RESEARCH

Novelty and Cultural Evolution in Modern Popular Music

Katherine O'Toole* and Emőke-Ágnes Horvát

*Correspondence:
kmarie.otoole@gmail.com
Northwestern University, 2240 Campus Drive, 60208 Evanston, USA
Full list of author information is available at the end of the article

Abstract

The ubiquity of digital music consumption has made it possible to extract information about modern music that allows us to perform large scale analysis of stylistic change over time. In order to uncover underlying patterns in cultural evolution, we examine the relationship between the established characteristics of different genres and styles, and the introduction of novel ideas that fuel this ongoing creative evolution. To understand how this dynamic plays out and shapes the cultural ecosystem, we compare musical artifacts to their contemporaries to identify novel artifacts, study the relationship between novelty and commercial success, and connect this to the changes in musical content that we can observe over time. Using Music Information Retrieval (MIR) data and lyrics from Billboard Hot 100 songs between 1974-2013, we calculate a novelty score for each song's aural attributes and lyrics. Comparing both scores to the popularity of the song following its release, we uncover key patterns in the relationship between novelty and audience reception. Additionally, we look at the link between novelty and the likelihood that a song was influential given where its MIR and lyrical features fit within the larger trends we observed.

Keywords: Cultural Novelty; Computational Methods; Quantitative Analysis; Computational Social Science

Introduction

When NWA dropped their hit song, 'Straight Outta Compton', it was one of the hottest new tracks of 1988, but in fact, a key component of the song hearkened all the way back to 1969. By incorporating a sample of the famous 'Amen Break' drum solo from The Winstons' song 'Amen Brother', the song is an example of the way that musical traits can persist even as they undergo change and reinvention. This juxtaposition highlights the paradox that make culture so fascinating; it provides us with a foundation of established aesthetics and practices to draw from, even as it continues to change and evolve. This dynamic balance between established norms, and the introduction of novelty provides a rich area of inquiry for cultural analysis that looks not only at the impact of novelty on patterns of commercial production and consumption, but also at how the introduction of novel creative artifacts drives cultural evolution [1, 2, 3, 4, 5, 6]. With the rise of digital media, information about the consumption and production of cultural artifacts is available to us at unprecedented scale. In addition, the digitization of artifacts allows us to apply computational analyses to better understand the often nebulous concepts of creativity and novelty, and unlock insights into their effects on cultural change. With music in particular, the availability of digital data, and advances in computational

methods of audio analysis have made it possible to investigate these questions at scale. The ubiquity of popular music also means that established markers of success, such as Billboard charts, are based on the opinions of a large population. Additionally, while music styles vary across different genres and cultures, music is a 'human universal' found in virtually all societies, and displays organizational properties that we can track over time in order to analyze patterns of cultural change [3, 1].

Currently it is possible to extract quantitative metrics from large music data sets using Music Information Retrieval (MIR) software. This data, referred to as audio features, or audio descriptors, is information that can be extracted from audio signals, and can be roughly classified as low-level and high-level features. Low-level features directly describe the audio signal data, for example spectral descriptors, while high-level features typically describe more holistic information about the song such as key, energy level, or danceability [7, 8]. Previous work has demonstrated that these high level MIR features provide accurate and robust data for modeling musical preference [7, 9], along with comparisons of content similarity that inform automatic genre classification [10, 11]. This has enabled researchers to contextualize songs within the larger musical ecosystem they exist in, and to identify long term trends in how genres and styles evolve over time [12, 1, 3, 2, 4, 5]. This data is also used by streaming services such as Spotify to develop music discovery tools and generate recommendations for users. In addition to MIR data, word and document embeddings of lyrics have also been shown to be a rich source of data for music content analysis, including genre and mood classification [13, 14, 15, 16, 17]. As with MIR features, these vector embeddings can be used to evaluate the similarity between lyrics of different songs [18, 19, 20].

In combination with data on commercial success and popularity, MIR and lyric data has enabled researchers to examine how the novelty of songs correlates with their success. For both MIR features and lyrics, the novelty of a song relative to its peers has been found to play a role in determining its cultural success, with the most popular songs demonstrating an optimal level of differentiation that allows them to stand out without being perceived as too dissimilar [21, 22, 23]. Identifying these patterns of optimal differentiation in consumer preferences is important for both the music industry at large, and for development of recommender systems [24]. However, previous work has still looked at MIR and lyrical novelty separately, and there is a lack of understanding as to how the relationship between these two dimensions of novelty might affect listeners perception of overall song novelty, and the success of the song. Work in genre and mood classification has shown MIR and lyric data to be complementary, with the inclusion of both sets of features having a positive impact on classification accuracy [25, 16, 13, 26]. Since both of these components contribute to the overall perception of the song's mood and genre, we propose to study whether there is also a relationship between a song's MIR novelty and its lyrical novelty, and if that relationship influences its performance on the Billboard Hot 100 chart. Additionally, we consider an alternative definition of success in terms of how likely it was that the song exerted some degree of stylistic influence on the cultural ecosystem. This has been done in previous studies with classical music by tracking the reappearance of specific motifs or harmonic patterns, however the granularity required in this type of analysis makes it difficult to scale

[4]. By using MIR and lyric features though, it is possible to perform this type of analysis at scale by using the similarity measures between a song and later releases to determine the likelihood that the song in question was influential [27]. In doing so, we can examine whether a song's novelty and initial success affect its likelihood of being influential in the long term, and in doing so, gain insight into how the introduction of novel attributes fuels ongoing creative evolution in modern popular music.

In this paper, using MIR and lyric feature data from Billboard Hot 100 songs between 1974-2013, we calculated novelty scores for each song relative to its genre and release year, and compared these to the total number of weeks the song spent on the Hot 100 chart. We found that the novelty scores at which optimal differentiation occurred were quite similar for both MIR and lyrics, and the most successful songs where those that were optimally differentiated for both. When looking at the probability of a song being influential, we also observed optimal differentiation occurring with respect to the novelty scores. Additionally, we found that there was no correlation between the time the song spent on the chart, and its probability of being influential. Rather, we found that for different novelty scores, the amount of time the song spent on the chart affected its likelihood of being influential. By utilizing computational data and methodology to extract high level patterns of change within the musical ecosystem, this research highlights the importance of considering alternate metrics for evaluating success when studying cultural artifacts by providing insight into how novelty affects both short and long term performance of cultural artifacts.

Data and Methodology

Our data comprises songs from the Billboard Hot 100 chart, which tracks the 100 most popular songs, based on Nielsen radio play scores, for each week [22]. This data set allowed us to limit our analysis to only popular music that was most representative of the prevailing cultural space at each point in time. The data set included song genre from Discogs.com, the position and total number of weeks each song spent on the chart, and high level MIR feature data from The Echo Nest which was derived by Music Information Retrieval (MIR). A discription of the MIR features can be found in Table 4 in the Appendix. The subset of the data used in our analysis consisted of 14,248 songs that were on the Billboard Hot 100 chart between 1974-2013, encompassing 3,973 unique artists and bands across 643 different record labels and 17 genres.

We acquired text of the lyrics for each song from a variety of online sources with our custom scraping tools. We cleaned, preprocessed, and tokenized lyrics using the Gensim simple preprocessing utility [28]. We then trained a Doc2Vec model which had a vector of 100 dimensions and iterated over the training corpus 40 times [19]. The minimum word count was set to 2 in order to discard words with a single occurrence. This model was then used to generate a feature vector for the lyrics of each individual song. Unlike the MIR features, the lyric features do not map to concrete concepts, however all together they define a vector space where we can compare how relatively similar the contents of two documents are by looking at how close their vectors are. This is an extension of word embedding vectors, where

we would see that the vectors for the words 'sad' and 'morose' would be closer to one another than the vectors for the words 'sad' and 'happy'.

