

Black Magazines as a Mirror to 1960-70s America: A *machine learning perspective*

Kushal K Dey kkdey@uchicago.edu Department of Statistics, University of Chicago

Lei Sun sunl@uchicago.edu Department of Statistics, University of Chicago

Lauren M. Jackson lmjackson@uchicago.edu Department of English,
University of Chicago

Abstract

In many ways, the turbulent times of the 1960s and early 1970s America, though fraught with social unrest, foreign crisis, and cultural revolution, played a great role in shaping a generation. In addition to the many significant developments of the time, this era was also a witness to changing African American society, lifestyle and culture. This evolution was reflected in African American magazines and publications of the time.

In this paper, we use machine learning and natural language processing tools to analyze two leading African American magazines of the time – Negro Digest (which later changed its name to Black World) and Ebony – for a period of 16 years from 1961 to 1976. We observe primarily one major shift in media narrative during this time period that coincides with the end of the civil rights movement in 1968. Narratives connecting African Americans to race, discrimination, protests and movements, which are more pronounced during the civil rights era gradually make way for narratives related to African American art, music and culture post 1968. We also show how the contextual relatedness of words changes across time and how the gradual shift from “negro” to “black” as the more popular identity term impacts the media narratives.

Introduction

Since the latter half of the 20th century, news media has served to not only reflect and represent society’s morale, but must be understood as an active participant in the social and political climate. As the times change, media narratives change and vice-versa. These narratives are not simply determined as matter of

two-party politics and a wider, more involved consideration of the discourse that shapes our society is required to feel out their contours and index evolutions in public opinion. Besides the usual social and political influences, one issue of key interest is the evolution of *blackness* as a staple idiom with African American or Black American publications.

In an attempt to address this issue, we focus on a time period in American history that has had monumental influence in shaping up Black American society and culture of today – the 1960s and early 1970s. This was a period of great social and political turmoil. The freedom struggle of the Black Americans seeking equal rights and social justice gained nationwide attention – a movement popularly known as the civil rights movement. With the assassination of the movement's greatest pioneer Dr. Martin Luther King in 1968, this movement eventually entered a new phase with the rise of the Black Power and the Black Arts movements. This period was also marked by foreign crisis in the form of the Vietnam war (1961-1975) and the many anti-war movements it spawned and the rise of a new counterculture and the hippie revolution in late 1960s and early 1970s. A detailed timeline of the events of the time period of our study is provided in Supplementary Figure 1.

In this paper, we attempt to understand how the social and political movements and developments of this era shaped the media narratives through the lens of two leading monthly African American publications of the time, *The Negro Digest* - which changed its name to *Black World*, and *Ebony*. However subjectively comparing and contrasting these magazines and their changing narratives over time is no easy task. To this end, we employ state-of-the-art machine learning and natural language processing tools to objectively quantify the differences in narratives between these magazines and also for the same magazine portal over time.

Our analysis points to one major shift in media narrative happening around the time of Dr. King's assassination in 1968 that marked the end of the civil rights movement. The narrative during the civil rights movement era (1961-1968) appears to be driven more by social and political issues focusing on racism, discrimination, protest movement and freedom struggle. The post 1968 narrative however is driven by narratives related to the Black arts movement and also narratives that talk about social and economic development. Also, during this era, we see a gradual decline in the identity word "negro" which is largely substituted in later years by the identity word "black". This shift in identity reveals interesting examples of narratives that are contextually more (less) similar to one identity term compared to the other, or are not affected by this shift at all.

Methods and Materials

Founded originally in 1942 by the Chicago based publisher John Johnson, the *Negro Digest* magazine was revived in 1961 and was consistently in circulation for a period of 16 years from 1961 to 1976. In May 1970, this magazine was re-named *Black World*. The *Ebony* magazine, also founded by John Johnson, was founded in 1945 and has been in continuous circulation since then up to this day. Both these magazines have been archived around 1960 onwards in Google Books. We extracted the text data from these archived versions of the magazines, for the time frame of 16 years (1961-1976) when both *Negro Digest/Black World* and *Ebony* were in circulation.

