the other hand, is a distinct yet complementary technique used in text analysis. It involves the condensation of a longer text into a shorter, coherent summary while retaining the essential information and meaning. Text summarization can be performed through extractive or abstractive methods.

Extractive summarization

Extractive summarization is a text summarization technique that aims to condense a document or text by selecting and extracting the most salient sentences or phrases from the original content (Erkan & Radev, 2004). In this approach, the summary is created by identifying and pulling sentences that are the most informative, coherent, and representative of the source text. These selected sentences are typically presented in the same order as they

appear in the original document, maintaining the document's structure and coherence.

Extractive summarization is often favored for its simplicity and transparency, as it avoids the generation of new content and relies on the inherent information present in the source text. It involves several key steps:

Sentence Scoring: To determine the importance of each sentence in the document, various scoring methods are employed. These methods may include sentence similarity calculations, keyword extraction, or other statistical measures. Common algorithms used for extractive summarization include TextRank (Mihalcea & Tarau, 2004) and LexRank (Erkan & Radev, 2004).

Sentence Selection: After scoring, the sentences with the highest importance scores are chosen for inclusion in the summary. The selection process ensures that the most relevant and significant content is retained.

Order Preservation: The selected sentences are typically presented in the same order as they appear in the original text, maintaining the flow and coherence of the document.

Extractive summarization is particularly useful when the primary goal is to provide a concise overview of the main ideas and key points in a document while avoiding the generation of new content. It is commonly applied in news summarization, content extraction from web pages, and document indexing.

Abstractive summarization

Abstractive summarization, in contrast to extractive summarization, involves the generation of new content that is not present in the original document (Rush, Chopra, & Weston, 2015). It aims to produce a summary that not only condenses the information but also rephrases and restructures it to create a coherent and concise representation of the source text. This approach requires a deeper understanding of the content and the ability to generate human-like language.

The process of abstractive summarization typically includes the following steps:

Content Understanding: Abstractive summarization models, such as Sequence-to-Sequence (Seq2Seq) models (Sutskever et al., 2014) and Transformer-based models (Vaswani et al., 2017), use neural networks to understand the content of the source text. This involves encoding the input text into a fixed-length representation that captures its meaning.

Generation: After understanding the content, the model generates a summary by decoding the encoded representation into natural language. During this step, the model can paraphrase, rephrase, and synthesize information to create a coherent and concise summary.

Language Fluency: Abstractive summarization models aim to generate summaries that are grammatically correct and linguistically fluent. This requires the model to not only capture the information but also produce human-like language.

Abstractive summarization is valuable when the goal is to provide concise and coherent summaries that go beyond mere extraction of sentences. It is commonly used in applications such as document summarization, chatbot responses, and automatic content generation for news articles.

In summary, extractive summarization focuses on selecting and preserving important sentences from the source text, while abstractive summarization involves understanding the content and generating new, coherent language to create summaries. The choice between these approaches depends on the specific summarization task and desired output.

Using text analysis models on hip hop and rap lyrics

Topic modeling and text summarization are increasingly utilized in computational text analysis of hip hop and rap music as researchers recognize the value of distant reading techniques for this domain. A growing body of scholarship has demonstrated how these methods can provide

macroscopic perspectives on thematic trends and stylistic evolutions within rap lyrics over time.

Topic modeling algorithms like latent Dirichlet allocation (LDA) have been extensively applied for uncovering latent topics and themes frequently occurring across corpora of hip hop lyrics (Ohriner, 2019; Sula, 2018). For instance, Du (2019) employed LDA to extract prevalent hip hop topics related to money, romance, violence, and other dimensions, tracing how these topics distributed across different eras. Meanwhile, Mohr and Bogdanov (2013) proposed aLDA-inspired model tailored for music analysis. They demonstrated its effectiveness in extracting topical tags from lyrics while accounting for artist and album information too.

Beyond Identifying themes, text summarization techniques like TextRank have also shown utility for distilling key ideas from massive rap lyric collections into condensed summaries (Mihalcea & Tarau, 2004; Nasukawa & Yi, 2003). These summaries provide efficient overviews of thematic landscapes. Sentiment analysis is also sometimes applied to assess emotional valence across different periods (Koltcov et al., 2022). Overall, studies unanimously highlight how topic modeling and text summarization enable scalable computational analysis revealing linguistics patterns and cultural perspectives expressed through decades of hip hop music (Ohriner, 2019; Sula 2018). However, some caution that preprocessing steps like cleaning corpora can introduce biases, and human interpretation is still essential (Svendsen et al., 2021). Still, when thoughtfully deployed, these techniques unlock invaluable insights at macroscopic levels beyond traditional close reading.

2. Research Question, Hypothesis and Objectives

2.1. Research Question

How have the themes and topics in Kanye West's lyrics evolved over his musical career from 2002 to 2019?

2.2 Hypothesis:

The thematic content in Kanye West's lyrics has changed over time, reflecting different stages of his personal life and musical style.

2.3 Research Objectives:

- To perform a comprehensive topic analysis on Kanye West's lyrics from 2002 to 2019 using LDA algorithm
- To identify the prominent themes and topics within each album based on the lyrics
- To compare and visualize the similarities and differences in themes between albums over the years
- To connect the evolution of topics and themes with key events in Kanye's personal life and musical career
- To apply text analysis techniques to gain insights into Kanye's creative evolution through his discography spanning almost two decades

3. Methodology

3.1. Dataset Source

The dataset was downloaded from Kaggle from:
https://www.kaggle.com/datasets/convolutionalnn/kanye-west-lyrics-dataset

The original dataset was scraped from the Genius lyrics website by a Kaggle user named David Gottlieb. It contains lyrics from 8 of Kanye's studio albums and his first mixtape from 2002 to 2019. It does not include his latest album from 2021 and two other mix tapes that he published.

The dataset was uploaded to Kaggle as a single text file with no metadata. Just the raw lyrics for 129 songs.

To make the data more usable, we searched for and added metadata including:

  - Song title

  - Album name

  - Album release year

We organized this into an Excel spreadsheet with columns for:

  - Id

  - Song

  - Album

  - Year

  - Lyrics

3.2. The Dataset

The lyrics span Kanye's career from 2002 to 2019. The albums covered are:

- · Get Well Soon (2002) with 3 songs in the dataset.
- The College Dropout (2004) with 18 songs in the dataset.
- Late Registration (2005) with 22 songs in the dataset.
- Graduation (2007) with 14 songs in the dataset.
- 808s & Heartbreak (2008) with 11 songs in the dataset.
- My Beautiful Dark Twisted Fantasy (2010) with 14 songs in the dataset.
- Yeezus (2013) with 10 songs in the dataset.
- The Life of Pablo (2016) with 20 songs in the dataset.
- Ye (2018) with 7 songs in the dataset.
- Jesus is King (2019) with 10 songs in the dataset.

In total there are 129 songs in the dataset. The lyrics for each song are in one continuous block of text.

This provides a rich dataset of Kanye's discography over almost two decades.

3.3. Data Preparation and Cleaning Process

In the endeavor to prepare the acquired dataset for topic modeling and analysis, several essential steps were made to ensure data quality and semantic uniformity:

1. Data Quality Assurance

The initial phase of this process involved a thorough verification to identify any instances of null values or duplicates within the dataset which is imperative to maintain data integrity and analytical accuracy. No null values / duplicates were found in our dataset.

2. Textual Data Cleansing

Subsequently, a data cleansing function was thoughtfully developed to address textual components, particularly lyrics. This function was designed to improve the quality and consistency of the textual content for subsequent analysis.

a. Stop word Removal and Lemmatization

To enhance the coherence and analytical value of the textual data, the process of stop word removal and lemmatization was implemented. Stop words, which are frequently occurring words like articles and conjunctions, were systematically removed from the text. This was done to reduce the impact of words with minimal semantic significance, which could potentially distort the analysis.

Lemmatization, a linguistic technique, was used to simplify words into their base or dictionary forms. This approach helps to standardize words, ensuring consistency and enabling semantic equivalence among different word forms. Consequently, it improves the interpretability and comparability of the textual data.

2.Acronym Dictionary

To further improve semantic consistency, an acronym dictionary was introduced. This dictionary was used to replace commonly used acronyms with their full forms, promoting clarity and uniformity in meaning across the dataset.

3.Removal of Song-Structure Related Words

Recognizing that certain song-structure related terms, such as "chorus," "verse," and "intro," were prevalent and repetitive across lyrics, a deliberate decision was made to exclude these terms. Such exclusions were aimed at minimizing redundancy and focusing analytical efforts on the lyrical content itself.

4.Handling of Rap Vocals and Garbled Words

In the specific context of rap music, it was noted that there were instances of garbled words within the textual content. In order to retain a sharp analytical focus on semantically meaningful content, a manual effort was undertaken to identify and subsequently remove these garbled words from the lyrics. This action aligns with the objective of reducing redundancy and directing analytical attention toward the lyrical content itself, similar to the approach taken in section c.

In summary, this thorough data preparation and cleaning process followed established best practices in text preprocessing. It ensured that the dataset was well-prepared for subsequent analytical tasks, all the while maintaining its semantic consistency and overall quality.

