Latent topic and tempo differences between East and West Coast Hip-Hop

Ábel Boros[1], and Arthur Flexer[1]

[1] Institute of Computational Perception, Johannes Kepler University Linz, Austria

Abstract

Our work presents a computational approach to exploring differences between Hip-Hop music from the US East and West Coast during the so-called ”rap wars”. This period in time was characterized by an intense rivalry prevalent especially in the lyrics of the music. We built a representative dataset of music from respective eminent artists and used topic modeling to analyze song lyrics. The resulting topic of ”Vulgarity and Violence” is more attributed to West Coast artists and ”Street Life and Rhythm” more to the East Coast. In addition we show that tempo of the music is slower for the West Coast and in general slowing down over the observation period. Our quantitative results align with existing theories concerning spatiality of Hip-Hop culture.

Keywords: music, topic modeling, lyrics, tempo, Hip-Hop, text analysis, audio analysis

1 Introduction

It is an established hypothesis in Hip-Hop studies that spatiality, i.e. the geographical origin of artists, is an important ”organising principle of value, meaning and practice within Hip-Hop culture” [12]. In this work it is our goal to obtain and document a quantitative basis of differences in lyrical content and musical style between two geographically defined rivaling music sub-cultures: US East Coast and West Coast Hip-Hop.

Hip-Hop is a cultural phenomenon that originated in the mid 1970s among mainly African American, Latin American and Caribbean ethnic minorities in the South Bronx of New York, USA, consisting of at least four major pillars: disk jockeying (DJing), break dancing, graffiti art and rapping [3]. Our work presents a computational approach to two of these pillars, namely the lyrics of the rap songs and the tempo of the disk jockeyed music, during the period of so-called ”rap wars”.[1] The rap wars between 1986 and 1998 coincided with what became to be known as the ”golden years” [11] of Hip-Hop, which were marked by intense rivalry between East Coast and West Coast artists. By the late 1980s, West Coast artists began to challenge the East Coast’s dominance of mainly New York based artists. The epicenter for this rising challenge was Compton in Los Angeles, California, with the formation of *N.W.A (Niggaz Wit Attitudes)* and their groundbreaking album *Straight Outta Compton* released in 1988. The ensuing rivalry was first expressed via so-called ”diss-tracks”, which are songs that are specifically made to criticize, insult, or even blackmail competing artists with clever wordplay and rhymes. Although such controversies usually do not lead to physical confrontation, in case of the rap wars they evolved into actual street shootings and the murder of *2Pac* and *Biggie*, two iconic figures of the West and East Coast Hip-Hop scene.

Ábel Boros, and Arthur Flexer. “Latent topic and tempo differences between East and West Coast Hip-Hop.” In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 1195–1205. https://doi.org/10.63744/XvZKQm15UbPN.

[1] For more information see: `https://hip-hop-music.fandom.com/wiki/East_Coast-West_Coast_hip_hop_rivalry`

In order to explore and empirically document differences in East and West Coast Hip-Hop we discuss related work in section 2, we present a carefully curated dataset of representative songs in section 3, present methods of topic modeling and tempo estimation in section 4, give our results in section 5 and conclude in section 6.

2 Related work

Music lyrics are capable to communicate emotional expressions, topics, and stories which can greatly influence a listener's impression of songs. The field of lyrics-related studies has been termed "lyrics information processing" (LIP) in a comprehensive survey [25], explicating that it shares core technologies with both natural language and music information processing. Popular application areas of LIP include lyrics structure analysis (e.g. rhyme scheme identification [1] or verse-bridge-chorus labeling [17]) and lyrics semantic analysis, including estimation of mood or emotion of lyrics [9; 15; 19], or modeling different topics in lyrics [16; 21; 23], e.g. concerning how typical they are for certain genres [2]. In a rare example of a large scale topic modeling study [8] on 124,288 lyrics from the metal music genre it was shown that, even in such a niche genre, topics are wide-ranging from love&romance and epic tales to dystopia, satanism, battle and madness.

