Dynamic Influence Ranking Algorithm Based on Musicians' Social and Personal Information Network

Yiming Liu [1], Longxin Wang [2], Yunsong Jia [1], Ziwen Li [1] and Hongju Gao [1,*]

[1] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; liuyiming@cau.edu.cn (Y.L.); jiayuns@cau.edu.cn (Y.J.); ziwen2049@163.com (Z.L.)

[2] College of Engineering, China Agricultural University, Beijing 100083, China; wlx0228abc@163.com

* Correspondence: hjgao@cau.edu.cn

Abstract: Social influence analysis is a very popular research direction. This article analyzes the social network of musicians and the many influencing factors when musicians create music to rank the influence of musicians. In order to achieve the practical purpose of the model making accurate predictions in the broad music market, the algorithm adopts a macromodel and considers the social network topology network. The article adds the time decay function and the weight of genre influence to the traditional PageRank algorithm, and thus, the MRGT (Musician Ranking based on Genre and Time) algorithm appears. Considering the timeliness of social networks and the continuous development of music, we realized the importance of evolving MRGT into a dynamic social network. Therefore, we adopted audio data analysis technology and used Gaussian distance to classify and study the evolution of music properties at different times and different genres and finally formed the dynamic influence ranking algorithm based on musicians' social and personal information networks. As a macromodel heuristic algorithm, our model is explanatory, can handle batch data and can avoid unfavorable factors, so as to provide fast speed and improved accuracy. The network can obtain an era indicator DMI (Dynamic Music Influence) that measures the degree of music revolution. DMI is the indicator we provide for music companies to invest in musicians.

Keywords: influence ranking; PageRank algorithm; Gaussian distance; the similarity of music genre; HP filter

1. Introduction

The music market is an integral part of the cultural market, with a large audience and promising development prospects. A music company that invests in a potential musician and purchases a potential music record can earn considerable income. Conversely, investing in an influential musician will result in losses. The sales potential of music records is closely related to the musicians and their attributes. The market needs a quantitative indicator to measure the influence of musicians, and the indicator of musical influence is a topic worthy of study.

The market can usually judge the popularity of musicians. Musicians have many fans, and the popularity of previously released records will naturally have a higher market value. However, market judgment is lagging. When a musician is not famous or has not entered a market, the market value of the musician is challenging to evaluate. Based on the fact that it is difficult to evaluate the influence of musicians, we started the research of this article.

First, the creation of a musician will be affected by many factors, such as the musician's natural qualities, the current era, the genre, and the degree to which others influence it. These factors can be divided into social and non-social. For example, the background of the times, the prevalence of the genre a musician belongs to, and the degree of influence by others can be classified as social factors. Here, these social factors can be evaluated by constructing a reasonable social network for musicians. The division of music genres and a

series of indicators that measure the degree of change of music genres can be classified as social factors. This can be evaluated through the existing massive data sets.

This article first proposes a social network influence model for musicians to measure social factors. There are two roles of mentor and student in the social network of musicians: music influencer and music influenced person. According to the different genres and ages of musicians, we can construct a weighted network with parameters. When the weight of each edge and the edge that each node links in and out can be determined, we improved the PageRank algorithm that Google prospects and proposed the MRGT (Musician Influence Ranking based on Genre and Time) model.

We used Gaussian distance to analyze the similarity of genres. The similarity index was used to analyze the difference in music similarity within and between genres. Then, we used the data set to judge whether the similarity index was reasonably established. The similarity within the genre was much higher than that between the genres, reflecting the division's correctness. The distance index was a non-social index. The distance of each period can be intuitively discovered through the HP filtering process to measure the degree of change in music with the times. Only when the degree of change is higher or lower than this threshold can it be approximated that a sudden change has occurred. Changes in the times in social factors that were difficult to quantify achieved quantification through distance. Then, we added this quantitative indicator to the influence model of musicians' social networks. Then, we added this quantitative indicator to the influence model of musicians' social networks, called DMRGT (Dynamic MRGT). In this way, we obtained the value of influence in the musicians' social network based on a specific period, which is the so-called "musician potential". Music companies can invest by referring to this "musician potential" value. In order to enable readers to understand the context of the article more intuitively, the flow chart for the procedure of our study is shown in Figure 1.
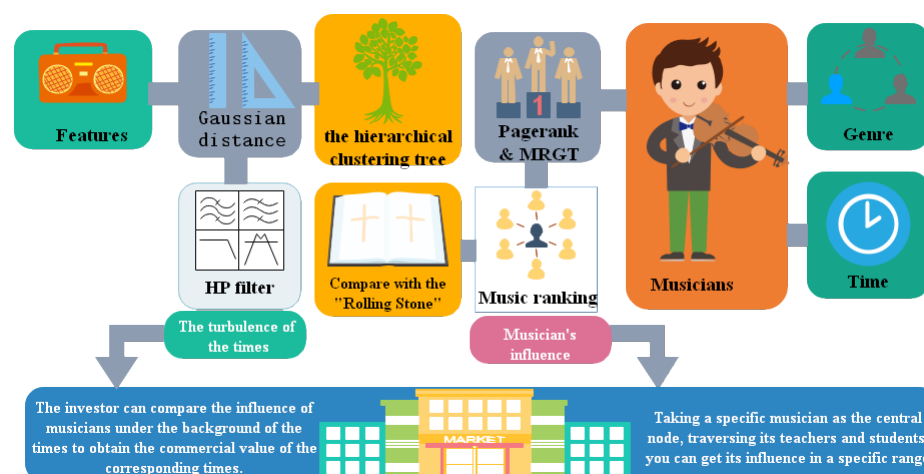

Figure 1. Flow chart.

