Spring 5-7-2020

Innovation Dynamics of Cultural Production: Evidence in Rap Lyrics

James M. Zumel Dumlao
University of San Francisco, jmzumeldumlao@gmail.com

Junjie Lei
University of San Francisco, jlei6@dons.usfca.edu

Emeka Nwosu
University of San Francisco, eenwosu@dons.usfca.edu

Li Yu Oon
University of San Francisco, loon@dons.usfca.edu

Tsai Ling Jeffrey Wong
University of San Francisco, twong23@dons.usfca.edu

See next page for additional authors

Author
James M. Zumel Dumlao, Junjie Lei, Emeka Nwosu, Li Yu Oon, Tsai Ling Jeffrey Wong, James Rising, and Jesse Anttila-Hughes

Innovation Dynamics of Cultural Production: Evidence in Rap Lyrics

Thesis Submission for the Masters of Science Degree
in International and Development Economics

James M. Zumel Dumlao[1], Junjie Lei[2], Emeka Nwosu[2], Li Yu Oon[2], Tsai Ling Jeffrey Wong[2], James Rising[3] & Jesse Anttila-Hughes[4]

7 May 2020

Abstract: Culture is a driving force in organizing the structure of societies, and is conjoined with economic development. However, quantifying the impact of culture is difficult. Culture manifests itself in cultural production, through art, performance, music, etc. Innovation and influence in cultural production industries partially determines product quality. Using techniques from the "digitized humanities", we agnostically identify informational distance to describe the spatiotemporal dynamics of innovation and influence in Rap music lyrics. Rap emphasizes lyricism and hometown pride more than other genres of popular music, and is interesting as a globally impactful manifestation of the racially segregated labor market in the U.S. Resources and production are not spread evenly within an economy. Geographic clustering of economic activity is well discussed across the social sciences. Although first discussed in relation to manufacturing, urban agglomeration has been observed empirically for both cultural production and innovation in general. We find that number of Rappers and maximum novelty scale with overall and Black/African-American population at the MSA-level, suggesting an increasing return to maximum novelty through greater chance of recombination. Rapper population is predicted by Black/African-American population, while measures of song quality are predicted by total population. This paper contributes a novel dataset and application of the methodology to economic questions of cultural production.

[1]M.S. IDEC candidate, Economics Department, University of San Francisco [2]M.S. Applied Economics graduate student, Economics Department, University of San Francisco [3]Assistant Professorial Research Fellow, Grantham Research Institute on Climate Change and the Environment, The London School of Economics and Political Science [4]Associate Professor, Economics Department, University of San Francisco

1. Introduction

Culture is a driving force in organizing the structure of economies. It is the foundation of communication, trust, and coordinated equilibria, among many other facets of society (Alesina & Giuliano 2015). Economic development and culture are conjoined. However, measuring culture and its effects remains difficult. Culture exists both as an abstract collection of norms and values, and also as a pragmatic way to organize society (Patterson 2015). Therefore, we can study the manifested products of culture— art, dance, music, etc. to gain a partial estimate of the effect of culture. Cultural products contain within them declarative, procedural, and evaluative domains of knowledge structures, which binds them with culture as a constituted body of knowledge.

One obvious phenomenon of economic development is urban agglomeration, first observed by Marshall (1890) in manufacturing. Krugman (1991) later formalized geographic clustering, and attributed the core-periphery pattern to minimizing transportation costs between linkages in industries. Urban agglomeration has also been found empirically for cultural production industries (Currid 2007; Hellmanzik 2010; Lloyd 2002; Markusen and Schrock 2006; Mitchell 2019) and for innovation in general (Bettencourt et al. 2007; Carlino and Kerr 2015). The value of cultural products, to individuals and society, is difficult to determine (Throsby 1994), but quality is in part attributed to product innovation and influence. We can begin to understand the contribution of cultural production in economic development by describing how the industry is organized geographically. Answers to questions of cultural product quality are typically discourse-based. We aim to bring quantitative methodology to these questions of quality. Specifically, we look at the cultural production industry of Rap music and describe how innovation and influence in rappers' lyrics moves over space and time.

In this paper, we present a novel dataset collected through a Census of artists from Wikipedia categorical pages and a Corpus of lyrics from Genius. We use a combination of generative topic modelling and a measure of informational distance to quantify each song's lyrical novelty (innovation), transience, and resonance (influence). Resonance is defined as the difference between novelty and transience (Barron et al. 2018). These measures are matched to Rap artists' location of origin. With this novel application of "digitized humanities" methodology, we ask a number of questions about the cultural production process and its economics. Theoretical findings from the scaling theory literature provide some first-order directions for our analysis (Bettencourt, et al. 2007; Gomez-Lievano, Patterson-Lomba, & Hausmann 2016; Weitzman 1998). We find that innovation and influence in Rap music are not related after a low novelty threshold. Novelty has experienced sharp booms and busts, but has generally increased over time; influence exhibits a U-shaped pattern over time, declining in the 2000's and rising in the 2010's. Rapper population is predicted by Black/African-American population, while measures of song quality are predicted by total population. Lastly, we find that number of Rappers and maximum novelty and resonance scale with total and Black/African-American population at the MSA-level. This suggests that the mechanism for agglomeration in cultural production is increasing returns to maximum novelty and resonance by way of greater chance of recombinatory interactions.

Following this introduction, Section 2 reviews the literatures from various disciplines on cultural production and agglomeration. Section 3 describes the process of constructing our novel dataset, and gives an introduction to the natural language processing and information theoretic methods that are used in our analysis. The results are shown in Section 4. Finally, Section 5 discusses the results and their implications.

2. Literature Review

2.1 Culture and Cultural Production

It is important to ground our conversation of culture in the long-standing literature from cultural sociology and economics. Culture is amorphous and difficult to define. Edelmann and Mohr (2018) chronicles the origin and progression of the sub-field of Sociology concerned with culture. Because culture is particularly the realm of associated meaning, studies of culture mostly involved qualitative, interpretive methods; quantitative studies have become more popular over time. The primary theoretical paradigm with which Sociology approaches culture is looking at the relationship between a "cultural matrix" and a "social matrix". For example, the review features a study by Frederic Godart using two-mode networks to look at the ties between styles (cultural matrix) and fashion houses (social matrix).

Patterson (2015) uses the same theoretical approach of two-component culture. The first component is that of constituted cultural knowledge, which encompasses the evaluative (norms, values), procedural, and declarative functions of culture. The second component is made of the cultural pragmatics, involving context and knowledge production. His theory of culture helps us formalize the relationship between culture and cultural production, "Cultural knowledge is, at minimum, shared meanings about the world… The meanings may be internally represented or embodied, or they may be externally represented in artifacts, signifying events and practices" (8). An important facet of constituted cultural knowledge is the distinction between espoused and experiential values. Espoused values are those that are presented to others, while experiential values are those which drive individuals' actions. The espoused values of one might influence the experiential values of another (Patterson 2015, 12). The meanings encapsulated in cultural artifacts, then reenter the abstract of culture through individual consumption. He explains, "[Culture], however externally represented and phenomenologically experienced as objectively real (Zucker 1977) must in the final analysis be processed in the minds of people" (15). Culture evolves, partially through this mechanism of transmission of current espoused values and future experiential values. Young (2015) speaks to this idea through the concept of "Norm Entrepreneurs". He is particularly referring to individuals who are positions to set public examples, such as religious leaders or village elders. He goes on to note that these folks have much to lose if their efforts fail. However, artists could be considered norm entrepreneurs that have an incentive to innovate and push the culture in a new direction. Here sociological and economic theory synthesize to formalize the relationship between culture and cultural production industries.

While largely considered the domain of other social sciences, economists have theorized about the role of culture in our economic lives. Alesina and Giuliano (2015) give a review of the economic literature regarding culture and institutions. Culture teaches members of a society how things are done. It informally governs perceptions of trust, the place of the individual in a collective, morality, attitudes toward work, etc. A thread of literature views culture as the basis for social cooperation in general (Richerson, et al. 2016; Santos, Santos and Pacheco 2018; Smith, et al. 2017; Steward, Parsons and Plotkin 2016). Considering that human cooperation itself should not be taken as a universal outcome and is a basis for our fundamental understanding of economic behavior, culture and economics are inseparable. Incorporating culture into economic thinking cannot be ignored.

Most notably, game theoretic models are often used to describe the evolutionary dynamics of cultural traits (Bisin and Verdier 2011; Grief 1994; Young 2015). Conventions come about through repeated interaction among individuals, and social-proximity produced norms lead to competing local equilibria. These norms can also cascade into "universally shared social conventions" (Centola and Baronchelli 2015, 1990). Whether these cultural traits persist or change is addressed in Giuliano and Nunn (2017). Pulling from evolutionary anthropology, they form a model showing how cultural reliance is determined by an endogenous process in an information imperfect environment. The authors show that culture tends to persist in societies in stable climates, whereas variable climates make cultural change more likely; the conventions of the previous generation are useful to the current generation if their environments are similar.

The artifacts of culture have been the interest of economists for a long time, but its formal analysis is relatively new. Considered to be the first paper in cultural economics, Baumol and Bowen (1966) theorize a "cost disease" associated with the high human capital investments, and low productivity gains, of performers. This leads them to forewarn the arts' perpetual reliance on public subsidies. The addictive quality of music is formalized in Stigler and Becker (1977). Throsby (1994) represents a landmark collection of the cultural economics theory up to that point. Demand, production, and labor markets in the art world are all quite different from traditional markets studied in economics. However, art is not excluded from the realm of economic analysis. We understand cultural products to be the manifestations of culture, which also contribute to the evolution of cultural traits a la Patterson (2015). Artists take influence from work that has come before and contribute to the art that comes after.

In Rap music, lyrics and their meaning are essential to judging a product's quality. Sociology has thoroughly explored the generation and propagation of symbols in music. Peterson and Berger (1975) put forth an influential theory of cyclical symbol production in popular music. Their analysis shows that music follows a punctuated equilibrium dynamic. However, this is debated in Dowd (2004), which supports an open systems model. With the advent of decentralized production, symbol concentration no longer has a cyclical effect on talent entering in the industry. Overall, there is some level of optimal differentiation in the musical and lyrical makeup of popular songs (Askin and Mauskapf 2017). Perhaps our novel dataset and use of methodology can provide evidence in favor of one side of this debate. The dynamics of musical symbols is a key piece to our study of Rap because consumers of music identify with what they listen to (J. S. Park 2015; Martinez 1997; Wang 2016). Martinez (1997) pulls from a theory of oppositional culture, that marginalized groups in America use their cultural heritage to resist oppression, to discuss Rap music in particular as a method of resistance. The symbols in Rap lyrics have meaningful impact on the culture of Black communities in America. Culture influences the organization of institutions and individual actions; therefore, cultural products might have some direct effect on economic development.

2.2 Industrial and Innovation Agglomeration

Marshall (1890) is credited with first writing on the economic phenomenon of resources accumulating in urban clusters, that "the mysteries of the trade become no mysteries; but are as it were in the air" (198). He identified three attributes of an agglomeration economy: 1) a local pool of skilled labor, 2) local supplier linkages, and 3) local knowledge spillovers. This observation has been formally modelled from a number of angles. Marshallian external economies provide an intuitive way of explaining agglomeration. However, Abdel-Rahman and Fujita (1990) propose an endogenous model of agglomeration with Chamberlinian monopolistic competition[1] to give this idea formal clarity. An even simpler model, presented in Krugman (1991), is highly influential in our understanding of how economic geography forms into clusters. He proposes two regions and two industries; agriculture has constant returns and is fixed to a region, while manufacturing sees increasing returns and can be in either region. Following his theory of labor movement between regions, the increasing-returns characteristic of manufacturing lends itself to geographic concentration. This model could be applied to other industries that seem to follow increasing-returns, namely those involving human capital (Romer 1990). Porter (1998, In Press) reports that industries of varying knowledge intensities in economic clusters have an advantage in competing through productivity. The benefits of urban agglomeration for industry are complicated by the findings of Potter and Watts (2011). They bring together the literatures of how agglomerations form and how agglomerations evolve over time. What they find is that agglomeration economies have a life cycle, first with increasing returns, then decreasing economic performance. I imagine we will find a similar boom-bust dynamic in our analysis of creativity in Rap music, where a region will emerge as up-and-coming, then become stale over time.

Central to my thesis is connecting the literatures of cultural production and urban agglomeration. Many sociologists have taken an interest in how artists interact with urban environments. There seems to be a coevolution of urban agglomeration, and the attraction of artists, a relationship that benefits economic and artistic ends both. The benefits of agglomeration (knowledge spillovers, thicker market matching, etc.) can be applied to innovation in technology (Carlino and Kerr 2015). This is found for cultural producers, empirically, in multiple case studies and applied economic studies (Hellmanzik 2010; Lloyd 2002; Markusen and Schrock 2006; Mitchell 2019). Currid (2007), who studied the art community in New York City, suggests that

[1] This model is especially fitting in studying cultural production. Chamberlinian monopolistic competition describes a market where every firm has monopoly power, but monopoly profits are zero due to entry. A song is the legal intellectual property of the label, but other labels present close substitutes.

clustering is essential to the process of promoting one's work in the art market. Artistic agglomeration seems to have both a creative and political purpose.

All this becomes especially powerful when considering a model of recombinant growth as found in Weitzman (1998). He brought up that "an expansion process based on untried combinatoric reconfigurations of existing elements is generically more powerful than exponential growth" (Weitzman 1998, 337). Weitzman's theory can easily be applied to cultural production. With artists pulling from multiple influences and recombining them into new products, the potential for growth is limited by a cost of R&D. Agglomeration is determined by the ability to lower costs in resources (Krugman 1991). Therefore, cultural producers cluster to increase their ability to build new concepts from old ideas. This theory is consistent with that of endogenous growth as explained in Romer (1994). Human capital and innovation play a central role in driving growth, and have beneficial spillovers that should be encouraged when possible. In relation to cultural production, this implies that there is some optimal level of clustering wherein innovation is made stronger by proximity, but some level of dispersion is needed to maintain a diversity of thought. With this theoretical foundation, we should expect to see patterns of agglomeration in the novelty and transience of Rap music.