The estimation of the global tempo of a music recording in beats per minute (BPM) is usually defined as "estimating the frequency with which humans tap along to the beat" [22]. It starts with beat tracking which identifies the temporal positions of beats within an audio signal. These positions correspond to a perceived regular pulse that aligns with the rhythm of the music [10]. Historically, beat-tracking techniques have relied on detecting onsets—beginnings of musical events or notes—and subsequent post-processing steps to refine the placement of beats. Modern algorithms distinguish actual beats from other musical events by filtering out non-relevant onsets and employing probabilistic models or machine-learning techniques plus postprocessing to obtain final predictions. Many approaches also adapt dynamically to tempo fluctuations, ensuring robust performance even in compositions with varying rhythmic structures. Although research on improvements continues [13], the `madmom` python library [6] is still considered state-of-the-art and will be used for our computations related to beat tracking and tempo estimation.

3 Data

In order to compile a set of representative artists and their albums we utilized the renowned "Rolling Stone" magazine's lists of 100 best Hip-Hop songs from the East Coast[2] and the West Coast[3] published in 2023 covering songs from 1979 to 2020. In order to focus on the historical period of greatest rivalry in the "golden era" of Hip-Hop (see section 1), we kept only albums between 1986 and 1998 from these lists. To ensure a comprehensive dataset, we additionally included all albums produced by the listed artists between 1986 and 1998, even if the albums are not part of the top 100 lists.

To obtain the lyrics, we scraped the artist's pages from Genius,[4] a well-known platform for song lyrics and annotations. For audio extraction, we utilized the `yt-dlp`[5] library to acquire the audio files from YouTube. After completing our data collection, we acquired 1364 lyrics and audio files, representing 16 East and 16 West Coast artists with 698 (East) and 666 (West) songs. The distribution of songs per Coast and year is shown in figure 1, highlighting peak periods of album releases from the early to mid 1990s. The list of artists, album and songs can be obtained from our

[2] `https://www.rollingstone.com/music/music-lists/best-east-Coast-rap-songs-1234737704`
[3] `https://www.rollingstone.com/music/music-lists/best-west-Coast-hip-hop-songs-1234712968`
[4] `https://genius.com`
[5] `https://github.com/yt-dlp/yt-dlp`

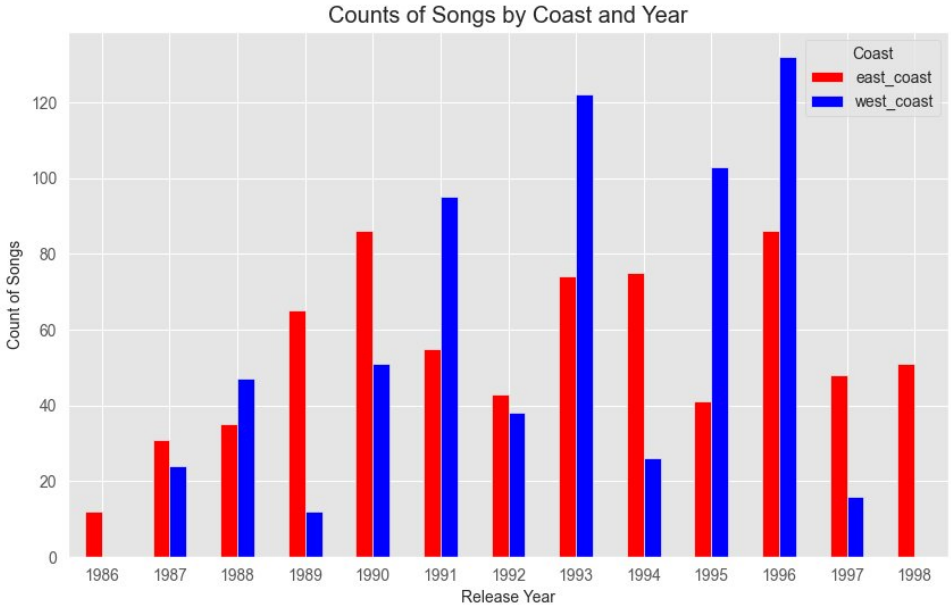

Figure 1: Annual distribution of songs: East Coast (698) vs. West Coast (666)

repository.[6]

4 Methods

4.1 Lyrics Preprocessing

Hip-Hop lyrics often include unique linguistic structures, extensive use of slang and regional dialects, which necessitate a robust and adaptable text-cleanup pipeline. The preprocessing steps we implemented are as follows:

1. Text normalization: We removed special characters and extra symbols that do not contribute to the semantic content, like brackets, parentheses, hashtags, quotations, marks, apostrophes, repeated punctuation, and multiple whitespaces after each other.