For each magazine, we first downloaded its Google Book version in Portable Document Format (.pdf) and then used Optical Character Recognition (OCR) to extract the text from these files into a .txt format. This was followed by manual inspection to test for the OCR quality, which was found to be reasonably high. Next,

we converted all the text to lowercase letters and removed any special characters and numbers from the extracted text. Some of the bigrams, say *civil rights*, were then converted into single word (*civil_rights*) so that we can demonstrate how this phrase as a whole shapes the narrative instead of its component words (*civil* and *rights*). Other examples of phrases that were converted into single words were *martin luther king*, *dr king*, *black power*, *black panther*, *malcolm x*. We aggregate the monthly processed text data into a single text data for each year. This processed yearly text data is ultimately used as input in all our statistical modeling frameworks. See Figure 1 for an illustration of the data collection and pre-processing steps.

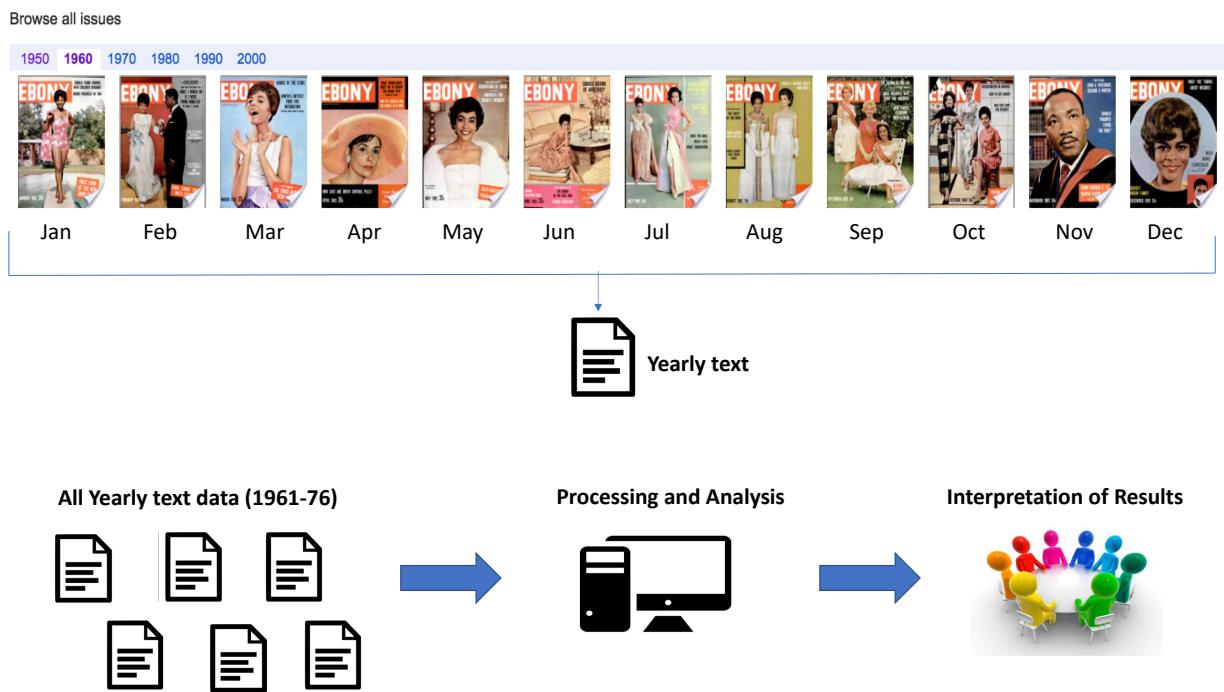


Figure 1: A demo of the entire workflow involving data collection, processing and analysis of the magazine data. The monthly issues of the magazines for a year are extracted from Google Books, then aggregated into a single yearly text. The yearly texts for 16 years of study (1961-1976) are processed and analyzed.

The statistical methods applied for defining narratives and identifying narrative shifts can be categorized into two groups – ones that are based on word frequency changes over time and the others that account for changes in contextual similarity and co-occurrence patterns of different words over time.

Word Frequency Analysis

For this approach, we first compile the frequencies of occurrence of each word in the yearly text data of each magazine. This can be processed into a matrix of counts $C = ((c_{ng}))$ for each magazine, with the rows representing the year of publication, the columns representing the words in the vocabulary and each entry c_{ng} denoting the number of times word g occurs in the yearly text of year ($1960+n$).

