

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 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 were 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.

Related Literature

Novelty Metrics

The word novelty is used to describe ideas or artifacts which are new, original, and in some way dissimilar and different to what came before [28, 29]. The production of novelty is important for innovation. Thus, understanding how novelty occurs is a salient question across many domains [30], including the sciences [29, 31, 32], and creative industries such as music [33], film [34], fashion [6], and literature [35]. Novelty can be evaluated in terms of how similar or dissimilar an artifact is when compared to other artifacts within the larger cultural context, which allows examining its relationship to the cultural space it is embedded in [36].

One way this can be achieved is by constructing feature representations that capture information about the key attributes of the artifacts. By representing each artifact based on set of features, this allows us to map individual artifacts to a shared multidimensional feature space and compare them to one another based on their relative positioning. It is then possible to use a distance metric as a way of measuring how similar or dissimilar artifacts are from one another [22, 21, 23, 35, 19, 18, 37, 20].

Previous work with music similarity has utilized MIR features to model individual songs as feature vectors due to the ability of MIR features to capture perceptually relevant audio information that has been validated against human perceptions of audio similarity [22, 21, 38].

This feature representation approach can also be applied to textual data. Previous research into lyrical novelty used Latent Dirichlet Allocation (LDA) to identify latent topics based on word co-occurrence, and represent individual songs based on

their topic composition [23]. A similar approach using LDA has also been applied to research into fan fiction, with the Jensen-Shannon Distance between the topic representations of artifact being used to calculate their relative similarity [35]. The development of word embedding models means that text can also be represented as a feature vector. These models use textual training data to analyze word usage and map each word to a position in multidimensional feature space based on the context in which they are used, allowing us to compare the similarity of words by compare the relative positions of their vectors to one another [19].

For example, the vectors for the words 'sad' and 'morose' would be closer to one another in the vector space than the vectors for the words 'sad' and 'happy'. This can also be extended to map longer text inputs such as paragraphs or entire documents to single vectors. This approach has been leveraged in previous work on the analysis of patent novelty, where document embedding models have been used to generate feature vector representations for patents, allowing the cosine similarity between them to be calculated [20]. There are also domain specific models that can be used for document embeddings, such as BioBERT, which as has similarly been used to assess the relative novelty of PubMed Articles by generating document feature vectors [37].

The benefit of this approach to calculating novelty is that mapping artifacts to a shared feature space allows us to contextualize our measurement of novelty within the larger domain context. A challenge with measuring novelty is that what is considered novel is always changing, and to assess whether or not an artifact is novel, we must contextualize it by comparing it to its contemporaries [39, 40, 12]. By mapping artifacts to a shared feature space and using distance metrics to evaluate similarity, we are able to identify novel artifacts that are informed by this context, without first needing to identify specific markers of novelty.

Novelty and Success

Previous research from psychological studies of culture has suggested that the novelty of cultural artifacts impacts how favorably they will be perceived by audiences [23]. At the individual level, the relationship between subjective novelty and enjoyment has been modeled as an inverse U-curve [41, 42]. In this model, objects that are too familiar or too novel will be less successful, with there being a 'comfort zone' that describes the desired amount of novelty. When considering large scale consumption, we also see that competition for audience attention means that artifacts needs some degree of novelty to stand out, however audiences are also shown to be averse to very high levels of novelty as well [43]. In studies of scientific research, a bias against highly novel work has been observed, with very novel work being less likely to be initially recognized and successful, even in cases where high levels of success are achieved in the long term [44]. Multiple sociological studies exploring this idea across other domains have also found that there appears to be a certain degree of novelty that allows individual artifacts to stand out from their peers, referred to as 'optimal differentiation' [22, 21]. The idea behind optimal differentiation in the music industry is that although songs must be similar enough to previous work to maintain cohesion in the cultural schema, there must be the introduction of new elements that innovates on the established genre norms and sets them apart, without straying too far out of that comfort zone.

Within the music industry, there are of course many factors that influence the likelihood of a song's success. While it is impossible to control for all of these, previous work has demonstrated that the relationship between an artifact's novelty and their likelihood of success is still significant. Previous research on lyric differentiation found that the degree of differentiation had a significant effect on the ranking of a song on the Billboard digital downloads list even when controlling for amount of radio airplay, artist, and specific lyrical topics [23]. Additionally, research on MIR feature differentiation also found that the effect of differentiation on the amount of time a song spent on the Billboard Hot 100 chart remained significant when controlling for artist popularity in terms of how many times the artist had previously charted, genre preferences, and variations in amount of institutional support that artist received based on whether there were with a major or independent music label [22, 21]. In this paper, we build on this previous research to examine whether we can observe a relationship between lyric novelty and MIR novelty at the individual song level, as well as whether the relationship between these different types of novelty also correlates with patterns of commercial success.

Cultural Evolution

We can also think of success from the perspective of impacting cultural evolution. We know that over time, musical styles and genres evolve, and their defining characteristics change. As novel artifacts are introduced into the wider cultural ecosystem, they bring new ideas and creative perspectives, which may or may not be incorporated into the existing stylistic norms [36]. Studies have shown that we can quantitatively track this evolution over time by analyzing changes to the presence and frequency of musical features over time [1, 2, 3, 4]. We can therefore see whether the features of a given artifact become more or less prominent in the style as a whole over time. If later artifacts are very similar to the artifact in question, this tells us that many of the artifact's features have been incorporated into the stylistic norms, and therefore the artifact is more likely to have been stylistically influential. 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]. Although novelty plays an important role in fueling stylistic evolution, we are lacking empirical evidence about the correlation between the degree of novelty in an artifact, and how likely it is that the artifact will be influential.