Rap is essentially urban (see background in Appendix A). Therefore, understanding the urban agglomeration literature helps us take a scaling theory approach to questions of cultural production. Scaling theory comes out of the idea that the size of a system, whether it be an ecosystem, firm or city, determines particular outcomes of the system. Sharp increases in CEO pay can be largely explained by firm size (Gabaix & Landier 2008). The size of cities across the United States follows a tight Zipf's Law[2] distribution (Gabaix 1999). Bettencourt, et al. (2007) examines multiple types of resources in cities of various population size. Fitting the data to power law scaling equations[3], they find that characteristics can be grouped by their linearity across urban systems. Infrastructure resources like gasoline stations and road surface are sublinear, while housing, employment, and consumption are linear. Most interestingly, new patents, wages, GDP, and crime show superlinear relationships with city size and are "unique social characteristics with no equivalent in biology and are the quantitative expressions that knowledge spillovers drive growth, that such spillovers in turn drive urban agglomeration" (Bettencourt, et

[2] Zipf's Law is where the size and rank of an object has a nearly perfect inverse relationship. Gabaix (1999) shows that the line made by relating log-rank and log-population fits this description.

[3] They estimate $Y(t) = Y_0 N(t)^{\beta}$, where Y denotes some material resource, N(t) is city size at time t, and β describes a particular dynamic across urban systems.

al. 2007). We can take this same approach to identify whether cultural production is infrastructure (sublinear), satisfies a human need (linear), or social currency (superlinear). If theories of culture align with this scaling theory, we should expect that cultural production is grouped with information and exhibits a superlinear relationship across urban systems. Gomez-Lievano, Patterson-Lomba, and Hausmann (2016) propose a theoretical basis for empirical findings of scaling in urban phenomena. They view urban phenomena as the culmination of a number of factors, some of which are only present as population size increases. The prevalence of a phenomenon is a matter of the number of people seeking to participate in the phenomenon and the number of factors that the city system has present. While surprisingly simple, the theory is able to predict phenomena with only the population of a particular city. Theory of urban agglomeration and scaling helps guide our description of cultural production dynamics of innovation.

3. Data and Methodology

3.1 Data

Textual data has been historically underutilized by economists due to an emphasis on mathematical and quantitative estimation. However, a wide range of research questions would be better answered using data from written text, especially once text is digitized. Gentzkow, Kelly, and Taddy (2019) provide a review of how textual data has been used in applied research. A major characteristic of using text as data is that it is high in dimensionality, with each word being its own variable to be counted among documents in a corpus of text. There are a variety of techniques to reduce this dimensionality, including topic modelling which we use in this paper.

One of the contributions of this paper is the production of a novel dataset. The dataset includes measurements of novelty, transience, and resonance for Rap music lyrics from 1994 to 2019. But before we could use our methodology to obtain those measures, we had to build a Corpus through a multi-step process. The following subsections details each step of that process, and shares some potential issues with the validity of the dataset.

3.1.1 Census of American Rappers

In order to generate the Rap lyrics corpus, we first had to take a census of all the rappers in America. There exists no comprehensive list of all artists in the genre. Therefore, we considered multiple Internet sources such as the Billboard Rap charts, AllMusic, and Wikipedia for use in this study. Due to barriers of data availability, we chose to construct our census by combining all the entries under two categories in Wikipedia: 20th- and 21st-Century American Rappers. Among all the categories in Wikipedia that list articles of Rap artists, these two combined narrowed the analysis to rappers whose main work is in the United States.

There are a number of artists that were cleaned out from this census. The main reasons for removal were 1) not being a rapper, 2) working primarily outside the U.S., or 3) being primarily a producer of Rap music. One exception to the second rule is Drake, based in Ontario, Canada, who is regarded as one of the most important artists in the history of the genre. There were also some influential rappers that were not in the list that were added in post. We do not claim this census to be totally comprehensive, but it is complete enough to carry on with analysis without significantly changing findings. Our Census contains 2,095 rappers after removing entries based on the criteria above.

Locational origin data for rappers in the census was manually gathered using Wikipedia articles. Origin is often stated in the article's biographical section, but differs in meaning across articles. It can refer to the rapper's place of birth, where they started their career, or where they first made a significant contribution to the industry. Some articles did not explicitly state an artist's origin. If the artist was part of a Rap group at any point in their career, we would impute the origin of the group. For articles where information was ambiguous, we used our discretion to single out an origin. Overall, we wanted to capture where rappers pulled most of their influence from, and where their work contributed to the local Rap culture. We exploit this unique feature of Rap as a genre to spatially visualize the measures of innovation and influence.

3.1.2 Lyrics Corpus

Having established rapper census, we sought to generate a corpus of lyrics in every song from every artist. Again, there is no existing comprehensive list of all Rap songs from the beginning of the genre. A complete corpus could be constructed by manually pulling the discographies of each artist, but this was deemed infeasible and the gains to doing so would be slim. We argue that the slight incompleteness of this corpus does not significantly hinder our ability to describe the spatiotemporal dynamics of novelty and transience.

Our Rap lyrics corpus consists of lyrics found on Genius. Genius uses crowdsourcing to gather lyrics of songs, and also has users comment on the meanings of lyrics. Users can up and down vote interpretations to ensure the highest quality information is the first available. Established in 2009, the digital media company first focused on Hip Hop/Rap music, as the meanings of lyrics in the genre are often discussed and studied among fans. Genius boasts a userbase of 100 million, with 2 million of those contributing to the body of "the best lyrics and knowledge database on the internet—that's over 25 million songs, albums, artists, and annotations" (Genius, "About Genius").

We queried the Genius API using a Python text wrapper (Miller 2017). A list of rappers, generated from our Census, was fed into the wrapper. The API was queried for one artist at a time, retrieving each entry listed under that artist. In all, we captured 117,425 entries for our Corpus. A critical feature of the analysis is that all songs analyzed should be in English. To filter out non-English entries, we used a dictionary of popular English vocabulary[4] (Mathews 2017) and found the proportion of recognized English in each song's lyrics. Entries with less than 50 percent recognized English were removed, resulting in 111,435 entries remaining. These remaining entries are referred to as our Corpus of Rap lyrics.

3.1.3 Dataset Generation Process and Caveats

We then merged the origin locational data from the Census to the songs in the Corpus by matching on artist name to make later geographical analysis possible. This further reduced the number of entries to 92,263. The drop in entries is likely due to the Genius API returning an entry for the wrong artist based on the similarity of a Census rapper's name to some irrelevant artist. Stop words[5] (Rehurek & Sokja 2010) and words with less than 3 characters were cleaned out to reduce computational burden. At this point, we also removed any entries with missing release date data to allow for sorting, reducing the Corpus to 57,724 entries. A major concern at this step in the process was introducing or exacerbating a bias toward more recent entries. As Genius was established in 2009, our dataset is heavily unbalanced over time as evidenced by Figure 2. Another source of this bias is that we require song lyrics to be digitized, meaning pre-Internet music is less likely to be included. However, a first-glance of those songs removed for lack of missing release date data leads us to believe that more recent music was missing the data, rather than older music. Omitting these entries, we argue, has no effect on, or may even work in opposition to, our recent music bias. Finally, entries with less than 300 characters after cleaning

[4] This dictionary of 25,322 unique words was compiled by statistically analyzing a sample of 29 million words used in English TV and movie scripts.
[5] Removing stopwords (pronouns, propositions, etc.) is a common practice in textual data (Barron, et al. 2018).

were removed. This is to filter out any near empty entries, as 300 characters is only about 48 words. This left us with 31,962 entries in our final iteration of the dataset.

There were a number of choices that we made during this dataset generation process that could jeopardize the validity of our results. The first major choice was to source our Census and Corpus from crowdsourced information databases. The accuracy of our origin locational data relies on the accuracy of Wikipedia articles. However, these sources allow us to collect a more comprehensive corpus of Rap lyrics than what was previously available. Our process also involved many filters of cleaning that resulted in attrition at each step. Going from 117,425 to 31,962 entries is a loss of nearly 75 percent of the entries. It is difficult to know if this attrition significantly changes the results, but most of the removed entries were likely just bycatch of our scraping methods. Regardless, the volume of lyrical data gathered through this method is so large that a systematic analysis with typical qualitative methodology would not be feasible.

On another note, collaboration among artists is central to Rap culture and found frequently within our Corpus. If we were concerned about quantifying exactly how novel each rapper was, we might want to separate the lyrics in a feature into separate entries labeled under each respective artist. Doing this would add dimensionality to our corpus that we believe is unnecessary. Leaving features intact slightly changes the interpretation of quantified novelty. Rather than capturing individual contributions to the genre, our measure will capture the impact of all projects that artists were involved in creating.

Appendices B and C present some robustness checks of our novel dataset. Looking at maximum novelty and resonance among the most productive artists and across MSA's, we find that the dataset matches our prior expectations with regard to artist and space. We also looked at averages, but results were lukewarm.

3.1.4 MSA Population Data

A major aspect of our analysis involves looking for scaling relationships. Urban scaling is evidenced by an exponential relationship between some economics phenomenon and city size, requiring a cross-section of multiple cities. We use 2018 MSA population statistics from the United States Census Bureau American Community Survey 1-year estimates. This is the year with the greatest number of entries in our dataset. The data provides demographic data for 519

MSA's. The average population size is 570,025.6, with a minimum of 55,414 and a maximum of 20 million.

Our research setting of Rap music originated in Black/African-American urban communities, so this data is of particular interest in our analysis. Three MSA's reported no data for this community. The mean Black/African-American population size in 2018 MSA's is 84,446.3, with a minimum of 330 and maximum of 3,735,537. The rich variation in city size allows us to examine relationships between cultural production measures and city scale.

3.2 Methodology

This paper follows a thread of literature in the "digitized humanities" that is concerned with patterns of novelty and transience in text. One of the first uses of the apposite methodology looked at the reading list of celebrated scientist Charles Darwin (Murdock, Allen, & DeDeo 2015). Researchers found that his early readings were close in subject matter, while readings later in life were conceptually further apart using measures of cognitive surprise. We model our methodology closely to that employed in Barron, et al. (2018). They use a combination of topic modelling and Kullback-Leibler Divergence to look at how ideas were created, propagated, or disappeared in the parliamentary assembly during the early days of the French Revolution. This paper is interested in using their methods to measure innovation and influence in Rap music.

3.2.1 Latent Dirichlet Allocation

Lyrics refer to a countless number of ideas, but many fit into similar themes. For that reason, we use coarse-graining to reduce dimensionality in our textual data. Latent Dirichlet Allocation (LDA) is a generative topic model. Blei, Ng, and Jordan (2003) propose this model as a solution to a trade-off problem between efficient processing and the preservation of relationships in textual data. Before introducing the machine learning process and statistics behind LDA, the authors state that "the basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words" (996).

LDA assumes that documents in a corpus are generated in a three-step process. First, the sequence of words, N, is chosen according to a Poisson distribution. This is reasonable to assume in text because over the vector space of all words, a majority are likely to be used with low or zero frequency. Next, a random variable, θ, is chosen which follows a Dirichlet distribution. Then for each of the N word sequences, a topic is chosen, z, according to a multinomial distribution with θ as its parameter, and a word, w, is chosen by multinomial probability conditioned on z. In other words, a document is created out of random distributions of words in topics in sequences revolving around this Dirichlet random variable. A threatening assumption of LDA is that the order of words in each document do not matter. While this assumption may hold easily in text like speeches or news, for music and art the order of words may matter much more. For the sake of this analysis, we will maintain this assumption. LDA is just one of multiple topic models that help reduce the dimensions in analyses of text corpora.

LDA gives a distribution of topic probabilities in each document in a corpus. It uses the probability density function and joint distribution of topic mixtures to arrive at this model for the probability of a corpus:

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\,\theta_d \tag{1}$$

The probability of documents in a corpus D, conditional on corpus-level parameters α, β, is made of the product of the marginal probabilities of single documents. Document-level variables θ_d are sampled once per document, and word-level variables z_{dn}, w_{dn} are sampled once for each word in each document. Figure 3, taken from Blei, Ng, and Jordan (2003), graphically represents this process of probability generation.

Used on our Corpus of Rap songs, LDA agnostically coarse grains high dimensional documents of words into distributions of topic probabilities. For this analysis, we have elected to specify a number of topics K=100. This means there are 100 topic groupings, and each song is represented as a probability distribution over 100 topic bins. In Appendix D, we attempt to recover the meanings of some categories by looking at specific songs with high weighting in the same topic. This "exemplar" method of topic interpretation is used by Barron et al. (2018), who later compile the most probable words from each topic. Now that each song is transformed into a probability distribution over 100 topics, we can compare those distributions using the following measure.

3.2.2 Kullback-Leibler Divergence

Our primary metric for innovation is the novelty of lyrical topics in a song. Novelty could be measured in a number of ways. For example, we could analyze the number of first appearances of particular n-grams across songs. However, for a more rigorous and agnostic quantification of novelty, we look to Kullback and Leibler (1951). Kullback-Leibler Divergence (KLD) is a measure of distance between two distributions of information[6]. When applied to human cognition, it can be interpreted as a measure of surprise. Formally described, the KLD metric is calculated by

$$KLD(p|q) = \sum_{i=1}^{N} q(x_i) \log_2 \frac{q(x_i)}{p(x_i)} \tag{2}$$

where $p(x)$ is the distribution an optimal learner is trained on, and $q(x)$ is a newly introduced distribution.

The literature that our study's methodology is inspired by uses KLD on a corpus, where a document's novelty is defined as its informational distance from the documents preceding and its transience is its informational distance from documents after. Barron, et al. (2018), defined resonance as novelty less transience and can be thought of as a measure of influence. Measures of novelty, transience, and resonance are difficult to interpret, but can be compared in relation to each other. High novelty signals innovation in the corpus, and high resonance is evidence of lasting impact.