3.4. Data Exploration
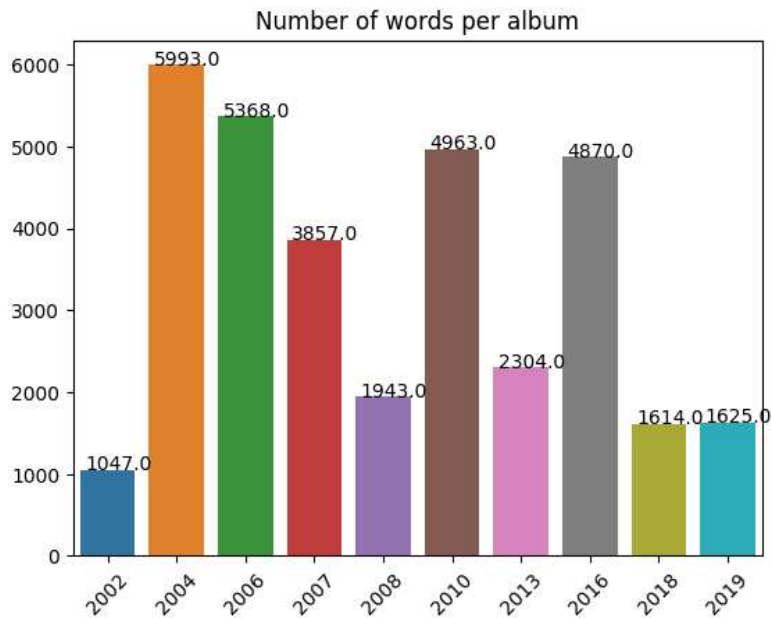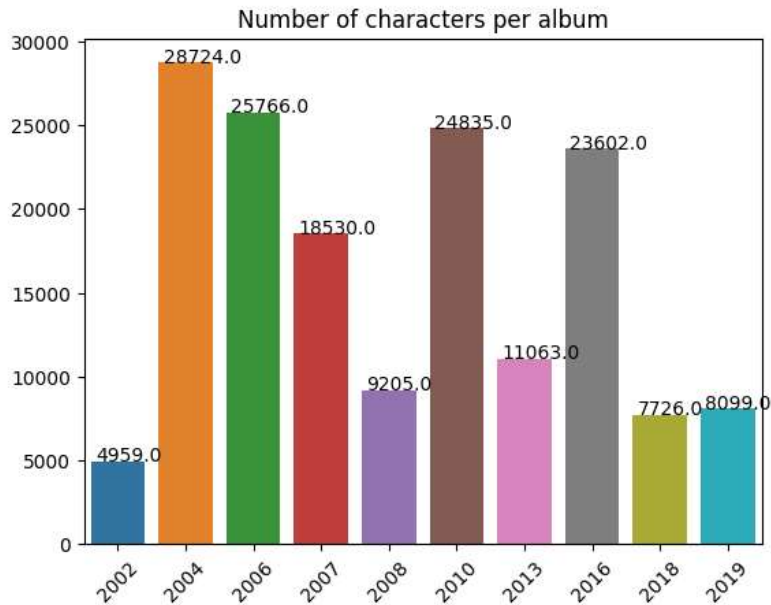
1. **Word Cloud and Common Words per Album Analysis:**
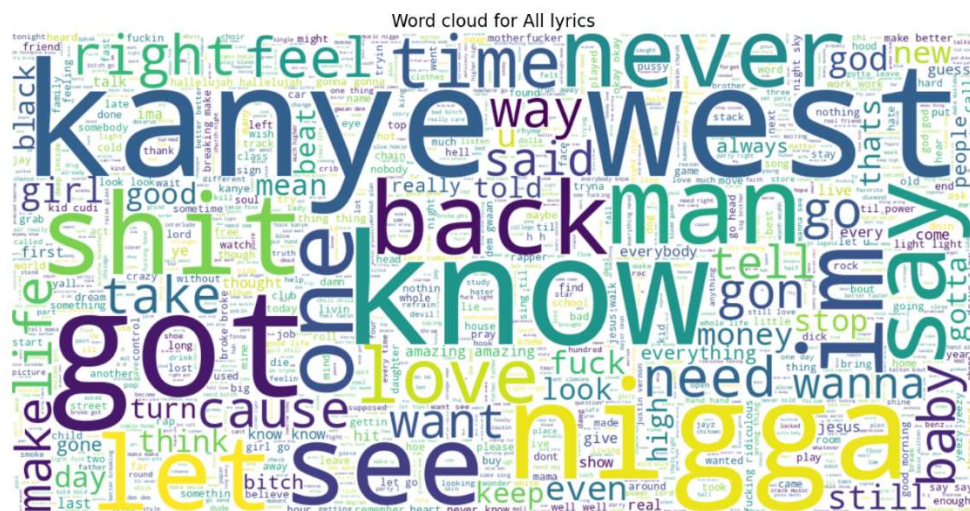
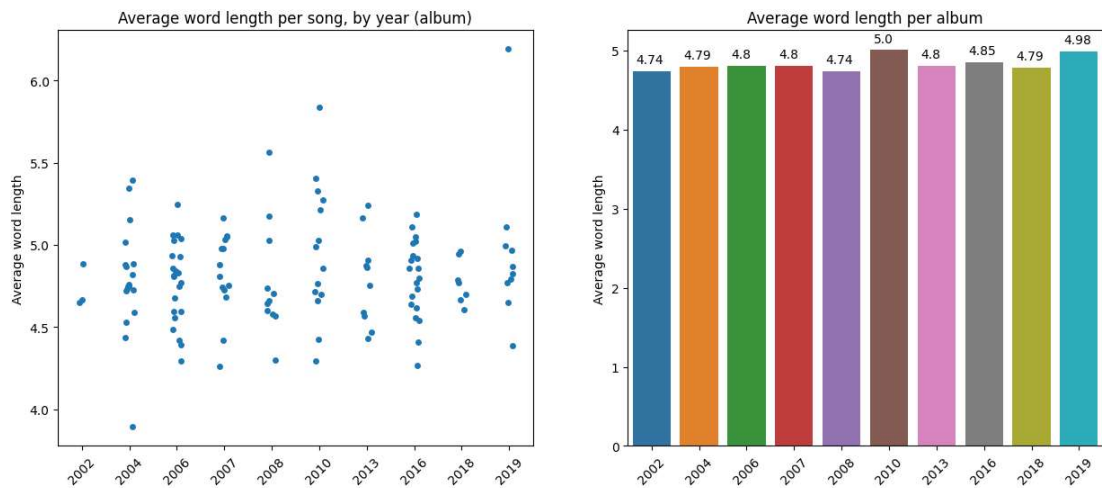|	2002	2004	2006	2007	2008	2010	2013	2016	2018	2019
0	jesus	get	get	like	know	know	nigga	kanye	know	hallelujah
1	two	like	got	know	amazing	kanye	bitch	know	love	sing
2	walk	know	west	cannot	love	west	know	feel	get	work
3	right	man	kanye	get	never	like	go	like	nigga	hand
4	go	gonna	like	good	keep	see	dem	west	make	lord
5	word	thing	make	never	like	let	like	get	still	every
6	like	got	broke	west	stop	im	gwaan	nigga	see	need
7	still	well	better	kanye	thing	light	god	god	like	god
8	god	kanye	know	got	kanye	get	cannot	need	day	jesus
9	way	nigga	ever	light	see	nigga	kanye	love	kanye	come

o Our data exploration began with the creation of a comprehensive word cloud, encompassing all lyrics of all albums. This initial step aimed to offer a broad overview of the predominant themes evident in Kanye West's life, as conveyed through his song lyrics.

o Notably, the largest words in the word cloud prominently featured Kanye West's full name. Given Kanye's tendency to refer to himself in the third person in his songs, this observation shed light on his self-centered lyrical persona.

o Another word of significance in the word cloud was "nigga," a derogatory term associated with African-American identity. This observation enriched our perspective, especially when considering Kanye's life experiences as a Black individual raised by a single mother in Chicago.

1. **Analysis of Text Characteristics per Album:**

To explore whether the textual patterns aligned with shifts in Kanye's personal life, including changes in family, marital status, and personal

health, we analyzed various text characteristics for each album. We examined metrics such as the character count, word count, and average word length for each album. However, contrary to our expectations, we did not discern substantial textual shifts that indicated thematic changes between albums or time periods. Themes such as religion, narcissism, and certain derogatory words remained consistent across most albums.

Number of characters per album

Number of words per album

Average word length per song, by year (album)

Average word length per album

1. **N-Gram Analysis:**

   As part of our data exploration, we conducted bi-gram and trigram analysis for each album.

   - Upon commencing these analyses, we encountered a particular challenge. The outcomes frequently displayed instances where the same word was repeated consecutively. We attributed this phenomenon to the characteristic writing style of rap music. However, recognizing that such repetition lacked significant semantic meaning, we implemented a function to eliminate consecutively repeated words before conducting the n-gram analysis.

   - Noteworthy findings emerged from the n-gram analysis:
     o In eight out of ten albums, the most frequently recurring word pair was Kanye West's full name. This observation emphasized Kanye's frequent use of the third person when referring to himself in his songs. Furthermore, even in the two remaining albums, his full name appeared among the top ten most frequent word pairs, albeit in lower positions.

     o Furthermore, the n-gram analysis uncovered an uncommon occurrence of profanity in the year 2010. This discovery necessitates further investigation in conjunction with the outcomes from the topic modeling phase of our research.