Novelty Scores

As stated above, the lyric features represent dimensions that define a vector space in which we can compare the similarity of sets of lyrics based on the relative positions of their lyric feature vectors. Using the MIR features, we can do the same, and define a vector space that allows us to compare aural similarity between songs based on the relative positions of their MIR feature vectors. We can then group the songs by genre and release year and look at where their vectors are distributed within the lyric vector space, and within the MIR vector space. This allows us to assign an MIR novelty score and a lyric novelty score to each individual song, based on the average distance between it and the other songs in the subset within the MIR vector space, and within the lyric vector space. We choose to only generate within-genre novelty comparisons, as the stylistic variation between genres means that cross-genre comparisons would not give us a good measure of novelty. Additionally, this allows us to track how the genre as a whole is moving through each of these vector spaces over time, which represents the stylistic evolution the genre undergoes. As knowing the magnitude of movement through the space is important for understanding this, we opted to use a distance metric when quantifying the change in average genre positioning, and the individual song novelty scores, as opposed to cosine similarity which was used in previous studies.

For the lyric vector distances, we calculated the average feature vector for each genre-year group, and measured the Euclidean distance between this average and each song within the distribution. For the MIR vector distances, we followed the same approach, but with a slight variation. With MIR data, some feature values are used as input for determining the values of other features, for example when calculating the danceability of a song, tempo and valence values are included. As a result, we must account for co-variances among MIR features by taking the Mahalanobis distance between each song and the rest of its genre-year distribution [29]. The Mahalanobis distance is calculated using the covariance matrix of the data, so that it can account for the fact that the variance across the different principal components is not the same, due to the dependencies between matrix columns. For a given song described as its MIR feature vector $\vec{x} = (x_1, x_2, x_3, \ldots, x_N)^T$ and a set of songs with mean MIR features $\vec{\mu} = (\mu_1, \mu_2, \mu_3, \ldots, \mu_N)^T$ and a covariance matrix S, the Mahalanobis distance D_M of song x to the set can be calculated as follows:

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}.$$

In order to be consistent when calculating the within-genre distances for different years, a covariance matrix was generated for each genre using data from all years, and used in the distance calculations, rather than generating a covariance matrix for each genre-year subset. Because covariances between MIR features mainly vary across genres, but are fairly consistent within genre overtime, this allowed us to make sure that the normalization applied by the Mahalanobis distance calculation

was consistent across all the subsets of a genre, allowing us to compare different time periods.

This yields a distribution of MIR vector distances and a distribution of lyric vector distances for each of the genre-year subsets. How novel a song's MIR features and lyric features are is dependent on how much higher or lower the vectors' distances are when compared to the mean distance for their respective distributions. For example, two songs might have the same distance between their MIR vectors and the average vectors for their subset, but relative to the overall distance distributions for each subset, one song might actually be much more novel than the other. Because each of these distributions will have variations in the mean distance and total range of distances, if we want to compare song novelty between different distributions, we need to use a novelty score which takes the MIR vector distance and the lyric vectors distance of each song, and normalizes it relative to the distribution it is in. We do this by calculating the z-score for each individual vector distance. The z-score tells us the relative positioning of a vector distance within a distribution by subtracting the mean distance of the distribution, and then dividing the difference by the standard deviation of the distribution. In doing so, the z-score tells us how many standard deviations from the mean that particular value is, which indicates how novel the vector distance is, and allows us to compare it to novelty scores drawn from different distributions.

Influence Probability

In order to evaluate the likelihood that a song was influential, we have to compare it to the cultural ecosystem at later points in time. Although it is not possible to prove a causal relationship, measuring the degree of similarity between a cultural artifact and other artifacts produced at a later time is a standard approach for inferring potential influence [27]. To compare a song's similarity to songs released in later years, we can use the same approach that we took for calculating the novelty scores. However, instead of comparing the song to songs in the same genre-year subset, we compare it to songs in the same genre, but in a later year. For example, if a Rock song was released in 1982, we would compare its feature vectors to those of Rock songs that were released in 1984, 1985, 1986 and so on. This gives us the relative novelty score for that song, because it is relative to the particular year we are comparing it to. We can then determine whether or not it was probable that a song was influential or not based on whether its relative novelty score is higher or lower than its novelty score for the year it was released. If it is lower, that means that in later years, the average features of songs in that genre became more similar to those of the song we are considering, indicating that the song was potentially influential. The greater the decrease in relative novelty, the more likely this is. Therefore, we can assess the relative likelihood that a song was influential by calculating its change in relative MIR novelty, and its change in relative lyric novelty.

Not all of the genres in our dataset had songs included on the Hot100 chart for every year within the time period we looked at, meaning there were a large number of relative novelty scores that could not be calculated. Because of this, we limited our analysis to the 5 genres with the greatest number of songs; Rock, R&B, Rap, Country, and Pop. For each song within these genres, we compared it to the genre-year subsets of the subsequent ten years following the song's release. The comparison

song was included in the subset, and we calculated the Euclidean distance for the lyric vector distances, and the Mahalanobis distance for the MIR vectors distances. Again, in order to account for variations in the mean distance and total range of distances within different genre-year distributions, we calculated the z-score for the comparison song's MIR distance and lyric distance. This yielded a relative novelty score that we could then compare to the relative novelty scores for other years, and the initial novelty score.

Results

In comparing the MIR and lyric novelty score distributions across all years and genres, we found that the MIR novelty distribution had a greater positive kurtosis and a greater positive skew than the lyric novelty distribution (see Figure 1 top plot and Table 1). This tells us that there is a greater range in the above average novelty scores occurring within the MIR distribution. Although we can observe that the median value for the MIR novelty distribution, -0.21, is slightly lower than the median value for the lyric novelty distribution, -0.09, a one-way ANOVA test confirms no significant difference between the MIR novelty distribution and the lyrics novelty distribution (F=6.11e-30, p=1.0). These trends held true when analyzing the novelty distributions within individual genres (Figure 1 bottom plots and Table 1).

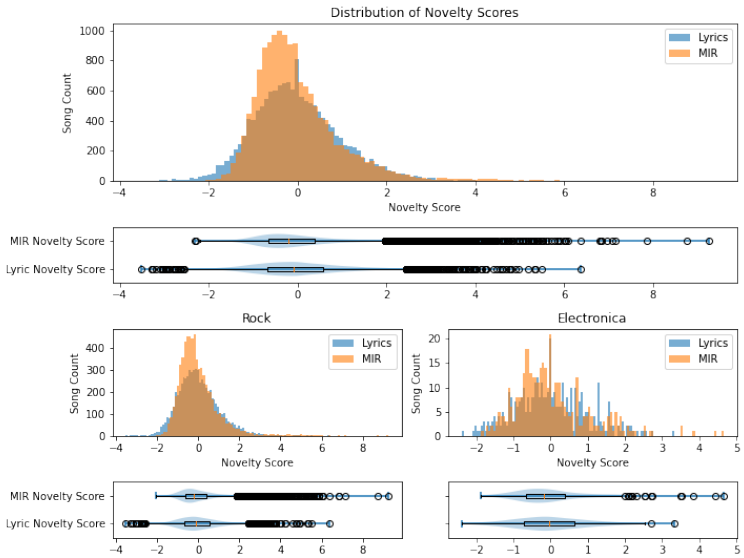

Figure 1 Distribution of MIR novelty scores and lyric novelty scores. Top plot displays the distributions for all songs within the dataset. The bottom two plots display the distributions for all songs within the Rock genre and Electronica genre, respectively.