We chose a scale w=1000, meaning that each song, represented as a distribution of topic probabilities after passing through LDA, will be compared to the topic distributions of the previous 1000 songs to measure novelty and the following 1000 songs to measure transience. This process is done systematically, beginning with the 1001st song and ending with the 1001st to last song in the corpus. Doing so reduces the number of songs in the analysis from 31,962 to 29,962 because 1000 songs on either end of the corpus will not have enough material to be compared with on either side. The dataset used in the following results consist of these 29,962 songs (by 1,133 unique artists) and their KLD measures grouped at different levels.

4. Results

4.1 Summary Statistics

There are a number of interesting first-order stylized facts that we find from simply looking at our dataset in various forms. Figure 1 is a time series of songs' release years in our dataset. Note that the years on each end of the distribution, 1994 and 2009, do not contain all the songs that were found in that year. They were partially chopped due to our chosen window of 1000 songs during the KLD process. The modal year among songs in the dataset is 2018. This led us to use 2018 cross-sectional data for scaling analysis.

[6] Interestingly, Mahalanobis' generalized distance (used commonly in econometrics for Covariate Matching techniques) is a special case of this informational divergence. KLD is also used in model selection through the Akaike Information Criterion (DeDeo 2018).

Figure 2 provides a spatial summary of our data. Each circle on the map represents a unique artist origin location. The size of the circle visualizes the number of songs attributed to that city, while the opacity of the circle conveys the number of unique artists hailing from that origin. We find evidence of agglomeration in this visualization, with clustering in the New York City, Southern California, and Atlanta areas. As these urban clusters correspond with three distinct cultural hubs of Hip-Hop, we find that this map strengthens the validity of our data.

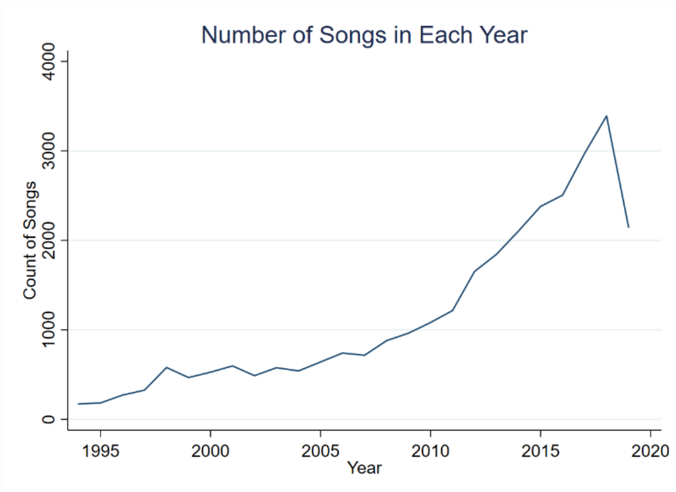

Figure 1: Number of Songs from Each Year in Corpus

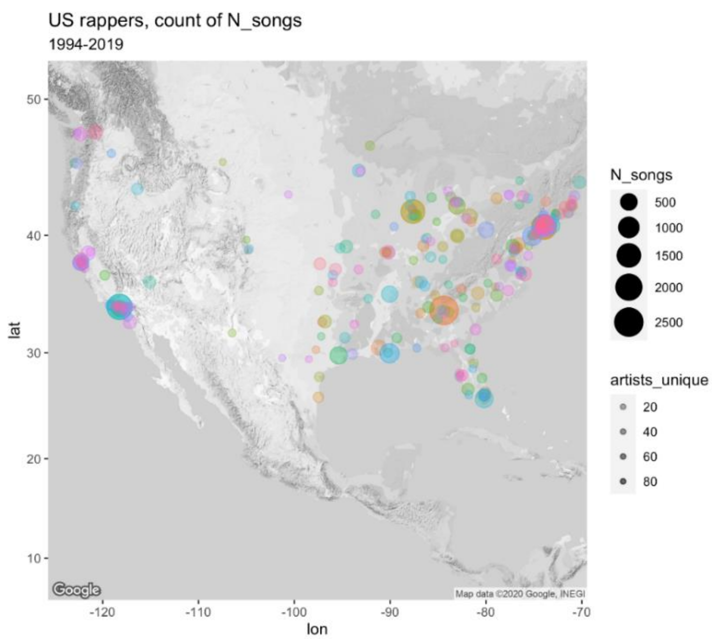

Figure 2: Visualization of Origin Locations, Hawaii, Alaska, and Puerto Rico not pictured

Next, we look at maximum novelty and resonance, which we refer to as maximum novelty and resonance. We are interested in capacities rather than averages, because they represent the highest quality product the system can produce. Figure 3 shows time series of maximum novelty and resonance. We find that maximum novelty experiences peaks and troughs, but generally increases over time. Maximum resonance displays a U-shape over time, starting at its height in the mid-1990's, bottoming out in the early-2010's, and rebounding in recent years. We discuss the implications of these patterns in Section 5.

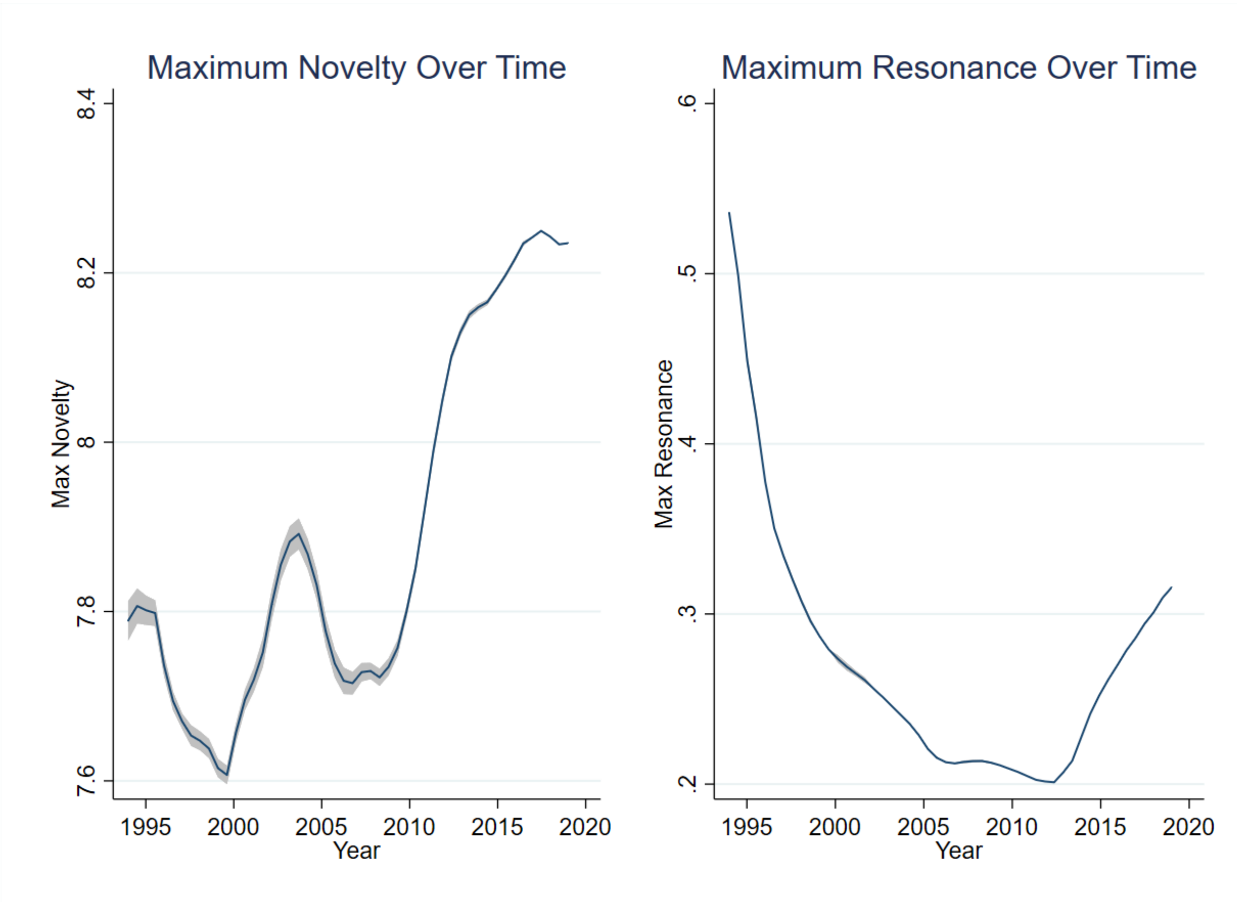

Figure 3: Maximum Novelty and Resonance over time, smoothed over 2-year periods

Turning our perspective back on the dataset as a whole, we look at corpus-wide relationships between our three measures obtained through LDA and KLD. Figure 4 plots the novelty versus transience. We find a tight, direct correlation between novelty and transience. Figure 5 plots novelty versus resonance. Contrary to our expectations, we find that novelty and resonance show no significant relationship after a low novelty threshold. The lasting impact of a song is negative below this threshold.

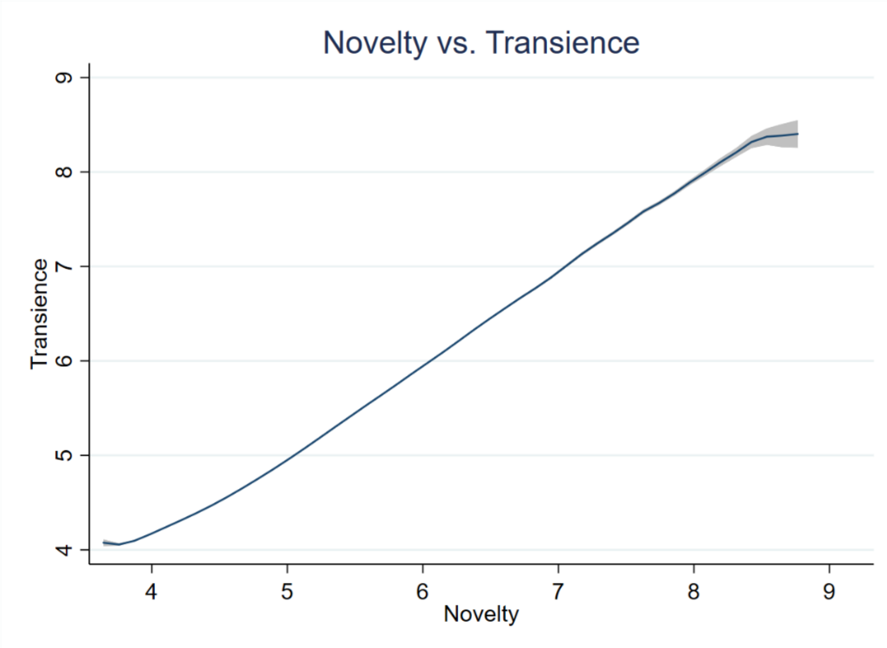

Figure 4: Novelty vs. Transience

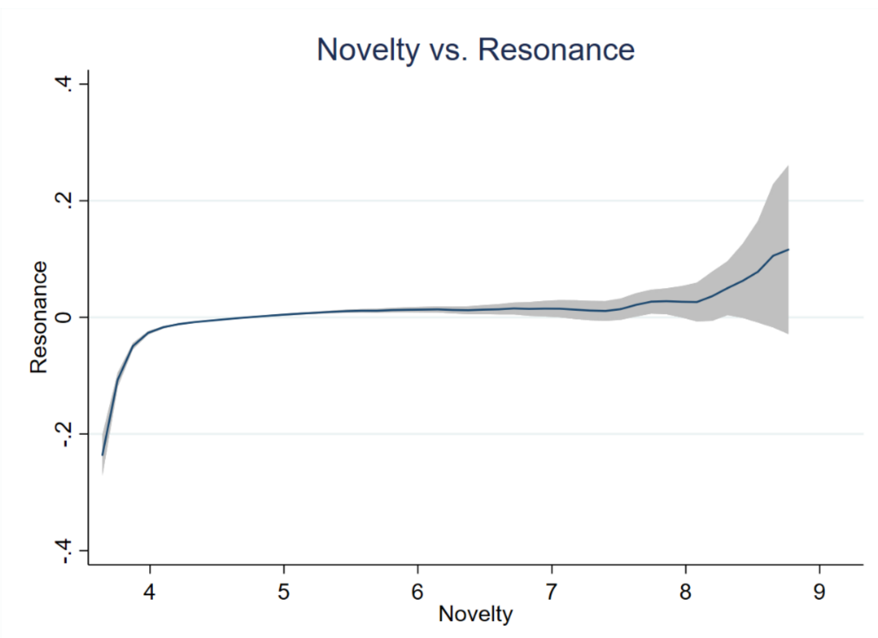

Figure 5: Novelty vs. Resonance

We conclude our presentation of summary statistics with some description of the population data gathered by the United States Census Bureau ACS given in Table 1. A total of 516 MSA's were accounted for in 2018. There is a wide variation in population sizes across these statistical areas, with the minimums and maximums separated by several orders of magnitude.

We also found that Black and African-American population is heavily correlated with total population with a .9112 correlation coefficient.

Table 1: Summary Statistics for ACS 2018 Population Data

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
TotalPop	516	572,881	1.457e+06	55,414	1.998e+07
BlackPop	516	84,446	267,512	330	3.736e+06

4.2 MSA-Level Scaling Analysis

In the interest of bringing together the data gathered with methodology from the digitized humanities with questions of economic importance, we seek an answer to the question of urban agglomeration in the cultural production industry of Rap music. First, we find evidence of clustering in the producers of Rap, rappers. The number of unique artists from each origin location in our full dataset were compared to cross-sectional population estimates for 2018. Figure 6 shows that the number of rappers in an MSA scales positively with both total and Black and African-American populations across city systems. This makes intuitive sense, as a higher population means there is a higher number of individuals inclined to rap. The nature of this relationship is explored in scaling regressions reported in Table 2. The estimated coefficients can be interpreted as the power value of an exponential equation[7]. Here it is revealed that the scaling relationship between the stock of rappers and population is sublinear. Interestingly, the number of rappers in a city is better predicted by the Black/African-American population size, rather

[7] Refer to footnote 3 for further clarification.

than the total population. This aligns with the historical understanding that Rap music has roots in African oral tradition.