Research Questions

Based on the above gaps in the literature, we aimed to answer the following research questions:

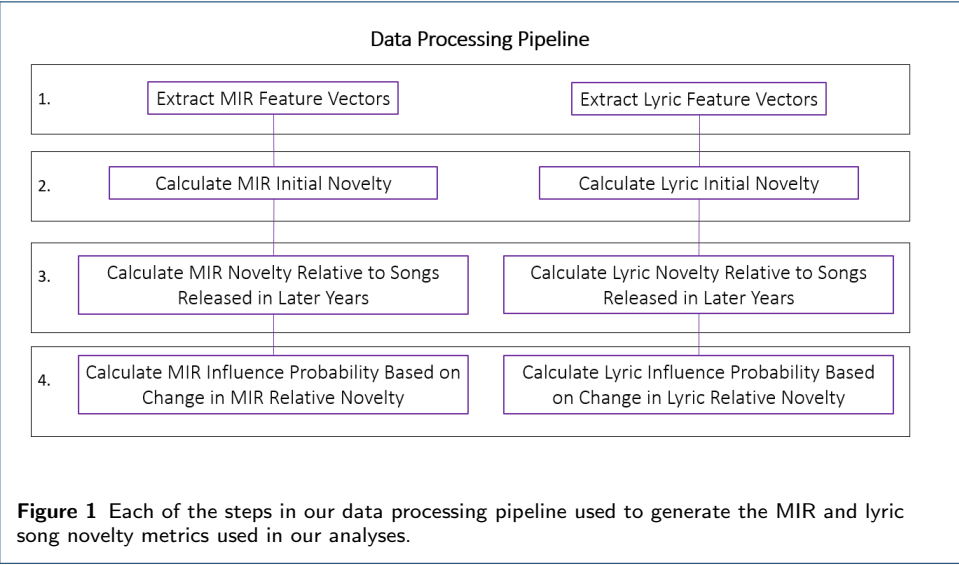
- **RQ 1:** What is the relationship between a song's MIR novelty and its lyric novelty?
- **RQ 2:** Does the relationship between a song's MIR novelty and lyric novelty impact its likelihood of success?
- **RQ 3:** Does the MIR novelty and/or lyric novelty of a song impact the probability of the song being influential?
- **RQ 4:** How does the amount of change over time to the average MIR features compare to the amount of change over time to the average lyric features?

Data

Our data comprises songs from the Billboard Hot 100 chart, which tracks the 100 most popular songs in the United States for each week based on Nielsen radio play scores, physical and digital music sales, along with streaming figures. It therefore serves to track the commercial success of individual songs. For the purposes of our analysis, we measured the success of each song at the time of their initial release based on how many weeks it had been included on the Hot 100 Chart. More time spent on the chart was therefore indicative of higher degrees of success. The Billboard Hot 100 chart is an industry standard for measuring song popularity, and has been used in numerous studies on popular music due to their reliable insight into the most popular American music at a given time [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 total number of weeks each song spent on the chart, and MIR feature data. We acquired text of the lyrics for each song from a variety of online sources with our custom scraping tools. 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.

Methodology

In Figure 1 we have included an illustration of the data processing pipeline used to generate the metrics used for our analyses.



- 1 Extract MIR and Lyric Features: MIR and lyric features represent dimensions that define an MIR feature space and a lyric feature space, respectively. The feature values we derived for each song provide us with a vector that maps that song to these multi-dimensional feature spaces. This allows us to compare the aural and lyrical similarity of sets of songs based on the relative positions of their MIR feature vectors, and the relative positions of their lyric feature vectors.

- 2 **Calculate Song Novelty:** Since the novelty of a song is assessed relative to the other songs released in the same time period, we can group the songs based on the year they were released. We also 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. We can then group the songs by genre and release year and calculate an MIR novelty score and a lyric novelty score for each individual song, based on the average distance between the song's vector and the other song vectors that were released in the same year and genre. For example, when calculating the novelty of a rock song released in 1985, we would compare it to the other rock songs that were also released in 1985. The farther away an individual song's vector is from the average position of the song vectors in that subset, the more novel that song is.
- 3 **Calculate Song Relative Novelty:** Once we have computed the initial novelty score of a song, relative to the year and genre it was released in, we can calculate how novel it is compared to songs that are in the same genre, but were released in a later year. This gives us a score which shows the relative novelty of the song when compared to a given year. For example, if a Rock song was released in 1982, we could calculate its relative novelty with respect to 1985 by finding the average distance between its feature vectors, and the feature vectors of all the Rock songs that were released in 1985.
- 4 **Calculate Influence Probability:** We can calculate the change in relative novelty by subtracting the song's relative novelty score from its initial novelty score. If there is an increase in relative novelty, it is not likely that the song was influential. If there is a decrease in relative novelty, it is more likely the song was influential.

Feature Extraction

The MIR features used in our analysis consisted of quantitative data for 13 high level MIR features which were derived from The Echo Nest using their Music Information Retrieval (MIR) system. A description of the MIR features can be found in Table 5 in the Appendix.

The lyrics features were generated using a document embedding system. We cleaned, preprocessed, and tokenized the lyrics using the Gensim simple preprocessing utility [45]. 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 100 length 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 feature space where we can compare how relatively similar the contents of two documents are by looking at how close their vectors are.

Novelty Scores

To generate our novelty scores, we calculated the distance between each individual song vector, and the other song vectors within the same genre and release year.

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. This allows us to track how much the average position of a genre's feature vectors is changing over time for both MIR features and lyric features, in addition to measuring the individual song novelty scores.

For the lyric vector distances, we used Euclidean distance, as there was no significant covariance between any of the individual features that comprised the feature vectors. For each song, the lyric feature vector can be written as:

$$\vec{l} = (l_1, l_2, l_3, \dots, l_{100})$$

For each genre-year group of songs, we can then take the component-wise average across all the individual song vectors to calculate the average feature vector for that genre-year group, which can be written as:

$$\vec{\mu} = (\mu_1, \mu_2, \mu_3, \dots, \mu_{100})$$

We then calculate the Euclidean distance between this average vector, and the individual song vectors for each song within the genre-year group:

$$d(\vec{l}, \vec{\mu}) = \sqrt{(l_1 - \mu_1)^2 + (l_2 - \mu_2)^2 + \dots + (l_{100} - \mu_{100})^2}.$$

For the MIR vector distances, we followed a similar approach. With MIR data however, some feature values are used as input for determining the values of other features, which means that there are covariances between the features. For example when calculating the danceability of a song, tempo and valence values are included. As a result, we cannot use the Euclidean distance, and instead need to use the Mahalanobis distance. The Mahalanobis distance is similar to Euclidean distance, but is calculated using the covariance matrix of all of the feature vectors, so that it can account for any dependencies between the features, and scale the distance accordingly.