2. Tokenization: The cleaned text was segmented into individual tokens or words using the `nltk` word tokenizer[7] for easier manipulation and analysis of the text data.

3. Number removal: Tokens containing numerical characters were removed, because they typically do not add meaningful context within the lyrics-related content.

4. Stopword removal: We removed an extended list of stopwords[8] that carried little or no semantic information.

5. Slang and colloquial expression handling: Recognizing that hip-hop lyrics heavily feature slang and colloquial language, we relied on manual fine-tuning to bring these terms to a

[6] See section 3 here: `https://github.com/CPJKU/thesis_abel_boros/blob/main/Master_Thesis_Appendix_Abel_Boros.pdf`
[7] `https://www.nltk.org/api/nltk.tokenize.word_tokenize.html`
[8] See section 1 here: `https://github.com/CPJKU/thesis_abel_boros/blob/main/Master_Thesis_Appendix_Abel_Boros.pdf`

common form. Many words or phrases have the same meaning but are expressed in slightly different formats due to regional or cultural variations. By mapping different slang terms to the same words, we improved the consistency of our corpus (see the full list of slang mappings at our repository[9]). Specific dictionaries[10] were helpful in understanding these expressions accurately. See table 1 for examples of slang terms and their mappings.

Slang	**Replacement**
niggas	nigga
yo	yeah
doggz	dog
wanna	want to
lemme	let me
imma	i am going to
aks	ask
bizzle	bitch
muthaphuka	motherfucking
LA	Los Angeles
Cali	California

Table 1: Examples of slang mappings

On top of these preprocessing steps, we used the `OCTIS`[11] library for lemmatization and further punctuation removal and stopword elimination. Finally we filtered out words that appeared in less than 5% or more than 85% of the documents, helping us to avoid both overly rare and overly common terms. Table 2 gives the statistics of tokens in the lyrics corpus before and after preprocessing, showing e.g. that the number of unique tokens decreases from 31,622 to 938 through preprocessing.

preprocessing	**# lyrics**	**# tokens**	**average # tokens per lyric**	**# unique tokens**
before	1,368	792,599	579.4	31,622
after	1,368	243,134	177.7	938

Table 2: Statistics of lyrics corpus before and after preprocessing

4.2 Topic modeling

For topic modeling, a standard unsupervised method is latent Dirichlet allocation (LDA) [5], which tries to model latent topics in large amounts of text. LDA's basic entity are documents (in our case lyrics of one song) which are modeled as probability distributions over word occurrences. The assumption is that documents (lyrics) contain text about different topics which manifests itself in usage of different words. These latent topics are then also modeled as probability distributions over word occurrences, since different topics will use different words related to their different content. LDA then finds a probability distribution of topics across documents, trying to optimize separation of documents over topics. The assumption here is that different documents contain information about rather different (and few) topics, but of course overlaps are allowed.

[9] See section 2 here: `https://github.com/CPJKU/thesis_abel_boros/blob/main/Master_Thesis_Appendix_Abel_Boros.pdf`
[10] `https://www.dreadpen.com/hip-hop-slang-dictionary`
[11] `https://github.com/MIND-Lab/OCTIS`

While LDA is effective for uncovering general themes in a text corpus, it operates in a so-called bag-of-words regime treating words as isolated entities without considering the context in which a word appears. The more recent approach [4] of Contextualized Topic Modeling (CTM) is based on word embeddings, which model how words are related in different contexts. This makes CTM especially useful for analyzing lyrics, where slang, metaphors, and double meanings are common.