Next we employ an unsupervised clustering framework that can identify narrative groups purely based on the word frequency matrix C for each magazine. The model we propose here is popularly known as topic models in natural language processing (Blei et al). This model fits a multinomial model to each row of the matrix C .

$$(c_{n1}, c_{n2}, \dots, c_{nG}) \sim \text{Mult}(c_{n+}, p, p_{n2}, \dots, p_{nG})$$

where

$$p_{ng} = \sum_{k=1}^K \omega_{nk} \theta_{kg} \quad \text{with} \quad \sum_{k=1}^K \omega_{nk} = 1 \quad \text{and} \quad \sum_{g=1}^G \theta_{kg} = 1$$

Here K is the number of topics or *frequency narratives* and θ_{kg} is the relative contribution of word g to the topic k , whereas ω_{nk} is the proportional contribution or grade of membership of the topic k in the yearly text for year ($1960+n$). The estimated

ω_{nk} are plotted in using a stacked bar chart, also called a STRUCTURE plot (Pritchard). The estimated θ_{kg} values can used to identify the top words that distinguish one topic from the rest using the method proposed in Dey et al 2017 (*ExtractTopFeatures()* in R package **CountClust**) [ref].

Word Context Analysis

The word frequency approach does not account for the semantic structure of the sentences and the contextual relationship between words. For example, words like *music* and *songs* are contextually close as they will likely have similar words as neighbors where they occur in the text. A machine learning approach that can quantify this contextual relationship between each pair of words is *word2vec* (ref). As the name suggests, this method projects each word to a vector in a D dimensional space (D user specified) using a neural network modeling framework. The projection mechanism entails if two words are contextually close, the angular distance between their projections would be proportionately close and hence the cosine of the angular distance can be used as a measure of the contextual similarity of the word pairs.

Here we fit the word2vec model on the text data for each year of a magazine to estimate the contextual (cosine) similarity of word pairs for that specific year. Then for each pair of words or word sets, we investigate the trends of contextual similarity over time. This approach lets us identify word pairs whose contextual similarity increases or decreases or stays at the same level over time.

For our application, we used the Continuous Bag of Words (CBOW) version of the word2vec model. The dimension of projection space was chosen to be 100, and the context for each word was defined by 10 words immediately around it in the processed text (without special characters and numbers). Words that occur less than or equal to 3 times in the yearly text were removed from consideration.

Results

Figure 1 shows the trends in relative frequencies of the identity words “black(s)” and “negro(s)” in the two magazines. We observe that “negro” as an identity term gradually falls out of favor while “black” becomes the more pronounced identity word in the post civil-rights era – a shift that has been extensively studied already (ref).

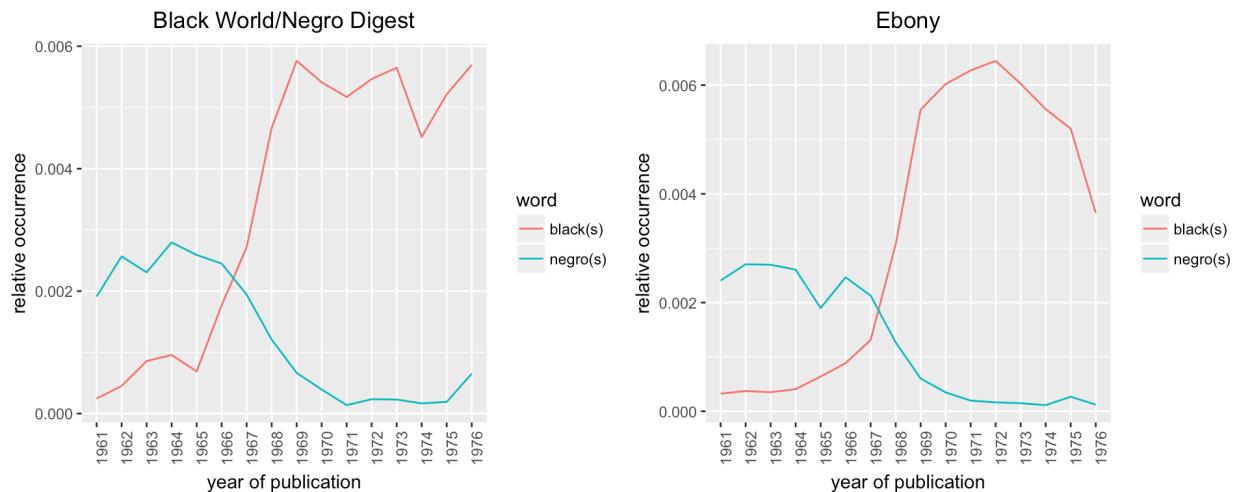


Figure 2: Yearly trends in relative occurrences of the words black(s) and negro(s) for the two magazines – Black World/Negro Digest and Ebony. The identity “negro” appears to fall out of favor as “black” becomes the more popular identity word.