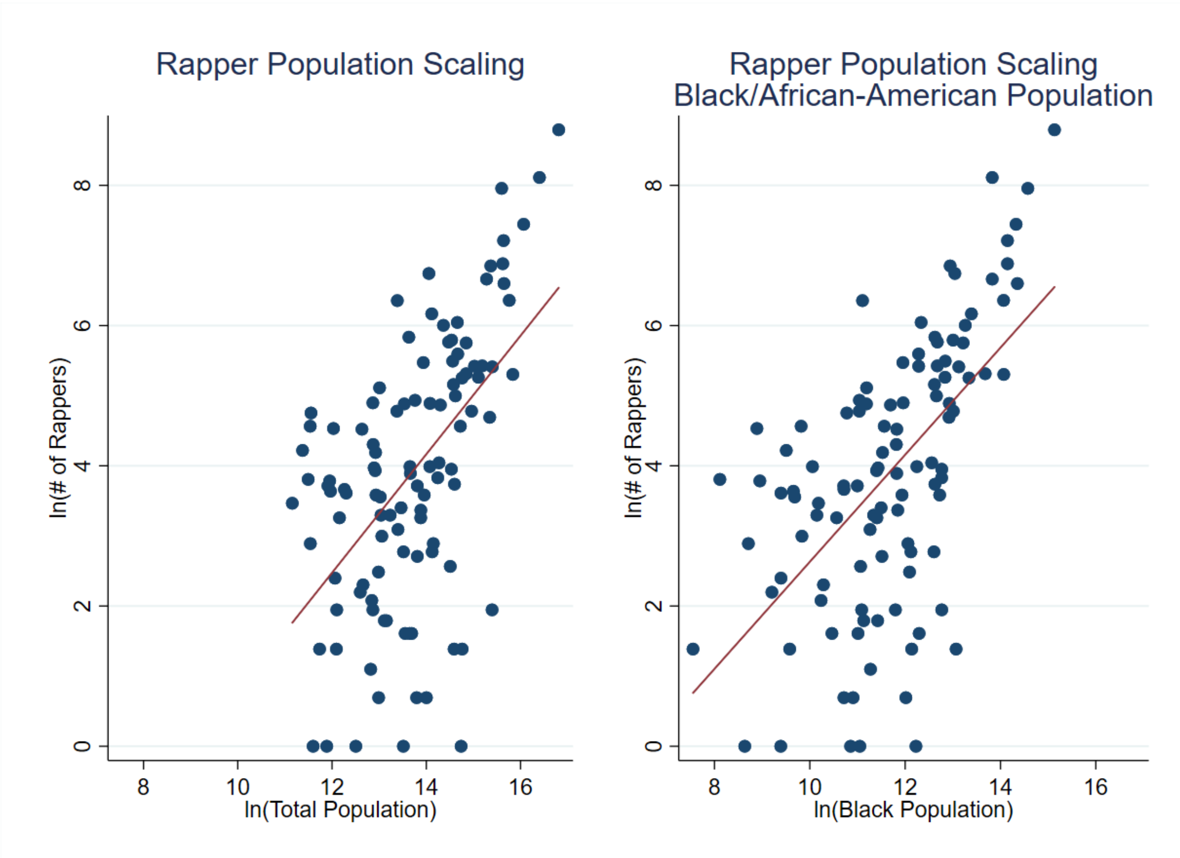

Figure 6: Does the stock of Rappers scale with population size?

Table 2: Rapper Population Scaling Regressions

VARIABLES	(1) ln(# of Rappers)	(2) ln(# of Rappers)	(3) ln(# of Rappers)
ln(TotalPop)	0.846***		0.172
	(0.131)		(0.250)
ln(BlackPop)		0.764***	0.639***
		(0.1000)	(0.195)
Constant	-7.672***	-5.006***	-5.912***
	(1.795)	(1.188)	(1.862)
Observations	108	108	108
R-squared	0.302	0.359	0.362

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, we look for scaling relationships between maximum novelty and resonance and city size. This time, we take a subsection of the data from the modal year of 2018, and compare the capacities across each origin location with the population estimates from 2018. Figures 7 and 8 confirm that a positive scaling relationship exists for max novelty and resonance, respectively. Note that we cannot take the natural log of resonance because the variable could take on negative values. In Table 3 we explore the nature of these relationships, again with scaling relationships. We find that maximum novelty and resonance scale extremely sublinearly with population size. Also, both max novelty and max resonance are better predicted by total population over Black and African-American population across the urban systems in our analysis.

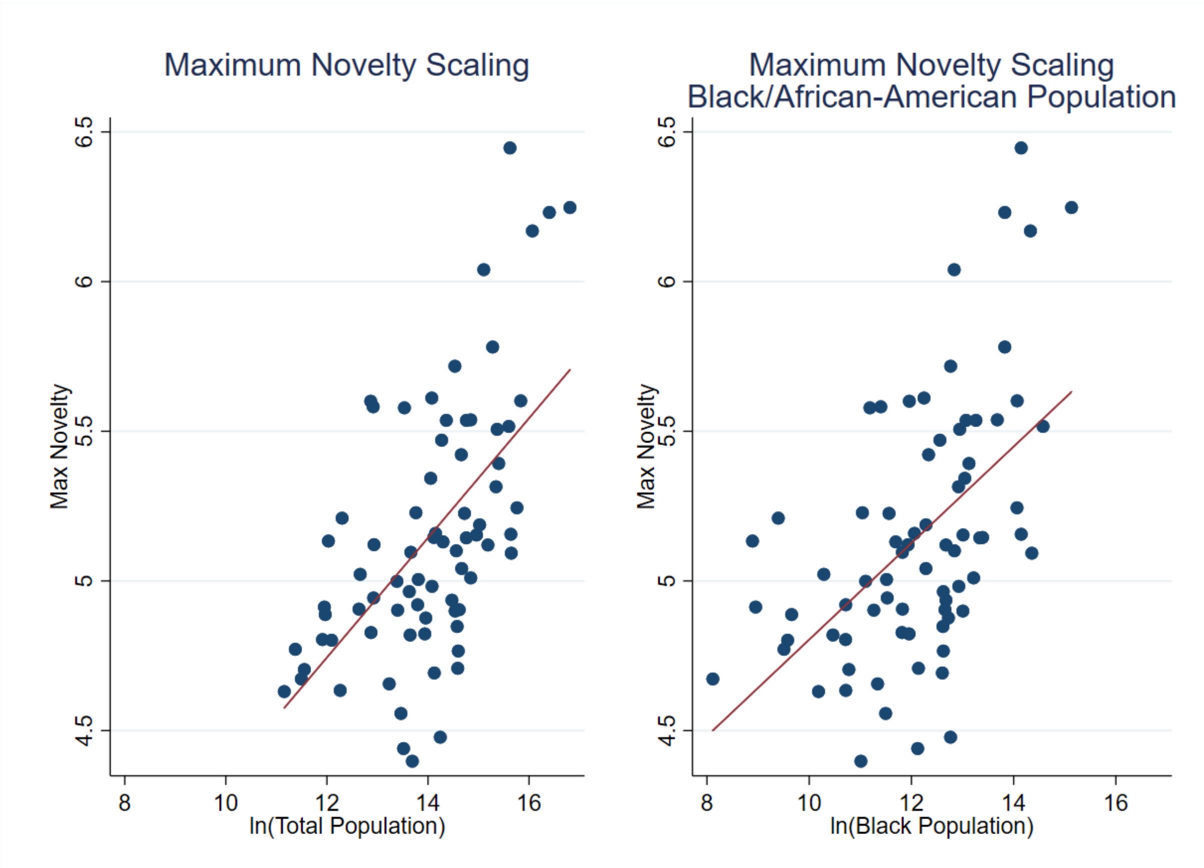

Figure 7: Does Max Novelty scale with population size?

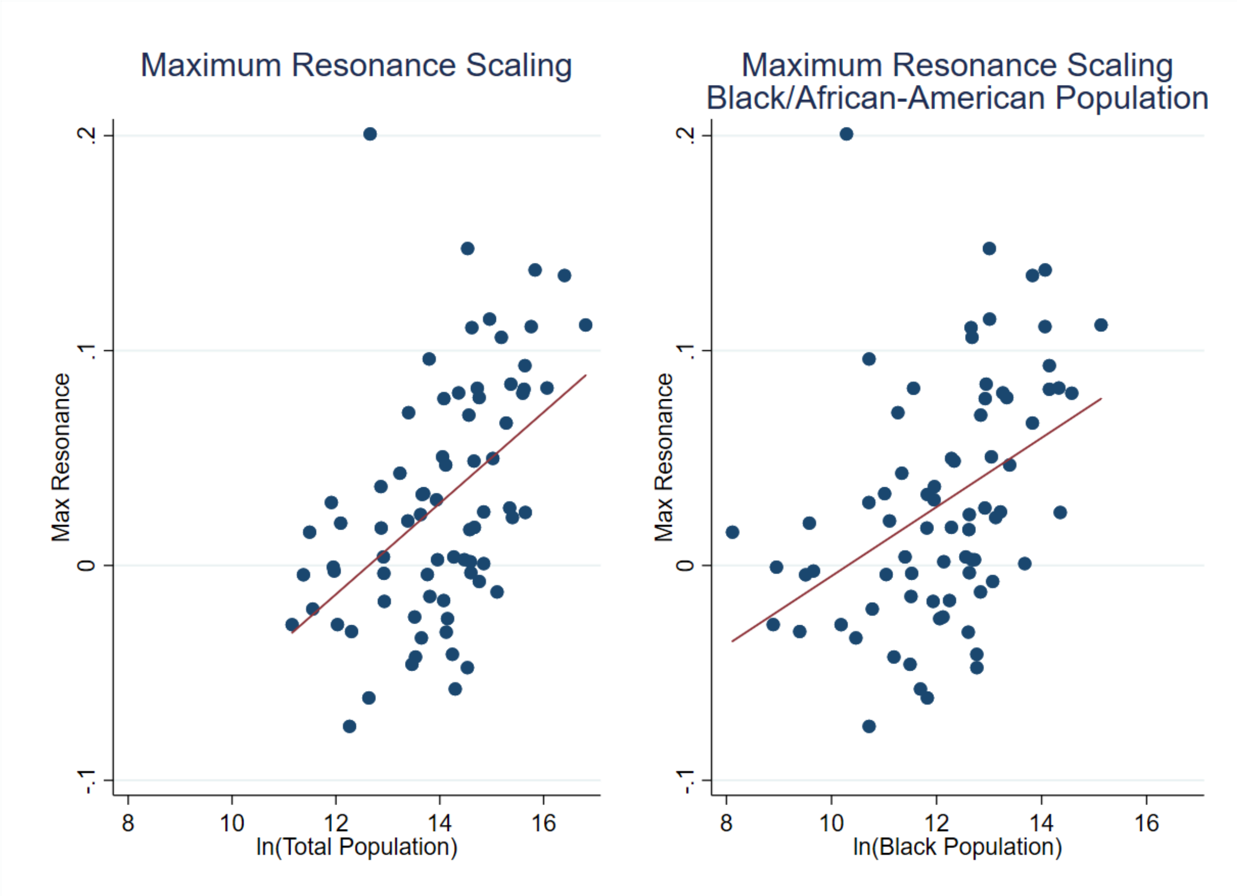

Figure 8: Does Max Resonance scale with population size?

Table 3: Maximum Novelty and Resonance Scaling Regressions

VARIABLES	(1) Max Novelty	(2) Max Novelty	(3) Max Novelty	(4) Max Resonance	(5) Max Resonance	(6) Max Resonance
ln(TotalPop)	0.200***		0.178***	0.0212***		0.0239***
	(0.0319)		(0.0597)	(0.00395)		(0.00801)
ln(BlackPop)		0.161***	0.0216		0.0161***	-0.00263
		(0.0312)	(0.0520)		(0.00394)	(0.00760)
Constant	2.346***	3.192***	2.390***	-0.268***	-0.166***	-0.273***
	(0.434)	(0.367)	(0.445)	(0.0563)	(0.0491)	(0.0554)
Observations	75	75	75	75	75	75
R-squared	0.346	0.290	0.347	0.232	0.173	0.233

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results from this descriptive analysis find positive scaling relationships for rapper population, maximum novelty, and maximum resonance. In the following section, we discuss the implications of our findings, issues with this study, and the possibilities for future research.

5. Discussion

Returning to Figure 3, we discuss the implications and drivers of these patterns of maximum novelty and resonance over time. First, the cyclical pattern found in maximum novelty is reminiscent of the business cycle, found in the market as a whole. This is evidence that cultural production is indeed an industry with similarities to other traditionally studied industries, like manufacturing. This pattern does not appear when looking at average novelty, suggesting that cultural production is aimed at pushing boundaries and achieving maximum novelty, rather than sustaining a high average. We attempt to reconcile the patterns of maximum novelty in Figure 9 by marking some turning points with the Billboard top Rap song from that year. Maximum resonance, however follows a U-shape. It is possible that the upturn at the later end of the time series is a result of our unbalanced panel of songs. Our KLD process compared a song's distribution of topics to the distribution of topics over the previous and following 1000 songs, meaning the dispersion in the distribution before any given song would be greater than in the distribution after that song. In other words, a song will be further in time from the 1000th song before than it is from the 1000th song after. This likely results in biasing novelty upward and transience downward. Such an effect would be strongest near the modal year of 2018. Thus, we see an upward-biased resonance in the 2010's. Still, we can attempt to fit this U-shape pattern to eras in Hip-Hop history, which we do in Figure 9. The highest point of influence corresponds to the Golden Age of Lyricism in the mid-1990's; the early 2000's see an massive drop in influence, and corresponds with the Bling Era, a sound which we rarely hear today; lastly, the early 2010's signals the rise of the Trap Sound in Hip-Hop beats, a sound which has become ubiquitous across all genres of popular music.

