Understanding Taylor Swift's Popularity: A Data-Driven Approach to Music Analysis

Raquel Chaves Martínez[a], **Alessia Saccardo**[a], **and Paula Gambús i Moreno**[a]

This manuscript was compiled on December 11, 2024

Taylor Swift's music and public persona have had a profound cultural impact, making her a subject of interest in network analysis and sentiment studies. Her extensive discography, including re-released "Taylor's Version" albums, provides a rich dataset for exploring lyrical patterns and thematic evolution. Previous studies have examined the cultural significance of her music, but the underlying network dynamics of song relationships remain underexplored. This study addresses the question of how lyrical similarity and thematic elements connect songs within Taylor Swift's discography. Using a network analysis approach, songs were represented as nodes, with directed edges linking earlier songs to later ones based on shared tf-idf-derived keywords. Community detection via the Louvain algorithm and album-based groupings revealed distinct clusters, with the Louvain partition yielding nine communities and a modularity score of 0.47, indicating strong thematic coherence.

Our findings highlight significant overlaps between album-based and Louvain communities, with most groups sharing common lyrical themes while a subset displayed unique patterns. Sentiment analysis of song lyrics revealed predominantly neutral to positive tones, with variations reflecting emotional highs and lows. Centrality measures did not strongly correlate with Spotify streaming numbers, suggesting that lyrical structure alone does not determine popularity.

This study demonstrates demonstrates how network analysis can uncover patterns of influence and thematic cohesion in music, offering a deeper understanding of Taylor Swift's enduring appeal.

Taylor Swift | Network analysis | Lyrics analysis |

Significance Statement

This study investigates the relationships and thematic patterns within Taylor Swift's music through network and sentiment analysis. By representing songs as a network and analyzing their lyrical similarities, this research highlights how shared thematic elements connect tracks across different albums, including re-released "Taylor's Version" ones. Sentiment analysis reveals the emotional tones of her songs, while centrality measures offer insights into how network structure influences their popularity. The findings provide a deeper understanding of Swift's cultural impact and offer a novel approach for analyzing music through network-based methods. This research is valuable for exploring how emotions and network dynamics influence the broader landscape of popular culture, offering insights into how these factors shape the impact of artists like Taylor Swift.

Taylor Swift's influence extends far beyond the realm of music, spreading culture, economy and even politics. She has broken records with her tours, impacted economic trends and even influenced elections. Moreover, prestigious universities have recognized her cultural significance by offering courses on her work (1). In 2024, she was named Spotify's top artist for the second year in a row, while simultaneously having three of her albums rank among the top 10 most-streamed globally.(2).

On her career as a well-known pop and country singer, Taylor Swift has produced an abundance of highly acclaimed music. She has received a total of 14 Grammy Awards from 52 nominations and is the first solo female artist to win Album of the Year twice (3). Her music has attracted such popularity because of its treatment of a wide variety of interpersonal relationships.

In recent years, Swift has re-released several of her albums, labeled as "Taylor's Version." This decision was driven by disputes over the ownership of her original recordings. By re-recording and re-releasing these albums, she regained control over her music, ensuring that the profits and rights remained with her (4). All this process was very successful for her because the support that she received from her fans was huge.

In previous work, such as the E. Dubrofsky's article (5), the cultural and social dimensions of her public persona were examined. It highlights how Taylor Swift navigates and constructs her identity within the public sphera, emphasizing the interplay between her image and her music.

Our study aims to uncover patterns in Taylor Swift's music, providing insights into the lyrical and thematic elements that contribute to her widespread popularity. Thanks to the use of Network Analysis (6) and by examining the centrality measures of the songs and their correlation with Spotify streaming numbers, we seek to understand the underlying factors that drive the success of her re-released tracks and overall musical repertoire.

The sentiment expressed in Taylor Swift's songs is a critical aspect of her widespread appeal. Her lyrics often explore a wide spectrum of emotions, from joy and love to hearbreak and sadness. Analyzing the sentiment of her songs could

Author affiliations: [a]Technical University of Denmark

allow to go deeper into how these tones resonate with her audience and popularity. Therefore, it could be possible to uncover ways in which her songs connect with her listeners on a personal level (7).

This study began by examining the data to uncover relationships between Taylor Swift's songs. By constructing a network based on song similarity, patterns and connections were identified that highlighted thematic and stylistic clusters within her music. This network-based approach facilitated the visualization of the structure of her repertoire and emphasized songs with significant roles.

The emotional tone of the lyrics was then examined, focusing on how different sentiments resonate with listeners. This analysis provided insights into the emotional depth of her music and its connection to her widespread appeal.

