

How Rap Music Reflects Our Society

Jaehyuk Park¹, Fabio Rojas², and Yong-Yeol Ahn¹

¹ School of Informatics and Computing, Indiana University, Bloomington, USA

² Department of Sociology, Indiana University, Bloomington, USA

From novel to music, cultural contents have reflected contemporary social issues and cultural codes [1]. Among them, rap music is focused by social scientists due to its historical background and demographic features of its major audience. Although part of its cultural characters are originated from traditional African music, the basic style of current rap music is constructed from 1970's among black youth in Bronx area of New York [3]. Its historical background makes rap music keeping its position as one of the most political music genres [7] as well as functioning as a channel for minority to make public aware of their despair and anger by narrating their personal experiences directly [4][2].

However, previous studies used to analyze only small number of songs manually selected by researchers so have a limitation to generalize their results. Furthermore, since rap songs used in previous researches also used to be selected based on popularity, such as *Billboard's* top chart songs, the rap songs by independent labels or artists — called underground rap music — have not been reflected enough, despite those songs are more self-expressional and less affected by commercial pressures.

In this paper, we analyze over 66,000 rap lyrics to show how rap music has reflected contemporary social and economic issues. Rap lyrics are collected from *The Original Hip Hop Lyrics Archive (OHHLA)*, which is one of the biggest user-generated rap lyrics website. We also collected basic metadata, such as artist name, album name, and song title. We then used the API to obtain the release year of Spotify's 3,863 albums, which account for 34% of whole lyrics.

To analyze the trends of specific topics over time, we selected five words for each of three categories — common, crime, and economy. The words “god”, “friend”, “neighborhood”, “life”, and “hiphop” are selected for common category, while “gun”, “shot”, “prison”, “drug”, and “murder” are selected for crime category. For economy category, “money”, “job”, “buy”, “rent” and “business” are selected. Then for each of these initial words, we generate word set consist of 20 most similar words calculated by the word2vec model, which embeds each word into a vector space [5]. Our word2vec model successfully identify idiosyncratic usage of slang. Non-standard English words meaning “money” such as “dough”, “cheddar”, “scrilla” and “loot” are well detected by word2vec algorithm.

Based on the rap lyrics that we have release year information, we calculate the portion of each word set over time from 1990 to 2015. The timeseries of word set shows its association with corresponding social and economic events. For example, police-related words are more frequently used in rap lyrics during the year when important protests on racial discrimination were held (Figure

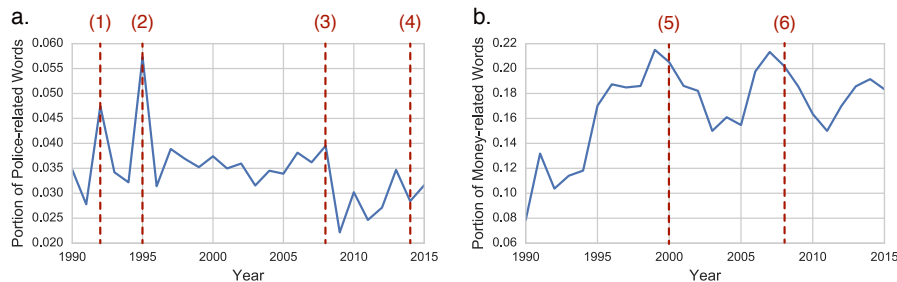


Fig. 1: Portion of police-related (a, left) and money-related words (b, right) in rap lyrics from 1990 to 2015. Dotted red lines represent the years of social events related to each topic. (1) Los Angeles Riots (1992), (2) Million Man March (1995), (3) Barack Obama elected as America’s first black president (2008), (4) Death of Michael Brown in Ferguson, Missouri (2014), (5) The collapse of Dot-com bubble (1999 – 2001), (6) Subprime mortgage crisis (2007 – 2009)

1a). Similarly, money-related words are more frequently used before the external economic shocks (Figure 1b).