The two most widely used metrics for evaluating topic modeling are topic coherence and topic diversity [20]. A high topic coherence score suggests that the words which are representative for a topic are likely to appear together in real-world contexts. One way to compute coherence is to use a sliding window across documents (lyrics) and provide a score based on how often the topic's words co-occur within documents. Topic diversity measures the degree of uniqueness across topics in a model. It evaluates how different the generated topics are by considering the overlap of top representative words between different topics. High topic diversity suggests that the model has successfully identified distinct topics, whereas low diversity indicates significant redundancy. Topic diversity is calculated as the proportion of unique words across topics to the total number of words across all topics.

4.3 Tempo estimation

For tempo estimation we relied on the `madmom`[12] software's `RNNBeatProcessor` routine employing multiple recurrent neural networks followed by `TempoEstimationProcessor`, which outputs a histogram of candidate tempi in beats per minute plus their probablities (BPM, P). The histogram typically shows two dominant peaks, namely the fundamental period T and its first harmonic $2T$ (or, equivalently, $T/2$). To decide which of the peaks is the correct one is known as the half/double–time disambiguation problem. We retained the two most-probable candidates and applied a simple rule:

- Let $b_{\text{low}} < b_{\text{high}}$ be the BPMs of the two peaks.
- If their likelihoods differ markedly ($|P_{\text{low}} - P_{\text{high}}| > \tau = 0.15$), we keep the one with the larger likelihood.
- Otherwise we adopt b_{low}, the slower value, because it corresponds to the bar-level pulse and keeps all Hip-Hop tracks in the stylistically typical range of roughly 70–110 BPM.

The resulting tempo is stored for further analysis.

5 Results

To gain insights into the thematic content of East and West Coast lyrics, we applied and compared the Latent Dirichlet Allocation (LDA) model with that of the Contextualized Topic Modeling (CTM) approach. For both LDA and CTM, the number of topics k is assumed to be known a priori, therefore we compared models with $k = 2, 3, 4$. We computed multiple models (36 for LDA and 18 for CTM) and provide average and best results in table 3. We then ranked the models first by topic coherence score and next, in the case of ties, by topic diversity. Please note that although higher coherence values indicate more clearly separated topics [18], quantitative metrics should always be complemented with a qualitative evaluation [7], in our case an interpretation of the most salient words for each topic.

Overall speaking, all CTM models outperformed all LDA models both in terms of coherence and diversity, irrespective of the number of topics chosen with CTM achieving top average coherence of 0.28 at k equal 2 and 4 (diversity of 1.00 and 0.94) vs. LDA's top average coherence

[12] `https://pypi.org/project/madmom`

model	# models	k	ave. coherence	ave. diversity	best coherence	best diversity
LDA	36	2	0.19 ± 0.0024	0.80 ± 0.0000	0.19	0.80
LDA	36	3	0.19 ± 0.0025	0.60 ± 0.0000	0.19	0.60
LDA	36	4	0.20 ± 0.0033	0.54 ± 0.0110	**0.21**	**0.55**
CTM	18	2	0.28 ± 0.0270	1.00 ± 0.0000	**0.34**	**1.00**
CTM	18	3	0.27 ± 0.0255	0.99 ± 0.0183	0.33	1.00
CTM	18	4	0.28 ± 0.0284	0.94 ± 0.0411	0.34	1.00

Table 3: Comparison of LDA and CTM models. Shown are number of models computed, number of topics k, average coherence and diversity $\pm$ standard deviation, coherence and diversity of best performing models. Overall best performances for LDA and CTM are shown in bold face.

