Genre-conformity in the topics of lyrics and song popularity

Anna Aljanaki
University of Tartu
aljanaki@gmail.com

Abstract

The genre of a song defines both musical (rhythmic, timbral, performative) aspects of a song, but also the themes of lyrics and the style of writing. The audience has certain expectations as to emotional and thematic content of the genre they listen to. In this paper we use Music4All database to investigate whether breaking these expectations influences song popularity. We use topic modeling to divide song lyrics into 36 clusters, and apply tag clustering to separate the songs into 15 musical genres. We observe that in some genres (metal, hip-hop) lyrics are mostly written in specific topics, whereas in other genres they are spread over most topics. In most genres, songs that have lyrics that are not representative of the genre, are more popular than songs with lyrics that are more typical for the genre.

1 Introduction

For different listeners, different aspects of a song might be important: rhythm, lyrics, timbre, or epoch that a recording comes from (Huang et al., 2023). Moreover, these principles might vary for different genres and even countries (Schedl et al., 2020). Lyrics seem to be a universally important aspect of a song for many listeners, influencing perceived emotion, enhancing experience and shaping preferences (Alinka Greasley and Sloboda, 2013).

However, lyrics are an often overlooked aspect in music information retrieval. In a recent survey on content-driven music recommendation, 70% of the studies used audio signal in content based music recommender system, while only 30% used any embedded metadata, including lyrics (Deldjoo et al., 2024).

The genre of a song plays an important role in shaping the content of its lyrics. For instance, pop music typically focuses on the topics of romantic love and heartbreak. In hip-hop and rap genres, it is typical for the lyrics to contain slang and obscene vocabulary, and cover topics of social justice and politics.

In (Tsaptsinos, 2017) the songs were classified by genres using lyrics with high accuracy, showing that specific words can be indicative of the genre with a high certainty.

In this paper we will apply topic modeling to song lyrics, and answer the following research questions:

1. How specific are lyrics topics in various genres?
2. Does genre-conformity of the lyrics influence song popularity, and how?

2 Related Work

Lyrics, including topic modeling approaches, have been successfully used for music recommendation (Vystrcilová and Peka, 2020; Patra et al., 2017; Jang et al., 2019; Sasaki et al., 2014). Song popularity prediction was attempted both from audio content (Lee and Lee, 2018) and from lyrics (Martín-Gutiérrez et al., 2020). In (Agatha et al., 2024), lyrics emotional content was estimated using Sentence-BERT transformer, and song popularity was predicted based on that.

3 Data

In this research we will use the lyrics and listening history from music4all database (Santana et al., 2020). This database contains 109269 songs, out of which 91% contain lyrics, and the rest are instrumental. The songs come with more than 5 million listening events, generated by 14127 users.

Out of the songs that have lyrics, 84% are in English, and the rest of the songs are in 44 different languages. For our purposes we do not need instrumental songs and their listening history, hence we discard them. We translate the lyrics in foreign languages into English using Google Cloud Translation API[1]. Some of the English songs contained mixed English and Korean text, and occasional Korean words were translated as well. The cleaned dataset is available at the project repository:[2].

[1] cloud.google.com/translate
[2] github.com/aljanaki/lyrics_topic_analysis

Genre	Amount of songs	Representation kurtosis	Non-representative topics / representative topics	P-value
Rock	22427	-0.99	0.94	0.11
RnB	5140	1.98	0.99	0.96
Punk	3849	-0.6	0.93	0.39
Pop	19555	2.17	**0.75**	**0.01**
New age	1389	14.08	**1.26**	**0.05**
Jazz	959	2.4	**1.45**	**0.05**
Industrial	6033	1.00	1.16	0.06
Hip-hop	4129	27.27	**0.73**	**0.00**
Hardcore	4332	0.18	**1.39**	**0.00**
Funk	5610	11.27	1.08	0.46
Folk	7726	-0.76	1.06	0.38
Electronic	5834	2.65	1.13	0.42
Death metal	6802	8.84	**1.34**	**0.00**
Blues	1089	4.69	1.06	0.75
Black metal	1019	14.34	**1.28**	**0.00**

Table 1: Statistics per genre. Amount of songs: how many songs in the dataset were in that genre. Representation kurtosis: kurtosis computed on song amounts in lyrics topics. Non-representative/representative topics: average play counts ratio in less-representative for this genre divided by average play counts in topics more representative for this genre. P-value: result of the t-test for play counts comparison between these groups.