For each song, the MIR feature vector can be written as:

$$\vec{m} = (m_1, m_2, m_3, \dots, m_{13})$$

We again take the average of all the song vectors within the genre-year group in order to yield the average MIR feature vector:

$$\vec{\mu} = (\mu_1, \mu_2, \mu_3, \dots, \mu_{13})$$

We then calculate the covariance matrix, C for the MIR feature vectors, which gives us the covariance between each pair of MIR features included in our feature vectors. 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 over time, 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.

The Mahalanobis distance D_M between an individual song \vec{m} and the rest of the song vectors in the same genre-year group can then be calculated as follows:

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

Calculating the individual song distances yields a distribution of MIR vector distances and a distribution of lyric vector distances for each of the genre-year subsets. In Table 1 we have included the total number of songs for each year, as well as the average MIR and lyric vector distances across all genres for each year.

Table 1 Number of songs per year and average MIR and lyric vector distances across all genres

Year	# of Songs	Avg. MIR Vector Distance	Avg. Lyric Vector Distance
1974	464	3.45	13.45
1975	495	3.39	13.11
1976	468	3.47	13.51
1977	412	3.42	13.27
1978	422	3.33	13.78
1979	435	3.30	14.18
1980	428	3.36	13.52
1981	376	3.25	13.40
1982	397	3.32	13.62
1983	427	3.23	13.83
1984	414	3.23	14.24
1985	391	3.09	14.20
1986	384	3.11	13.99
1987	381	3.06	14.04
1988	371	3.03	13.86
1989	382	3.08	14.09
1990	353	3.03	14.06
1991	353	3.15	13.95
1992	344	3.21	14.12
1993	331	3.23	14.64
1994	318	3.33	14.36
1995	339	3.18	14.38
1996	304	3.22	14.06
1997	329	3.27	14.10
1998	330	3.36	14.23
1999	313	3.38	14.40
2000	313	3.28	14.57
2001	297	3.34	14.86
2002	291	3.29	14.68
2003	302	3.41	14.60
2004	303	3.36	14.68
2005	341	3.41	14.28
2006	361	3.31	14.52
2007	337	3.38	14.88
2008	387	3.31	14.76
2009	429	3.30	14.71
2010	474	3.34	14.71
2011	487	3.27	14.65
2012	369	3.17	14.75
2013	174	3.08	14.38

In order to compare song novelty between songs from different years and genres, we then normalize each song's vector distance relative to the mean distance and standard deviation of the distances for all the songs within the same genre-year

subset. This allows us account for variations in the distributions of distances within each genre-year subset. 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.

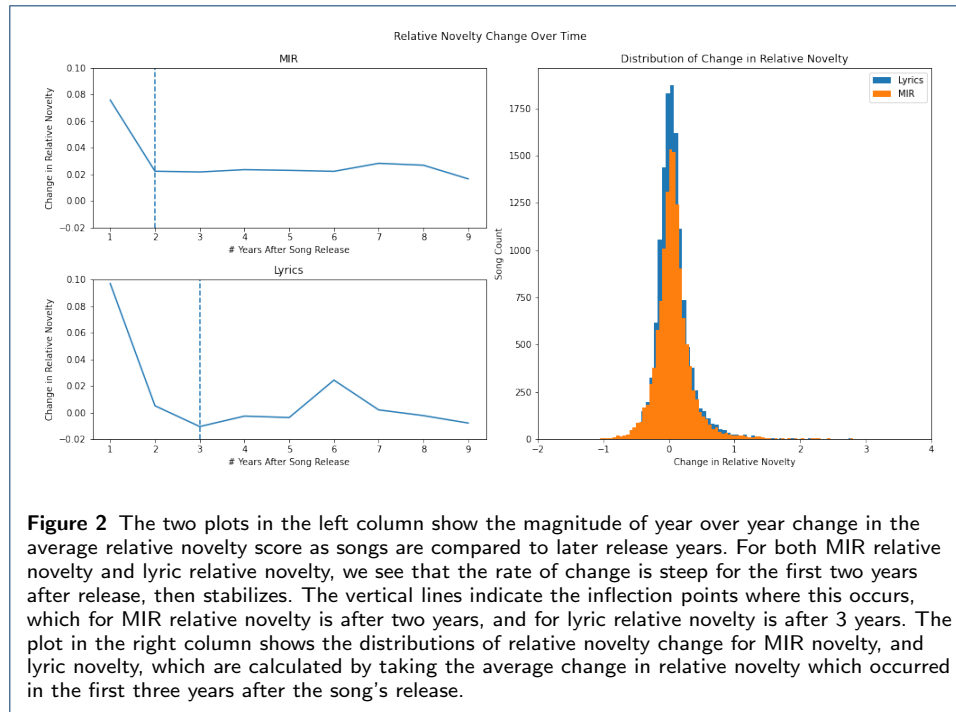
Relative Novelty Scores

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. 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, but instead of comparing the song to songs in the same genre-year subset, we compare it to songs in the same genre, but released in a later year.

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 6 genres with the greatest number of songs; Rock, R&B, Rap, Country, Pop, and Electronica. For each song within these genres, we compared it to the genre-year subsets of the subsequent ten years following the song's release. For example, if a Rock song was released in 1982, we would compare its feature vector to the average feature vector of Rock songs that were released in 1983, 1984, 1985 and so on. Using the same process as we used for calculating the initial novelty scores, we calculated the Euclidean distance for the lyric vector distances, and the Mahalanobis distance for the MIR vector 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 each of the 10 relative vector distances that had been calculated for each song. For each relative vector distance, this was done using the mean and standard deviation of the distances in the genre-year subset that had been used for that specific relative comparison. This yielded relative novelty scores that we could then compare to the relative novelty scores for other years, and to the initial novelty score.