Table 1 Novelty Score Distributions

Modality	Skew	Kurtosis
MIR Novelty - All Genres	1.88	6.53
Lyric Novelty - All Genres	0.60	1.13
MIR Novelty - Rock	2.21	8.88
Lyric Novelty - Rock	0.69	1.48
MIR Novelty - Electronica	1.32	2.82
Lyric Novelty - Electronica	0.33	-0.05

Initial success

We incorporated song commercial success data to determine whether these variations in the novelty distributions of the two modalities were indicative of differences in how they impacted the likelihood of a song becoming popular. Using total number of weeks on chart as the metric for song success, we found that similar to previous findings, the most popular songs had a degree of optimal differentiation both for MIR novelty, and for lyric novelty [22, 23] (see Figure 2 top row). We looked at the relationship between novelty and success for each modality separately, and using the Hotelling T2 test, found no statistically significant difference between the joint distribution of total weeks on chart with respect to MIR novelty, and the joint distribution of total weeks on chart with respect to lyrics novelty (F=3.07e-30, p=1.0). This was also found to be the case at the genre level as well (Figure 2 bottom row). Specifically, Hotelling T2 test found no significant difference between the MIR joint distribution and the lyrics joint distribution for either Rock (F=7.03e-30, p=1.0) or Electronica (F=5.44e-31, p=1.0).

Given that the songs with the most success fell into a rather narrow range of novelty values, we performed a Kernel Density Estimation analysis to estimate both the MIR novelty score and lyric novelty score which had the highest probability of being in the top 85th, 90th, and 95th percentile of total weeks on chart.

For both the MIR-total weeks joint distribution and lyrics-total weeks joint distribution, the Python library scikit-learn was used to generate a Kernel Density Estimation using a Gaussian mixture model and a bandwidth of 0.3 [30]. The KDE was used to generate probability scores for hypothetical pairings of novelty scores and total weeks on chart, which indicated the likelihood that a song with the given novelty score would be on the chart for the given number of weeks. This was done for 250,000 individual generated data points that were equally distributed across 500 unique values in the range of -1 to 1, which represented novelty values, and across 500 unique values in the range of 20 to 76, which represented the top 85th percentile of total weeks on chart. For each novelty value, we took the summation of the generated probability scores to calculate the relative probability that a song with that amount of novelty would reach anywhere within the top 85th percentile of total weeks on chart. This process was then repeated for the top 90th percentile of total weeks on chart, and the top 95th percentile of total weeks on chart.

For each of these, we can see in Figure 3 that for both MIR novelty and lyric novelty, the probability of success increases as the novelty score increases, until a certain point at which it peaks and then because to decrease again. The novelty score for this peak value that we have estimated in our analysis indicates the degree of optimal differentiation that is most likely to help them succeed. Below this, the song is likely to be too similar to stand out from other songs, while above this,

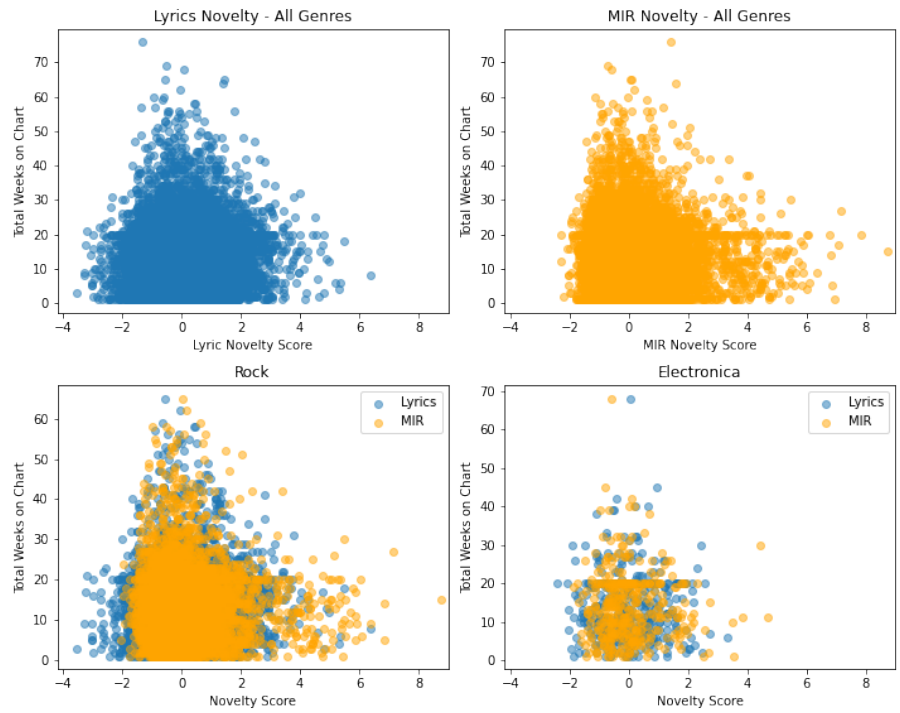

Figure 2 Joint distributions of song novelty scores and total number of weeks the song spent on the chart. Top row shows the joint distributions for all genres. Bottom row shows the overlay of MIR-total weeks joint distribution and lyrics-total weeks joint distribution for Rock, and Electronica. In each distribution we can see that the highest number of total weeks on chart occur within a certain range of novelty scores.

it starts to diverge too much from what the audience expects. While the novelty scores of a song cannot be used to predict exactly how successful it will be, songs with novelty scores close to our estimates will have a greater chance of achieving high levels of success than songs with novelty scores that are higher or lower.

We found that for both lyrics and MIR, the novelty scores which had the highest probabilities of success for each of these performance tiers was just slightly lower than the mean novelty scores of the population. We also found that across the three total week ranges we tested, the MIR novelty score was consistently slightly lower than the lyric novelty score (see Figure 3 and Table 2). Additionally, as the analysis narrows from the top 85th percentile to only the top 95th percentile, we also see that that the MIR novelty score increases, while the lyric novelty score decreases, causing the difference between them to grow smaller.

Table 2 Novelty Score with Highest Success Probability

Total Weeks Range	85th Percentile	90th Percentile	95th Percentile
MIR Novelty Score	-0.35	-0.33	-0.27
Lyric Novelty Score	-0.14	-0.17	-0.18

When examining the relationship between MIR novelty and lyric novelty of individual songs, we did not find a significant correlation between the two (Pearson correlation test r=-0.01, p=0.10). They appear to be independent of one another, with no consistent patterns found in the relationship between the MIR novelty

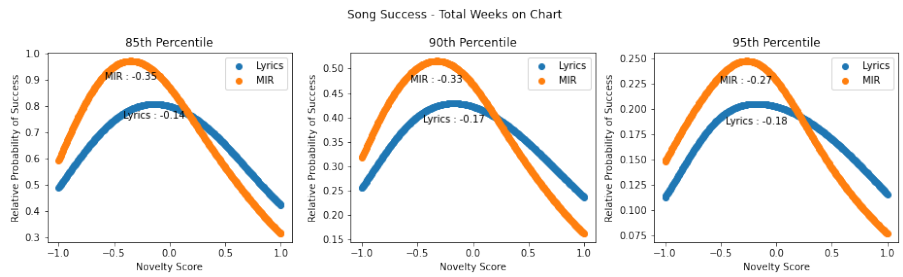

Figure 3 Relative probability of success in reaching the 85th, 90th, and 95th percentile of total weeks on chart across a range of MIR novelty scores, and across a range of lyric novelty scores. Since the relative probability values are a summation of probability estimates of a generated data set, they are dependent on the total number of data points generated for our sample, and should not be treated as an absolute probability value.

score and the lyric novelty score of a given song. As a result, a set of songs with the same MIR novelty score might have a wide variation in their lyric novelty scores, and vice versa. Given that this is the case, we wanted to explore whether different combinations of MIR novelty scores and lyric novelty scores would impact a song's probability of success.