3.1 Defining genres

In Music4all database, genre labels have a very large cardinality (there are 853 different genres). Most of these genres are represented only by a few songs, whereas each song is usually annotated with several genres, which were scraped from a website by the dataset creators. In order to make analysis by genre possible, we clustered these fine-grained genres using the following process:

1. We computed genre co-occurrence matrix C on the song by genre matrix.
2. Based on matrix C, we computed genre pairwise cosine similarity matrix S.
3. Next, we applied hierarchical clustering of the rows (genres) of S.

In this way, we were able to reduce 853 sparsely used genres to just 15 genre clusters. Clusters were labeled manually by selecting the most frequent parent genre in the cluster. E.g., a cluster containing 'avant-garde black metal', 'greek black metal' and 'usbm', along with 23 other similar tags, was named 'black metal'. Unfortunately, two of the 15 clusters, which we named *new age* and *industrial*, were rather eclectic. 21% of songs in the dataset fell into rock cluster, next by popularity were pop, electronic and folk. The smallest genres are blues, black metal and jazz, which had a little over 1000 songs each.

3.2 Lyrics topic modeling

We applied topic modeling to lyrics of songs using Bertopic[3] approach:

1. We computed sentence embeddings using a MPNet sentence transformer from HuggingFace[4]. The embeddings were computed on a complete lyrics text, treating it as a single paragraph of text, creating 768-dimensional embedding vector that represented each song.
2. We applied dimensionality reduction on these embeddings with UMAP, extracting 50 components using cosine similarity.
3. The dimensionally reduced embeddings were clustered using K-means with $k = 40$.

We also experimented with HDBSCAN and BIRCH clustering algorithms, but they did not result in satisfactory clusters. K-means was able to create clusters of roughly equal size and not as many outliers as HDBSCAN.

We inspected the topics in various ways (extracting influential words with TF-IDF, BERT, and