Furthermore, some of word sets have significant correlations with related social statistics. We calculate the correlation between the FBI’s annual number of reported crime in Uniform Crime Report (UCR) by FBI¹ and each of word set in common and crime categories. The result shows that the portion of life-related words ($r=-0.87$, $p<0.01$) and shot-related words ($r=0.77$, $p<0.01$) are significantly correlated with UCR statistics (Figure 2a and 2b, respectively). In rap lyrics, the word “life” is mostly used as a positive meaning similar to profession, philosophy, and existence. Thus, the negative correlation implies that there are less positive and constructive words about life in rap lyrics when the number of crime is increased. In a similar manner, the positive correlation with shot-related words implies that more violent words are appeared in rap music, when the number of crime is increased.

Some words in economy category also show significant correlations with corresponding social statistics. We calculated the correlation between the words in economy category and the annual poverty rate provided by U.S. Bureau of the Census². Among the five words, the portions of money-related words ($r=-0.54$, $p<0.01$) and rent-related words ($r=-0.58$, $p<0.01$) show significant correlations. The correlation of rent-related words with the poverty rate is presented in Figure 2c. Both negative correlations can be interpreted as that when the economic growth is declined so poverty rate is increased, money-related and rent-related words become less likely to be mentioned in rap lyrics. It is also matched with our

¹ <https://www.ucrdatatool.gov/>

² <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-people.html>

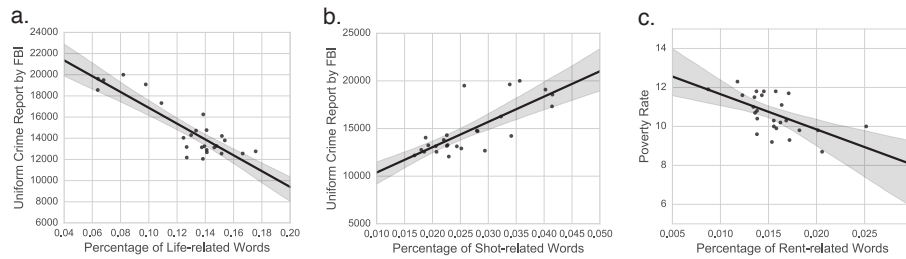


Fig. 2: Correlation between social statistics and portion of word set from 1990 to 2015. **a.** Total crime in Uniform Crime Report and the portion of life-related words, **b.** Total crime in Uniform Crime Report and the portion of shot-related words, **c.** Poverty rate and the portion of rent-related words

previous analysis that the mention of money-related words is increased around external economic shocks.

Despite of its limitation in the selection of the words in each category, this paper has three important meanings to enlighten how rap music reflects our society. First, by utilizing large number of rap lyrics, it shows more generalized result compared to the previous results induced from small samples. Second, by including majority of underground rap music, which are more self-expressive and free from the influences from big labels, this research reveals the potential value of underground rap lyrics for future cultural studies. Third, the selection of similar words using word2vec helps future linguistic analyses on rap lyrics. Since analogical and slang words have prevented researchers analyzing rap lyrics with computational tools, word2vec can opens new possibility to data-driven researches on large number of rap lyrics.

References

1. Freeman, M.L., Valentine, D.P.: Through the eyes of Hollywood: images of social workers in film. *The Social worker* 49(2), 151–161 (2004)
2. Gardstrom, S.C.: Music exposure and criminal behavior: Perceptions of juvenile offenders. *Journal of Music Therapy* 36(3), 207–221 (1999)
3. Kahan, J.B.: Bach, beethoven and the (home) boys: censoring violent rap music in america. *S. Cal. L. Rev.* 66, 2583 (1992)
4. Krohn, F.B., Suazo, F.L.: CONTEMPORARY URRAN MUSIC : Controversial Messages in Hip-Hop and Rap Lyrics. *ETC: A Review of General Semantics* 52(2), 139–154 (1995)
5. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems*. pp. 3111–3119 (2013)
6. Rose, T.: "Fear of a Black Planet": Rap Music and Black Cultural Politics in the 1990s. *Journal of Negro Education* 60(3), 276–290 (1991)