For this, we generated a Kernel Density Estimation for the joint distribution which included both lyric novelty and MIR novelty, along with total weeks on chart. This was then used to calculate the probability scores for 1,000,000 equally distributed generated data points having lyric novelty scores between -1 to 1, MIR novelty scores between -1 to 1, total weeks on chart in the 90th percentile, between 22 to 76. For each unique pair of MIR and lyric novelty scores, we took the summation of the generated probability scores to calculate the relative probability of a song with those scores being in the top 90th percentile of total weeks on chart.

We see in Figure 4 that the highest relative probability of success occurs when a song is close to the optimal novelty values for both lyrics and MIR. As we move outward from this area where both novelty scores are close to optimal, the radial pattern of the gradient indicates that variance in the probability of success is equally affected by both variance in the MIR novelty score, and variance in the lyric novelty score. Additionally, given that the gradient is roughly equal for points that have the same distance from this optimal center point, this tells us that the proportional relationship of MIR novelty to lyric novelty is not an explanatory variable, but rather it is the combined total distance from the optimal center that impacts a songs probability of success.

Whether the point lies above or below the optimal value for either novelty score does not change the relationship, which tells us that having a higher than optimal novelty score for one modality can't be 'balanced out' by a lower than optimal novelty score for the other modality. If that were the case, and it were only about reaching an optimal value for the average of the the two novelty scores, then we would expect to see, for instance, a song with an MIR novelty score of -0.22, 0.1 higher than optimal, and a lyric novelty score of -0.27, 0.1 less than optimal, to have an equal probability of success as a song with the optimal values of an MIR novelty score of -0.32 and a lyric novelty score of -0.17. In the contour map we see that the

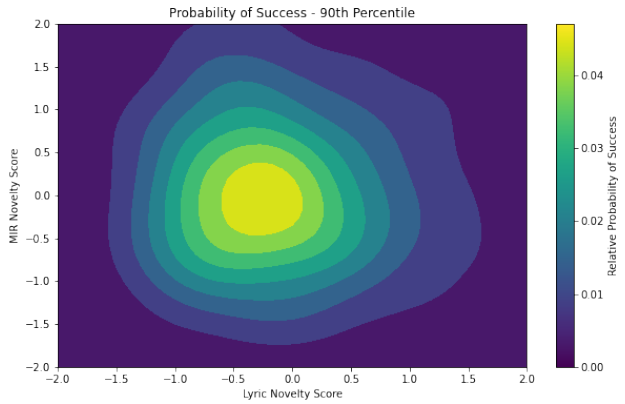

Figure 4 Relative probability of success in reaching the 90th percentile of total weeks on chart for different combinations of MIR and lyric novelty scores. We see that the probability decreases as either score moves away from the value that provides the optimal degree of differentiation within its modality. Since the relative probability values are a summation of probability estimates of a generated data set, they are dependent on the total number of data points generated for our sample, and should not be treated as an absolute probability value.

MRI novelty score that optimizes success probability does not change for different values of lyric novelty, and vice versa, the lyric novelty score that optimizes success does not change for different values of MIR novelty. However, we also observe that the combined distance of both novelty scores from their respective optimal values will impact a song's probability of success, indicating that even though a song's lyric novelty and MIR novelty are independent of one another, their deviations from the optimal values will have an additive effect on the overall perception of song novelty as it relates to optimal differentiation.

Influence Probability

Since we evaluated the relative likelihood that a song was influential by calculating the change in both its MIR and lyric relative novelty, our first step was to determine whether the rate at which relative novelty changed was consistent over time. Taking the average change in relative song novelty in the years following its initial release, we found that the rate of change for MIR relative novelty plateaued after 2 years, and the rate of change for lyric relative novelty plateaued after 3 years (see Figure 5 left plots). Because of this, we decided to only consider the relative novelty change that occurred in the 3 years following release, and used the average of those three values.

It is worth noting that after reaching the inflection point, the average year over year change in relative novelty for MIR stays positive, meaning that on average, the MIR features of a genre will tend to become less similar to those in previous years as the gap in time increases. The average year over year change in relative novelty for lyrics, however, is for the most part negative, suggesting that song lyrics within a genre tend to be more similar to those of songs released in previous years. When comparing the distribution relative novelty change for both modalities, we see that both follow a normal distribution, with a one-way ANOVA test showing

no statistically significant difference between them (F=0.05, p=0.83). However, the distribution of lyric relative novelty change shows a larger positive kurtosis and skew than that of the MIR relative novelty change distribution, which is the opposite of the trend we observed between MIR and lyrics in the initial novelty score distributions (see Figure 5 right plot and Table 3). Additionally, while the difference is not statistically significant, we observe greater variance between the two relative novelty change distributions than we do between the initial MIR and lyric novelty score distributions.

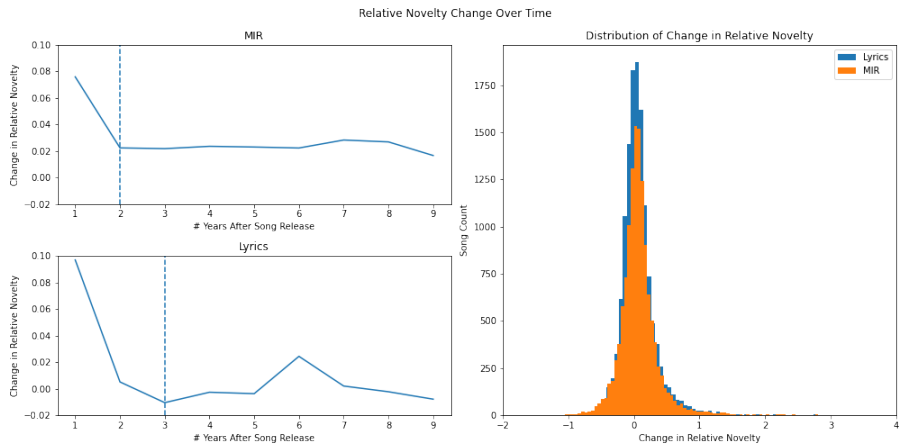

Figure 5 The two plots in the left column show the magnitude of year over year change in the average relative novelty score as songs are compared to later release years. For both MIR relative novelty and lyric relative novelty, we see that the rate of change is steep for the first two years after release, then stabilizes. The vertical lines indicate the inflection points where this occurs, which for MIR relative novelty is after two years, and for lyric relative novelty is after 3 years. The plot in the right column shows the distributions of relative novelty change for MIR novelty, and lyric novelty, which are calculated by taking the average change in relative novelty which occurred in the first three years after the song's release.

Table 3 Relative Novelty Change Distributions

Modality	Skew	Kurtosis
MIR Relative Novelty	6.77	123.95
Lyric Relative Novelty	15.18	450.16

Since a greater decrease in a song's relative novelty indicates a higher likelihood that the song was influential, scores in the bottom 10th percentile represent high performers for this metric. To determine whether a song's initial novelty scores had any correlation with how its relative novelty changed over time, we generated a Kernel Density Estimation for the joint distribution between MIR novelty scores and MIR relative novelty change, as well as for the joint distribution between lyric novelty scores and lyric relative novelty change. Using the same procedure as for the initial success KDE, we found that for both lyrics and MIR, the novelty scores that correlated with the highest probabilities of seeing a large decrease in relative novelty were below the average novelty score of the population, and slightly lower than than the optimal differentiation novelty values we estimated for initial success (see Figure 6). Again, we see that the optimal MIR novelty score is slightly lower than the optimal lyric novelty score.