Finally, the relationship between streaming popularity, sentiment, and critical reception was investigated. This analysis explored how emotional tones and external evaluations influence the success of her songs, offering a deeper understanding of the factors driving her impact as an artist.

Results

Preliminary analysis. After the Data collection, an analysis was performed to know the characteristics of the dataset and extract some information from it. With a clearer understanding of Taylor Swift's background, different plots according the Spotify streams, the genre, the album and the track name, were generated to understand how her songs impacted to the society. The results showed notable trends.

Regarding to the number of streams, the original and Taylor's Version albums have similar popularity. At the same time, the "Live" albums got the lowest stream numbers. Additionally, in 2020 there is a stand out with high streaming counts due to the release of *Folklore* and *Evermore*. In 2017 there is also a peak due to the release of *Reputation* after a three-year gap. Finally, in 2023 she re-released two of her original albums which lead to rekindled engagement.

Furthermore, the genre analysis confirms the dominance of Country Pop in Taylor Swift's discography, even though she also sings Synth Pop and Pop Rock. Album length tends to be similar across her repertoire, except for the *Stadium Tour Surprise Song Playlist*, which has significantly more songs due to its comprehensive nature.

Initially, the dataset contained duplicates that were straightforward to identify and remove, as they appeared in the dataset with identical `track_name` entries. However, during the computation of cosine similarity based on TF-IDF scores, additional duplicates were discovered. These were songs that had been republished under different names, often because they were included in different albums. These duplicates were identified by their similarity score of 1, indicating that their lyrics were exactly the same. To address this, the corresponding rows were merged. During the merging process, each field was compared: if the values were identical, one was retained; if they differed, all unique values were combined into a list.

Network creation. To construct the network, songs were represented as nodes, with directed edges connecting them based on their lyrical similarity. Each node in the network was assigned attributes including release date, album, Spotify streaming numbers, and genre. A directed edge was added between two songs if they shared at least 3 out of their top 20 most common words, determined using TF-IDF scores. The direction of the edges captured the chronological relationship between songs, linking earlier releases to later ones. This design was guided by the hypothesis that earlier songs might have influenced or inspired subsequent releases, allowing potential patterns of lyrical or thematic evolution to be observed over time.

The threshold of three common words for edge creation was determined through trial and error. This approach ensured the resulting network had a sufficient number of edges to enable meaningful analysis while avoiding an overly dense or sparse graph.

Community analysis. Communities were identified using both the Louvain algorithm and album-based groupings on the same subset of the network, which included all songs from albums with Metacritic reviews. This approach ensured consistency with subsequent analyses. The Louvain community partition, applied to the giant connected component, produced 9 distinct communities with a modularity score of 0.47, indicating well-defined and consistent groupings, even though this may not appear evident from Figure 2. Each community was named after the song with the highest degree within it.

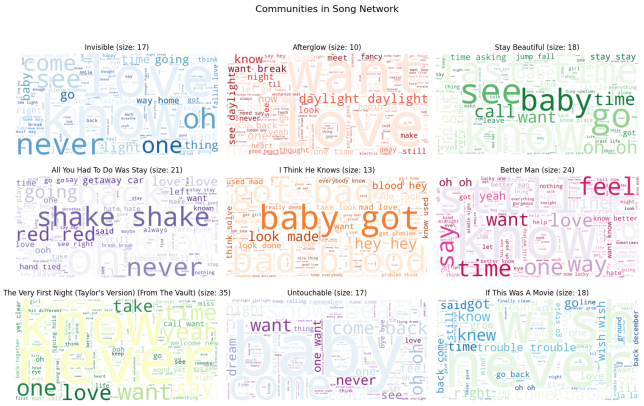

Fig. 1. Word clouds of Louvain communities in the Taylor Swift song network, named after the highest-degree song. Word size reflects frequency, highlighting recurring themes like *love*, *know* and *never* with distinct patterns in each community.

In contrast, modularity could not be calculated for the album-based communities due to the presence of overlapping communities. Analysis of word clouds and the most common words for each community revealed a similar thematic distribution across both partitioning methods. For instance, in the word cloud generated for the Louvain partitions (see Figure 1), most communities shared recurring common words such as *know*, *love*, and *baby*, prominently featured in communities like *Invisible*, *Afterglow*, *Stay Beautiful*, *Better Man*, *The Very First Night*, *Untouchable*, and *If This Was a Movie*. Conversely, the final two communities, *All You Had to Do Was Stay* and *I Think He Knows*, exhibited distinct common words, including *Shake Shake*, *Red Red*, and *Bad Blood*. These variations are likely due to the repeated use of certain terms within these communities, making those words disproportionately prominent.