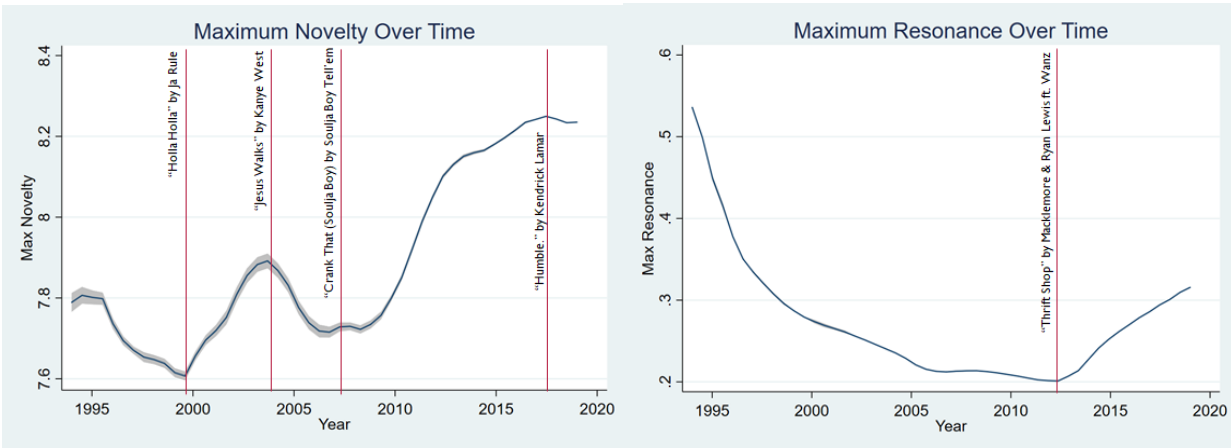

Figure 9: Maximum Novelty and Resonance over time with selected events

There are several issues with this study that threaten the validity of our data and findings. We only analyze innovation and influence in Rap lyrics and their topics, but innovation in music is multidimensionality (instrumentation, production, etc.). KLD could theoretically be applied to distributions of rhythmic timing, song structure, or other forms of innovation, but we looked at lyrics as textual data, which is the context LDA and KLD has been used in our review in the literature. Our pre-processing data generation relies upon digitized, crowdsourced informational databases. There may be systemic biases on Wikipedia and Genius that influence which rappers and songs are digitized. Furthermore, we experienced heavy attrition throughout the cleaning process, possibly introducing and exacerbating systematic bias in our data. With regards to our processing methodology of LDA and KLD, a critical assumption that the order of words does not matter threatens the validity of our use of LDA. Rap lyrics, like poetry, are often imbued with wordplay and multiple meanings based on the intentional structures of words. We acknowledge this as a major weakness in our application of the methodology. Lastly, common discourse surrounding the quality of Rap music assumes that innovation and influence go together. We did not find this relationship in our data, only that there is a novelty threshold below which resonance is negative.

There are countless application possibilities for the data gathering method presented in this paper. Some avenues for further research include combining cultural production data with other data sources to perform associative or even causal studies, examining cultural evolution at the individual-level a la Murdock, Allen, and DeDeo (2017), or analyzing political speech for questions of political economy a la Gentzkow and Shapiro (2010). We hope to see this methodology applied in new and interested contexts in the near future.

This paper attempts to bridge methodology from the digitized humanities and questions from economics. Our application of LDA and KLD on Rap lyrics has garnered interesting results speaking to typically discoursed-based discussions on the genre, and to cultural production as an economic industry. We find that novelty is not strongly related to resonance, contrary to the novelty bias found in political speech by Barron, et al. (2018). However, we do find evidence of scaling for both the stock of Rappers in a city and their innovation and influential capacities. The former aligns with findings from Bettencourt, et al. (2006), in that a sublinear scaling relationship is associated with quantities of infrastructure. However, we find heavily sublinear scaling relationships for maximum novelty and resonance, which runs contrary to the superlinear relationships Bettencourt, et al. (2006) attribute to innovation phenomena. We attribute this to our atypical measurement of novelty and resonance. Where innovation phenomena like patents and R&D spending are absolute quantities, the measures attained through KLD are relative. Thus, they may not follow the same universal rules that physical proxies in urban systems do. Together, these scaling relationships suggest a particular mechanism for agglomeration in cultural production. We infer that urban clusters provide an increasing return to maximum novelty and resonance for artists by way of an increasing stock of artists with which to recombine existing ideas (Weitzman 1998). This mechanism would align with the scaling theoretical model from Gomez-Lievano, Patterson-Lomba, and Hausmann (2016), which predicts that economic phenomena are a result of the number of participants and the number of factors a city provides. Our findings are further complicated, in that rapper population is tied closer to Black/African-American population, while song quality is tied closer to total population. This suggests that the people of color in a city are the factors of production, while the total population is the number of market participants for Rap music. The total population may be more integral to song's quality in terms of production, distribution, and potential audience, than the communities of the artists' alone.

Works Cited

Abdel-Rahman, Hesham, and Masahisa Fujita. 1990. "Product Variety, Marshallian Externalities, and City Sizes." *Journal of Regional Science* 165-183.

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature* 898-944.

Ashabi, Mahsa. 2019. "Captive Nations: Measuring Economic Growth on Native American Reservations in California." *Master's Theses. 1195.* 1-42. https://repository.usfca.edu/thes/1195/.

Askin, Noah, and Michael Mauskapf. 2017. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." *American Sociological Review* 82 (5): 910-944.

Barron, Alexander T. J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo. 2018. "Individuals, institutions, and innovation in the debates of the French Revolution." *PNAS* 4607-4612.

Baumol, William J., and William G. Bowen. 1966. *Performing Arts, The Economic Dilemma: a study of problems common to theater, opera, music, and dance.* Cambridge, Mass.: M.I.T. Press.

Beighey, Catherine, and N. Prabha Unnithan. 2006. "Political Rap: The Music of Oppositional Resistance." *Sociological Focus* 133-143.

Bettencourt, Luis M. A., Jose Lobo, Dirk Helbing, Christian Kuhnert, and Geoffrey B. West. 2007. "Growth, innovation, scaling, and the pace of life in cities." *PNAS* 7301-7306.

Bisin, Alberto, and Thierry Verdier. 2011. "Chapter 9: The Economics of Cultural Transmission and Socialization." In *Handbook of Social Economics*, by Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 339-416. Amsterdam: North-Holland.

Blanchard, Becky. 1999. "The Social Significance of Rap & Hip-Hop Culture." *Poverty & Prejudice: Media and Race* (Ethics of Development in a Global Environment (EDGE)) n.a.

Blei, David M., Andrew Y. Ng, and Michael L. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 993-1022.

Boldrin, Michael, and David K. Levine. 2005. "The economics of ideas and intellectual property." *PNAS* 1252-1256.

Carlino, Gerald A., and William Kerr. 2015. "Agglomeration and Innovation." *Bank of Finland Research Discussion Paper No. 27* 1-65.

Centola, Damon, and Andrea Baronchelli. 2015. "The spontaneous emergence of conventions: An experimental study of cultural evolution." *PNAS* 112 (7): 1989-1994.

Ciccone, Antonio, and Giovanni Peri. 2006. "Identifying Human-Capital Externalities: Theory with Applications." *The Review of Economic Studies* 381-412.

Currid, Elizabeth. 2007. "How Art and Culture Happen in New York." *Implications for Urban Economic Development* 73: 454-467. Accessed November 28, 2019.

DeDeo, Simon. 2018. "Information Theory for Intelligent People." *n.a.* 1-15.

Dowd, Timothy J. 2004. "Concentration and Diversity Revisited: Production Logics and the U.S. Mainstream Recording Market, 1940-1990." *Social Forces* 82 (4): 1411-1455.

Edelmann, Achim, and John W. Mohr. 2018. "Formal studies of culture: Issues, challenges, and current trends." *Poetics* 1-9.

Egas, Martijn, Ralph Kats, Xander van der Sar, Ernesto Reuben, and Maurice W. Sabelis. 2013. "Human cooperation by lethal group competition." *Scientific Reports* 1-4.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 736-767.
Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased so Much?" *The Quarterly Journal of Economics* 49-100.
Gachter, Simon, and Jonathan F. Schulz. 2016. "Intrinsic honesty and the prevalence of rule violations across societies." *Nature* 496-499.
Genius. n.d. *About Genius.* Accessed February 16, 2020. https://genius.com/Genius-about-genius-annotated.
Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 35-71.
Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 535-574.
Giuliano, Paola, and Nathan Nunn. 2017. "Understanding Cultural Persistence and Change." *NBER Working Paper No. 23617* 1-53.
Gomez-Lievano, Andres, Oscar Patterson-Lomba, and Ricardo Hausmann. 2016. "Explaining the prevalence, scaling and variance of urban phenomena." *Nature: Human Behaviour* 1-6.
Grief, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *The Journal of Political Economy* 912-950.
Hellmanzik, Christiane. 2010. "Location matters: Estimating cluster premiums for prominent modern artists." *European Economic Review* 199-218.
Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 483-499.
Kullback, S., and R. A. Liebler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 79-86.
Lloyd, Richard. 2002. "Neo-Bohemia: Art and Neighborhood Redevelopment in Chicago." *Journal of Urban Affairs* 517-532.
Loersch, C., and N. L. Arbuckle. 2013. "Unraveling the mystery of music: music as an evolved group process." *Journal of Personality and Social Psychology* 777-798.
Markusen, Ann, and Greg Schrock. 2006. "The Artistic Dividend: Urban Artistic Specialisation and Economic Development Implications." *Urban Studies* 1661-1686.
Marshall, Alfred. 1890. *The Principles of Economics.* London: Macmillan.
Martinez, Theresa A. 1997. "Popular Culture as Oppositional Culture: Rap as Resistance." *Sociological Perspectives* 40 (2): 265-286.
Mathews, Dolph. 2017. "List of the Most Common English Words." February 9. https://github.com/dolph/dictionary.
Michalopolous, Stelios, and Melanie Meng Xue. 2019. "Folkore." *NBER Working Paper Series* NBER Working Paper No. 25430: 1-77.
Miller, John W. 2017. "LyricsGenius: a Python client for the Genius.com API." August 21. https://github.com/johnwmillr/LyricsGenius.
Mitchell, Sara. 2019. "London calling? Agglomeration economies in literature since 1700." *Journal of Urban Economics* 16-32.
Murdock, Jaimie, Colin Allen, and Simon DeDeo. 2017. "Exploration and Exploitation of Victorian Science in Darwin's Reading Notebooks." *arXiv Cognition* 117-126.

Park, Judy Soojin. 2015. "Searching for a Cultural Home: Asian American Youth in the EDM Festival Scene." *Dancecult: Journal of Electronic Dance Music Culture* 7 (1): 15-34.
Patterson, Orlando. 2014. "Making Sense of Culture." *Annual Review of Sociology* 1-30.
Persaud, E. Jerry. 2011. "The Signature of Hip Hop: A Sociological Perspective." *International Journal of Criminology & Sociological Theory* 626-647.
Peterson, Richard A., and David G. Berger. 1975. "Cycles in Symbol Production: The Case of Popular Music." *American Sociological Review* 40 (2): 158-173.
Porter, Michael E. 1998, In Press. "Clusters and the New Economics of Competition." *Harvard Business Review*, November-December n.a.: n.a.
Potter, Antony, and H. Doug Watts. 2011. "Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle." *Journal of Economic Geography* 417-455.
Rehurek, Radim, and Petr Sokja. 2010. "Software Framework for Topic Modelling with Large Corpora." *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.* Valletta, Malta: ELRA, May 22.
Reich, Michael, David M. Gordon, and Richard C. Edwards. 1973. "Dual Labor Markets: A Theory of Labor Market Segregation." *The American Economic Review* 359-365.
Richerson, Peter, Ryan Baldini, Adrian V. Bell, Kathryn Demps, Karl Frost, Vicken Hillis, Sarah Mathew, et al. 2016. "Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence." *Behavioral and Brain Sciences* 39: 1-68.
Romer, Paul M. 1990. "Endogenous Technological Change." *The Journal of Political Economy* s71-s102.
Santos, Fernando P., Francisco C. Santos, and Jorge M. Pacheco. 2018. "Social norm complexity and past reputations in the evolution of cooperation." *Nature* 242-245.
Smith, Daniel, Philip Schlaepfer, Katie Major, Mark Dyble, Abigail E. Page, James Thompson, Nikhil Chaudhary, et al. 2017. "Cooperation and the evolution of hunter-gatherer storytelling." *Nature Communications* 1-9.
Steward, Alexander J., Todd L. Parsons, and Joshua B. Plotkin. 2016. "Evolutionary consequences of behavioral diversity." *PNAS* E7003-E7009.
Stigler, George J., and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum." *The American Economic Review* 67 (2): 76-90.
Throsby, David. 1994. "The Production and Consumption of the Arts: A View of Cultural Economics." *Journal of Economic Literature* 32 (1): 1-29. Accessed November 28, 2019.
Wang, Juan. 2016. "Classical Music: a Norm of "Common" Culture Embedded in Cultural Consumption and Cultural Diversity." *International Review of the Aesthetic and Sociology of Music* 47 (2): 195-205.
Weitzman, Martin L. 1998. "Recombinant Growth." *The Quarterly Journal of Economics* 332-360. Accessed November 28, 2019.
Young, H. Peyton. 2015. "The Evolution of Social Norms." *Annual Review of Economics* 359-387.

Appendix

Appendix A: Background on Rap Music and Justification of Research Context

As this is the first paper to our knowledge that focuses on Rap music in the Economics literature[8], it is important to provide a background on its attributes as an industry. It is generally agreed that the genre was formed in 1973 in the South Bronx neighborhoods of New York City. Hip Hop remained in New York for the next few years, until in 1979, Sugarhill Gang released the radio hit "Rapper's Delight", the first commercial success for Rap music.