Word	Topic 1	Topic 2	Topic 3	Topic 4
Word 1	go	nigga	nigga	one
Word 2	know	fuck	bitch	know
Word 3	rhyme	shit	know	go
Word 4	make	motherfucker	fuck	time
Word 5	come	ass	want	man
Word 6	let	go	shit	love
Word 7	say	know	make	come
Word 8	man	come	go	see
Word 9	back	back	see	live

Table 4: Most salient words for LDA model at $k = 4$

of only 0.20 at k equal 4 (diversity of 0.55). The higher coherence and diversity scores are also evident when exploring the resulting topics from the best performing LDA and CTM models. The nine most salient terms for each topic of the these models are given in tables 4 and 5. Looking at the best performing LDA results in table 4, which was achieved at k equal 4 (coherence 0.21, diversity 0.55), there is an overlap between topics 1 and 4, which seem to deal with various aspects of personal life. Topics 2 and 3, which are also overlapping, contain a range of vulgar words which are typical for the genre Hip-Hop. Looking at table 5, CTM provided significantly clearer topics for their best performing model at k equal 2 (coherence 0.34, diversity 1.00), clustering the lyrics into two distinct themes which we named: topic 1 - Street Life and Rhythm (artistic, rhythmic, musically driven), and topic 2 - Vulgarity and Violence (street-oriented aggression). Whereas topic 2 is again about the prevalent vulgarity and aggression in Hip-Hop lyrics, topic 1 is more concerned with the musical and cultural side of the genre. Comparing this to an LDA result at k equal 2 also given in table 5, topic 2 features vulgar and aggressive terms comparable to the CTM result, but the salient terms of topic 1 appear to be relatively generic with terms like ”go”, ”make”, or ”man” being about everyday life rather than anything specific to Hip-Hop culture. By being able to model the context of words in lyrics rather than treating them as isolated occurrences, the CTM approach seems to clearly outperform LDA.

Next we explored whether there is a difference of the contribution of East and West Coast lyrics to the two topics. For each lyric of a song, the CTM model gives a probability for how strongly it belongs to each topic. These probabilities are summed up for each topic for the East and West Coast contribution, normalized so that the four quota sum up to 100%, and depicted as four bars in figure 2. The results show that topic 1 is more prevalent among East Coast artists (26% vs. 14%), while topic 2 is more dominant in the lyrics of West Coast artists (25% vs. 34%). This suggests that East Coast artists tend to focus more on lyrics-related complexity and musicality, while West Coast

	CTM		LDA	
Word	**Topic 1**	**Topic 2**	**Topic 1**	**Topic 2**
Word 1	rhyme	nigga	go	nigga
Word 2	cuz	shit	know	fuck
Word 3	soul	ride	make	shit
Word 4	rhythm	dick	one	bitch
Word 5	stage	bitch	come	know
Word 6	lyric	die	man	motherfucker
Word 7	music	big	time	go
Word 8	cut	street	let	want
Word 9	funky	man	want	ass

Table 5: Most salient words for CTM and LDA models at $k = 2$

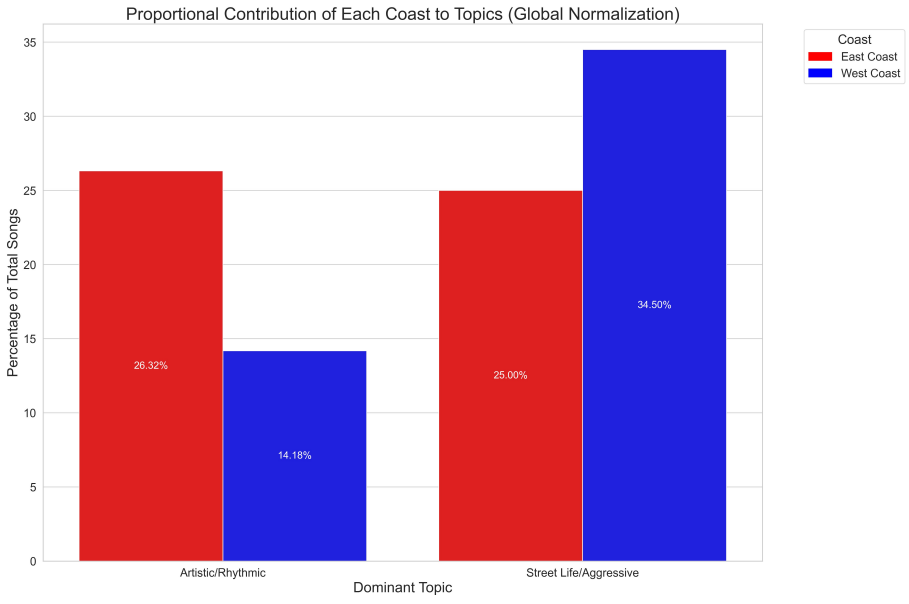

Figure 2: Proportional contribution of East (red bars) and West (blue bars) Coast lyrics to topic 1 (left side) and 2 (right side)