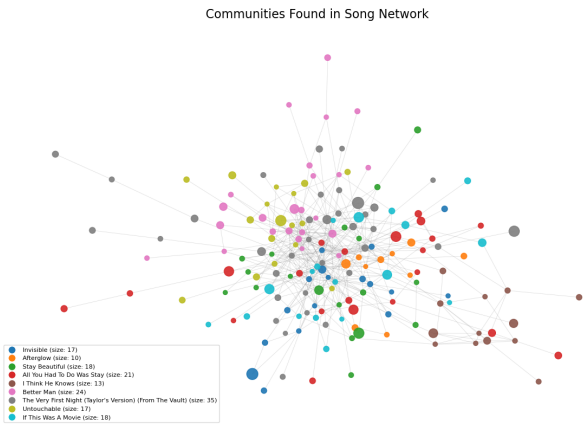

Fig. 2. Communities in the song network, with nodes representing songs and edges denoting lyrical similarity. Node colors indicate Louvain-detected communities, and node sizes are proportional to Spotify streams. The legend lists community names and sizes.

Centrality analysis. To understand the relationship between the structural properties of Taylor Swift's songs in a network and their popularity on Spotify, an analysis focusing on the centrality measures and stream numbers was performed to determine whether network centrality could serve as indicators of a song's popularity. For this reason, the correlation between these measures and the stream counts was calculated and the pairplot in Figure 3 was generated.

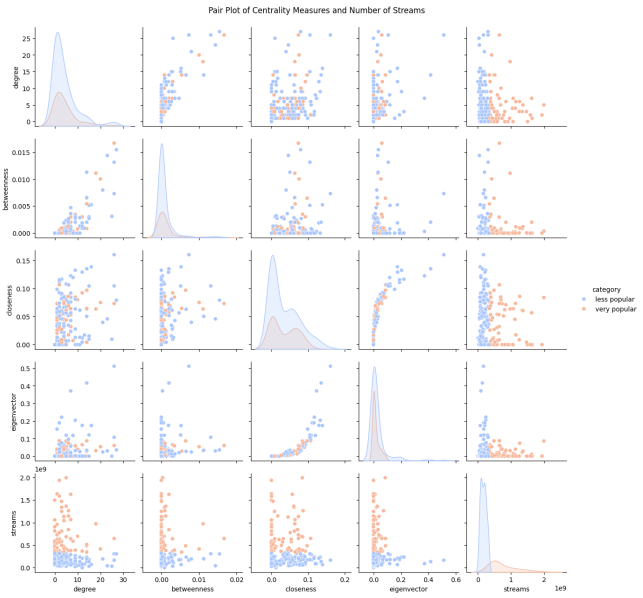

Fig. 3. Pairplot of centrality measures and Spotify Streams of Taylor Swift's songs. The two different categories were created according to the mean and the stream counts.

The obtained correlation reveals no direct relationship between centrality measures and the number of streams. However, certain centrality metrics exhibit a strong correlation, likely because connected songs frequently act as bridges and maintain closer relationships with other songs within the network. Nevertheless, these structural properties do not translate to higher stream counts.

In Figure 3 it can be observed that the degree centrality shows a slight negative correlation with streams, indicating that more shared words does not indicate higher popularity. Moreover, betweenness and eigenvector centralities are low, so they do not impact streams significantly. Finally, closeness centrality shows a weak clustering with higher streams, suggesting a slight advantage for more central songs. Overall, centrality measures do not strongly correlate with stream numbers, suggesting that other factors such as lyrical content, marketing and fan engagement play more significant roles in a song's popularity.

Sentiment analysis. Taylor Swift is known for writing songs that deeply connect with listeners, often exploring love, heartbreak, and personal growth. A sentiment analysis of her song lyrics was performed to understand the emotions behind her storytelling.

To do so, a function to calculate the sentiment of the lyrics based on the **LabMT** (8) wordlist was built. More details regarding the calculation of the sentiment can be found in Sentiment scores and categories. The overall sentiment scores for each song were calculated, revealing values ranging from 4.6 to 6.2, with an average of 5.5. These results suggest that the songs generally convey a neutral tone, leaning slightly toward a more positive sentiment. This indicates that while her lyrics explore various emotions, they often maintain an uplifting or balanced mood.

To further analyze the impact of happy, neutral, and sad words, the percentage of each category was calculated for the 15 most popular songs based on Spotify streams. The results are presented in Figure 4.