Rap began as simple rhymes spoken over a DJ's beats, hyping the crowd and lauding turntablists' "skillz" (Blanchard 1999). Originally called MC's (or emcees), rappers were inspired by radio DJ's who would rhyme to introduce the next song. In fact, the earliest song in our Corpus is a short rhyme from a radio DJ. Rappers compete to innovate on rhymes and flows[9], and topics shifted from the ability of the turntablist and recognizing friends in the audience to experiences of the rapper and their community. Storytelling was an essential component in separating Rap as its own art form.

Blanchard (1999) points to previous forms of Black culture in explaining the function of Hip Hop for young, urban, working class African-Americans

> Hip-hop music originated from a combination of traditionally African-American forms of music—including jazz, soul, gospel, and reggae... While rap's history appears brief its relation to the African oral tradition, which provides rap with much of its current social significance, also roots rap in a long-stranding history of oral historians, lyrical fetishism, and political advocacy.

Two aspects of African oral tradition, *nommo* and *griots*, connect Rap to the African-American community. *Nommo* is a philosophical concept from the Malian Dogon ethnic group that describes an "animative ability of words and the delivery of words to act upon objects, giving life" (Blanchard 1999). *Griots* were African oral historians who were tasked with preserving knowledge of tribal history and genealogy (Persaud 2011). Rap is also connected to rhyming games, which were played in resistance to slavery, "Rhyming games encoded race relations between African-American slaves and their white masters in a way that allowed them to pass the scrutiny of suspicious overseers. Additionally, rhyming games allowed slaves to use their creative intellect to provide inspiration and entertainment" (Blanchard 1999). Rap quickly became a form of resistance in its own right, using rhymes to criticize racist political systems like police

[8] Jaynes (1998) and Caplan & Cowen (2004) do make mention of Rap music, but only in passing and as part of larger ethnographic datasets.

[9] Flow is the interaction of rhyme and rhythm, generally referring to how lyrics are delivered over a beat.

brutality, educational bias, and mass media misrepresentation (Beighy & Unnithan 2006). Persaud (2011) describes Rap music as essentially political in nature, stating, "Hip Hop emerged in direct response to ruling class power. It is an evolutionary form of the cultural resistance through language and music that Blacks have been using for generations" (629-630).

Rap music is severely understudied in the economic literature. However, the genre comes with multiple advantages in understanding innovation among cultural producers. Firstly, Rap is fascinating from an economic perspective. In undergraduate economics, we are taught of a unified labor market with a fixed supply of labor that is the entire labor force. Reality is more complicated than this. The distribution of labor market options is undoubtedly tied to the distribution of roles in a social hierarchy. A theory of segregated labor markets was first introduced by Reich, Gordon and Edwards (1973). American workers, divided by race, sex, etc., are subject to "different working conditions, different promotional opportunities, different wages, and different market institutions" (359). In other words, not everyone can have the same jobs, and different jobs come with different environments and compensation. This is the concept of primary and secondary markets. Primary labor markets are those that we traditionally attribute to ideals of success in the United States. They have stable, skilled jobs, upwards mobility, and good wages. However, secondary labor markets discourage stability, have low average wages, high turnover rates, and few opportunities for promotion. The authors acknowledge that racial minorities are often relegated to secondary markets, or subordinate or independent primary sub-markets, due to prejudice and labor market institutions.

We find that this perfectly describes the context in which Rap music is created. When members of the Black community view their prospects on the labor market, they see a secondary labor market. Drug dealing, sports, and Rap all represent stereotypical avenues to financial success in impoverished Black communities. The first risks social or institutional punishment. The latter two require high levels of human or social capital investment, but may be desired in comparison to other unstable jobs in the secondary labor market. It becomes clear why rappers may choose their career in music when considering the theory of superstars from Gabaix and Landier (2008). The appeal of being at the top of a power law distributed industry is hard to deny. A paucity of viable options pushes folks into less certain career paths. Viewing the labor market and other conditions in marginalized communities in the United States in such a way opens up opportunities to study the developed nation with development economics perspectives like in Ashabi (2019).

Secondly, research on the impact of cultural values on institutions have direct analogs in Rap. A cultural legacy of distrust and weak institutions damages intrinsic honesty in the present day (Gachter and Schulz 2016). Such an environment was manifested in America by slavery and continued racism. The emphasis of "realness" among rappers may be an attempt to protect from the dishonesty among community members springing from this history. Michalopolous and Xue (2019) follow a pattern theory of culture— that "all parts of culture are related and reflect the same values and beliefs" (12)— to look at how folklore influences cultural values. Lyrical storytelling is essential to Rap, and may be viewed as a mechanism of cultural transmission across space and through time. Bowles (2012) states that conflict is essential to the formation of societies and the overturning of inequality within communities. This can help explain the prevalence of "beef[10]" in Rap. Both storytelling and conflict are studied as the basis for human cooperation at large (Egas, et al. 2013; Loersch and Arbuckle 2013; Smith, et al. 2017). When applying the economic impacts of culture to the cultural product of Rap music, we can appreciate theory from the Sociological literature on this topic (Beighey & Unnithan 2006; Martinez 1997). Rap comes from a minority culture that reinforces resistance against continued historical oppressions. It is this through line of cultural production, evolution, and resistance that places the study of Rap in within the development economics literature.

More formally, this paper views American Rap as a sort of folkloric literature in itself. Rappers are the authors of lyrics collected into songs, which contains messages encouraging or discouraging certain cultural traits in its listeners. They collaborate, often within regions of origin, to recombine cultural products in a novel way. Therefore, urban clusters should exhibit the highest levels of novelty because artists' proximity allow for more frequent and intense recombinations.

Here we have established the setting of this study. Rap music presents a fascinating context through which to view the consequences of cultural production. Further, it lends itself to textual analysis because of the lyrical emphasis of Rap in comparison to other popular music genres, rapper origin is often claimed with pride in songs, and rappers from the same city share comradery in their music through features. Out of all the genres of popular music, we find Rap to be the ideal site of this study. However, there is one major weakness associated with choosing this setting. Music has several dimensions for innovation aside from lyrics, including flow, beats,

[10] Beef is essentially a feud among rappers. It is the inspiration for diss tracks, which aim to lyrically disarm the opponent, and even results in physical violence.

sampling and flipping previous music, etc. Our methods will not capture innovation in these dimensions. Therefore, this analysis will only look at where the most novel and influential lyricists are from, not necessarily the most innovative overall.

Appendix B: Maximum Novelty and Resonance by Rapper, Top 50 of the 250 Most Productive Rappers
We restricted our dataset to the 250 most frequent unique rappers (Barron et al. 2018), then ordered by maximum novelty and resonance. This acts as a tangential robustness check on the validity of our combined LDA and KLD methodologies. Essentially, we check that the most productive, innovative, and influential rappers according to our methods match common understandings of the genre's landscape.

Artist Name	Max Novelty
Mac Miller	8.52
Nas	8.39
21 Savage	8.36
Nicki Minaj	8.32
Joe Budden	8.19
Mistah F.A.B.	8.14
Styles P	8.11
Jarren Benton	8.09
Mary J. Blige	8.06
Open Mike Eagle	7.98
Young Scooter	7.94
Lil Durk	7.85
NF	7.85
Papoose	7.84
Bow Wow	7.82
DaBaby	7.81
Kid Cudi	7.77
J. Cole	7.76
Playboi Carti	7.76
Lil Tracy	7.75
Smoke DZA	7.7
Colt Ford	7.67
Black Milk	7.62
Charles Hamilton	7.61
Young Dolph	7.61
Lil Reese	7.61
Mac Lethal	7.58
Ice Cube	7.56

Dizzy Wright	7.55
Skeme	7.54
Freddie Gibbs	7.52
Young Thug	7.52
Fetty Wap	7.49
N.O.R.E.	7.47
Z-Ro	7.44
2 Chainz	7.44
MC Lars	7.42
Lil Xan	7.41
Ab-Soul	7.38
Skyzoo	7.37
Pitbull	7.37
Chris Webby	7.35
Akon	7.35
Dave East	7.35
Drake	7.34
Lil Dicky	7.33
Timbaland	7.31
Juvenile	7.3
Busta Rhymes	7.29
French Montana	7.29

Artist Name	Max Resonance
MC Eiht	0.733
Master P	0.574
Nas	0.41
Russ	0.408
Mary J. Blige	0.408
French Montana	0.372
Killah Priest	0.341
Big K.R.I.T.	0.329
Fat Joe	0.325
Spice 1	0.324
Scarface	0.321
Sage Francis	0.314
Gucci Mane	0.308
Aesop Rock	0.303
LL Cool J	0.294
Gunna	0.292

Mac Lethal	0.289
AZ	0.284
Method Man	0.281
Ghostface Killah	0.28
The Notorious B.I.G.	0.273
Azealia Banks	0.27
C-Rayz Walz	0.263
Chino XL	0.262
Xzibit	0.262
Krayzie Bone	0.257
Lil Durk	0.252
Busdriver	0.251
South Park Mexican	0.249
N.O.R.E.	0.248
DMX	0.247
CeeLo Green	0.244
Mac Dre	0.244
Silkk the Shocker	0.243
Charles Hamilton	0.24
Smokepurpp	0.235
Wyclef Jean	0.233
Ras Kass	0.231
Diddy	0.23
Starlito	0.227
Missy Elliott	0.226
Lil Wayne	0.226
Nelly	0.22
Busta Rhymes	0.217
Comethazine	0.213
Z-Ro	0.213
Shy Glizzy	0.211
Kodak Black	0.211
Mac Miller	0.21
MF Grimm	0.209

Appendix C: Maximum Novelty and Resonance by MSA, Top 10

In discourse surrounding Rap music, New York City, Southern California, Atlanta, and Chicago are celebrated as highly innovative and influential scenes for the genre as a whole. We present these tables as a robustness check that our methodology matches the discourse spatially. These tables partially match our prior expectations.

MSA	Max

	Novelty
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	8.132026
Atlanta-Sandy Springs-Alpharetta, GA Metro Area	8.073611
Boston-Cambridge-Newton, MA-NH Metro Area	7.992399
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	7.977849
Charlotte-Concord-Gastonia, NC-SC Metro Area	7.810751
Fayetteville, NC Metro Area	7.760603
New York-Newark-Jersey City, NY-NJ-PA Metro Area	7.628901
Riverside-San Bernardino-Ontario, CA Metro Area	7.414847
Bainbridge, GA Micro Area	7.312618
Syracuse, NY Metro Area	7.245005

MSA	Max Resonance
Atlanta-Sandy Springs-Alpharetta, GA Metro Area	0.4084246
New York-Newark-Jersey City, NY-NJ-PA Metro Area	0.3715656
Evansville, IN-KY Metro Area	0.31642
Los Angeles-Long Beach-Anaheim, CA Metro Area	0.292172
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	0.2515152
Houston-The Woodlands-Sugar Land, TX Metro Area	0.2490117
St. Louis, MO-IL Metro Area	0.2133138
Miami-Fort Lauderdale-Pompano Beach, FL Metro Area	0.2107557
Boston-Cambridge-Newton, MA-NH Metro Area	0.2033038
Pittsburgh, PA Metro Area	0.2021501

Appendix D: Topic Classification

Latent Dirichlet Allocation was used on our dataset of Rap lyrics to transform each song into a probability distribution of 100 topic bins. LDA uses machine learning to agnostically weight these songs for each topic. To pull out the associations that are meaningful to us, we find exemplar songs and attempt to classify their general topic (Barron, et al. 2018). Exemplar songs are defined here as those with the five largest weights in each topic bin. Below are our manually classified topics and the exemplar songs' title, artist, and probability weight. Song titles frequently contain profanity, which we choose not to censor unless such language is racially charged.

Topic 1: The Downsides of Gangbangin'

- "The Message" by Nas (.576)
- "N****s" by The Notorious B.I.G. (.573)

- “Rikers Island” by Troy Ave (.553)
- “Ruffest Ryders” by DMX (.547)
- “N.Y. State of Mind, Pt. III” by Nas(.543)

Topic 2:

- “Mama Africa” by Akon (.746)
- “Puppy Love” by Nate Dogg (.738)
- “Intro (Being Myself)” by Juvenile (.726)
- “Send Your Love” by Krizz Kaliko (.627)
- “Love No Limit (Puff Daddy Mix)” by Mary J. Blige (.580)

Topic 3: Spanglish

- “Mami” by Vico C (.941)
- “Bikini Weather/ Corazon en Afrika” by Princess Nokia (.879)
- “Take It Off (Spanglish Version)” by Lil Jon (.791)
- “Run, Come, Follow Me” by Daddy Yankee (.720)
- “Take It Off” by Lil Jon (.643)

Topic 4:

- “Seeing Double” by 21 Savage (.918)
- “Juicy” by Doja Cat (.755)
- “Crack” by 2 Chainz (.723)
- “Bojangles (Remix)” by Pitbull (.720)
- “Bojangles” by Pitbull (.643)

Topic 5:

- “Posse on Broadway (Anime remix)” by Sir Mix-a-Lot (.885)
- “Cock the 9” by Eazy-E (.859)
- “Glock With A Dyck” by Plies (.767)
- “Endonesia” by Chuck D (.712)

- “Rock Bottom” by Joyner Lucas (.659)

Topic 6:

- “Molly” by Tyga (.866)
- “MY SAMSUNG” by Lil Tracy (.773)
- “Foreigns” by Chief Keef (.761)
- “Venezuelan Foreign” by Comethazine (.712)
- “harmony” by Lil Tracy (.637)

Topic 7:

- “Where Y’all At” by Nas (.895)
- “Okey Dog” by Murs (.775)
- “Really From Da Hood” by Plies (.596)
- “Responsibilities” by Don Trip (.573)
- “STUPID (Clean)” by Ashnikko (.530)