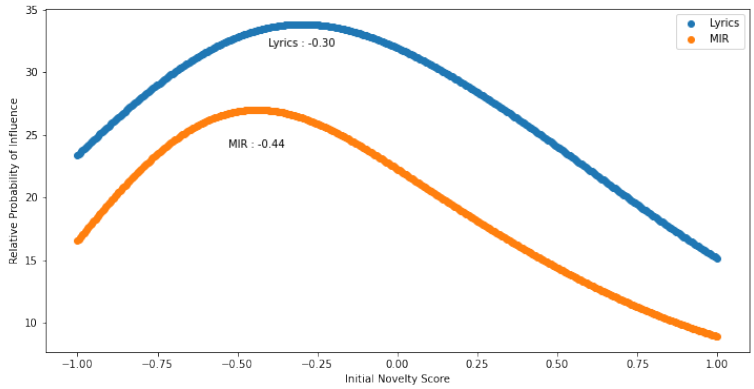

Figure 6 Relative probability of a song being in the bottom 10th percentile for relative novelty change based on initial novelty scores. We see that the probability decreases as either score moves away from the value that provides the optimal degree of differentiation within its modality. Since the relative probability values are a summation of probability estimates of a generated data set, they are dependent on the total number of data points generated for our sample, and should not be treated as an absolute probability value.

Because optimal differentiation is about the relationship between cultural artifacts and their contemporaries, the fact that we see a degree of optimal differentiation in the relationship between artifacts released at different points in time is a new finding. While one possible explanation is that the initial popularity of songs that are optimally differentiated makes them more likely to be influential, we did not find any correlation between total number of weeks on chart, and either lyric or MIR relative novelty change. To investigate this, we ran an analysis to examine the relationship between total weeks on chart and influence probability when controlling for limited ranges of novelty scores. Running two analyses, one for MIR novelty and MIR relative novelty change, and one for lyric novelty and lyric relative novelty change, we considered songs that fell within three different ranges of novelty scores; the bottom 10th percentile, the top 10th percentile, and the 10th percentile centered around the novelty score with the highest probability of maximizing total weeks on chart. Within each range, the songs were then grouped by the number of weeks they had spent on the chart. An aggregated influence probability was calculated for each week in the range of 0 to 35 by calculating the percentage of songs in that grouping whose relative novelty had decreased. Additional details for this process can be found in the Appendix.

We found that, relative to the song's novelty score, the influence probability varied with increases in the amount of time the song spent on the chart. In Figure 7 we see that regardless of the modality or the novelty grouping, influence probability initially increases with more time spent on chart, then at a certain point, peaks and starts to decrease. For lyric novelty, we see a peak occurs at the same time for all three novelty bins, at roughly 10-15 weeks. Beyond that, both the bottom 10th percentile and the top 10th percentile see a decrease, although the rate of decrease for influence probability appears to be more pronounced for the songs in the top 90th novelty percentile. For lyric novelty in the optimal range, we do see a second peak

around 25-30 weeks, however beyond that the influence probability again drops. We see that regardless of the total amount of time on chart, higher lyric novelty scores are correlated with lower influence probability. For MIR novelty, we see in Figure 7 that the impact of additional weeks on the chart is different depending on the song's novelty score. Songs in the bottom 10th percentile see influence probability peak at 10 weeks, and then consistently decrease with additional time on the chart. For songs in the optimal zone, however, there is a consistent increase in influence probability until roughly 25 weeks, at which point the influence probability quickly decreases. For songs in the top 90th percentile, we see a steeper increase in influence probability which lasts until roughly 30 weeks before hitting the peak and then decreasing. Here we also see that in contrast to the pattern observed for lyric novelty scores, higher MIR novelty scores are correlated with higher influence probability regardless of the number of weeks spent on the chart. We typically associate increased exposure with greater success, for both short term commercial success, and also for being influential within the creative space. It would seem intuitive that greater exposure would lead to a greater probability of exerting influence, as more people hear and become familiar with the cultural artifact in question. However these results suggest that while that is sometimes the case, there is also an optimal amount of exposure, which can vary depending on the novelty and modality of the attributes being considered.

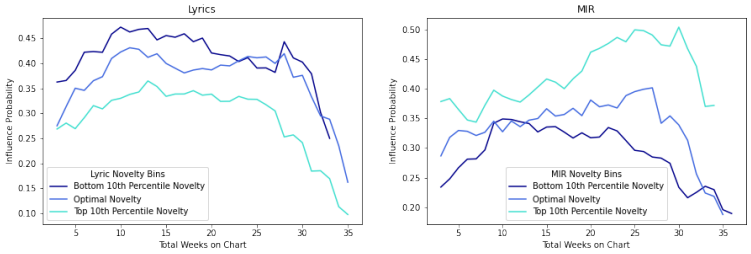

Figure 7 Controlling for initial novelty, the influence probability increases at first with additional weeks on chart, then after a certain amount of time it peaks and begins decreasing. For lyric novelty, the peak occurs at roughly 10 weeks. For MIR novelty, the effect of increased time on chart varies depending on the initial novelty score. For higher initial novelty scores, we see the peak influence probability occur after a longer period of time on the chart.

Although we previously found no correlation between individual songs' initial lyric novelty and initial MIR novelty, we did find that there was a small but significant correlation ($r=0.22$, $p< 0.001$) between the average change in lyric relative novelty and the average change in MIR relative novelty (see Figure 8). Additionally, we observed that for songs with low lyric novelty scores, higher MIR novelty scores had a slight positive effect on influence probability, while for songs with high MIR novelty scores, higher lyric novelty scores also had a slight positive impact on influence probability. A possible explanation is that these combinations of high and low novelty scores impact how memorable a song is, even if they are more likely to hurt the songs initial success probability. However, it is not clear why we observe these interaction patterns only for songs at the extreme ends of the novelty score ranges, and these findings highlight an avenue for further research into how

the attributes of cultural artifacts impact their likelihood of influencing the larger cultural ecosystem.

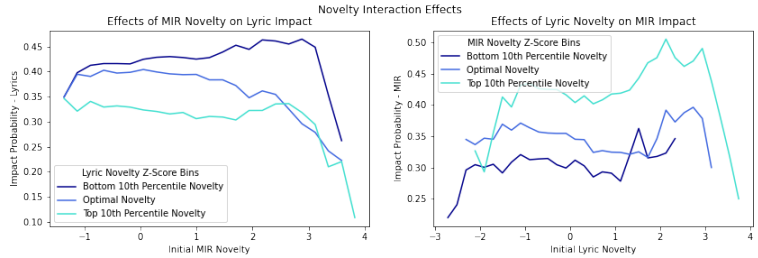

Figure 8 Controlling for initial novelty, we see that different combinations of high and low MIR and lyric novelty scores have different impacts on the influence probability.

Long Term Feature Vector Trends

In order to delve into why the effect of exposure was so different for lyric novelty than for MIR novelty, we examined whether there were any long term trends in how the average feature vectors within each genre were changing over time. Since our novelty score only compares an individual song to songs released in the same year, this allows us to consider the song's novelty within a broader context by evaluating the relationship between individual song novelty, and what is happening within a genre over a longer time period. In order to contextualize the relative novelty change of individual songs with respect to the evolution of their genre over time, we considered the position of each genre within the MIR and lyric vector spaces, and whether and how it shifted over time. For each genre, we calculated both the average MIR feature vector and the average lyric feature vector for each year, and performed a 2 component Principal Component Analysis (PCA) to compress the dimensionality of the vectors. This allowed us to create a 2-dimensional representation of how the genre was moving through each of the vector spaces over time.