We applied the topic model to the counts matrix C for each magazine separately . To remove the clear identity bias in the

narratives as evident in Figure 1, words like *black(s)*, *negro(s)*, *afro-american* etc were removed from the analysis. Figure 2 shows results from topic model fits with K=2 on yearly texts of each of the two magazines. In both the magazines, the pre-1968 (civil-rights era) issues show high membership in one topic/narrative whereas the post 1968 issues show high membership in the other, topic.

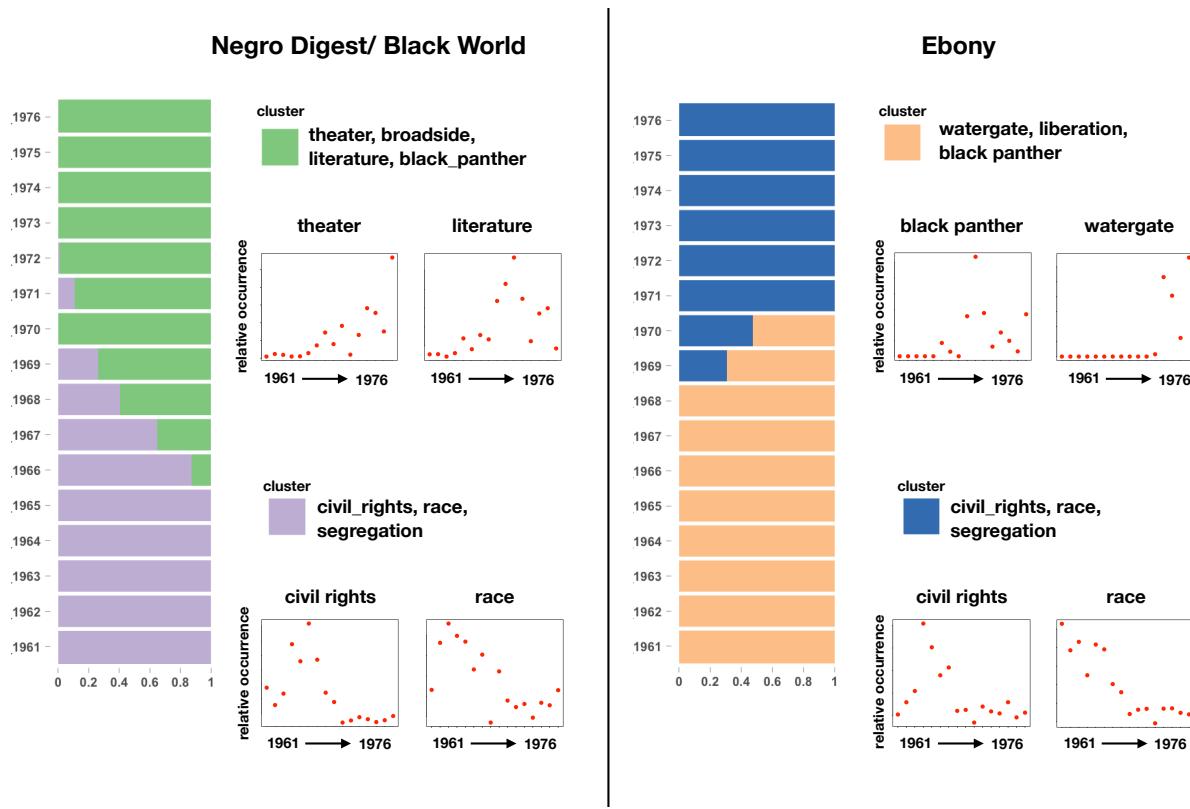


Figure 3: Topic model fits, with K=2, of the yearly text data from Negro Digest/ Black World (left) and the Ebony (right) magazines. The top distinguishing words for each narrative cluster are reported along with scatter plots of relative use of some of these words. The trends in scatter plots are a validation of why these words were picked by our clustering method as major driving forces behind narrative shifts.