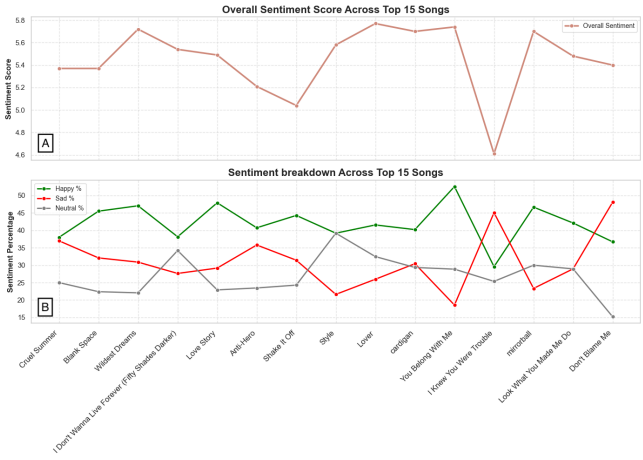

Fig. 4. **A:** Sentiment scores of the top 15 most-streamed songs on Spotify; **B:** Percentage breakdown of sentiment categories (happy, neutral, sad) in the top 15 most-streamed songs on Spotify

As it can be seen, songs like *Lover* and *Wildest dreams* have high sentiment scores, matching with high happy words percentage values, reflecting their optimistic and empowering nature. In contrast, *I knew you were trouble* has the lowest sentiment score, which aligns with a higher sad words percentage and lower happy percentage, reflecting its melancholic tone. This reflects Swift's own description of the song, highlighting her frustration with herself for overlooking warning signs in a relationship, which ultimately

led to heartbreak and regret (9). Neutral sentiment stays consistent, while shifts in happy and sad percentage match the highs and lows of overall sentiment, highlighting how these emotions shape the mood of each song. These neutral sentiments likely reflect lyrical complexity, subtle storytelling, or instrumental arrangements that introduce moderation and refinement.

Popularity and Critical Reception.

Spotify Streams vs. Sentiment Scores. A study to determine if there is a correlation between the sentiment of Taylor Swift's songs lyrics and their streaming popularity on Spotify was performed. To do so, both the total sentiment score and the percentage of happy, neutral and sad words in each song were taken into account.

Since the results obtained using the total sentiment score were not relevant, a second analysis taking into account the word type percentages was performed. In this last case, the findings determined that the sentiment alone is not a strong predictor for a song's popularity due to the presence of outliers and wide dispersion in the plots. However, it could be seen that a weak positive correlation is present relating the streams with "Happy" and "Sad" word percentages, and that songs with higher neutral sentiment percentages might have a slight tendency to have fewer streams.

Spotify Streams vs. Metacritic Scores. Here, the relationship between the popularity of Taylor Swift's albums (measured by Spotify streams) and their critical reception (reflected by Metacritic scores) were investigated. After web-scraping and calculating the average score for each album and summing the streams of each album, the correlation was performed.

The results revealed no significant correlation indicating that critical acclaim does not strongly influence the streaming popularity of her albums.

Original vs. "Taylor's Version" Albums. Furthermore, to inquire more into the topic, a comparison study between the original albums and their re-released "Taylor's Version" ones was done. The analysis showed that the re-released albums generally received higher average Metacritic scores suggesting more favorable critical reception. Despite the original albums having higher total Spotify streams, likely due to their longer presence in the market, the "Taylor's Version" albums still achieved substantial streaming numbers.

All these results reflect a strong engagement from listeners and highlights the success of Taylor Swift's re-release strategy, demonstrating the loyalty and support of her fan base.

Discussion

The analysis of Taylor Swift's music provided insights into the structure and themes of her discography. A network based on lyrical similarities revealed clusters of songs with shared thematic or stylistic elements, emphasizing the interconnectedness within her work. Sentiment analysis further highlighted her exploration of a broad emotional spectrum, with a tendency toward slightly positive tones, which may contribute to her widespread appeal.

No significant correlation was observed between network centrality, sentiment scores, or Metacritic reviews and the streaming popularity of her songs or albums. These findings suggest that external factors such as marketing efforts, fan engagement, and personal connections with listeners likely play a more significant role in driving popularity than structural or critical metrics.

The examination of "Taylor's Version" re-releases demonstrated their success in garnering listener support, despite the originals having been available for longer. This outcome underscores the strong loyalty of Swift's fan base and the effectiveness of her re-release strategy in reclaiming artistic ownership.