Bigram (Top 10 most common consecutive word pairs per album)

	2002	2004	2006	2007	2008	2010	2013	2016	2018	2019
0	jesus walk:15	kanye west:31	kanye west:60	kanye west:40	kanye west:25	kanye west:59	kanye west:26	kanye west:57	kanye west:15	til power:22
1	two word:11	set party:14	make better:23	cannot study:28	wrong thing:16	fucking ridiculous:20	dem gwaan:18	night sky:17	make mistake:10	sing til:19
2	still nowhere:10	party right:13	breaking make:22	good morning:18	never know:14	run away:15	gwaan dem:12	deep inside:16	mistake girl:10	power lord:16
3	nowhere go:10	workout plan:12	go head:18	good life:15	cannot stop:11	know see:14	send bitch:12	feel fadin:12	girl still:10	lord come:15
4	kanye west:6	raise glass:11	drive slow:18	cannot tell:14	heartless could:10	ima need:13	bitch go:12	real friend:11	still love:10	kanye west:15
5	throw hand:6	kanyes workout:10	girl go:17	better faster:13	keep love:9	really care:13	let go:10	lookin church:11	love way:6	come sing:12
6	go still:5	im tellin:9	gotta leave:16	make stronger:13	love lockdown:9	want see:12	go send:10	church night:11	shit could:6	everything need:10
7	mo def:5	tellin fall:9	leave get:16	tell nothing:13	worry bout:9	need see:12	justin vernon:8	god dream:9	could get:6	ant clemons:8
8	walk god:4	said want:9	get girl:16	feel like:11	real boy:9	see fucking:12	fall love:8	feel like:9	get menacin:6	need oo:8
9	god show:4	one thing:9	music nigga:14	never told:11	could heartless:8	fucking hand:12	new sve:7	yes deep:9	menacin frightenin:6	sing every:6

○

Bigram (Top 10 most common consecutive word pairs per album)

	2002	2004	2006	2007	2008	2010	2013	2016	2018	2019
0	still nowhere go:10	set party right:13	breaking make better:17	cannot tell nothing:13	heartless could heartless:8	ima need see:12	dem gwaan dem:12	lookin church night:11	make mistake girl:10	sing til power:19
1	nowhere go still:5	im tellin fall:9	gotta leave get:16	cannot study cannot:10	loving way wanted:8	need see fucking:12	gwaan dem gwaan:12	church night sky:11	mistake girl still:10	til power lord:16
2	go still nowhere:5	care people say:8	leave get girl:16	study cannot study:10	worry bout wrong:8	see fucking hand:12	send bitch go:10	yes deep inside:9	girl still love:10	power lord come:14
3	jesus walk god:4	ive workin graveshift:8	get girl go:16	good morning good:9	bout wrong thing:8	air really care:12	bitch go send:10	deep inside feel:7	shit could get:6	lord come sing:9
4	walk god show:4	workin graveshift made:8	girl go head:16	morning good morning:9	wrong thing wrong:8	get much higher:11	go send bitch:10	wanna feel liberated:6	could get menacin:6	come sing til:8
5	god show way:4	graveshift made shit:8	drive slow homie:13	make stronger need:9	thing wrong thing:8	fuck light fuck:11	let go let:6	bad bitch equinox:6	get menacin frightenin:6	everything need oo:8
6	show way devil:4	made shit wish:8	cmon homie major:13	work harder make:9	like til never:8	light fuck light:11	go let go:6	know right freak:6	menacin frightenin find:6	need oo everything:6
7	way devil tryna:4	shit wish could:8	go head get:12	harder make better:9	til never know:8	play blame game:11	fall love bound:6	party please baby:6	frightenin find help:6	oo everything need:6
8	devil tryna break:4	wish could buy:8	head get gotta:12	make better faster:9	stop vision gotta:7	much higher high:10	bound fall love:6	please baby party:6	get genie bottle:6	need sing til:5
9	tryna break jesus:4	could buy spaceship:8	get gotta leave:12	better faster make:9	vision gotta get:7	shit fucking ridiculous:10	know new sve:5	deep inside get:6	nothing hurt anymore:5	come sing every:4

3.5. Topic Modeling And Text Summarization Models

In order to find the themes in Kanye West's albums throughout the years using machine learning methods we applied LDA topic modeling and text summarization models.

Latent Dirichlet Allocation (LDA) Model:

LDA is applied on lyric corpus to discover album-level topics. The topics and top keywords are extracted and analyzed.

How it works:

- Treats each document as a mix of topics.
- Topics are distributions over words.
- Iteratively updates topic distributions per document and word distributions per topic to find the optimal model that best represents the data.

Parameters:

- num_topics = 10 (number of topics to discover)
- passes = 10 (number of times to iterate over entire corpus)

- alpha and eta are hyperparameters that control sparsity of distributions. 'Auto' sets them automatically based on a number of topics.

Output:

- lda_model - Trained LDA model
- topics - List of top keywords for each inferred topic with weights
- perplexity - Measure of how well the model predicts sample data. Lower is better.
- coherence - Measure of topic interpretability. Higher is better.

Text Summarization:

Summarization is applied on lyrics for each album to get an overview of topics and content. Summaries are analyzed along with LDA topics.

How it works:

- Splits text into sentences/lines
- Ranks sentences by importance using metrics like word frequency, POS tags, sentence position etc.
- Selects top ranked sentences as summary.

Parameters:

- num_lines = 5 (number of lines to include in summary)

Output: Concise summary extracting key points from lyrics.

3.6. Finding The Themes

In order to find the theme we extracted topic keywords and generated word clouds for every Album to visualize the topics.

Word Cloud for Year 2006

Word Cloud for Year 2007

Word Cloud for Year 2008

Word Cloud for Year 2010

Word Cloud for Year 2018

Word Cloud for Year 2019

Most noticeable are years 2019 and 2008 that has only few words that are very frequent. Indicating that these years have less themes with few frequent key words.

We computed topic similarity between albums using cosine similarity of keyword vectors. The cosine similarity score ranges from 0 to 1, with 0 indicating that the two vectors are orthogonal (completely different) and 1 indicating that the two vectors are identical Then we Visualized the similarity in a heatmap.

Similarity rates between the key words vectors were found to be very small.

After looking at the topic modeling results, the key words and the analysis we performed on the topics and summarized data we started to define themes.

- Some themes were recognized based only on the topic modeling data frame for example: topic 5 for year 2018 contains euphemisms that were found using the n_grams like: "get_rubbin_lamp_get" , so sex was found as a theme.
- We looked at nouns that were present in the topics like "war" in topic 1 in 2002 and at nouns that were found in the analysis like: way, time, hater, life etc…
- The presence of the word nigga was cataloged by as "black" since only afro American rappers can use this word and it indicates talking to or about an afro American person.
- In addition, we grouped similar words into a unifying theme:
  - The word Jesus, God, almighty, Pray etc were all called God, The words fuck, bitch, shit etc were all binned into swear words theme. Words that indicate fear and the words mind, real, head were categorized as mental health. The words daughter, father and family were grouped to family theme. The words broke and

money was binned into a money theme. The words girl and love were binned into a love theme. The words people and everybody were binned to people theme. The words light, morning, new were binned into hope. Words marking music like beat, rap or artists names were cataloged as hip hop.

- o Any mention of West or Kanye was categorized as a Kanye West theme.
- o Words that were very frequent across topics and summaries were also defined as a theme for example the words never, wrong and need.

We have tried to define words both by looking at each year separately and by looking at all the data as a whole to keep consistency that will enable us to analyze the evolving themes throughout the years.

The themes:

2019 god

2018 family, god, way, black, swear words, Kanye west, love, mental health, sex, hip hop

2016 family, time, black, God, Kanye west, money, people, hope

2013 Kanye west, black, swear words, hope

2010 love, mental health, black, Kanye west, swear words, hope

2008 life, wrong, hope, black, Kanye west

2007 haters, life, show, people, hope, black, never

2006 love, money, Kanye west, swear words, black

2004   life, money, people, hip hop, school, child, Kanye west, swear
        words, work, need

2002   way, God, people, swear words, war, hip hop

4. The results

The themes were written into an excel file called:

topics_and_summarized_text_combined_manualthemes
The file contains : Year, five topics that were generated earlier, text
summarization that was generated earlier and themes that were written by us
after looking at the models results.

Word Cloud of Theme Frequencies

Theme Counts (Overall)
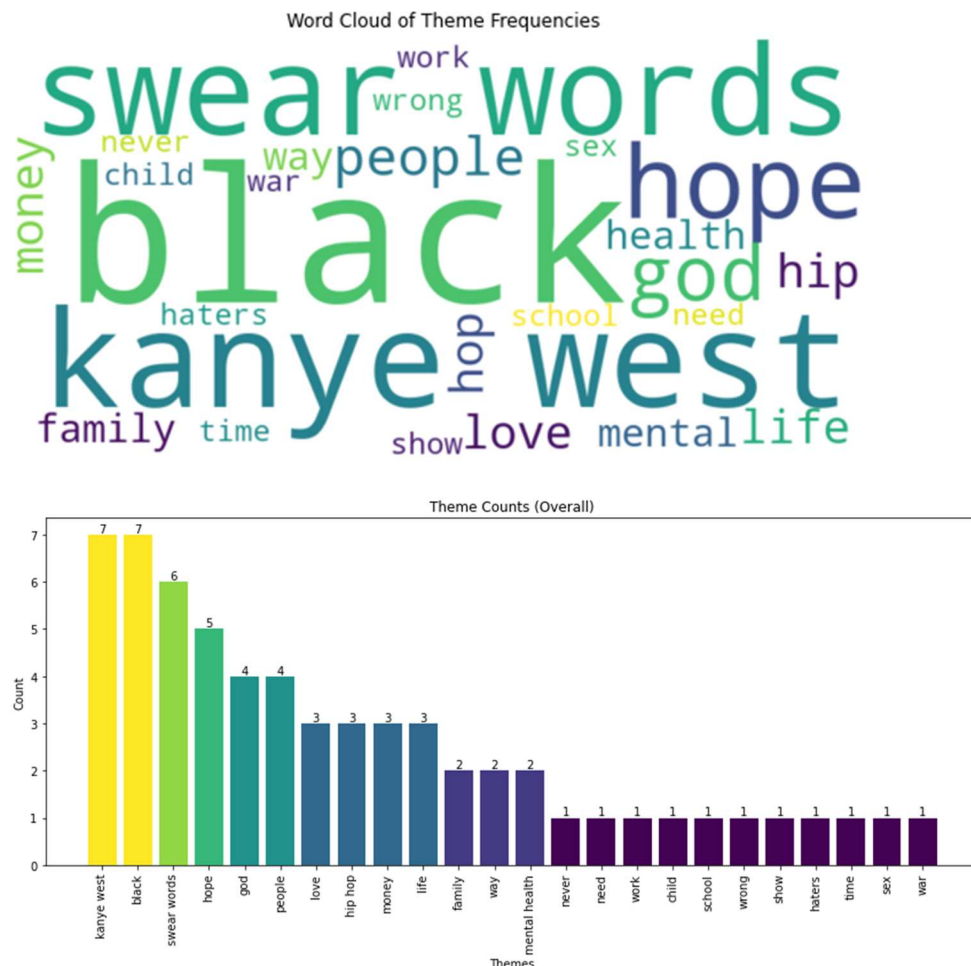
All Themes with Overall Counts:

    Themes: "kanye west" and "black", Overall Count: 7

    Themes: "swear words", Overall Count: 6

    Themes: "hope" overall Count: 5

Themes: "god","people" Overall Count: 4

Theme: "love","hip hop","money", "life" Overall Count: 3

Theme: "family","way","mental health" Overall Count: 2

Theme:

"never","need","work","child","school","wrong","show","time","sex","war"

Overall Count: 1

The most prevalent themes are references to himself ("kanye west") and racial identity ("black"), appearing in 7 albums each. This highlights Kanye's focus on his persona and lived experience as a Black man.

Profanity ("swear words") appears consistently, in 6 albums. This reflects Kanye's explicit lyrical style and unfiltered self-expression.

"Hope" emerges as a theme across 5 albums, capturing Kanye's messages of perseverance and optimism amidst adversity.

Spirituality ("god") and relatability ("people") arise as themes in 4 albums each, signaling Kanye's exploration of faith and desire to connect with audiences.

Core themes like "love", hip hop culture ("hip hop"), wealth ("money"), and lived experience ("life") surface in 3 albums apiece.

His family life ("family") and evolving worldview ("way") appear in 2 albums, coinciding with major life events.

Deeply personal themes like mental health ("mental health") are rarer, only in 2 albums.

Many idiosyncratic themes like "never", "need", "work", "child", "school" etc. appear only once, capturing specific phases.

Themes per years:

Theme: "kanye west", Years: [2004, 2006, 2008, 2010, 2013, 2016, 2018]

Theme: "life", Years: [2004, 2007, 2008]

Theme: "mental health", Years: [2010, 2018]

Theme: "money", Years: [2004, 2006, 2016]

Theme: "black", Years: [2006, 2007, 2008, 2010, 2013, 2016, 2018]

Theme: "school", Years: [2004]

Theme: "time", Years: [2016]

Theme: "hip hop", Years: [2002, 2004, 2018]

Theme: "never", Years: [2007]

Theme: "people", Years: [2002, 2004, 2007, 2016]

Theme: "need", Years: [2004]

Theme: "haters", Years: [2007]

Theme: "family", Years: [2016, 2018]

Theme: "show", Years: [2007]

Theme: "hope", Years: [2007, 2008, 2010, 2013, 2016]

Theme: "way", Years: [2002, 2018]

Theme: "sex", Years: [2018]

Theme: "child", Years: [2004]

Theme: "swear words", Years: [2002, 2004, 2006, 2010, 2013, 2018]

Theme: "love", Years: [2006, 2010, 2018]

Theme: "work", Years: [2004]

Theme: "god", Years: [2002, 2016, 2018, 2019]

Theme: "wrong", Years: [2008]

Theme: "war", Years: [2002]

In 2019, the sole theme of "God" reflects his public embrace of Christianity culminating in the "Jesus is King" album after his mental health struggles.

In 2018, "family" themes correlate to his stability with Kim Kardashian and their kids, while "mental health" connects to his bipolar diagnosis.

In 2016, "family" also mirrors his satisfaction with married life and children around The Life of Pablo era.

In 2013, abrasive themes like "swear words" matched his aggressive creativity during the Yeezus phase.

In 2010, "mental health" foreshadows troubles to come on the introspective My Beautiful Dark Twisted Fantasy album.

In 2008, "wrong" and "hope" fit his regret about past misdeeds during the reflective 808s & Heartbreak album.

In 2007, "haters" and "show" reflected his drive to prove critics wrong at his pop culture peak with Graduation.

In 2004, "school" and "work" themes fit Kanye's journey to stardom after producing for years while striving to make his own album.
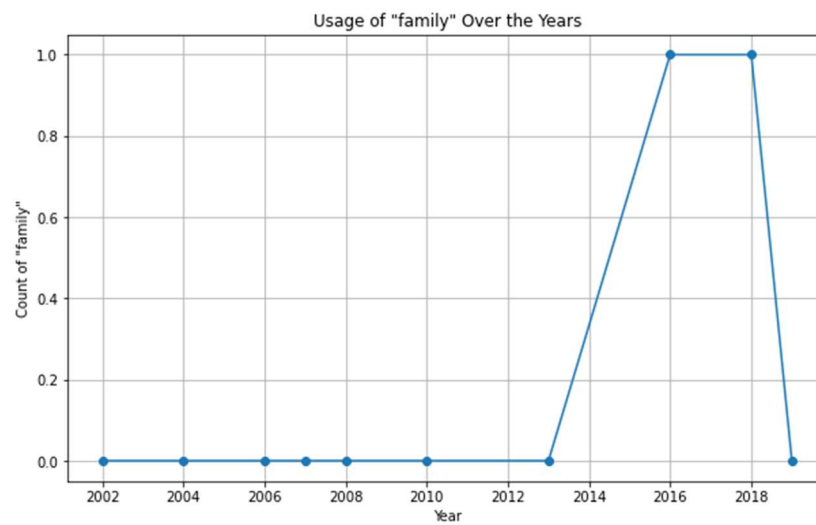
Themes evolution throughout the years:
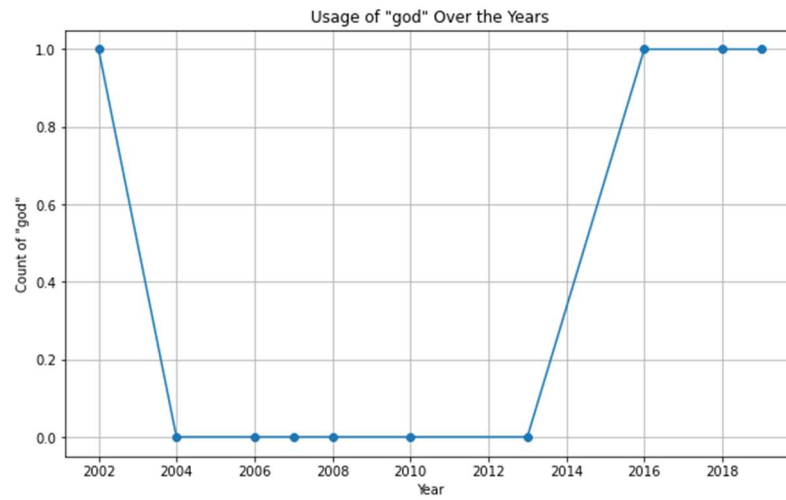
Theme: "god"

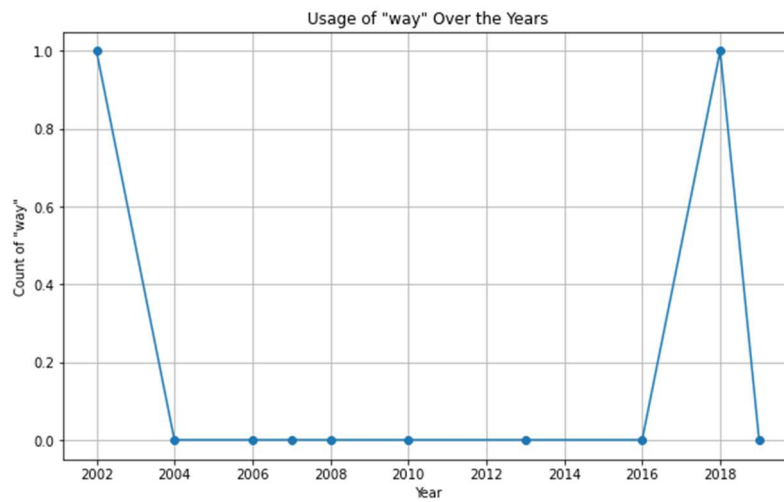Year: 2002,2016,2018,2019

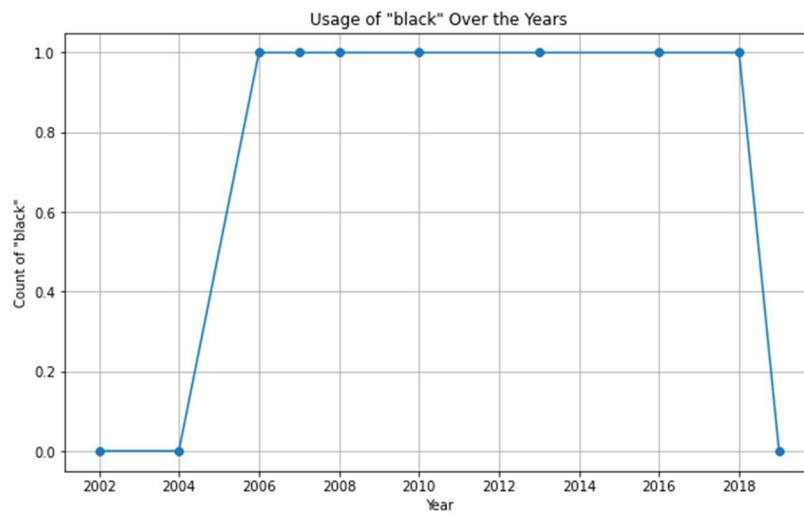Overall Count: 4

Theme: "family"