Influence Probability

We can determine whether or not it was probable that a song was influential or not based on whether its relative novelty score had increased or decreased in relation to its initial novelty score. Since we evaluate the relative likelihood that a song was influential by calculating the change in both its MIR and lyric relative novelty, we first determined 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 2 left plots). Because of this, we decided to only consider the relative novelty change that occurred in the 3 years following a song's release by taking the



average of the relative novelty scores for those first three years, and subtracting the song's initial novelty score. Songs released in 2011 or later were excluded since we did not have data for the full three years following those release years. For example, for Prince's, 'When Doves Cry', the average change in MIR relative novelty was an increase of 0.15, and for lyric relative novelty it was an increase of 0.12. Because the relative novelty increased, this tells us that between 1985 and 1987, the average MIR and lyric features of the Rock genre became more dissimilar to the features seen in 'When Doves Cry'.

0.1 Average Feature Change Over Time

In addition to the initial novelty and relative novelty of individual songs, we also consider the magnitude of both MIR and lyric stylistic change over time. Within our data set, this can be understood as changes to the average position of the genre's feature vectors in feature space over time. Since a snapshot of a genre's position at a given time is represented by a distribution of song feature vectors, to consider the novelty score within a broader context of the genre's movement in feature space, we examined the amount of feature variance seen within an individual genre over time for both MIR and lyric features. To perform this analysis, we assigned each song to a decade based on their release years, grouping them into ten-year intervals of 1974-1983, 1984-1993 and so on. Since the MIR and lyric feature vectors for each song provide us with an attribute-based representation of our data, we can compare how distinct the feature distributions of each class are from one another by training a decision tree classifier to predict the temporal class of a given song based on its feature values. Using the training data, a decision tree learns how to partition the feature space to best predict the temporal class of a song. The more distinct the area of feature space that each class inhabits, and the less overlap each has with

other classes, the more accurate the decision tree. As a result, a higher accuracy tells us that there is less similarity between the feature distributions of different decades.

For our classifier, we used the random forest classifier model from the scikit-learn library [46]. A random forest works by fitting multiple decision trees to the data, and averaging results to improve accuracy and avoid overfitting. For our model, we used 500 trees with no depth limit. For individual genres, we then compared the accuracy of the classifier in predicting the temporal class of individual songs when trained using the MIR features versus when trained using the lyric features. For each set of features, a cross-validation was run using a repeated K-fold with 5 splits and 5 repeats, allowing us to generate the distribution of accuracy scores across different train-test splits of the data.

Results

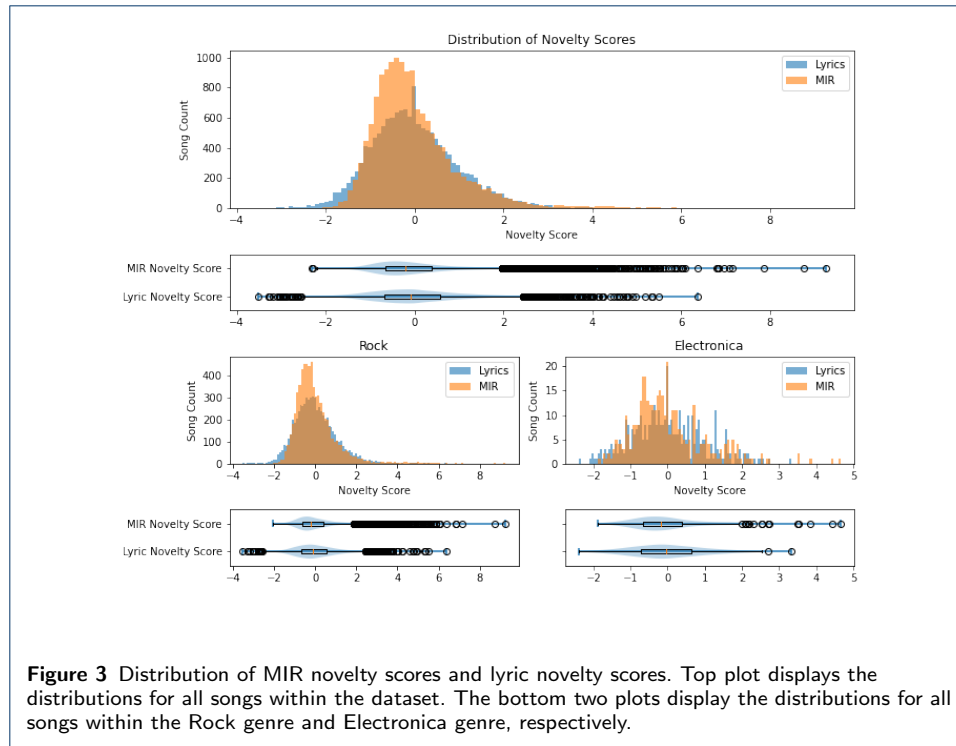
Relationship Between MIR Novelty and Lyric Novelty

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 3 top plot and Table 2). 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 3 bottom plots and Table 2).

Table 2 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

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 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.