Topic 8: Southerness

- “Like That” by Isaiah Rashad (.810)
- “Bubba Gump” by Yung Baby Tate (.738)
- “Ballad Of The Bass (My Sub V)” by Big K.R.I.T. (.732)
- “Red White Blue & Blessed” by Colt Ford (.659)
- “Intro” by Yelawolf (.618)

Topic 9:

- “Teeny Bopper crack Whore” by Porcelain Black (.599)
- “Anttila Ambrus” by Astronautalis (.446)
- “Roman’s Revenge Alternate Version (I FEEL PRETTY)” by Nicki Minaj (.423)
- “Roman’s Reveng” by Nicki Minaj (.407)
- “No Jumper Cables” by Aesop Rock (.405)

Topic 10:

- “Immutable” by Shai Linne (.631)
- “Still the Same” by YFN Lucci (.572)
- “Make It Rain” by Fat Joe (.547)
- “Make It Rain (Trina Remix)” by Fat Joe (.532)
- “Sabbath” by Chuuwee (.524)

Topic 11:

- “Ready” by Tate Kobang (.693)
- “Bird Is the Word” by Ski Mask the Slump God (.661)
- “Feel The Bern (When You Pee)” by Rucka Rucka Ali (.643)
- “Turn It” by Dre Murray (.601)
- “Turn It to a Rave” by Waka Flocka Flame (.595)

Topic 12: Exceptionally Frequent Use of the N-Word

- “N**** Whut” by Redman (.641)
- “For my Lil n****s” by Boosie Badazz (.508)
- “My N****” by YG (.505)
- “Amber Alert” by Young Buck (.501)
- “N**** N****” by 50 Cent (.497)

Topic 13:

- “Never Leave Me Alone” by Nate Dogg (.735)
- “Home” by Illa J (.672)
- “Danger Zone” by Big L (.657)
- “Troubled Boy” by Kid Cudi (.599)
- “Don’t Go” by Mary J. Blige (.566)

Topic 14:

- “Milky Cereal” by LL Cool J (.818)

- "Old Man" by Masta Killa (.756)
- "Dennehy" by Serengeti (.724)
- "Mushrooms And War" by Charles Hamilton (.551)
- "Fat Belly" by Homeboy Sandman (.521)

Topic 15:

- "My Friend (Jack Daniels)" by Jon Bellion (.718)
- "Best Friend" by 50 Cent (.672)
- "Best Friend (Remix)" by 50 Cent (.611)
- "Tell Me a Secret" by Ludacris (.543)
- "Tell Me (unreleased)" by Boogie (.487)

Topic 16:

- "Blood" by El-P (.772)
- "I'll Be Good" by Hodgy (.640)
- "Left, Right" by YG (.631)
- "Who's There?" by Hopsin (.622)
- "Magic Walk" by Georgia Anne Muldrow (.601)

Topic 17:

- "Hollow Tips (Freestyle)" by 21 Savage (.866)
- "Slidin" by Blueface (.801)
- "Uncle Phillip" by Blueface (.677)
- "LA to NY" by Desiigner (.668)
- "Never Heard" by 03 Greedo (.650)

Topic 18:

- "Party No Mo'" by Ludacris (.782)
- "Jump" by Smokepurpp (.762)
- "Party Train" by Redfoo (.722)

- “Party Started” by Wyclef Jean (.690)
- “Tippin’ In Da Club” by Nelly (.683)

Topic 19: Drug Dealer as a Persona

- “Wrap Me in Your Arms” by NF (.772)
- “Maniac Drug Dealer III” by Lil Ugly Mane (.634)
- “Birth To A Drug Dealer” by Stitches
- “Drug Dealer (Remix)” by Stitches
- “Drug Dealer” by Stitches (.591)

Topic 20: Christian Religion

- “Our Treasure Is Christ” by Shai Linne (.818)
- “The Temple” by Timothy Brindle (.817)
- “Behold the Lamb (Live)” by Passion (.817)
- “Worship God” by Shai Linne (.779)
- “Disciple” by Nas (.775)

Topic 21:

- “No Bleedin” by Vince Staples (.870)
- “Lot to Learn (Music Video Version)” by Luke Christopher (.668)
- “Higher Learning (Intro)” by Dizzy Wright (.598)
- “Noticeably Negro” by Serengeti (.538)
- “Hoes Study Me But They Can’t Graduate” by Princess Vitarah (.534)

Topic 22:

- “Ankle” by Ski Mask the Slump God (.712)
- “Phenomenon” by LL Cool J (.574)
- “Say Less” by K. Sparks (.557)
- “I Like Head” by YG (.555)
- “Heaven” by Nelly (.534)

Topic 23:

- “REVOFEV” by Kid Cudi (.832)
- “Just Touched Down” by Big K.R.I.T. (.717)
- “Favor for a Favor” by Nas (.712)
- “No More” by Swoope (.518)
- “Kelly Rowland” by MadeinTYO (.510)

Topic 24:

- “Bang” by Rye Rye (.629)
- “Kill Yo Self” by Bodega Bamz (.608)
- “She Wildin’” by Fabolous (.608)
- “Change Gon’ Come (50 Shots)” by Papoose (.601)
- “Stomp!” by Shaggy 2 Dope (.555)

Topic 25: Nightlife

- “Last Night (Christoph Andersson Remix)” by G-Eazy (.785)
- “Bright Lights Bigger City” by CeeLo Green (.785)
- “Partytime” by Missy Elliott (.764)
- “Dolla Slice” by Dai Burger (.558)
- “Checked” by Russ (.535)

Topic 26: Luxury Consumerism

- “Fashion Week” by Rico Nasty (.703)
- “Phantom” by Smokepurpp (.701)
- “Audemar” by Young Thug (.667)
- “FIJI” by Lil Pump (.658)
- “Youngest Flexer” by Lil Pump (.645)

Topic 27: Aggressive Gun Violence (Old School)

- “Buckin’ Em Down” by LL Cool J (.773)
- “We Come Strapped” by MC Eiht (.701)
- “N****z That Kill (Endolude)” by MC Eiht (.699)
- “Set Trippin’” by MC Eiht (.673)
- “Clip & the Trigga” by Spice 1 (.637)

Topic 28:

- “Freaks” by French Montana (.796)
- “MVP” by Charles Hamilton (.739)
- “Freaks (Hip-Hop Remix)” by French Montana (.707)
- “Raising Hell (Justin Caruso Remix)” by Kesha (.688)
- “Wiggy” by South Park Mexican (.684)

Topic 29:

- “Tired” by Deniro Farrar (.857)
- “Steelspitting(mp3codec error included)” by Lil Ugly Mane (.801)
- “Til I Die Remix” by Upchurch (.612)
- “I.K.E.A” by XV (.529)
- “Movin” by Lil Pump (.453)

Topic 30: Dreams and Aspirations

- “Banged and Blown Through” by Saul Williams (.636)
- “A Dream” by Mary J. Blige (.565)
- “The World Is Yours” by Nas (.530)
- “Take the World” by Jon Connor (.510)
- “Destiny” by NF (.496)

Topic 31: Smoking Marijuana

- “Smokin Killah” by Webbie (.812)
- “Blueberry Yum Yum” by Ludacris (.802)

- “Marley and Me (Remix)” by Smoke DZA (.801)
- “In the Cut” by Wiz Khalifa (.727)
- “Kush Ups” by Project Pat (.702)

Topic 32:

- “Clap Your Hands” by LL Cool J (.731)
- “Conscious” by Charles Hamilton (.710)
- “Letter to the Firm (Holy Matrimony)” by Foxy Brown (.658)
- “The Bullpen (Live)” by Dessa (.533)
- “The Bullpen” by Dessa (.508)

Topic 33:

- “Leaving Interscope” by 50 Cent (.608)
- “Self Made” by Philthy Rich (.501)
- “MTV Intro” by 50 Cent (.476)
- “The Break Up” by Spose (.476)
- “DJ DiBiase Intro” by Big K.R.I.T. (.411)

Topic 34:

- “True Religion Fein” by Chief Keef (.733)
- “Genaside” by K.A.A.N. (.559)
- “Juice (Kito Remix)” by Lizzo (.549)
- “Juice” by Lizzo (.543)
- “Two Pairs” by MC Jin (.536)

Topic 35:

- “On All My Mommaz” by Mistah F.A.B. (.903)
- “Push Back (Remix)” by Christopher Martin (.820)
- “Push Back” by Christopher Martin (.760)
- “The Tale of Mr. Morton” by Skee-Lo (.648)

- "Freaky Gurl" by Gucci Mane (.602)

Topic 36:

- "BLACK BALLOONS | 13LACK 13ALLOONZ (Like A Version)" by Denzel Curry (.769)
- "Beach Boy" by Imani Coppola (.701)
- "Grow" by Cozz (.648)
- "Too Much" by Sims (.593)
- "Closer" by Mike Stud (.513)

Topic 37:

- "Type of Way (Remix)" by Ludacris (.635)
- "Type Of Way" by Rich Homie Quan (.584)
- "Still Brazy" by YG (.563)
- "NCS" by K.A.A.N. (.530)
- "No Service" by Futuristic (.473)

Topic 38:

- "Unbelievable" by The Notorious B.I.G. (.694)
- "Pistol Smoke" by Ghostface Killah (.648)
- "Shout Out to My Set" by Gucci Mane (.641)
- "I'm Coolin" by Fat Trel (.588)
- "Notorious Thugs" by The Notorious B.I.G. (.567)

Topic 39:

- "What Is Evil" by Esham (.693)
- "Cypher with Self" by Canibus (.687)
- "My Queen" by Joyner Lucas (.659)
- "4 Business Benefits of Installing Queue Management Systems" by Pete Rock (.645)
- "One in a Million" by K.A.A.N. (.626)

Topic 40:

- “WITCH N****S GON CAST SPELLS 2 DIS” by Lil Tracy (.847)
- “Unforgettable (Mariah Carey Acoustic Remix)” by French Montana (.773)
- “Unforgettable (StarFox Remix)” by Swae Lee (.762)
- “Unforgettable (Mariah Carey Remix)” by French Montana (.711)
- “That’s Detroit” by King Gordy (.693)

Topic 41:

- “All of Us” by Talib Kweli (.628)
- “All About That Rape” by Rucka Rucka Ali (.570)
- “Black (Single Version)” by Buddy (.557)
- “Super AIDS” by Rucka Rucka Ali (.547)
- “100 Black Coffins” by Rick Ross (.542)

Topic 42:

- “Bombay” by Timbaland (.796)
- “Reputations On The Line” by Kevin Gates (.752)
- “Pressed for Time (Crossed My Mind)” by Mick Jenkins (.614)
- “Elevate” by Wrekonize (.516)
- “Mischief on My Mind” by YNW Melly (.485)

Topic 43:

- “Modern Warfare” by Hasan Salaam (.812)
- “Chillin’ Like A Villain” by Indo G (.618)
- “Tonight” by Kid Ink (.453)
- “Wings” by Vic Mensa (.399)
- “Jump Out” by Gunplay (.395)

Topic 44: Referring to Women as Baby

- “Baby Doll” by Mickey Avalon (.794)

- “Baby” by LL Cool J (.622)
- “Can You Feel It” by Sean Kingston (.544)
- “Baby Girl” by Bryce Vine (.536)
- “I Need a Girl, Pt .3” by Tyga (.527)

Topic 45:

- “Summa Dis” by Freddie Gibbs (.829)
- “Mook” by Bas (.798)
- “Boyfriend” by SahBabii (.619)
- “That’s Wusup” by Mac Dre (.580)
- “Nag Champa (Live at the North Sea Jazz Festival)” by Common (.557)

Topic 46: Self-Promoting Braggadocio (Old School)

- “Coldblooded” by Common (.829)
- “Me Versus Me” by Biz Markie (.654)
- “Afro Puffs” by The Lady of Rage (.635)
- “Afrika shox” by Afrika Bambaataa (.614)
- “Yo Vanilla” by Vanilla Ice (.602)

Topic 47:

- “Da Ntro” by DJ Jazzy Jeff (.538)
- “Gettaway” by Missy Elliott (.523)
- “Rhyme Doublin’ (Clean)” by C-Rayz Walz (.504)
- “Jewish Flow” by Lil Dicky (.468)
- “Rhyme Doublin’ (Clean)” by C-Rayz Walz (.457)

Topic 48:

- “Ricky” by Meek Mill (.655)
- “Cold Cruel World” by Boondox (.604)
- “Devils & Angels” by Freeway (.557)

- “Greatful” by Sevin (.524)
- “Keep The Devil Off” by Big K.R.I.T. (.510)

Topic 49:

- “Stop Schemin’” by Styles P (.671)
- “Dearly Departed” by Aaron Carter (.525)
- “Minimum Wage” by Taylor Bennett (.495)
- “Wu-Tang Forever” by Drake (.488)
- “It Ain’t a Problem” by Rick Ross (.459)

Topic 50: Reminiscing/Regrets

- “Nothing Lasts Forever” by J. Cole (.683)
- “Tear Us Apart” by Lil Tjay (.592)
- “I Remember” by Takeoff (.554)
- “I Remember” by Quando Rondo (.536)
- “Mad” by Hoodie Allen (.527)

Topic 51: Flocka

- “Bringing Gangsta Back” by Waka Flocka Flame (.891)
- “Gangsta Hop” by Waka Flocka Flame (.720)
- “DJEMBA FLOCAK FLAME$” by Waka Flocka Flame (.694)
- “Serena” by Akua Naru (.626)
- “MY” by Jay Park (.595)

Topic 52:

- “Get In Get Out” by Lil Jon (.727)
- “MTF” by Plies (.704)
- “2 Week Notice” by Jon Connor (.529)
- “Now” by Logic (.485)
- “MF” by Chief Keef (.461)

Topic 53: Knowledge of Relevant Information

- "What's Happenin'" by Webbie (.854)
- "That's What" by Chief Keef (.751)
- "I Know How to Self Destruct" by Trippie Redd (.475)
- "I Know" by Chad Jones (.457)
- "Silverlined" by Peace (.451)