Year: 2016,2018

Overall Count: 2
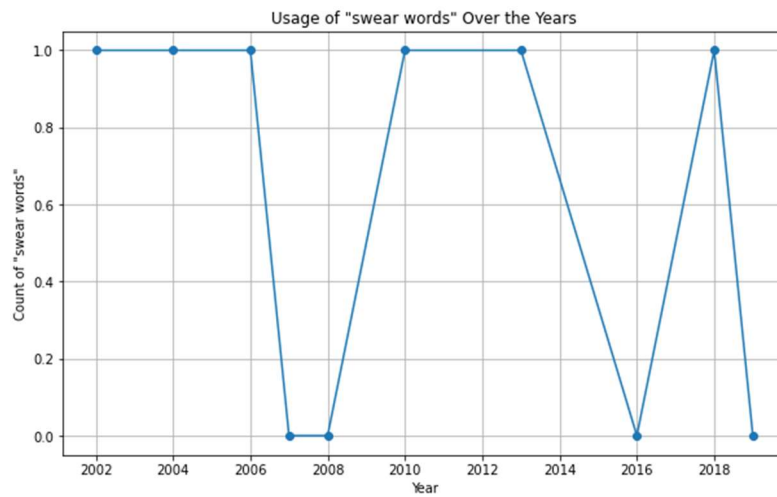
Year: 2002,2018

Overall Count: 2

Theme: "black"
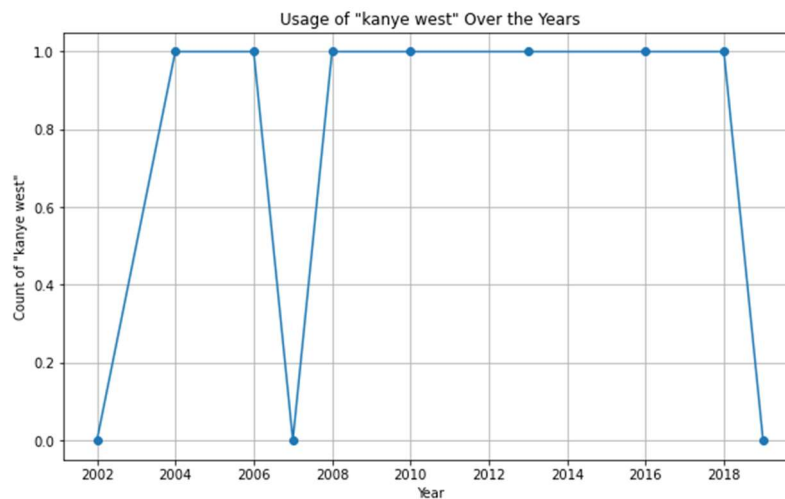
Year: 2006, 2007,2008,2010,2013,2016,2018

Overall Count: 7
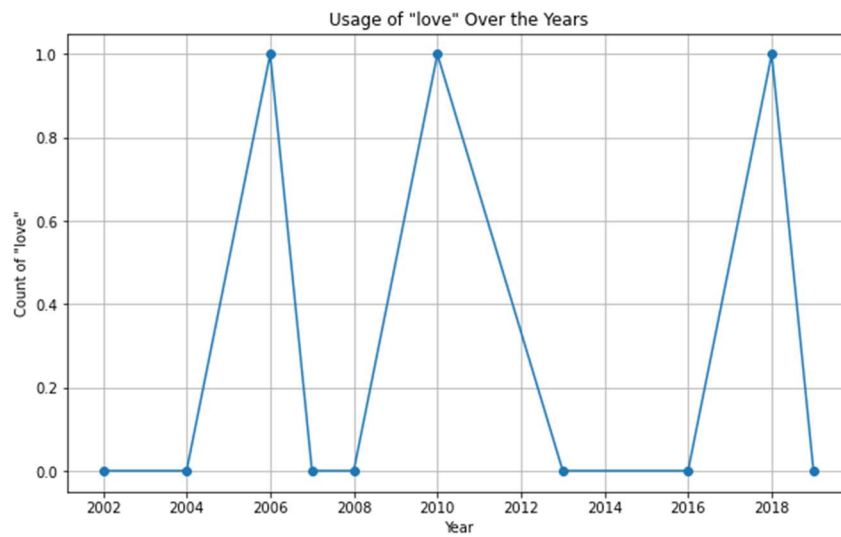
Theme: "swear words"

Year: 2002,2004,2006,2010,2013,2018

Overall Count: 6

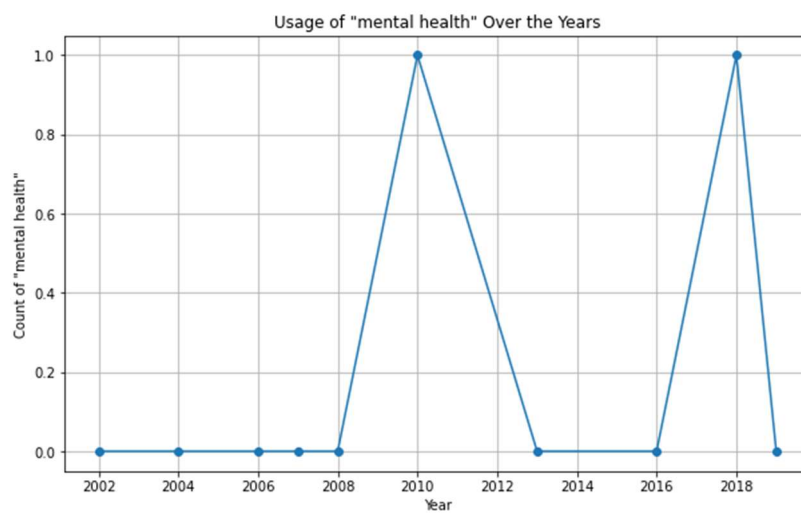Theme: "kanye west"

Year: 2004,2006,2008,2010,2016,2018
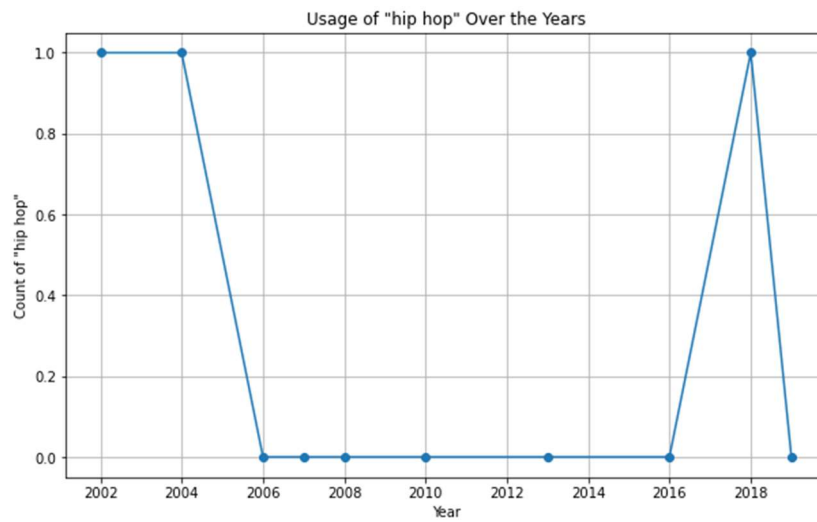
Overall Count: 7

Usage of "love" Over the Years

Theme: "love"

Year: 2006,2010,2018

Overall Count: 3

Usage of "mental health" Over the Years

Theme: "mental health"

Year: 2010, 2018

Overall Count: 2

Usage of "hip hop" Over the Years

Theme: "hip hop"