[3]https://maartengr.github.io/BERTopic/index.html

[4]https://huggingface.co/sentence-transformers/all-mpnet-base-v2

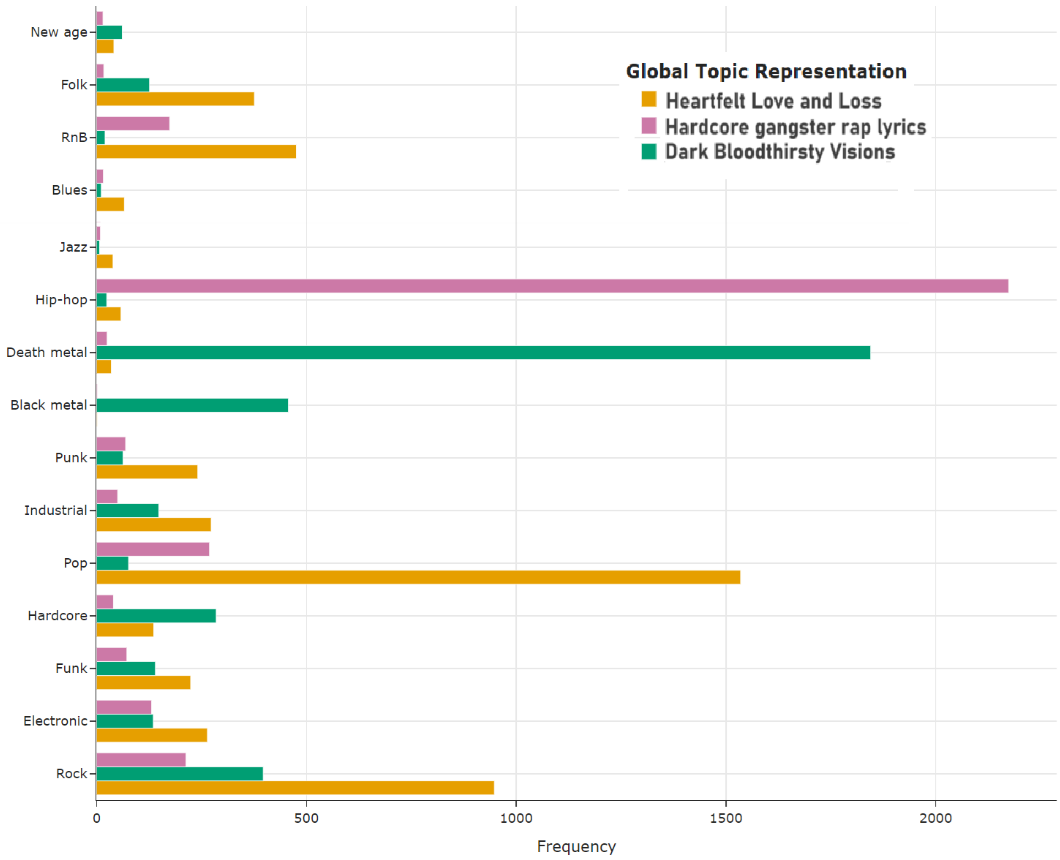

Figure 1: Distribution of three selected lyrics topics over all the genres.

using openAI to label the topics (see Figure 1)). The most popular topics were related to themes of love, unrequited love and breakup. There were also clearly separable topics with obscene lyrics, and lyrics with dark epic themes on death, war and adventure. Four topics were very small, containing less than 50 songs, and were removed, leaving 36 topics.

4 Results

In this section we describe the results that were computed on the following processed data: for each song, we determined a genre cluster that the song belongs to, a lyrics topic, and aggregated play counts for that song over a whole period reflected in the dataset.

4.1 Genre lyrics specificity

For each genre, we will compute how specific are the topics described by the lyrics of the songs to this genre. In order to do that, for each topic c we compute which percentage of the songs from genre g belong to this topic:

$$genre_t = \frac{|genre_g \cap topic_c|}{|topic_c|} \quad (1)$$

In such a way, for each genre, we obtain a vector of values. If lyrics from that genre only belong to a few topics, we will observe large kurtosis of this vector (such as for Hip-Hop or Funk). If the lyrics are spread uniformly across various topics, we will observe small kurtosis (such as for large diverse genres like Pop and Rock).

From Table 1 we can see that such genres as *death metal*, *funk* or *Hip-Hop* have large kurtosis and *rock*, *pop* or *industrial*, which contain a lot of songs and sub-genres, and therefore can either be difficult to define, or were eclectic to begin with, have small kurtosis and are spread over most topics.

4.2 Song popularity vs song lyrics genre conformity

Next, for each genre, we will compute whether conforming to the usual topics of this genre is beneficial for song popularity (increased play counts). The median play counts per topic vary between

8 for the least popular topic ("Inner Demons and Struggles") to 18 for the most popular topic ("Hardcore gangster rap lyrics"), with significant difference between topic play counts on a Kruskal-Wallis H test ($\chi^2(36) = 1902.45$, $p < 0.0001$).