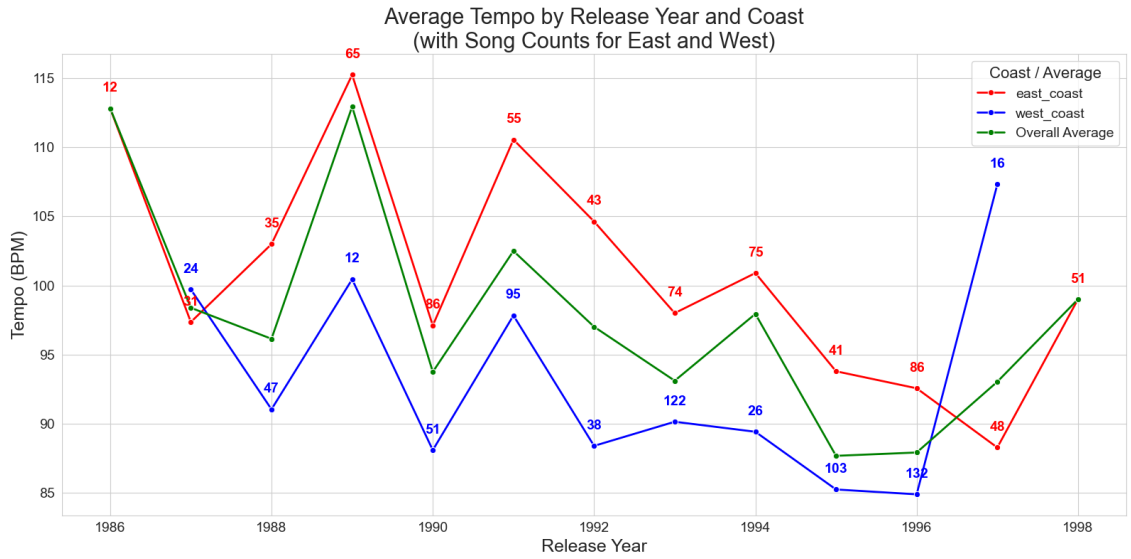

Figure 3: Average tempo in BPM from 1986 to 1998 for East Coast (red), West Coast (blue) and all artists (green)

artists emphasize themes related to street life and aggression. This is well in line with manually annotated results on a dataset of 340 Hip-Hop songs [14] showing that the amount of violent lyrics doubled from about 30% to 60% from the 1980s to the 1990s. For so-called "gangsta rap" songs the amount of violent lyrics even rose to 95%. Gangsta rap is a sub-genre of Hip-Hop romanticizing the outlaw and a "violent lifestyle of American inner cities afflicted with poverty and the dangers of drug use and drug dealing"[13]. Although originating on the East Coast it became a dominant form of Hip-Hop music especially on the West Coast in the 1990s, in line with the stronger attribution of topic 2 to West Coast artists.

To visualize the result of our tempo estimation in figure 3, we plot the average tempo per year (from 1986 to 1998) for the East and West Coast artists in red and blue and the average over all artists in green. In addition the figure contains the numbers of songs forming the basis for each depicted data point, thereby showing that for instance the start and end of the observation period has less data support with the years 1986 and 1998 not being based on West Coast artists at all.

The first observation is that the average tempo is decreasing considerably over most of the time period, from 113 BPM in 1989 down to 88 BPM in 1996. This trend does not hold before 1989 or after 1996, but as already mentioned there is little data support for these years. Since the decrease is visible for both East and West Coast artists we can state that tempo was gradually moving to a slower pace in the golden era of Hip-Hop era overall.