The pre-1968 narrative/cluster is similar for both the magazines with words like civil rights, race and segregation driving this

narrative. The post 1968 narrative/cluster in Negro Digest/Black World magazines appears to be greatly influenced by the Black Arts movement, with arts-related terms – theater, literature, broadside – showing up among the top distinguishing words for this narrative. The same, however, cannot be said about the post 1968 narrative of the Ebony magazine, which is driven more by the contemporary events. The Black Panther party, founded in late 1966, does get heavy mention in the post 1968 issues of both the magazines and hence shows up as a distinguishing term.

Supplementary Figure 2 shows the results of topic model fits on these magazines for higher values of K, K=3, 4, 5. For K=3, the years marking the transition from one cluster to another in Figure 3 now forms a separate narrative group of its own. For higher values of K, we start encountering narratives showing memberships in only one or two issues. These narratives are primarily driven by contemporary events and are hence not so interesting from our perspective.

For further validation of the above results, we combined the counts matrices C for the two magazines and then ran topic model for K=2 and K=3 on the combined data, see Supplementary Figure 3. For K=2, the two narratives separated out the two magazines, highlighting possible overall narrative biases between these portals. The top distinguishing words in the Ebony narrative are sports related – teams, coach, baseball, football, basketball – implying overall higher coverage of sports related news in Ebony compared to Negro Digest/Black World. The top distinguishing words in the Negro Digest/Black World narrative are arts related – literature, drama, poetry, arts, music – implying overall higher coverage of Black arts and culture in Negro Digest/Black World compared to Ebony. For K=3, the civil rights era issues of both the magazines formed one narrative cluster, while the post civil rights era issues for the two magazines formed two separate

clusters. This conforms with the results in Figure 3, where we show that the civil rights era issues in both magazines had a similar narrative focusing on socio-political issues like race, segregation, civil rights and discrimination, but post 1968, the Negro Digest/Black World narrative leant much more towards covering the Black Arts movement, compared to Ebony.

Next, we applied word2vec on yearly text for each of the magazines and computed cosine or contextual similarity between any pair of words for each year. This association score lies between -1 and 1- more positive (negative) the association score, more positive (negative) is the contextual relationship between the word pair. An association score close to 0 implies no contextual relationship between the word pair. If a word does not show up in the yearly text for a particular year, its association score with other words is taken to be 0.

We observe trends in the contextual similarity score between each pair of words across time. Supplementary Figure 4 plots this trend for the word pair *martin-luther-king* and *assassination*.

Expectedly, we find the association score between these two words to jump post 1968 after Dr. King's assassination- the jump is more clear in the Negro Digest/Black World magazine compared to Ebony.

Supplementary Figure 5 reports the word pairs with the highest median association score throughout the period of 16 years (1961-1976), as well as the word pairs with the highest difference in median association between the pre-1968 and post 1968 issues for the two magazines. Word pairs with consistently high association score throughout the 16 years of study period were socio-political in nature – (civil rights, movement), (power, structure), (political, power), (economic, social), (race, relations) etc. Word pairs with

consistently much higher association in pre-1968 issues than in post 1968 issues were indicative of narratives related to race, struggle and oppression. On the other hand, word pairs with consistently much higher association in post-1968 issues compared to pre-1968 issues were indicative of narratives related to socio-economic development - healthcare, education etc.