Future research could explore the influence of social media, live performances, and cultural trends in amplifying her impact as an artist. Additionally, analyzing the sentiment of web-scraped reviews and correlating them with Taylor Swift's songs could provide deeper insights. Another intriguing focus for future studies could involve investigating the evolution of Taylor Swift's music over the years, considering the release dates of her songs to uncover patterns and shifts in her musical journey.

Materials and Methods

Data collection.

Taylor Swift discography. The Taylor Swift dataset was downloaded from Kaggle the 6th of November of 2024. This dataset has 27 columns and 577 rows, with each row representing a song. The most important columns used are:

- `track_name`: Name of the song
- `track_musical_genre`: Genre of the song
- `album`: Name of the album
- `release_date`: Release date of the album
- `track_lyrics`: Lyrics of the song

Metacritic reviews. Metacritic is a website that compiles reviews for movies, television shows, games and music.

To analyze album reviews, data was scraped from Metacritic using two custom Python functions: `scrape_metacritic_reviews` and `scrape_metacritic_album`. The first function extracts critic reviews from a given album reviews page, taking the URL and the number of reviews to fetch as inputs. It uses the requests library to retrieve the webpage content and BeautifulSoup to parse it, extracting source names, numerical scores, and review bodies, which are returned as a list of tuples. The second function retrieves general album details from its Metacritic page, including the title, number of critics, and the review link. This function calls `scrape_metacritic_reviews` to gather detailed review data.

Centrality measures. Centrality measures are a vital tool for understanding networks because they calculate the importance of any given node in a network. There are four different metrics and all of them were calculated in this project:

- **Degree Centrality:** it represents the number of direct connections a node has.
- **Betweeness Centrality:** it measures the extent to which a node lies on path between other nodes, indicating its role as a bridge within the network.

- **Closeness Centrality:** it indicates how close a node is to all other nodes in the network, reflecting the speed at which information can spread from that node.
- **Eigenvector Centrality:** it measures a node's influence based on the connections of its neighbors, emphasizing the quality of connections

These four centrality metrics were calculated using functions from the NetworkX (nx) library.

Sentiment scores and categories. The sentiment score for the lyrics was obtained based on the scores in the LabMT wordlist. This list was developed by identifying the most common Twitter words and rating them by human evaluation, allowing the measurement of emotional content in text (8). To calculate the overall score of the song lyrics, a weighted average was used (10):

$$S = \frac{\sum_{i=1}^{n} f_i \cdot v_i}{\sum_{i=1}^{n} f_i}$$

Where:

- S is the weighted average
- f_i is the frequence of the ith word
- v_i is the sentiment value of the ith word
- n is the total number of unique words considered

To classify words into sentiment categories, the entire LabMT word set was analyzed, dividing its sentiment scores into three groups: sad, neutral, and happy. The 33rd and 66th percentiles of the score distribution were used as thresholds. Words with scores below the 33rd percentile were classified as "sad", those between the 33rd and 66th percentiles as "neutral", and those above the 66th percentile as "happy". Thresholds were calculated using NumPy's `percentile` function to ensure reproducibility.

GitHub repository. To access the full documentation of the project, access the GitHub repository following this link: Project repository

Contributions

	Alessia	Raquel	Paula
Data collection		x	
Preliminary analysis	x		x
Network creation	x		
Community analysis	x		
Centrality analysis			x
Sentiment analysis		x	
Correlation between sentiment and streams			x
Correlation between streams and critic scores		x	x
Report	x	x	x

References

1. BestColleges, These colleges have taylor swift classes (2024) Accessed: 2024-11-23.
2. Spotify Newsroom, Wrapped 2024: Los artistas, canciones, Álbumes, podcasts y audiolibros más escuchados de 2024 (2024) Accessed: 2024-12-04.
3. Taylor Swift Fandom, Grammy awards (2024) Accessed: 2024-11-23.
4. The Independent, Taylor swift: Timeline of her 'taylor's version' albums (2024) Accessed: 2024-11-23.
5. WG Staples, The vernacular and the politics of identity. *Surveillance & Soc.* **21** (2023) Accessed: 2024-11-23.
6. AL Barabási, *Network Science.* (Cambridge University Press), (2016).
7. RC Flynn, Info 3501: Taylor swift: An analysis of sentiment, variety, and album success (2024) Accessed: 2024-12-11.
8. PS Dodds, KD Harris, IM Kloumann, CA Bliss, CM Danforth, Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLOS ONE* **6**, 1–1 (2011).
9. Taylor Swift Fandom, I knew you were trouble (n.d) Accessed: 2024-12-10.
10. PS Dodds, CM Danforth, Measuring the happiness of Large-Scale written expression: Songs, blogs, and presidents. *J. Happiness Studies"* **11**, 441–456 (2010).