The second observation is that East Coast songs have an average higher tempo of 100.15 BPM compared to 90.16 BPM for West Coast songs. This is corroborated with an independent two-sample t-test comparing the tempo distributions of East and West Coast songs across all years. The result was statistically significant ($p < 0.05$) with a t-value of 7.107 at $698 + 666 - 2 = 1362$ degrees of freedom. This observation aligns with many West Coast albums being part of the "G-funk" subgenre [26], which started in 1992 with former *N.W.A* member's *Dr. Dre* realease of *The Chronic* including "Nuthin' but a 'G' Thang". This album was a huge commercial success and introduced a slower form of Hip-Hop which became very popular especially on the West Coast.

[13] `https://www.britannica.com/art/gangsta-rap`

6 Conclusions

Our work on East and West Coast Hip-Hop demonstrates that a computational approach is able to model and quantify differences in these musical sub-genres with results aligning with insights from musicological and sociological research. Both the fact that West Coast lyrics are more vulgar and the respective music slower in tempo have been discussed and acknowledged before (see e.g. [14] and [26]). Our work was able to corroborate these results concerning spatiality [12] of Hip-Hop culture in a quantitative way and on a larger data set than has been done before. Although our dataset of 1364 songs is not very large, our computational approach could of course be readily applied to bigger corpora of Hip-Hop music. In terms of construct validity [24], our dataset and experiments have been designed to be representative for eminent artists of the golden years of Hip-Hop from 1986 to 1998 by relying on an established source from music journalism. This could be extended to lesser known artists from that period or even be contrasted with results on Hip-Hop music not being associated with the East and West Coast rivalry.

Acknowledgements

This research was funded in whole by the Austrian Science Fund (FWF) [10.55776/P36653]. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

References

[1] Addanki, Karteek and Wu, Dekai. “Unsupervised rhyme scheme identification in hip hop lyrics using hidden Markov models”. In: *International conference on statistical language and speech processing*. Springer. 2013, pp. 39–50.

[2] Aljanaki, Anna. “Genre-conformity in the topics of lyrics and song popularity”. In: *Proceedings of the 3rd Workshop on NLP for Music and Audio (NLP4MusA)*. 2024, pp. 1–4.

[3] Alridge, Derrick P and Stewart, James B. “Introduction: Hip Hop in history: Past, present, and future”. In: *The Journal of African American History 90*, no. 3 (2005), pp. 190–195.

[4] Bianchi, Federico, Terragni, Silvia, and Hovy, Dirk. “Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2021, pp. 759–766.

[5] Blei, David M., Ng, Andrew Y., and Jordan, Michael I. “Latent dirichlet allocation”. In: *J. Mach. Learn. Res. 3* (2003), pp. 993–1022.

[6] Böck, Sebastian, Korzeniowski, Filip, Schlüter, Jan, Krebs, Florian, and Widmer, Gerhard. “Madmom: A new python audio and music signal processing library”. In: *Proceedings of the 24th ACM international conference on Multimedia*. 2016, pp. 1174–1178. DOI: 10.1145/2964284.2973795.

[7] Chang, Jonathan, Gerrish, Sean, Wang, Chong, Boyd-Graber, Jordan, and Blei, David. “Reading tea leaves: How humans interpret topic models”. In: *Advances in neural information processing systems 22* (2009), pp. 288–296.

[8] Czedik-Eysenberg, Isabella, Wieczorek, Oliver, Flexer, Arthur, and Reuter, Christoph. “Charting the Universe of Metal Music Lyrics and Analyzing Their Relation to Perceived Audio Hardness”. In: *Transactions of the International Society for Music Information Retrieval 7*, no. 1 (2024), pp. 129–143. DOI: 10.5334/tismir.182.

[9] Delbouys, Rémi, Hennequin, Romain, Piccoli, Francesco, Royo-Letelier, Jimena, and Moussallam, Manuel. “Music Mood Detection Based on Audio and Lyrics with Deep Neural Net”. In: *Proceedings of the 19th International Society for Music Information Retrieval Conference*. 2018, pp. 370–375.

[10] Dixon, Simon. “Evaluation of the audio beat tracking system BeatRoot”. In: *Journal of New Music Research 36*, no. 1 (2007), pp. 39–50. DOI: 10.1080/09298210701653310.