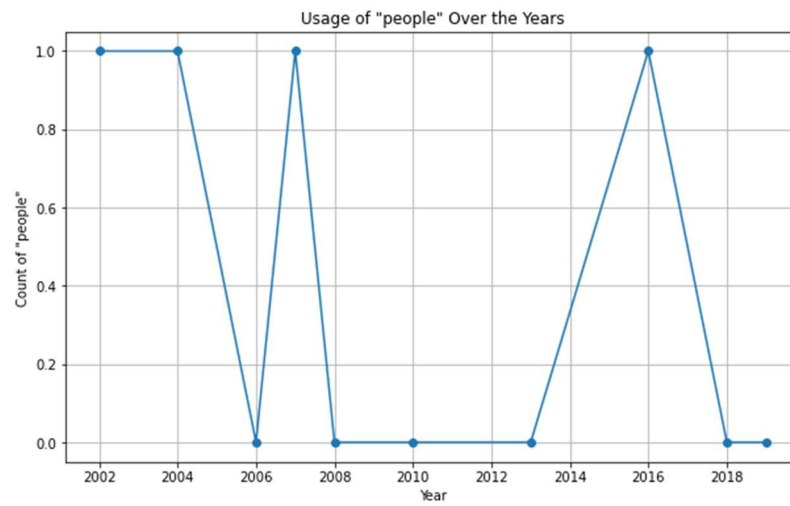Year: 2002, 2004,2018

Overall Count: 3

Usage of "money" Over the Years
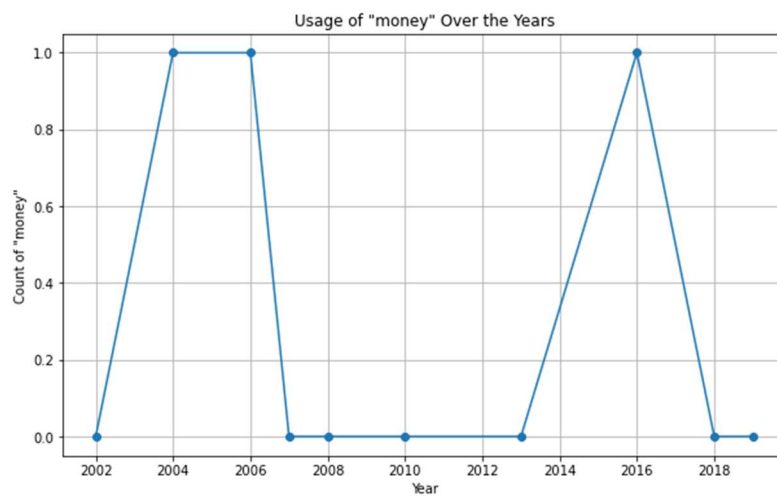
Theme: "money"
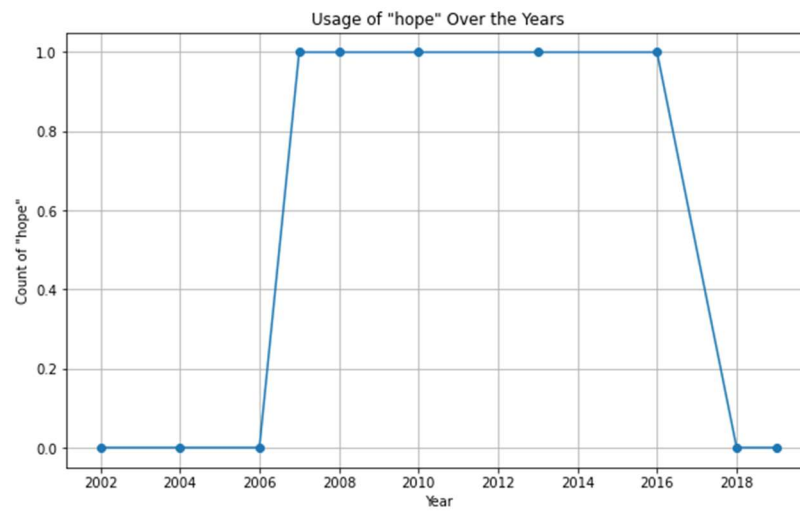
Year: 2004, 2006,2016

Overall Count: 3
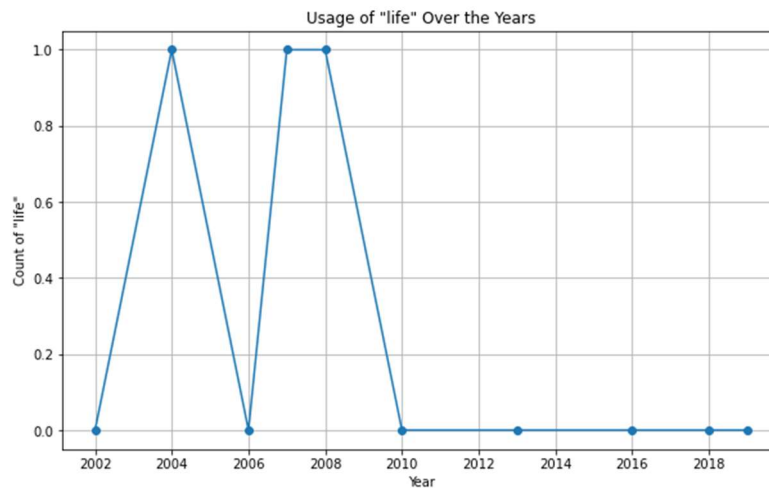
Theme: "people"

Year: 2002,2004,2007,2016

Overall Count: 4

Theme: "hope"

Year: 2007,2008, 2010, 2013, 2016

Overall Count: 5

Theme: "life"

Year: 2004,2007, 'life',2008

Overall Count: 3

- "God" appears consistently from early albums in 2002 to his faith-based album in 2019, highlighting his enduring relationship with spirituality.
- "Family" emerges later in 2016-2018, corresponding to his marriage and having children.
- "Black" identity is a regular theme across his career, peaking in 2010, reflecting his lived experience as an African American.
- Profanity ("swear words") arises often, except for his gospel album in 2019, indicative of his explicit style.
- Self-referential "Kanye West" themes persist through the years, underlining his narcissistic tendencies.
- "Love" surfaces sporadically in 2006, 2010 and 2018, associated with relationships.
- "Mental health" uniquely appears in 2010 and 2018, tying to his diagnosis and struggles.
- Core hip hop culture ("hip hop") and aspirations ("hope") arise in his early and later albums.
- Financial aspirations ("money") peaked in his early fame during 2004-2006.
- His focus on connecting with "people" declines over time.

Years differences and similarities


Yearly Theme Similarity Heatmap
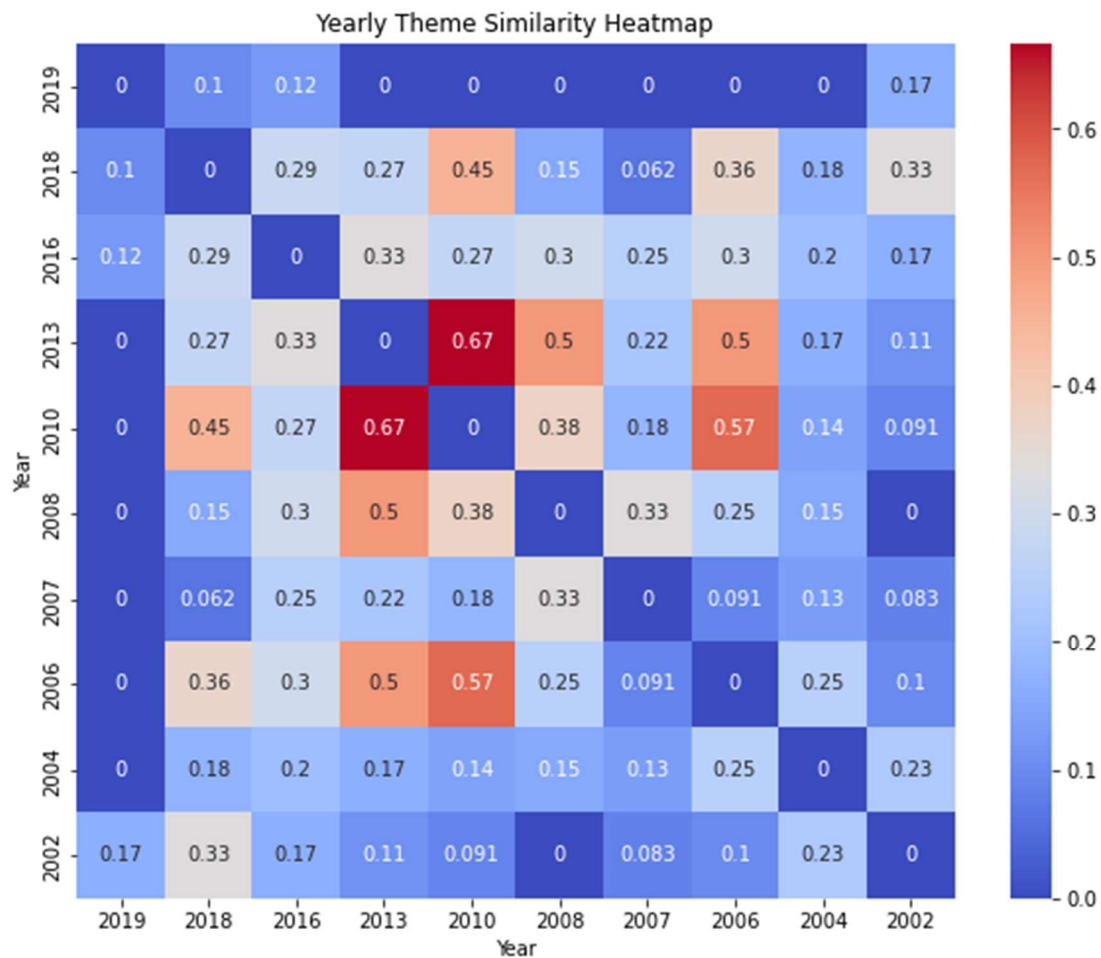
This plot shows Jaccard similarity between themes (sets of themes) for different years as a heatmap.

Notable similarities are between the years 2019 and 2004,2006,2007,2008,2010,2013. The only theme in 2019 is god and it appears as a theme also in these years .

The most different years are 2010 and 2013., noticeably different is also 2013 and 2008,2006 and 2010 and 2018.