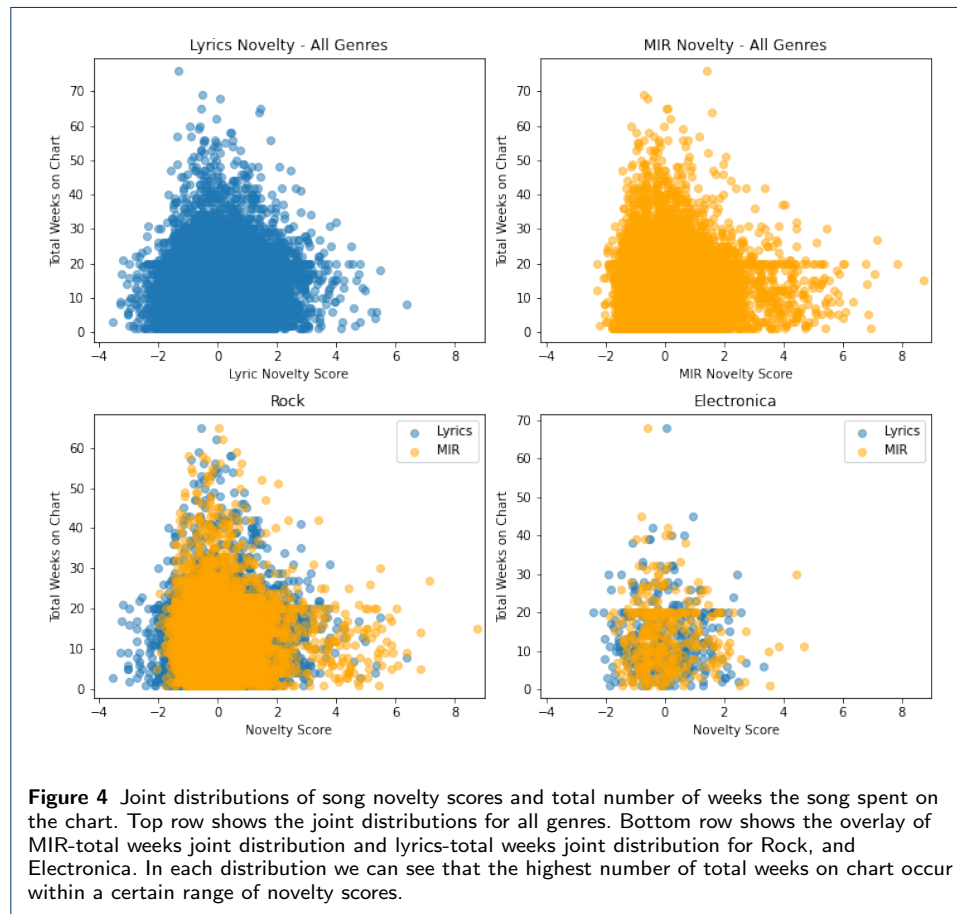
Initial Success

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 looked at the success of individual songs in relation to their MIR novelty score and the lyric novelty score. For example, Prince's 1984 song, 'When Doves Cry', spent 21 weeks on the chart, and had an MIR novelty score of -0.63, putting it at the 26th percentile, and a lyric novelty score of 0.48, putting it at the 68th percentile. This tells us that the MIR features of the song were less novel when compared to other rock songs in 1984, but that the lyric features were more novel.

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 4 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 4 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. [46] 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 5 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, 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 5 and Table 3). Additionally, as the analysis narrows from the top 85th percentile to only the top 95th percentile, we also see that the MIR novelty score increases, while the lyric novelty score decreases, causing the difference between them to grow smaller.

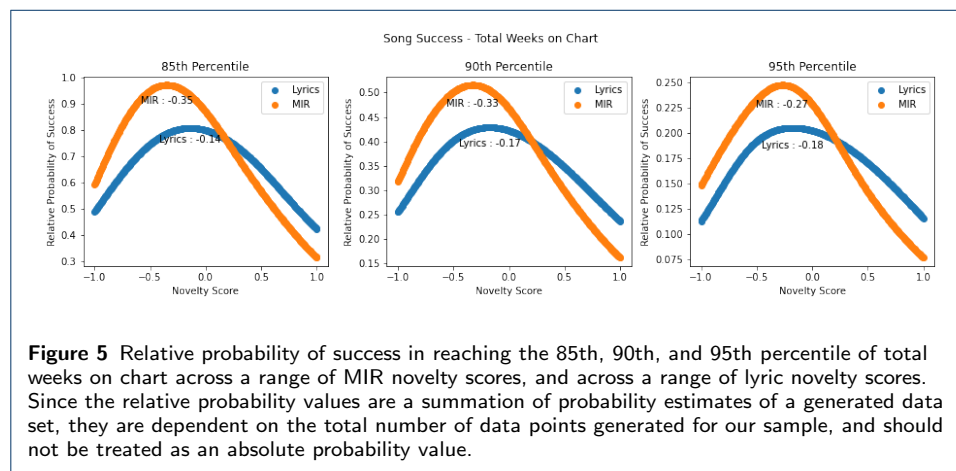


Table 3 Novelty Score with Highest Success Probability

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

Given that we did not find any consistent patterns in the relationship between the MIR novelty and lyric novelty of individual songs, 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 6 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

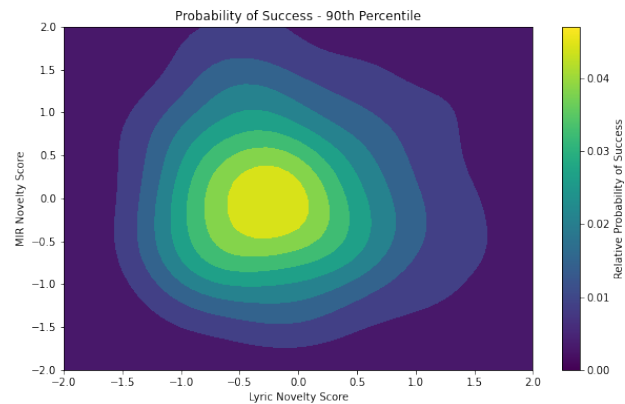


Figure 6 Relative probability of success in reaching the 90th percentile of total weeks on chart for different combinations of MIR and lyric novelty scores. We see that the probability decreases as either score moves away from the value that provides the optimal degree of differentiation within its modality. 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.