[11] Duinker, Ben and Martin, Denis. “In search of the golden age hip-hop sound (1986–1996)”. In: *Empirical Musicology Review 12*, no. 1-2 (2017), pp. 80–100. DOI: 10.18061/emr.v12i1-2.5410.

[12] Forman, Murray. “‘Represent’: race, space and place in rap music”. In: *Popular music 19*, no. 1 (2000), pp. 65–90.

[13] Foscarin, Francesco, Schlüter, Jan, and Widmer, Gerhard. “Beat This! Accurate Beat Tracking Without DBN Postprocessing”. In: *Proceedings of the 25th International Society for Music Information Retrieval Conference*. 2024, pp. 962–969.

[14] Herd, Denise. “Changing images of violence in rap music lyrics: 1979–1997”. In: *Journal of public health policy 30* (2009), pp. 395–406. DOI: 10.1057/jphp.2009.36.

[15] Hu, Xiao and Downie, J Stephen. “When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis.” In: *Proceedings of the International Society for Music Information Retrieval Conference*. 2010, pp. 619–624.

[16] Kleedorfer, Florian, Knees, Peter, and Pohle, Tim. “Oh Oh Oh Whoah! Towards Automatic Topic Detection In Song Lyrics.” In: *Proceedings of the 9th International Conference on Music Information Retrieval*. 2008, pp. 287–292.

[17] Mahedero, Jose PG, MartÍnez, Álvaro, Cano, Pedro, Koppenberger, Markus, and Gouyon, Fabien. “Natural language processing of lyrics”. In: *Proceedings of the 13th annual ACM international conference on Multimedia*. 2005, pp. 475–478. DOI: 10.1145/1101149.1101255.

[18] Mimno, David, Wallach, Hanna, Talley, Edmund, Leenders, Miriam, and McCallum, Andrew. “Optimizing semantic coherence in topic models”. In: *Proceedings of the conference on empirical methods in natural language processing*. 2011, pp. 262–272.

[19] Mishra, Vipul, Liew, Kongmeng, Epure, Elena V, Hennequin, Romain, and Aramaki, Eiji. “Are metal fans angrier than jazz fans? A genre-wise exploration of the emotional language of music listeners on Reddit”. In: *Proceedings of the 2nd Workshop on NLP for Music and Spoken Audio (NLP4MusA)*. 2021, pp. 32–36.

[20] Röder, Michael, Both, Andreas, and Hinneburg, Alexander. “Exploring the Space of Topic Coherence Measures”. In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, 2015, pp. 399–408. DOI: 10.1145/2684822.2685324.

[21] Sasaki, Shoto, Yoshii, Kazuyoshi, Nakano, Tomoyasu, Goto, Masataka, and Morishima, Shigeo. “LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics.” In: *Proceedings of the 15th International Society for Music Information Retrieval Conference*. 2014, pp. 585–590.

[22] Schreiber, Hendrik, Urbano, Julián, and Müller, Meinard. “Music Tempo Estimation: Are We Done Yet?” In: *Transactions of the International Society for Music Information Retrieval 3*, no. 1 (2020), pp. 111–125. DOI: 10.5334/tismir.43.

[23] Sterckx, Lucas, Demeester, Thomas, Deleu, Johannes, Mertens, Laurent, and Develder, Chris. “Assessing quality of unsupervised topics in song lyrics”. In: *European Conference on Information Retrieval*. Springer. 2014, pp. 547–552.

[24] Sturm, Bob and Flexer, Arthur. “A review of validity and its relationship to music information research”. In: *24th International Society for Music Information Retrieval Conference*. 2023.

[25] Watanabe, Kento and Goto, Masataka. “Lyrics information processing: Analysis, generation, and applications”. In: *Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA)*. 2020, pp. 6–12.

[26] Williams, Justin A. “’You never been on a ride like this befo’: Los Angeles, automotive listening, and Dr. Dre’s’ G-Funk’.” In: *Popular Music History 4*, no. 2 (2009). DOI: 10.1558/pomh.v4i2.160.