Strong similarities can be found between 2002 and 2010, 2007 and 2006 and 2007 and 2002.

Yearly Theme Similarity Dendrogram
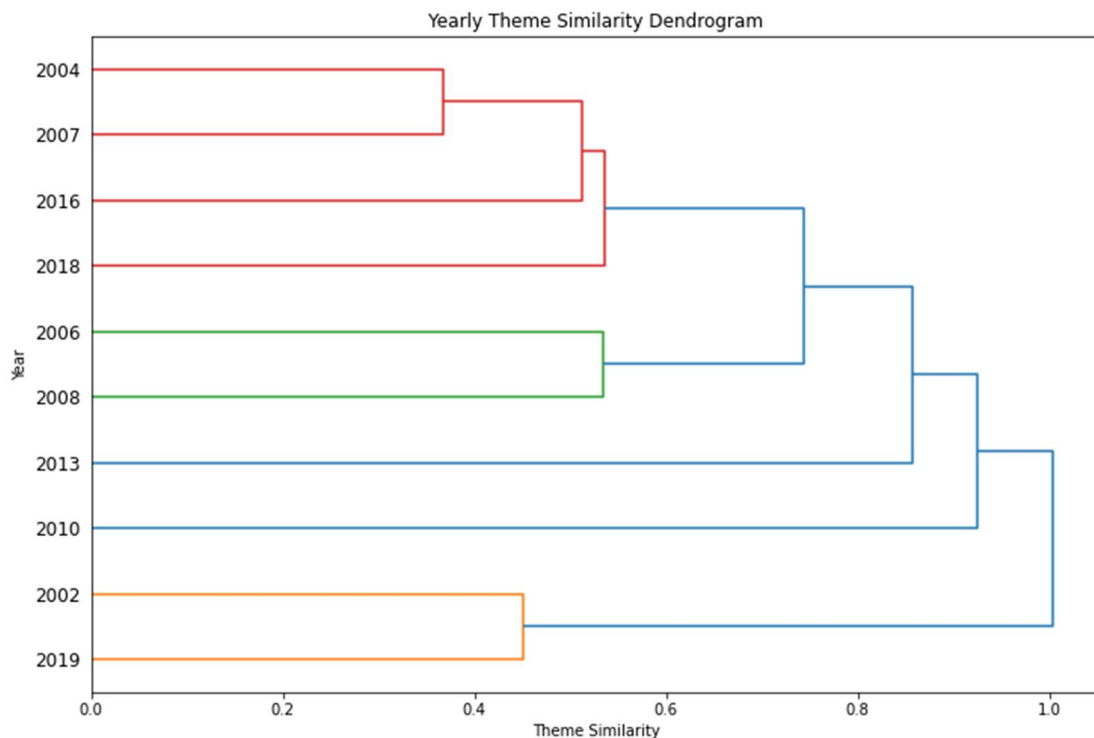
The dendrogram shows clearly the similarities between the years. 2019 and 2002 are clustered together, 2008 and 2006 are clustered together, 2004 and 2007 are clustered together. A weaker cluster but still notable is 2004,2007,2016 and 2018. 2010 is the most different year.

The high similarity between 2019 (only "God" theme) and years like 2004, 2006, etc. reflects the presence of the "God" theme across multiple albums, highlighting Kanye's enduring spirituality.

The dissimilarity between 2010 (introspective themes like "mental health") and 2013 (abrasive themes like "swear words") mirrors differences in his personal state of mind during those albums.

Similarity between 2002 (introspective themes like "God", "people") and 2010 (introspective themes like "mental health", "hope") aligns with these being personal reflection-focused albums.

Dissimilarity between 2013 ("swear words", "hope") and 2008 ("life", "hope") fits with 2013 having raw themes while 2008 was more uplifting during a positive relationship.

Clustering of 2019 and 2002 makes sense given both albums' focus on the "God" theme and faith.

2004-2007 cluster reflects relatable themes like "people", "life" during his early fame.

Loose 2004-2007-2016-2018 cluster fits with evolution to family themes like "love" in later years.

2010 being most distinctive aligns with unique personal themes like "mental health" that year.

The Most common pairings for themes:

Theme: "black"
  Pairing: "kanye west", Count: 4
  Pairing: "hope", Count: 3
  Pairing: "swear words", Count: 3

Theme: "family"
  Pairing: "black", Count: 2
  Pairing: "god", Count: 2
  Pairing: "kanye west", Count: 2

Theme: "god"
  Pairing: "hip hop", Count: 2
  Pairing: "kanye west", Count: 2
  Pairing: "people", Count: 2
  Pairing: "swear words", Count: 2

Theme: "hope"
  Pairing: "black", Count: 2

Theme: "kanye west"
  Pairing: "swear words", Count: 4
  Pairing: "hope", Count: 3
  Pairing: "black", Count: 2

Theme: "life"
  Pairing: "black", Count: 2
  Pairing: "hope", Count: 2
  Pairing: "kanye west", Count: 2
  Pairing: "people", Count: 2

Theme: "love"

Pairing: "black", Count: 2

Pairing: "kanye west", Count: 2

Pairing: "mental health", Count: 2

Pairing: "swear words", Count: 2

Theme: "money"

Pairing: "kanye west", Count: 2

Pairing: "people", Count: 2

Pairing: "swear words", Count: 2

Theme: "people"

Pairing: "hip hop", Count: 2

Pairing: "hope", Count: 2

Pairing: "swear words", Count: 2

Theme: "swear words"

Pairing: "hip hop", Count: 2

Pairing: "hope", Count: 2

Theme: "way"

Pairing: "hip hop", Count: 2

Pairing: "swear words", Count: 2

"Black" identity frequently intersects with his persona ("kanye west"), aspirations ("hope"), and explicit lyricism ("swear words").

"Family" themes often occur alongside spirituality ("god"), race ("black"), and self ("kanye west"), showing his core values.

"God" regularly pairs with hip hop culture ("hip hop"), his persona, relatability ("people"), and profanity, highlighting the duality.

His persona ("kanye west") regularly intertwines with profanity, aspirations ("hope"), and race ("black").

"Life" pairs with race, goals, persona, and relatability, underlying these as core themes.

"Love" links to race, persona, mental health, and explicit lyricism, showing its multilayered nature.

Wealth ("money") overlaps with self and others ("people"), underlining materialism.

Connecting with "people" intersects with hip hop, goals, and profanity.

His evolving "way" connects to hip hop origins and unfiltered self-expression ("swear words")

5. Final Discussion and Conclusions

Within the framework of this seminar, our primary objective was to conduct a thorough topic analysis of Kanye West's musical compositions.
To accomplish this goal, we collected data spanning the years 2002 to 2019, which included the following details by song: song name, album name, year and lyrics.

Our process began with a thorough literature review, covering Kanye West's personal and professional life, the transition from close to distant reading, and its connection to topic analysis. Lastly, we offered a detailed overview of our primary research focus: topic analysis and its complexities.

Upon recognizing the necessity of conducting our analysis by album (year) to establish a coherent timeline linking events in Kanye's life to his music, our journey began with a meticulous cleaning and preparation of our data. Subsequently, we delved into data exploration to extract initial insights and finally, we reached the phase of topic modeling, which stands as the focal point of our research.

Following our data exploration, several notable findings have emerged. The word cloud analysis prominently revealed Kanye West's recurrent self-reference through the extensive utilization of his full name. Also, the presence of the derogatory term "nigga," suggested a complex interplay between his lyrics and his identity as a Black individual raised by a single mother in Chicago.

In our analysis of text characteristics per album, we aimed to identify shifts in thematic content related to Kanye's personal life. However, we found

consistent themes like religion, narcissism, and certain derogatory words across most albums, challenging our expectations of textual variation. Additionally, our n-gram analysis revealed recurring patterns, with Kanye's full name as the most frequent word pair in most albums, reinforcing his self-centered lyrical style.

Our topic modeling showed the same. Unsurprisingly, the most prevalent themes in Kanye West's albums are references to himself ("Kanye West") and racial identity ("Black"), each appearing in 7 albums. This underscores Kanye's focus on his persona and his lived experience as an African American. Profanity ("Swear Words") also, consistently appears in 6 albums, reflecting Kanye's explicit lyrical style and unfiltered self-expression. Themes of "Hope" emerge across 5 albums, highlighting Kanye's messages of perseverance and optimism in the face of adversity. Spirituality ("God") and relatability ("People") appear in 4 albums each, indicating Kanye's exploration of faith and his desire to connect with his audience.

The themes also reflect different phases in Kanye's life: In 2019, the theme of "God" reflects his public embrace of Christianity, culminating in the "Jesus is King" album after his mental health struggles. In 2018 and 2016, themes related to "Family" correlate with his stability with Kim Kardashian and their children, while "Mental Health" connects to his bipolar outburst that were prevalent that year as indicated by his and Kim Kardashian twitter account. In 2008, "Wrong" and "Hope" fit his regret about past misdeeds during the reflective "808s & Heartbreak" album that was made a year after the Taylor Swift incident at the VMAs. In 2007, "Haters" and "Show" reflected his drive to prove critics wrong at his pop culture peak with "Graduation." In 2004, "School", "Work" and "child" themes fit Kanye's journey to stardom after producing for years while striving to make his own album and being young.