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 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

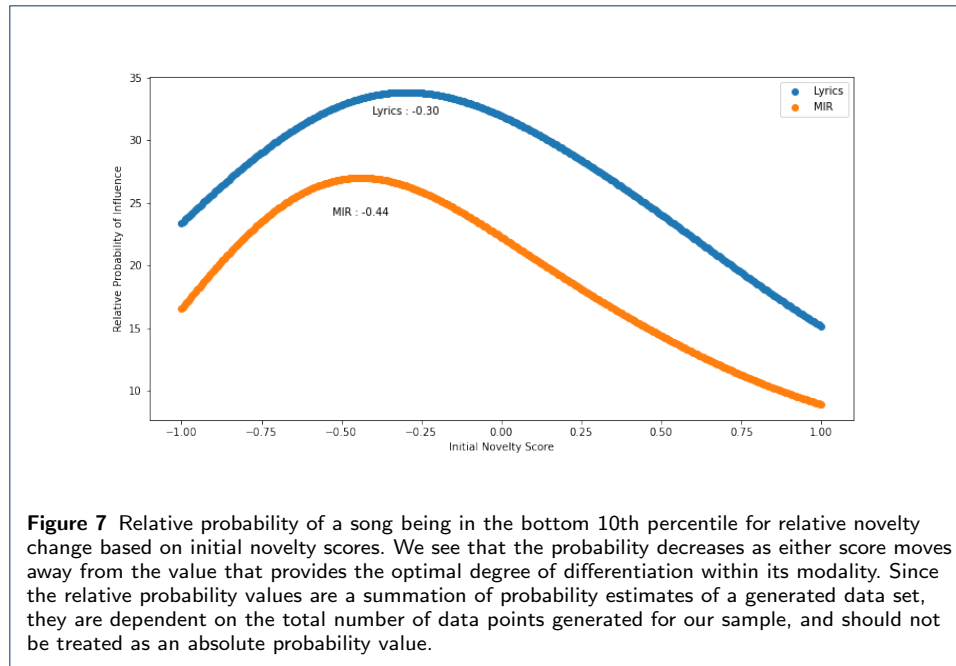
It is worth noting that when looking at the inflection points in Figure 2 that indicate a plateau in the rate of change of relative novelty, the average year over year change in relative novelty for MIR stays positive, meaning that on average, the MIR features of a genre will tend to become less similar 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 2 right plot and Table 4). 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.

Table 4 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 the optimal differentiation novelty values we estimated for initial success (see Figure 7). Again, we see that the optimal MIR novelty score is slightly lower than the optimal lyric novelty score.

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. For example, while both the 1998 Janet Jackson song 'Together Again' and the 2005 Kelly Clarkson song 'Since U Been Gone' were on the chart for a total of 46 weeks, the relative MIR novelty for 'Together Again' decreased by -0.21, while the relative MIR novelty for 'Since U Been Gone' increased by 0.22.

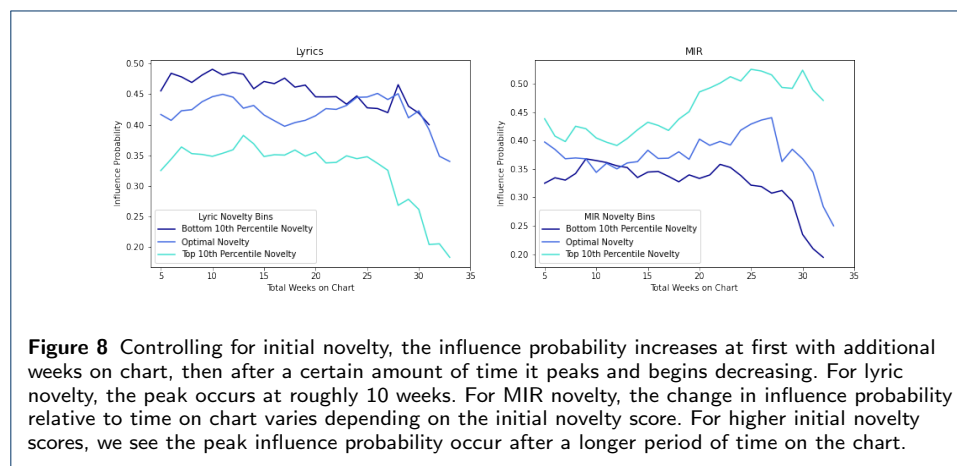


These results suggest that the amount of time a song has spent on the chart cannot be used to predict its likelihood of being influential.

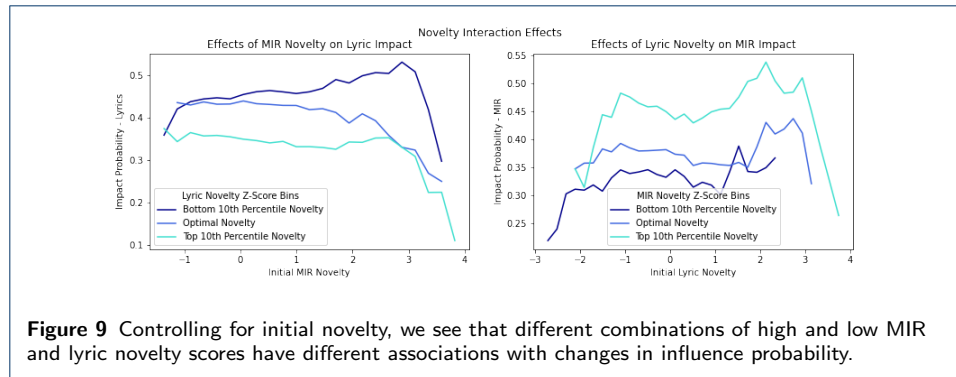
To investigate this, we ran an analysis to examine the relationship between total weeks on chart and influence probability when controlling for limited ranges of novelty scores. Running two analyses, one for MIR novelty and MIR relative novelty change, and one for lyric novelty and lyric relative novelty 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 8 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 8 that the relationship between the influence probability and additional weeks on

the chart varies 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 26-7 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.



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 9). Additionally, we observed that for songs with low lyric novelty scores, higher MIR novelty scores had a slight positive correlation with influence probability, while for songs with high MIR novelty scores, higher lyric novelty scores also had a slight positive correlation with 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.