We observed a stark difference between the movement of genres through the MIR vector space over time, and their movement through the lyric vector. In Figure 9, the average feature vectors are represented with a small point, while the shaded circle around them demonstrates how diffuse the range of individual song feature vectors was in the given year. Across all genres, we see that there is no significant movement of song distributions within lyric feature space. Although we observe slight shifts in positioning year over year, there is no long term directionality to this movement, and the distance that each distribution shifted from the previous year was typically less than 10 percent of the diameter of the previous year's distribution.

In contrast, within MIR feature space we see clear directionality in the movement of each genre over time. Here we observed that the magnitude of the shift in average positioning from one year to the next was frequently between 50-110 percent of the diameter of the previous years distribution. We can see examples in Figure 9 where some of the distributions for consecutive years do not have any overlap.

This explains why the average year over year change in relative novelty for MIR stays positive, while the average year over year change in relative novelty for lyrics

is for the most part negative. While the average distance between old and new songs in MIR feature space is always increasing due to directionality of genre evolution over time, old and new songs inhabit roughly the same area of lyric feature space, so when more songs are added to the same region over time, the overall density increases and the average distance between individual songs becomes smaller.

An interesting additional observation is that we see a similar contemporaneous pattern of movement occurring across each of the different genres, with the average vector position within the plots beginning on the far left, moving over to the far right, then moving back towards the center. Dimensionality reduction with PCA is done by identifying the principal components of a data set, which are orthogonal axes within the vector space represented by unit vectors. For a two dimensional PCA, the first vector captures the axis along which we see the greatest variance in our data, and the second vector defines the axis which is orthogonal to the first while maximizing the amount of variance captured, therefore it is expected that we would see more variation along one axis. However, although the magnitude of this movement varies between the genres, it suggests that there are long term trends in popular music that are not specific to the individual genres, but occurring globally within the modern popular music ecosystem.

Discussion

By utilizing computational methods to analyze cultural data at scale, this research contributes to our understanding of how novelty impacts the dynamics of cultural change over time. Our results highlight the ways in which stylistic change is conditioned on the previous state of the cultural ecosystem, and how this is impacted by the ways in which we perceive and evaluate different degrees of novelty and differentiation. Although the high-level and aggregate nature of our data does not enable us to create a prediction model for identifying hit songs in their unique context, our results contribute to understanding overarching patterns in how perceptions of novelty affect cultural evolution.

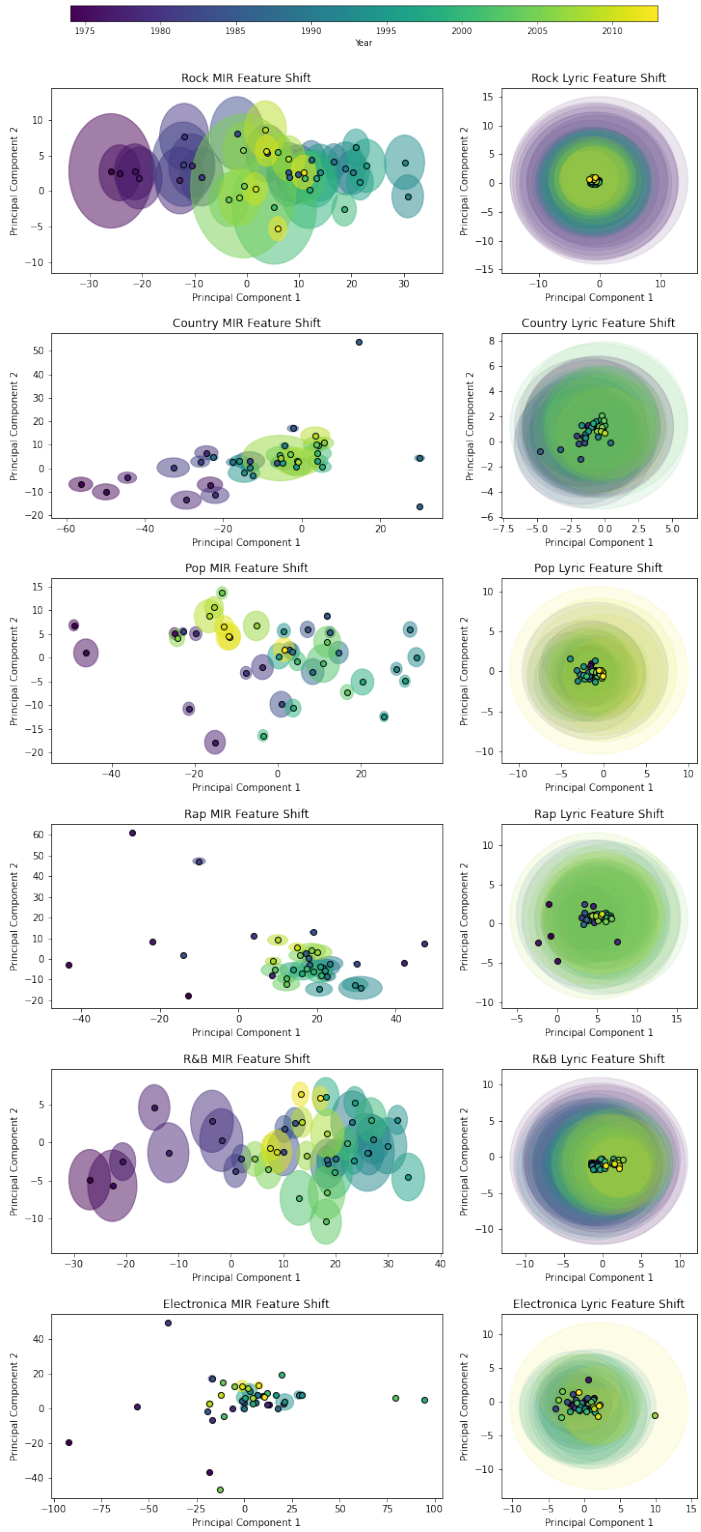

Figure 9 Average feature vectors for each year are represented with a small point. The size of the shaded circle in the corresponding color indicates the range of dispersion of the song feature vectors for that given year. Within the MIR vector space, we can observe directionality in the individual genre positioning as it moves over time. Within the lyric vector space, we see instead that the genre positioning stays fairly consistent throughout.

Novelty and Music Cognition

Our finding that there is no relationship between the lyric and MIR novelty scores of individual songs, and that the optimal novelty scores for each modality are also independent of one another is supported by previous work in music cognition, which finds evidence that music and lyrics are processed independently [31, 32]. This explains why the negative impact on success probability of an above optimal novelty score for one modality cannot be mitigated by the song having a below optimal score in the other modality. Our results do not provide any information which would allow us to evaluate the possibility of a causal relationship between the aforementioned music cognition data and the trends we have observed, however the observation of these connections between large scale phenomenon and cognitive processes that occur at the individual level suggest that this could be a productive area of interdisciplinary study. It is possible that research in the field of cognitive science could provide insights into cognitive perceptual processing, that could inform potential avenues of inquiry when investigating cultural trends and evolution.

Novelty and Exposure Effects

In our results, we observed that the relationship between a song's initial novelty score and its influence probability is affected by how many weeks it has spent on the chart. For songs with higher MIR novelty scores, we see that an increase in time on chart has a positive impact on influence probability within the time range we analyzed. This potentially explains why the distribution of MIR novelty scores is more heavily right-tailed than the distribution of lyric novelty scores, as the positive impact of increased exposure may cause more variation and range in the MIR novelty scores that end up being successful enough to reach the Billboard Hot100 chart.