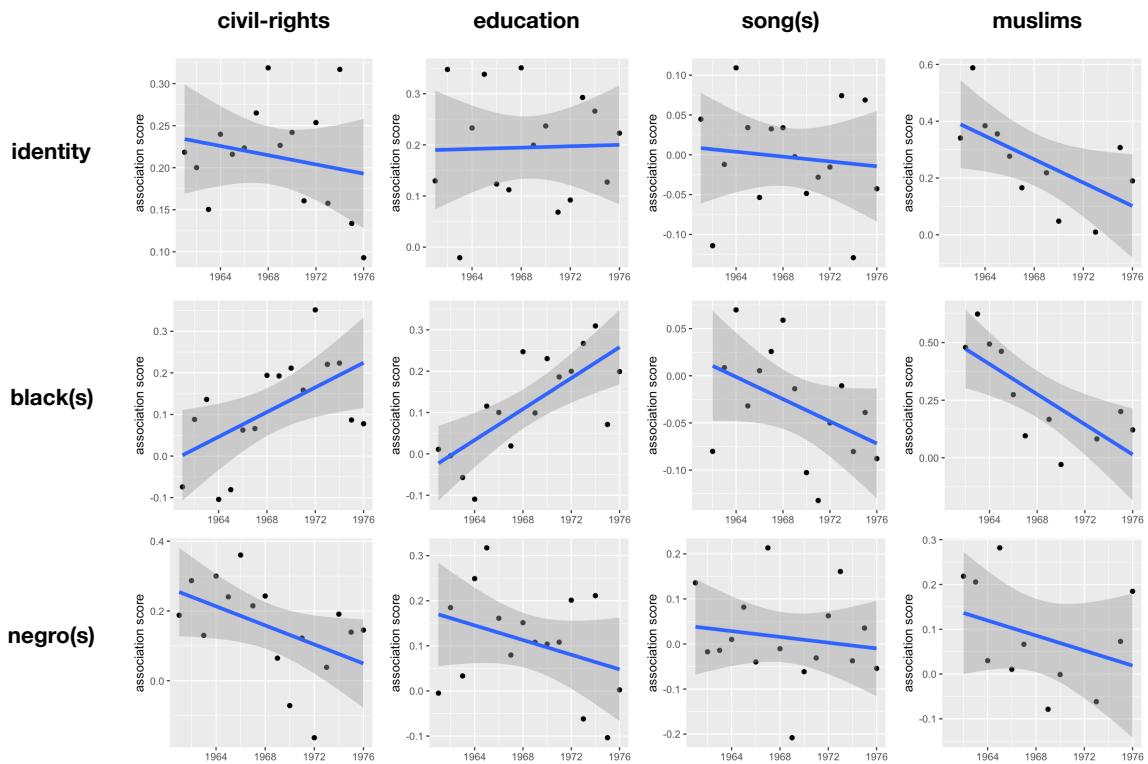
As evident from Figure 1, *black* substitutes *negro* as the more popular identity word in the later years of this period. We were interested in how this shift affects the association scores of these identity words with respect to other words and narratives. A word may show increasing trend of association with the identity word *black* as time progresses, but that could very well be a consequence of the fact that black is becoming more popular an identity word and hence is contextually more related to another random word. In order to address this issue, We rerun the word2vec model on a modified yearly text data where the words black and negro are replaced by one common term (say identity). We then compute the association score of each word with the word identity from the newly fitted model. These association score trends are then compared with that of the same word with *black(s)* and *negro(s)* from the previously fitted word2vec model on the original text data.

Figure 4 shows these association score trends for 4 words – *civil-rights*, *education*, *song(s)* and *muslim(s)*. For the last two words, association scores were computed for both the singular and plural forms and then averaged. *civil-rights* did not show much change in association score with the overall identity, but showed increasing (decreasing) trend with “*black(s)*” (“*negro(s)*”). *education* also showed a similar trend except for an increasing trend in association with the overall identity for the Ebony magazine. Many generic social, political and economic terms –

righters, votes, economy, leaders, opportunities, jobs, progress etc – showed trending similar to *civil-rights* or *education*.

However there were a few exceptions. Some words (e.g. – *song(s)*, *muslims*) showed opposite trending (see Fig 4) implying that they had strong contextual connection with identity “*black(s)*” even when “*negro(s)*” was the more popular identity word, and this association gets diluted over time as “*black*” became the more popular identity and started appearing in many other contexts. Besides *muslims*, other words showing this type of trending were found to be of two broad categories - arts (*songs, plays, writers, actress*) and aesthetics (*glory, beautiful*) related. Supplementary Figure 6 validates this finding by presenting some examples of the usage of the word “*black*” in the context of arts, aesthetics and in speaking of *muslims* (*islam*) in some of the early 1960s issues of the magazines, when “*negro(s)*” was the more popular identity word.