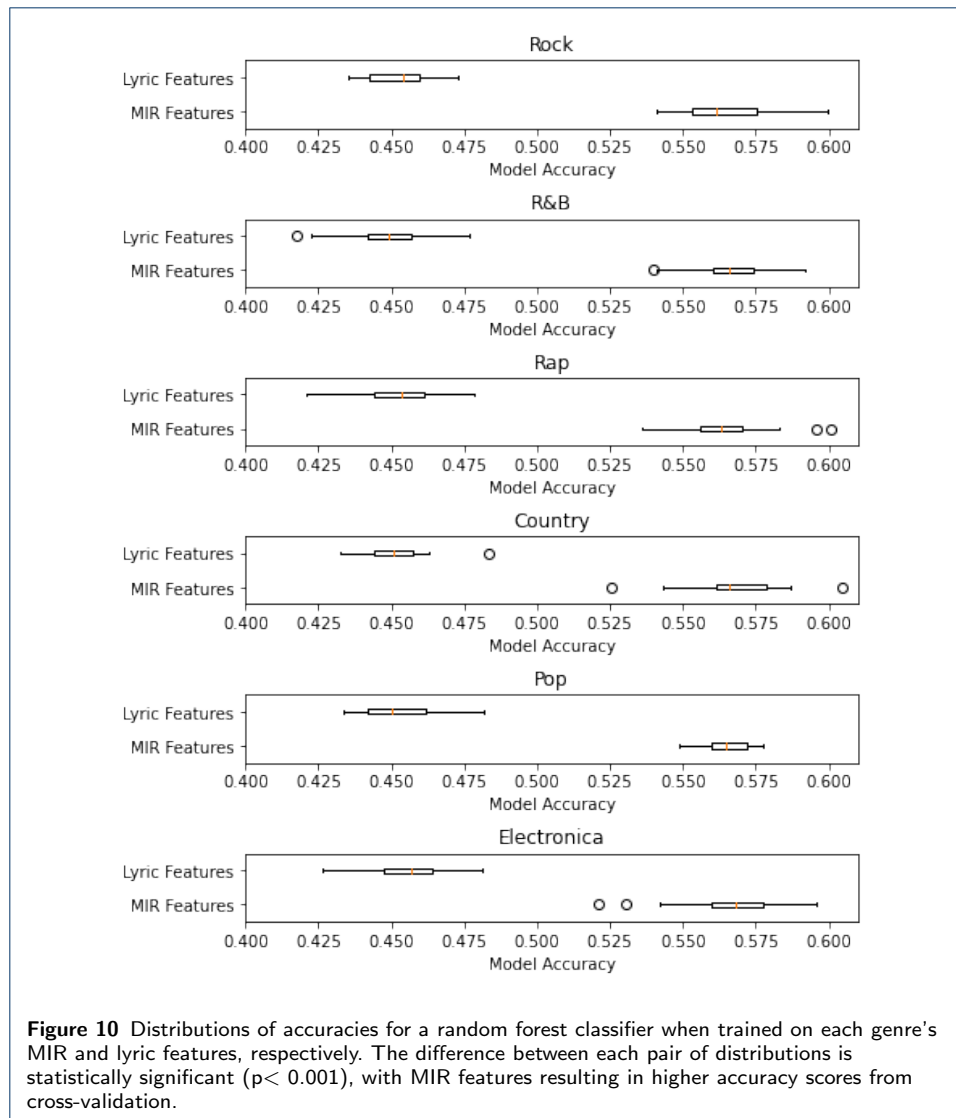


Feature Variance Over Time

In order to delve into why the effect of exposure was so different for lyric novelty than for MIR novelty, we examine the differences in the amount of change over time for the average MIR features compared to the amount of change over time for the average lyric features. In Figure 10, we compare the distribution of accuracy scores when using our random forest classifier to predict the temporal class of individual songs using the MIR features versus when using the lyric features. For each of the largest genres, we observed a statistically significant difference ($p < 0.001$) between the predictive power of the two distributions, with the MIR features consistently resulting in more accurate predictions. Given that the overall accuracy scores are fairly low, with the model trained on lyric features ranging between 42% and 48% accuracy and the model trained on MIR features ranging between 52% and 60% accuracy, this indicates that there is still overlap in the feature distributions of different decades. However, since our analysis demonstrates that there is more predictive power in the MIR features, this tells us that the distributions of MIR features from different decades are more distinct from one another in the feature space than are the distributions of lyric vectors for different decades. Given that we also found that the average year-over-year change in relative novelty for MIR stays positive, indicating that the average distance between old and new songs in MIR feature space is always increasing, we can infer that within the individual genres, there is long-term directionality to the genre's movement through MIR feature space. In contrast, the average year-over-year change in relative novelty for lyrics is, for the most part, negative, which when taken in conjunction with the lower classification accuracy scores when using lyric features, suggests that there is not a significant amount of directionality to the movement occurring within the lyric feature space.

Discussion

By utilizing computational methods to analyze cultural data at scale, this research contributes to our understanding of the relationship between novelty impacts the dynamics of cultural change within the context of the larger cultural ecosystem. Our results highlight 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, or causally attribute stylistic changes within a genre to the influence



of specific songs, our results contribute to understanding overarching patterns of novelty.

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 [47, 48]. 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, which could inform potential avenues of inquiry when investigating cultural trends and evolution.

Novelty and Exposure Effects

Although our data is not sufficient to investigate any possible causal relationship between initial novelty and influence probability, we observed that the relationship between a song's initial novelty score and its influence probability varies relative to 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 correlation with 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 Hot 100 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 [49]. 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 [50]. It is possible, then, that for high-novelty songs, the increased exposure due to more time spent on the chart may decrease the perceived novelty of the song, leading audiences to experience it as being closer to the optimal level of differentiation. For low-novelty artifacts, however, this decrease in perceived novelty could make them seem too familiar, hence we see that the amount of exposure which was beneficial varied depending on the initial novelty scores. For lyric novelty, we saw that increased time on the chart correlated with a decrease in influence probability, regardless of whether the lyric novelty score was lower or higher than optimal. Given that we observed significantly less variance over time for lyric features than we did for MIR features, it is possible that even songs with very high lyric novelty scores are not distinct enough to benefit from higher levels of exposure.