Previous work exploring the impact of repeated exposure to unfamiliar music and subsequent music preferences have found that this additional exposure increases the likelihood that the listeners will enjoy the music when they hear it again [33]. Additional research has also found that repeated exposure in the context of collective attention to news stories shared online, leads to novelty decay over time [34]. It is possible, then, that this decrease in perceived novelty may benefit artifacts that have above average novelty. In the case of songs with MIR novelty scores above the optimal value, a possible explanation for our observations may be that the increased exposure provided by more time spent on the chart causes the perception of its novelty to decrease, making it seem closer to the optimal values needed to increase influence probability. For artifacts already at or below optimal levels of differentiation, this decrease in perceived novelty would make them seem too familiar. This explains why the initial novelty of songs determined how much exposure was beneficial.

The question of why increased exposure was detrimental for lyric novelty even for songs in the top 90th percentile can then be explained by the long term genre trends we observed. Given that there were no significant changes in the relative positioning of each genre's average lyric features over time, it is possible that no matter how high a song's lyric novelty score is, its likely still within the same feature range of previous years' songs, and therefore still not novel enough to benefit from higher

levels of exposure. This suggests that when analyzing differences between cultural artifacts and their relationship to various metrics of success, it is important to draw a distinction between differentiation, which measures the amount of variation between the artifact and the other artifacts it is being compared to, and "true" novelty, which would consider the degree to which the artifact is introducing new material into the canon of its domain.

Our findings suggest that it is possible for an artifact to be 'overexposed', at which point in time the perception of novelty drops below an optimal level. Our results indicate that the amount of exposure it takes for this to occur is going to vary depending on the initial novelty of the artifact. This is an important consideration when modeling and predicting the dissemination of creative ideas and products, both in theoretical research, and in the development of practical applications, such as recommender systems. Additionally, this potential relationship between exposure and novelty perception suggests a possible mechanism by which the impact of social influence [35] affects an artifacts cultural success. In additional to providing social proof, song popularity may also impact perceived novelty by acting as a driver for greater exposure, thereby affecting the songs likelihood of being influential.

Implications for Recommender Systems

Understanding the degree to which the defining characteristics of musical genres change over time has applications for music recommendation software. There are limitations to traditional approaches of using historical data in predictions [36], and in order to avoid static behavior it is important for us to be able to identify what are the indicators that can help predict a preference shift [37]. If we can determine the degree to which more novel outliers indicate the evolution of a subset of music, we can take that into account when tracking an individual's music preferences, and better predict what they might like as their taste evolves. This is especially relevant for increasing the efficiency of recommender systems incorporating exploratory algorithms, as it can inform more directed exploration, as well as the ideal degree of novelty to incorporate with each round of exploration [38].

Limitations

As the Hot 100 data contains only a partial view of popular music, and its selection criteria has changed over time, future work could involve gathering a larger data set that would encompass a broader representation of modern music. For the purposes of this analysis, the Hot 100 songs served as a sample of the prevailing mainstream cultural trends in music. However, it is still a small sample of all the potential songs that could be included in the umbrella of modern popular music. Additionally, when grouping music by genre, we must acknowledge that genre classifications are inherently subjective. Genre labeling for this data set came from Discogs.com, which provides crowd-sourced data for songs, so the groupings provided do not necessarily represent an objective ground truth [39].

Conclusion

Utilizing MIR data to perform analysis at scale, we compared musical artifacts' relative novelty over time to identify consistent patterns in the dynamics of cultural

change. Our results showed evidence for both optimal differentiation in successful songs, and the conditioning effect of prior artifacts on stylistic change. By bringing in findings from sociology, cognitive science, and musicology to provide further insight into the impact of novelty on modern music evolution, our research provides quantitative methods that will enable media systems to track this organic evolution in a more informed manner.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
KO designed the study, ran all data analysis, and drafted the manuscript. EÁH provided feedback on study design and revised and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements
The authors would like to thank Miriam Alex for her assistance with data collection and cleaning, and would like to thank Michael Mauskapf and Noah Askin for sharing their unique data set.

Abbreviations
MIR, Music Information Retrieval; PCA, Principal Component Analysis

Availability of data and materials
The processed data that supports the findings of this study are available at:
`https://github.com/LINK-NU/Billboard-Hot-100-Music-Novelty.`

References
1. Mauch, M., MacCallum, R.M., Levy, M., Leroi, A.M.: The evolution of popular music: USA 1960–2010. Royal Society Open Science **2**(5), 150081 (2015). doi:10.1098/rsos.150081. Publisher: Royal Society. Accessed 2020-10-06
2. Weiß, C., Mauch, M., Dixon, S., Müller, M.: Investigating style evolution of Western classical music: A computational approach. Musicae Scientiae **23**(4), 486–507 (2019). doi:10.1177/1029864918757595. Accessed 2020-10-13
3. Serrà, J., Corral, A., Boguñá, M., Haro, M., Arcos, J.L.: Measuring the Evolution of Contemporary Western Popular Music. Scientific Reports **2**(1), 521 (2012). doi:10.1038/srep00521. Number: 1 Publisher: Nature Publishing Group. Accessed 2020-10-06
4. Bomin, S.L., Lecointre, G., Heyer, E.: The Evolution of Musical Diversity: The Key Role of Vertical Transmission. PLOS ONE **11**(3), 0151570 (2016). doi:10.1371/journal.pone.0151570. Publisher: Public Library of Science. Accessed 2020-10-06
5. Prockup, M., Ehmann, A.F., Gouyon, F., Schmidt, E.M., Celma, O., Kim, Y.E.: Modeling Genre with the Music Genome Project: Comparing Human-Labeled Attributes and Audio Features., Malaga, Spain, p. 7 (2015)
6. Klimek, P., Kreuzbauer, R., Thurner, S.: Fashion and art cycles are driven by counter-dominance signals of elite competition: quantitative evidence from music styles. Journal of The Royal Society Interface **16**(151), 20180731 (2019). doi:10.1098/rsif.2018.0731. Accessed 2021-03-19
7. Magron, P., Févotte, C.: Leveraging the structure of musical preference in content-aware music recommendation. CoRR **abs/2010.10276** (2020). arXiv: 2010.10276
8. Moffat, D., Ronan, D., Reiss, J.D.: An Evaluation of Audio Feature Extraction Toolboxes. In: Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, p. 7 (2015)
9. Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., Elowsson, A.: Using perceptually defined music features in music information retrieval. arXiv:1403.7923 [cs] (2014). Accessed 2020-10-06
10. Bertin-Mahieux, T.: Large-Scale Pattern Discovery in Music. PhD thesis, Columbia University (2013). doi:10.7916/D8NC67CT. `https://doi.org/10.7916/D8NC67CT` Accessed 2020-10-06
11. Lippens, S., Martens, J.P., De Mulder, T.: A comparison of human and automatic musical genre classification. In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, p. (2004). doi:10.1109/ICASSP.2004.1326806. ISSN: 1520-6149
12. Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., Komarova, N.L.: Musical trends and predictability of success in contemporary songs in and out of the top charts. Royal Society Open Science **5**(5), 171274 (2018). doi:10.1098/rsos.171274. Publisher: Royal Society. Accessed 2020-10-22
13. Mayerl, M., Votter, M., Zangerle, M.M.E.: Comparing Lyrics Features for Genre Recognition. In: Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA), pp. 73–77. Association for Computational Linguistics, Online (2020). https://www.aclweb.org/anthology/2020.nlp4musa-1.15.pdf
14. Mayer, R., Neumayer, R., Rauber, A.: Combination of audio and lyrics features for genre classification in digital audio collections. In: Proceedings of the 16th ACM International Conference on Multimedia. MM '08, pp. 159–168. Association for Computing Machinery, New York, NY, USA (2008). doi:10.1145/1459359.1459382. https://doi.org/10.1145/1459359.1459382 Accessed 2022-05-05
15. Hu, X., Downie, J.S., Ehmann, A.F.: Lyric Text Mining in Music Mood Classification. In: Proceedings of the 10th International Society for Music Information Retrieval Conference, Kobe International Conference Center, Kobe, Japan, p. 6 (2009)