Negro Digest/Black World



Ebony

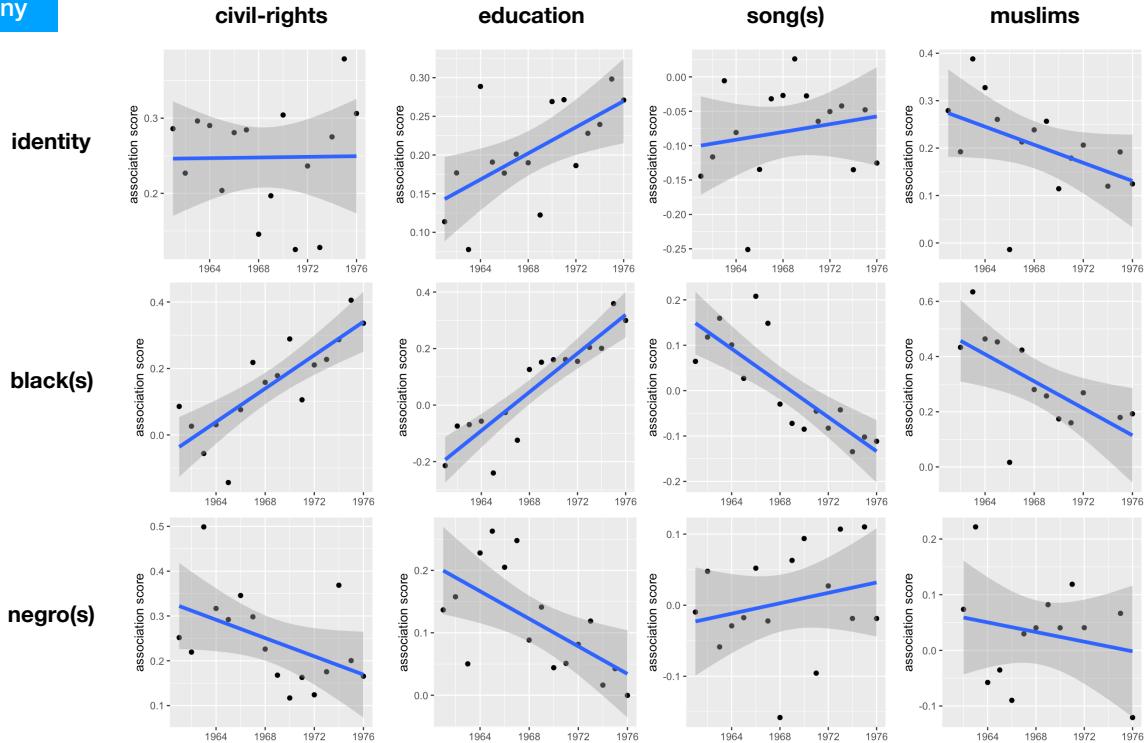


Figure 4: Trend plot of association of 4 words – *civil rights*, *education*, *song(s)* and *muslim(s)* with the generalized identity word, as well as separate identity words “*black(s)*” and “*negro(s)*”. The blue line and the band around it denote the average linear trend and the confidence interval for the trend. The trend plot of *muslim(s)* has some missing points, corresponding to years where the word *muslim(s)* did not occur in the text.

Conclusion and Future Works

1960-70s America is an extensively studied period in Humanities and here, we attempt to learn about this period through the media narratives as expressed in the two popular African American magazines of the time – Negro Digest/Black World and Ebony. Together with the shift in identity from “negro” to “black” from civil rights era to post civil rights era, we see a discrete change in media narratives. While the pre-1968 narratives were driven by socio-political issues of the time, like civil rights movement, race, segregation, discrimination etc, the post 1968 narrative was more influenced by the rise of the Black Arts movement and socio-economic developments. We also examined the trends of contextual relatedness between pairs of words and identified words that were most influenced by the shift in identity from negro to black. For example, we found several words, related to arts, aesthetics and in context of muslims or islam, that showed stronger connection with the word “black” when the latter was not so popular as identity word.

One future goal is to analyze the entire collection of Ebony magazines from 1950s up until present day, so as to be able to compare modern day narratives with those of the past and also track the changes in narratives over the years in the last 60-70 years. Also, in this paper, we focused on only two magazines from the 1960-70s era. We intend to use other resources – newspapers,

articles and music of the time to learn more about how different media portals covered the social movements of the time.

Digital humanities is a growing field and the computational and technological advances in the recent times allows us to handle extremely large scale data spanning years of study and decode them to learn the underlying structure. Our paper demonstrates how some of the state-of-the-art statistical and machine learning tools can be used to objectively analyze and interpret the media narratives through purely text based data. However, these methods are still limited in their abilities to glean information out of the texts, and the best approach is probably somewhere at the intersection of the subjective and objective approaches to learning.

The codes and data used in this paper for generating the figures are available on Github at : [Link](#).