Topic 54: Sexual Desirability of Women

- "Freaky Hoe" by Rick Ross (.915)
- "FMU" by Brooke Candy (.744)
- "FMU (Demo)" by Brooke Candy (.738)
- "2 Fat Fucks" by Big Scoob (.668)
- "Very Rare Supreme" by Kent Jones (.577)

Topic 55:

- "Riot" by Azealia Banks (.656)
- "Neon Skies" by Wrekonize (.644)
- "Phone Numbers – triple j Like A Version" by Dominic Fike (.612)
- "Murals" by Dumbfoundead (.612)
- "Questions" by Chaz French (.584)

Topic 56:

- "KARL" by Chanel West Coast (.674)
- "Dukes of Hazzard" by Upchurch (.599)
- "Outstanding" by MadeinTYO (.521)
- "Action" by Joey Fatts (.504)
- "DEMI LOVATO ON MY WRIST" by Lil Tracy (.434)

Topic 57: Social Isolation

- “Underwater FlyZone” by Trippie Redd (.650)
- “Lonely” by Yung Bans (.631)
- “Imaginary” by Peace (.586)
- “How You Feel” by Trippie Redd (.551)
- “Project Windows” by Nas (.544)

Topic 58: Male Ability to Collect Sexual Partners

- “Momma” by Kid Buu (.890)
- “So Many Hoes” by King Louie (.592)
- “10 Bad Bitchez” by SahBabii (.551)
- “Hoes On My Dick” by Kreayshawn (.537)
- “WONDO” by 6ix9ine (.530)

Topic 59:

- “Got Muscle” by GoldLink (.821)
- “We Hustle” by Freekey Zekey (.784)
- “The Christmas Song (Chestnuts Roasting On an Open Fire” by Mary J. Blige (.780)
- “Nada” by Emilio Rojas (.756)
- “Nada Remix” by Emilio Rojas (.735)

Topic 60:

- “Nobody Hates Nothin’” by King Gordy (.623)
- “My Money” by Nate Dogg (.553)
- “From Then Till Now” by Killah Priest (.533)
- “Zero” by MC Jin (.529)
- “SAD (Clean)” by XXXTENTACION (.436)

Topic 61:

- “Tell The Children” by Tink (.559)
- “Running the Streets” by Rick Ross (.537)
- “What U Doin’ Wid It” by Peewee Longway (.480)

- “Getting’ It” by Chingy (.476)
- “XXL Freshman Freestyle: Desiigner” by Desiigner (.468)

Topic 62:

- “Never Scared” by Bone Crusher (.742)
- “To Zion” by Lauryn Hill (.691)
- “Deaf” by Derek Minor (.580)
- “Haten” by Chief Keef (.579)
- “Air Force Ones (Remix)” by David Banner (.577)

Topic 63:

- “CALLIGRAPHY” by Saba (.656)
- “My Life” by Prozak (.631)
- “Up” by Sean C. Johnson (.573)
- “Schizophrenia” by Dreezy (.552)
- “Nowhere Fast” by Mary J. Blige (.508)

Topic 64: Drug Production and Trade (Cooking Up)

- “Boomin’” by 21 Savage (.767)
- “Cook” by Chief Keef (.751)
- “H2O” by 21 Savage (.687)
- “Packages” by Freddie Gibbs (.593)
- “Cray” by OJ da Juiceman (.576)

Topic 65:

- “Flooded” by MadeinTYO (.759)
- “Screaming” by Kirko Bangz (.692)
- “SADE IN THE 90s” by Qveen Herby (.616)
- “God Complex” by Sadistik (.611)
- “Yeah” by Don Trip (.608)

Topic 66:

- "Whip It" by Omarion (.879)
- "Clear As Day" by Tre Capital (.711)
- "Hummin'" by Marley Marl (.679)
- "Up Up & Away" by Kid Cudi (.668)
- "Speakers On Blast" by Skyzoo (.595)

Topic 67: Possession of Large Amounts of Currency

- "Hunnid Bands" by Wiz Khalifa (.698)
- "Hit the Lotto" by Chief Keef (.672)
- "Poppin' Rubber bands" by Wiz Khalifa (.634)
- "21 Thousand Units" by Twisted Insane (.633)
- "Fuck Up Some Commas (Remix)" by Lil Wayne (.611)

Topic 68:

- "Twerk It (Remix)" by Busta Rhymes (.660)
- "Good to Go" by Yelawolf (.634)
- "Hollis to Hollywood" by LL Cool J (.621)
- "Twerk It" by Busta Rhymes (.596)
- "Don't Reach For My Chain" by Young Scooter (.574)

Topic 69:

- "Flaw" by Plies (.737)
- "Fuckin N****z Bitches" by 21 Savage (.536)
- "Paid Bitches" by Kash Doll (.534)
- "My N****s" by DMX (.496)
- "Nathan" by Azealia Banks (.485)

Topic 70:

- "Southside Eagle (93 Bulls)" by Open Mike Eagle (.859)
- "Jarreau of Rap (Skatt Attack)" by Nas (.805)
- "Confrontation" by Mary J. Blige (.767)
- "My Language" by Macklemore (.635)
- "Tricky (Backstreet in Compton)" by Problem (.603)

Topic 71:

- "Live It Up" by Lil Durk (.867)
- "We Ride" by Gucci Mane (.824)
- "Ride Waves" by Young Scooter (.801)
- "Cybiko" by Princess Nokia (.631)
- "My Life" by Pitbull (.567)

Topic 72:

- "Elevate & Motivate" by Trippie Redd (.684)
- "Layup" by Big K.R.I.T. (.660)
- "Raincoat remix" by Nebu Kiniza (.649)
- "It Hit Different" by Uncle Murda (.518)
- "Raincoat" by Nebu Kiniza (.516)

Topic 73:

- "Leviathan" by G-Eazy (.664)
- "Leviathan (Hippie Sabotage Remix)" by G-Eazy (.580)
- "Put Me In Coach" by Dee-1 (.530)
- "Play Yaself" by French Montana (.526)
- "Wet Dreamz (Live)" by J. Cole (.508)

Topic 74:

- "Manipulator" by Kat Dahlia (.437)
- "Shop" by Cousin Stizz (.404)

- “Mouf” by Rocko (.365)
- “Glorified” by Andre Nickatina (.330)
- “I Can’t” by Trouble (.307)

Topic 75:

- “Professor Finessor” by Bali Baby (.840)
- “Hold It Back” by Oddisee (.632)
- “Back to the Otherside” by Kid Rock (.577)
- “El Tornado” by Jay Park (.572)
- “Diamonds” by Tedashii (.532)

Topic 76:

- “Dwellin’ in tha Labb” by JT the Bigga Figga (.743)
- “Pimp On” by Twista (.653)
- “Teenage Thug” by Nas (.546)
- “Intro” by JT the Bigga Figga (.523)
- “Playa Wit Game” by Master P (.507)

Topic 77:

- “Trilogy part Duh – The Street” by Jean Grae (.642)
- “On the Subject of Normal… (intro) by Adam WarRock (.621)
- “Double Katanas” by Jean Grae (.595)
- “The Unbeaten (Intro)” by Hell Razah (.569)
- “3%” by Hobo Johnson (.555)

Topic 78:

- “I Got” by Chief Keef (.773)
- “James Bond” by Lil Durk (.679)
- “Bitch I’m Posh” by Princess Nokia (.627)
- “Boss Ass Bitch (Remix)” by Nicki Minaj (.602)

- “Lil Bitch Remix” by Katie Got Bandz (.569)

Topic 79:

- “Deadman Walking” by Blaze Ya Dead Homie (.885)
- “Pull Up” by Chaz French (.604)
- “Snap Chat” by Khia (.578)
- “Let ‘Em Talk” by Kesha (.542)
- “Nada” by Iamsu! (.512)

Topic 80: Murder

- “Dome Split” by Twisted Insane (.744)
- “Murderers & Robbers” by Project Pat (.695)
- “Murder Rate” by Peewee Longway (.596)
- “Kill Bill” by Young Nudy (.499)
- “Hellter Skkklter (RADIO EDIT)” by Esham (.488)

Topic 81:

- “It’s All” by Colt Ford (.803)
- “Eye of the Tiger” by Ace Hood (.641)
- “N.A.S.A.” by Chris Rivers (.608)
- “Gray AREAS” by John Givez (.583)
- “Flying Away” by Mary J. Blige (.555)

Topic 82:

- “Beat It” by Sean Kingston (.736)
- “Nikes On My Feet” by Mac Miller (.707)
- “Four Leaf Clover” by Charles Hamilton (.672)
- “LEASH” by Lil Xan (.552)
- “No Surrender No Retreat” by Boosie Badazz (.537)

Topic 83:

- "Drop Girl" by Ice Cube (.831)
- "Houstatlantavegas" by Drake (.802)
- "Unzip Me" by Cazwell (.768)
- "Life Is a Gamble" by Big K.R.I.T. (.712)
- "Watch Yo Lady" by Ab-Soul (.585)

Topic 84:

- "Blow Me A Dub (Remix)" by Max B (.559)
- "Air Force Ones (Remix)" by Nelly (.505)
- "Versace" by Tyga (.469)
- "Air Force Ones" by Nelly (.469)
- "Throw It In The Bag" by Fabolous (.466)

Topic 85:

- "R.I.P Dolla" by Tyga (.698)
- "Minajatwa" by Lil Tracy (.567)
- "M.M.M. (Marilyn Maryland Marilyn)" by Trinidad James (.513)
- "John Madden (2010 Version)" by Spose (.470)
- "Before I Self Destruct – Thank You's" by 50 Cent (.469)

Topic 86:

- "Prank Calls" by Froggy Fresh (.517)
- "Mr. Scarface (Radio Version)" by Scarface (.497)
- "Lola from the Copa" by MC Lyte (.453)
- "I Tried" by Ray J (.438)
- "Not Sure Why I Came Back" by Blueprint (.437)

Topic 87: Spiritual Metaphor

- "Crusaids" by Killah Priest (.731)

- “Amplified Sample” by GZA (.729)
- “B.I.B.L.E. (Basic Instructions Before Leaving Earth)” by GZA
- “Islamophobic Lullabies” by Lowkey (.617)
- “Beyond the Galaxy” by Egyptian Lover (.567)

Topic 88: Nihlism

- “Nothing Matters” by Jay Park (.907)
- “Searching” by Mary J. Blige (.852)
- “Beautiful Day” by Mary J. Blige (.570)
- “Hate” by Mannie Fresh (.568)
- “Nothing Even Matters” by Lauryn Hill (.548)

Topic 89:

- “Round Em Up” by Canon (.776)
- “Money Over Bullshit” by Nas (.694)
- “Sound the Bells (Live)” by Dessa (.633)
- “Kitana” by Princess Nokia (.628)
- “Miss Camaraderie (Bon Vivant Remix)” by Azealia Banks (.582)

Topic 90: Odes to Women (Old School)

- “Girls” by Pitbull (.857)
- “Jealouso” by Pitbull (.759)
- “Fancy” by Drake (.715)
- “Girls” by Egyptian Lover (.642)
- “Jiggable Pie” by AMG (.631)

Topic 91:

- “Rock ‘N’ Roll” by Taylor Bennett (.712)
- “Blueberry Chills” by Chanel West Coast (.687)
- “0 to 100 (Freestyle)” by Chingo Bling (.549)
- “DR. PHIL” by Lil Pump (.542)

- “Beverly Hills” by Qveen Herby (.527)

Topic 92:

- “When I Get My Check ($, $, $)” by Lyrics Born (.884)
- “Burning Tears” by Ski Mask the Slump God (.813)
- “We Getting Money” by Meek Mill (.709)
- “Trap My Ass Off” by Waka Flocka Flame (.663)
- “Get Money” by Chief Keef (.643)

Topic 93:

- “Ram-Faced Boy” by Esoteric (.649)
- “Guilty Conscience” by 070 Shake (.589)
- “Palmreader” by Sadistik (.521)
- “Intro” by Anybody Killa (.520)
- “Angel” by Akon (.509)

Topic 94:

- “Each Tear” by Mary J. Blige (.857)
- “Over The Edge” by Akon (.690)
- “You’re the Only One (Feat. a.R. Flexx)” by George Moss (.622)
- “I’m Grateful” by Lazarus (.621)
- “When the Sun Goes Down” by Phora (.589)

Topic 95:

- “GODMODEGAME666 X TEKASHI 69” by 6ix9ine (.783)
- “Flossin” by Tyga (.699)
- “Bust It Open” by French Montana (.690)
- “Loaded” by Lil Uzi Vert (.610)
- “Stanky Leg Remix” by G-Eazy (.585)

Topic 96:

- “Introduce Me” by Lil Durk (.746)
- “Protect and Serve” by DaBaby (.587)
- “If I Could” by Lil Durk (.554)
- “I Could” by Lil Durk (.530)
- “Drippy” by Young Dolph (.526)

Topic 97:

- “Summer11” by Skizzy Mars (.616)
- “The Artful Dodger” by Mick Jenkins (.522)
- “Attitude” by Leikeli47 (.506)
- “Situation” by Young Scooter (.503)
- “Harambe Forever” by Rucka Rucka Ali (.477)

Topic 98:

- “Diggy Down” by LL Cool J (.829)
- “In A Minute Doe” by Gift of Gab (.827)
- “Slow Motion” by Juvenile (.655)
- “Settle Down” by Rico Love (.607)
- “Round & Round” by Iamsu! (.575)

Topic 99:

- “Hood Scriptures” by Foxy Brown (.791)
- “PUT DIS GUCCI PURSE AROUND YO SHOULDER” by Lil Tracy (.762)
- “Stupid” by Gucci Mane (.684)
- “Hillbilly” by Upchurch (.669)
- “D Rose” by Lil Pump (.665)

Topic 100:

- “Fix Ya Face” by Khia (.728)

- "Bake a Cake" by Lil Debbie (.501)
- "Maybach Music VI (Extended)" by Rick Ross (.477)
- "Early 20 Rager" by Lil Uzi Vert (.444)
- "Too Late" by Trippie Redd (.425)