16. Hu, X., Downie, J.S.: Improving mood classification in music digital libraries by combining lyrics and audio. In: Proceedings of the 10th Annual Joint Conference on Digital Libraries. JCDL '10, pp. 159–168. Association for Computing Machinery, New York, NY, USA (2010). doi:10.1145/1816123.1816146. https://doi.org/10.1145/1816123.1816146 Accessed 2022-05-05
17. McVicar, M., Giorgi, B.D., Dundar, B., Mauch, M.: Lyric document embeddings for music tagging, 10 (2021)
18. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs] (2013). arXiv: 1301.3781. Accessed 2020-12-08
19. Le, Q.V., Mikolov, T.: Distributed Representations of Sentences and Documents. arXiv:1405.4053 [cs] (2014). arXiv: 1405.4053. Accessed 2020-12-08
20. Whalen, R., Lungeanu, A., DeChurch, L., Contractor, N.: Patent Similarity Data and Innovation Metrics. Journal of Empirical Legal Studies **17**(3), 615–639 (2020). doi:10.1111/jels.12261. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jels.12261. Accessed 2020-12-04
21. Askin, N., Mauskapf, M.: What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. American Sociological Review **82**(5), 910–944 (2017). doi:10.1177/0003122417728662. Accessed 2020-10-13
22. Askin, N., Mauskapf, M.: Cultural Attributes and their Influence on Consumption Patterns in Popular Music. In: Aiello, L.M., McFarland, D. (eds.) (eds) Social Informatics. SocInfo 2014. Lecture Notes in Computer Science, vol. 8851, pp. 508–530. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13734-6_36
23. Berger, J., Packard, G.: Are Atypical Things More Popular? Psychological Science **29**(7), 1178–1184 (2018). doi:10.1177/0956797618759465. Publisher: SAGE Publications Inc. Accessed 2021-05-13
24. Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., Lalmas, M.: Algorithmic Effects on the Diversity of Consumption on Spotify. In: Proceedings of The Web Conference 2020, pp. 2155–2165. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3366423.3380281 Accessed 2022-05-06
25. Laurier, C., Grivolla, J., Herrera, P.: Multimodal Music Mood Classification Using Audio and Lyrics. In: 2008 Seventh International Conference on Machine Learning and Applications, pp. 688–693 (2008). doi:10.1109/ICMLA.2008.96
26. Neumayer, R., Rauber, A.: Integration of Text and Audio Features for Genre Classification in Music Information Retrieval. In: Amati, G., Carpineto, C., Romano, G. (eds.) Advances in Information Retrieval. Lecture Notes in Computer Science, vol. 4425, pp. 724–727. Springer, Berlin, Heidelberg (2007). http://link.springer.com/10.1007/978-3-540-71496-5_78
27. Saleh, B., Abe, K., Arora, R.S., Elgammal, A.: Toward automated discovery of artistic influence. Multimedia Tools and Applications **75**(7), 3565–3591 (2016). doi:10.1007/s11042-014-2193-x. Accessed 2021-09-08
28. Radim Řehůřek, Sojka, P.: Gensim: topic modelling for humans. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50. ELRA, Valletta, Malta (2010). https://radimrehurek.com/gensim/models/doc2vec.html Accessed 2020-12-08
29. De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L.: The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems **50**(1), 1–18 (2000). doi:10.1016/S0169-7439(99)00047-7. Accessed 2021-03-18
30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**(85), 2825–2830 (2011). Accessed 2022-05-31
31. Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., Requin, J.: Singing in the Brain: Independence of Lyrics and Tunes. Psychological Science **9**(6), 494–498 (1998). doi:10.1111/1467-9280.00091. Publisher: SAGE Publications Inc. Accessed 2022-05-05
32. Rigoulot, S., Armony, J.L.: Early selectivity for vocal and musical sounds: electrophysiological evidence from an adaptation paradigm. European Journal of Neuroscience **44**(10), 2786–2794 (2016). doi:10.1111/ejn.13391. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.13391. Accessed 2022-05-10
33. Peretz, I., Gaudreau, D., Bonnel, A.-M.: Exposure effects on music preference and recognition. Memory & Cognition **26**(5), 884–902 (1998). doi:10.3758/BF03201171. Accessed 2022-04-14
34. Wu, F., Huberman, B.A.: Novelty and collective attention. Proceedings of the National Academy of Sciences **104**(45), 17599–17601 (2007). doi:10.1073/pnas.0704916104. Publisher: Proceedings of the National Academy of Sciences. Accessed 2022-03-16
35. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science **311**(5762), 854–856 (2006). doi:10.1126/science.1121066. Publisher: American Association for the Advancement of Science Section: Report. Accessed 2021-01-14
36. Jung, S.-G., Salminen, J., Chowdhury, S.A., Ramirez Robillos, D., Jansen, B.J.: Things Change: Comparing Results Using Historical Data and User Testing for Evaluating a Recommendation Task. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. CHI EA '20, pp. 1–7. Association for Computing Machinery, New York, NY, USA (2020). doi:10.1145/3334480.3382945. https://doi.org/10.1145/3334480.3382945 Accessed 2022-06-14
37. Moore, J.L., Chen, S., Joachims, T., Turnbull, D.: Taste Over Time: The Temporal Dynamics of User Preferences. Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013, 6 (2013)
38. Xing, Z., Wang, X., Wang, Y.: Enhancing Collaborative Filtering Music Recommendation by Balancing Exploration and Exploitation. In: Proceedings of the 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan (2014)
39. Lorince, J., Zorowitz, S., Murdock, J., Todd, P.M.: The Wisdom of the Few? "Supertaggers" in Collaborative Tagging Systems. The Journal of Web Science **1** (2015). doi:10.1561/106.00000002. Accessed 2020-10-13

Appendix

Table 4 MIR extracted audio features provided by The Echo Nest API

Audio Feature	**Value Description**
duration_ms	The duration of the track in milliseconds.
key	The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
mode	Mode indicates the modality (major or minor) of a track. Major is represented by 1 and minor is 0.
time_signature	An estimated overall time signature of a track.
acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements.
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.
instrumentalness	Predicts whether a track contains no vocals. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content.
liveness	Higher liveness values represent an increased probability that the track was performed live.
loudness	The overall loudness of a track in decibels (dB).
speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.
tempo	The overall estimated tempo of a track in beats per minute (BPM).

Influence Probability Calculations

For the sake of reducing noise, we did not use the overall average of the relative novelty change for the aggregated influence probability value. Instead, two dummy variables were created, one corresponding with lyric relative novelty change, and one with MIR relative novelty change. For each song, if its change in lyric relative novelty was negative, the corresponding dummy variable was assigned a 1, and if positive, a 0. The same was done for MIR relative novelty change. When grouping the songs by novelty range and total weeks on chart, we then took the average of the appropriate dummy variable to calculate the probability that a song with those parameters would have a decrease in relative novelty, indicating a higher likelihood of being influential. Because the distribution of data meant that the size of each sample varied for the different numbers of total weeks on chart, the influence probability over the range of total weeks was calculated using a weighted rolling window average, with a window size of 8. Additionally, because there are a relatively few songs that spend more than 40 weeks on the chart, almost all of which fall into the optimal novelty bin for both modalities, we limited our analysis to the 0-35 week range.