In most genres, there is a long-tail distribution of non-representative topics, and just one or two most popular topics for each genre. We divide the topics into representative and non-representative topics in such a way, that both representative (containing more songs in that genre) and non-representative have not more than 10% difference in amount of songs. This boundary between representative and non-representative topics is different for each genre. For instance, for *black metal*, there is just one representative topic (**Dark bloodthirsty visions**) which contains more songs than all the rest of the topics combined. However, as we can see from Table 1, most often the most represented topics of a genre do not generate biggest play counts. For most genres, songs in topics that are less usual for this genre, receive more attention from the listeners and are listened to more often. For instance, for *hardcore* genre, the most common topic found in lyrics are **Lone wolf** and **Dark self-discovery**. The most listened songs in that genre are on the topics of **Mid-life crisis** and **Unhappy love**. There are some exceptions to this trend: for *Hip-Hop* and *Pop* songs, the most widespread topics are the most popular. For *Pop*, these are **Desire and longing**, **Reminiscing on past love**, and **Relationships and love**. For *Hip-Hop*, these are **Hardcore gangster rap lyrics**,**Mid-life crisis** and **Pep talk**. In some genres, there is no statistically significant difference in popularity between songs with genre-conforming lyrics and other songs.

5 Conclusion

In this paper we showed that artists writing lyrics vary in how much they restrict themselves to certain topics, depending on musical genre, with the most restricting genres being hip-hop, black metal and new age. Also, we showed that when the lyrics are written in a topic not representative for that genre it has a beneficial effect on popularity, for most genres, but not for hip-hop and rock.

References

Hana Agatha, Farica Putri, and Alethea Suryadibrata. 2024. Sentiment analysis on song lyrics for song popularity prediction using bert. *Ultimatics : Jurnal Teknik Informatika*, 15(2):99–105.

Alexandra Lamont Alinka Greasley and John Sloboda. 2013. Exploring musical preferences: An in-depth qualitative study of adults' liking for music in their personal collections. *Qualitative Research in Psychology*, 10(4):402–427.

Yashar Deldjoo, Markus Schedl, and Peter Knees. 2024. Content-driven music recommendation: Evolution, state of the art, and challenges. *Computer Science Review*, 51:100618.

Shuhua Huang, Chenhao Hu, Weiyang Kong, and Yubao Liu. 2023. Disentangled contrastive learning for knowledge-aware recommender system. In *The Semantic Web – ISWC 2023*, pages 140–158, Cham. Springer Nature Switzerland.

Sein Jang, Battulga Lkhagvadorj, and Aziz Nasridinov. 2019. Preference-aware music recommendation using song lyrics. In *Big Data Applications and Services 2017*, pages 183–195, Singapore. Springer Singapore.

Junghyuk Lee and Jong-Seok Lee. 2018. Music popularity: Metrics, characteristics, and audio-based prediction. *IEEE Transactions on Multimedia*, 20(11):3173–3182.

David Martín-Gutiérrez, Gustavo Hernández Peñaloza, Alberto Belmonte-Hernández, and Federico Álvarez García. 2020. A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, 8:39361–39374.

Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2017. Retrieving similar lyrics for music recommendation system. In *ICON*.

Igor André Pegoraro Santana, Fabio Pinhelli, Juliano Donini, Leonardo Gabiato Catharin, Rafael B. Mangolin, Yandre M. G. Costa, Valéria Delisandra Feltrim, and Marcos Aurélio Domingues. 2020. Music4all: A new music database and its applications. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 399–404.

Shoto Sasaki, Kazuyoshi Yoshii, Tomoyasu Nakano, Masataka Goto, and Shigeo Morishima. 2014. Lyricsradar: A lyrics retrieval system based on latent topics of lyrics. In *ISMIR*.

Markus Schedl, Christine Bauer, Wolfgang Reisinger, Dominik Kowald, and E. Lex. 2020. Listener modeling and context-aware music recommendation based on country archetypes. *Frontiers in Artificial Intelligence*, 3.

Alexandros Tsaptsinos. 2017. Lyrics-based music genre classification using a hierarchical attention network. *ArXiv*, abs/1707.04678.

Michaela Vystrcilová and Ladislav Peka. 2020. Lyrics or audio for music recommendation? *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*.