For lyric novelty, we saw that increased time on the chart correlated with a decrease in influence probability, regardless of whether the lyric novelty score was lower or higher than optimal. Given that we observed significantly less variance

over time for lyric features than we did for MIR features, it is possible that even songs with very high lyric novelty scores are not distinct 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 highlights the importance of considering the potential effect of social influence on not only the initial popularity of cultural artifacts, but also the longer term evolution of the cultural ecosystem. In Salganik et al's study on social influence in cultural markets, different social behaviors led to different patterns of success within an artificial music market, demonstrating the significant impact of social influence on what songs become popular [51]. While the effect of social influence was shown to be largely independent of the specific attributes of the individual songs, by impacting which songs become popular this will also impact the relative amount of exposure those songs will receive. As our analysis suggests that patterns of exposure may potentially impact the likelihood of an artifact exerting stylistic influence, suggests a possible mechanism by which the impact of social influence has a downstream effect on the stylistic evolution of the musical ecosystem.

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 [52], 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 [40]. 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 [53].

Limitations

It is important to note that there are many external factors that can affect whether or not a song become successful, which are unfortunately not captured in the scope of this data set. As a result, we cannot draw any inferences about direct causal relationships between novelty and the success metrics we examined. Additionally,

our data is lacking important controls that could be correlated with novelty and influence, due to the currently unavailability of such comprehensive data.

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 [54].

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

Availability of data and materials

The code that supports the findings of this study are available at:

<https://github.com/LINK-NU/Billboard-Hot-100-Music-Novelty>. Data available on request.

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. In: *Proceedings of the 16th ISMIR Conference*, 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. In: Proc. of the 15th International Symposium On CMMR, Online, p. 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. Uzzi, B., Mukherjee, S., Stringer, M., Jones, B.: Atypical Combinations and Scientific Impact. *Science* 342(6157), 468–472 (2013). doi:10.1126/science.1240474. Accessed 2022-06-08
29. Li, Y., Zhang, Y., Capra, R.: Analyzing information resources that support the creative process. In: ACM SIGIR Conference on Human Information Interaction And Retrieval. CHIIR '22, pp. 180–190. Association for Computing Machinery, New York, NY, USA (2022). doi:10.1145/3498366.3505817. <https://doi.org/10.1145/3498366.3505817> Accessed 2022-04-01
30. Liu, L., Wang, Y., Sinatra, R., Giles, C.L., Song, C., Wang, D.: Hot streaks in artistic, cultural, and scientific careers. *Nature* 559(7714), 396–399 (2018). doi:10.1038/s41586-018-0315-8. Number: 7714 Publisher: Nature Publishing Group. Accessed 2020-10-14
31. Shin, H., Kim, K., Kogler, D.F.: Scientific collaboration, research funding, and novelty in scientific knowledge. *PLOS ONE* 17(7), 0271678 (2022). doi:10.1371/journal.pone.0271678. Publisher: Public Library of Science. Accessed 2022-08-02
32. Shi, F., Foster, J.G., Evans, J.A.: Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks* 43, 73–85 (2015). doi:10.1016/j.socnet.2015.02.006. Accessed 2021-01-26

33. Miles, S.A., Rosen, D.S., Barry, S., Grunberg, D., Grzywacz, N.: What to Expect When the Unexpected Becomes Expected: Harmonic Surprise and Preference Over Time in Popular Music. *Frontiers in Human Neuroscience* **15** (2021). Accessed 2022-05-06
34. Sreenivasan, S.: Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords. *Scientific Reports* **3**(1), 2758 (2013). doi:10.1038/srep02758. Number: 1 Publisher: Nature Publishing Group. Accessed 2021-03-19
35. Jing, E., DeDeo, S., Ahn, Y.-Y.: Sameness Attracts, Novelty Disturbs, but Outliers Flourish in Fanfiction Online. arXiv:1904.07741 [cs] (2019). arXiv: 1904.07741. Accessed 2020-11-20
36. Park, D., Nam, J., Park, J.: Novelty and influence of creative works, and quantifying patterns of advances based on probabilistic references networks. *EPJ Data Science* **9**(1), 1–15 (2020). doi:10.1140/epjds/s13688-019-0214-8. Number: 1 Publisher: SpringerOpen. Accessed 2020-10-12
37. Liu, M., Bu, Y., Chen, C., Xu, J., Li, D., Leng, Y., Freeman, R.B., Meyer, E.T., Yoon, W., Sung, M., Jeong, M., Lee, J., Kang, J., Min, C., Song, M., Zhai, Y., Ding, Y.: Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology* **73**(8), 1065–1078 (2022). doi:10.1002/asi.24612. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24612>. Accessed 2023-01-01
38. Cheng, D., Joachims, T., Turnbull, D.: Exploring Acoustic Similarity for Novel Music Recommendation, 7 (2020)
39. Zangerle, E., Huber, R., Vötter, M., Yang, Y.H.: Hit song prediction: Leveraging low- and high-level audio features. *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, 319–326 (2019). doi:10.5281/zenodo.3258042. Accessed 2020-11-01
40. 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)
41. Berlyne, D.E.: Novelty, complexity, and hedonic value. *Perception & Psychophysics* **8**(5), 279–286 (1970). doi:10.3758/BF03212593. Accessed 2021-03-19
42. Chmiel, A., Schubert, E.: Back to the inverted-U for music preference: A review of the literature. *Psychology of Music* **45**(6), 886–909 (2017). doi:10.1177/0305735617697507. Accessed 2020-11-05
43. Chai, S., Menon, A.: Breakthrough recognition: Bias against novelty and competition for attention. *Research Policy* **48**(3), 733–747 (2019). doi:10.1016/j.respol.2018.11.006. Accessed 2023-01-01
44. Wang, J., Veugelers, R., Stephan, P.: Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* **46**(8), 1416–1436 (2017). Publisher: Elsevier. Accessed 2023-01-02
45. 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
46. 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
47. 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
48. 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
49. 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
50. 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
51. 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
52. 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
53. 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)*
54. 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 5 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.