Some themes evolve throughout Kanye's career: "God" appears consistently from early albums in 2002 to his faith-based album in 2019, highlighting his enduring relationship with spirituality. "Family" emerges later in 2016-2018, corresponding to his marriage and having children. "Black" identity is a regular

theme across his career, peaking in 2010, reflecting his lived experience as an African American. Profanity ("Swear Words") arises often, except for his gospel album in 2019, indicative of his explicit style. Self-referential "Kanye West" themes persist through the years, underlining his narcissistic tendencies. "Love" surfaces sporadically in 2006, 2010, and 2018, associated with relationships. "Mental Health" uniquely appears in 2010 and 2018, tying to his diagnosis and struggles. Core hip hop culture ("Hip Hop") and aspirations ("Hope") arise in his early and later albums. Financial aspirations ("Money") peaked in his early fame during 2004-2006. His focus on connecting with "People" declines over time.

The textual analysis clearly reveals that the themes and topics in Kanye West's lyrics have undergone significant evolution over his musical journey from 2002 to 2019. While certain core themes like self-reference, spirituality, profanity persist, Kanye's thematic focus has shifted across different phases of his personal life and career. Themes around family and mental health emerge later as he marries and has children, while explicit lyrics diminish with his faith album. Racial identity remains prevalent throughout. His early work centers on aspirational and materialistic topics which later give way to vulnerability. The changing patterns of topics quantitatively demonstrate how his lyrics act as a mirror to his life stages.

The hypothesis is supported by the computational text analysis conducted in this project. The identified topics and themes in Kanye's albums from 2002 to 2019 exhibit clear shifts that can be mapped to evolutions in his personal circumstances, mindsets, and musical experimentation over those years. The changing lyrical content across his discography closely mirrors his personal highs and lows, controversies, changing priorities, and musical growth spanning nearly two decades, quantitatively validating the hypothesis.

6.In summary

In summary, our seminar's extensive analysis of Kanye West's musical compositions has revealed compelling insights into the interplay between his life experiences and lyrical themes. Through meticulous data exploration and

text analysis models, we have gained a deeper understanding of his recurrent self-reference, the influence of his identity on his lyrics, and the consistency of thematic elements across his albums.

Distant reading through computational text analysis provides invaluable new perspectives for examining rap and hip-hop music and culture. By utilizing large digital corpora and algorithms like topic modeling, researchers can uncover key thematic trends and evolutions within hip hop lyrics over time and across regions. As Moretti (2013) argues, distant reading enables a "scaling up" of analysis to macroscopic levels impossible through traditional close reading of individual texts. This scalability allows exploration of linguistic patterns spanning decades of hip hop history and thousands of songs. Topic modeling in particular has shown immense promise for rap analysis by automatically detecting latent topics frequently occurring across a corpus of lyrics (Mohr & Bogdanov, 2013).

By exponentially expanding the analytical scope, distant reading and computational text analysis provide invaluable new tools for humanities researchers exploring hip hop as a vital cultural institution giving voice to marginalized communities. The macroscopic view reveals insights undetectable through a narrow textual focus. As this seminar has illustrated through modeling Kanye West's lyrics, these methods showcase immense potential for illuminating rap music's rich cultural legacy.

7. Appendix

Instructions to run the programs:

<u>Prerequisites:</u>

Python 3.

IDE for running Python programs.

Required Python packages installed - pandas, matplotlib, seaborn, gensim, sklearn, nltk.

Open the folder containing the programs and datasets. It will have:

- Program 1: kanye_topicmodeling.py
- Program 2: results visualizations.py
- Dataset 1: KanyeLyrics.xlsx
- Dataset 2: themes.xlsx

Instructions:

1. Open kanye_topicmodeling.py in your preferred IDE.
2. Install any missing imported python packages.
3. Set the path to dataset on your system by updating line 25:
4. data = pd.read_excel("path/to/KanyeLyrics.xlsx")
5. Run the complete kanye_topicmodeling.py to generate topics, word clouds, similarity analysis for the lyrics.
6. It saves the results to an excel file in the same folder.
7. After running kanye_topicmodeling.py, open results visualizations.py
8. in your IDE.
9. Update the dataset file path at line 22:
10. df = pd.read_excel("path/to/ themes.xlsx")
11. Run results visualizations.py to perform additional theme analysis on the lyrics over time.
12. It will print out all the theme analysis results.
13. Refer to the printed outputs and generated excel files to see the complete analysis from both programs.

8. Bibliography

Bennett, M. (2018). Kanye West, Donald Trump and the politics of Dragon Energy. Music & Politics, 12(2). https://doi.org/10.3998/mp.9460447.0012.204

Bernard, S. C. (2020). Kanye West's politics are as impulsive as his music. National Review. https://www.nationalreview.com/2020/07/kanye-wests-politics-are-as-impulsive-as-his-music/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993-1022.

Brice-Finch, J. (2009). Kanye West's construction of hip-hop masculinity. Popular Musicology Online. https://www.popular-musicology-online.com/issues/02/brice-finch.html

Caramanica, J. (2014). The collapse of Kanye West. Pitchfork. https://pitchfork.com/features/profile/9582-the-collapstation-of-kanye-west/

CBC. (2016, November 21). Kanye West hospitalized in Los Angeles after canceling U.S. concert tour. CBC News. https://www.cbc.ca/news/entertainment/kanye-west-hospitalized-1.3861292

Cole, M. (2016). Kanye West reveres and rages against the mothers who made him. Pitchfork. https://pitchfork.com/thepitch/1059-kanye-west-reveres-and-rages-against-the-mothers-who-made-him/

Elich, K. (2022, October 4). Kanye West criticized for wearing 'White Lives Matter' shirt at surprise Yeezy fashion show. CNN. https://www.cnn.com/2022/10/04/entertainment/kanye-west-white-lives-matter-shirt/index.html

ESPN. (2022, October 10). Kanye West to lose Adidas partnership after antisemitic messages. ESPN. https://www.espn.com/nba/story/_/id/34959763/kanye-west-lose-adidas-partnership-antisemitic-messages

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research, 22, 457-479.

Forbes. (2022). Kanye West. Forbes. Retrieved February 23, 2023, from https://www.forbes.com/profile/kanye-west/?sh=405b39c6ebd9

Genius. (2023). Kanye West Lyrics. Genius Media Group. https://genius.com/artists/Kanye-west

Guardian. (2020, July 21). Kim Kardashian asks for 'compassion' as Kanye West 'suffers bipolar episode'. The Guardian. https://www.theguardian.com/tv-and-radio/2020/jul/21/kim-kardashian-kanye-west-bipolar-episode

Jockers, M. L. (2013). Macroanalysis: Digital Methods and Literary History. University of Illinois Press.

Johns, T. (2019, May 7). Living with bipolar disorder: Kanye West brings attention to mental health struggles. USA Today. https://www.usatoday.com/story/life/2019/05/07/kanye-wests-bipolar-disorder-what-know-living-mental-health/1125541001/

Johnson, T. (2018, October 13). Kanye West goes on a rant in the Oval Office. Here's what happened. SFGate. https://www.sfgate.com/politics/article/Kanye-West-Oval-Office-rant-Trump-13291092.php

Kelley, C. (2022, November 20). Kanye West's bizarre behavior curbs, but doesn't end his cultural cachet. The Conversation. https://theconversation.com/kanye-wests-bizarre-behavior-curbs-but-doesnt-end-his-cultural-cachet-194248

Kubrin, C. E. (2005). Gangstas, thugs, and hustlas: Identity and the code of the street in rap music. Social Problems, 52(3), 360-378. https://doi.org/10.1525/sp.2005.52.3.360

McIntyre, H. (2009, September 14). Kanye West storms the VMAs stage during Taylor Swift's speech. Billboard. https://www.billboard.com/music/music-news/kanye-west-storms-the-vmas-stage-during-taylor-swifts-speech-1004077571/

McLeod, K. (2019). Exploring Kanye West's 'My Beautiful Dark Twisted Fantasy'. Pitchfork. https://pitchfork.com/features/article/exploring-kanye-wests-my-beautiful-dark-twisted-fantasy/

Moretti, F. (2013). Distant Reading. Verso.

Randall, M. (2009). Kanye West's vocal delivery: "Through the wire" and "touch the sky". Popular Musicology Online. https://www.popular-musicology-online.com/issues/05/randall.html

Recording Industry Association of America. (2022). RIAA Gold & Platinum. https://www.riaa.com/gold-platinum/?tab_active=top_tallies&ttt=TAA#search_section

TMZ. [TMZSports]. (2018, May 1). Kanye West says 400 years of slavery was a choice [Video]. Youtube. https://www.youtube.com/watch?v=s_M4LkYra5k

Twitter. [@kanyewest]. (2023). https://twitter.com/kanyewest

Zenger, A. A. (2012). Remembering tradition in the present: Kanye West's "Blood on the leaves". Journal of Religion & Society, 14. https://d1wqtxts1xzle7.cloudfront.net/34770728/Remembering_Tradition_in_the_Presen20160328-28155-1blyalf.pdf?1459195119=&response-content-disposition=inline%3B+filename%3DRemembering_Tradition_in_the_Present_Kan.pdf&Expires=1675929668&Signature=RF~1GYipIFdw6w3mm5FtiENzgY9jY3Tabg9Khcb-uY~uzRgaPQue~LtIBKfK6~vSXme~8gbxwa~Jgu~YTsgpXaKf~EcSUCfWy2xliEYYbpLIN~lZKXTD3su-VxTVDupIj~UoEAJRKYT7Pmqg6Y~0Oecz~CDNUXcxl~F-V

Zisook, B. (2022, August 5). How Kanye West Changed Hip-Hop. DJBooth. https://djbooth.net/features/2022-08-05-how-kanye-west-changed-hip-hop