The main contributions of the entire article are as follows:

- The DMRGT model proposed combines the advantages of macromodels, heuristic algorithms, and audio data analysis, adding the time decay function, the weight of genre influence, and audio data analysis influence factor to the traditional PageRank algorithm. Both social factors and non-social factors are included in the study.
- The DMRGT model is derived from the PageRank algorithm and can select nodes based on specific heuristic algorithm iterations, such as PageRank in this paper. The importance of nodes (musicians) can be calculated to measure the social influence from the social network data, which are social factors.
- A fast music similarity evaluation method based on Gaussian Distance and audio data analysis techniques is proposed in this paper in order to calculate music similarity, classify the genres, and extract music properties using massive music data, which are

non-social factors. An HP filtering process is used to measure the degree of change of music with the times.

- As a macromodel heuristic algorithm, the DMRGT model is explanatory, can handle batch data, and can avoid unfavorable factors, so as to provide fast speed and improved accuracy.

2. Related Work

Research on the market influence of musicians can start from their social networks. Morton and Kim noticed that a musician might directly influence another musician through direct and long-term personal interaction, but they may also be indirectly affected, e.g., hearing another musician's music in a coffee shop [1].

The analysis of a person's opinions, emotions, or behaviors influenced by others [2] is called social influence analysis (SIA) [3]. The main idea of SIA is how to quantify the influence of each user and how to identify the most influential users in social networks [4]. The market influence of musicians is an SIA issue. These models are usually divided into two categories: micromodels and macromodels. The micromodel focuses on the interaction of humans and examines the structure of the influence process [3]. Two famous influence diffusion models in this category are the independent cascade (IC) model and linear threshold (LM) model [5–8]. Kempe et al. develop a general model of diffusion processes in social networks that simultaneously generalizes the two models to explore the limits of models in which strong approximation guarantees can be obtained [5]. Most studies need to perform Monte Carlo (MC) simulation to evaluate the user influence in IC and LM models, which leads to tremendous computational costs [9]. Therefore, these approaches cannot achieve fast computing speed, and it is not suitable for large-scale music market assessments.

The macromodel posits that all users have the same attractiveness to information, the same propagation probability, and the same influence [10]. To find the most influential member groups in music social networks, a good starting point is ordinary social networks [11]. Most well-known models in this category are epidemic models, which are mainly used to model infectious disease spread. However, the macromodel ignores the topological characteristics of social networks [3]. The percentage of nodes in each class is calculated by the mean-field rate equations, which are too simple to depict such a complex evolution accurately [3]. Daley and Kendall study topological networks [12]. Some scientists study human behavior and influential diffusion mechanisms [13–15], but musicians' behavior mechanisms are not as widely available as big data platforms for the music market.

The influence maximization problem is to find a set of highly influential nodes that maximizes the influence propagation scale in the social network under a given diffusion model [16]. Kempe et al. [5] are the first to formalize the influence maximization as a discrete optimization problem and to prove the problem is NP-hard [16]. Greedy algorithms "greedily" select the active node with the maximum marginal gain towards the existing seeds in each iteration [3]. Using the optimal local solution can provide the maximum influence value of the node to approximate the optimal global solution. Many algorithms are proposed, including the climbing-up greedy algorithm [5], cost-effective lazy forward (CELF) method [17], NewGreedy and MixedGreedy algorithms [18], and upper bound-based lazy forward (UBLF) algorithm [19] et al. UBLF explored new upper bounds to significantly reduce the number of MC simulations and to discover the top k influential nodes in social networks [19]. Some of the algorithms used to study differences between individuals are based on these greedy algorithms [3]. A common limitation of these approaches is computational inefficiency on large networks [19].

Even with improved greedy algorithms, the running time is still large and may not be suitable for large social network graphs [20]. A possible alternative is to use heuristics [20]. Heuristic algorithms select nodes based on a specific heuristic, such as degree or PageRank, rather than calculating the marginal gain of the nodes in each iteration [3]. The efficiency

is achieved by trading off the accuracy. Pedro Cano et al. used PageRank to study the topological structure of music networks, and their analysis revealed the emergence of complex network phenomena in music information networks with artists as nodes and relationships as links. These attributes can provide some suggestions for searchability and possible optimization for designing a music recommendation system.

Audio data are relatively complex and contain rich structural information on multiple time scales. Second, the music itself continues to develop, and the artist, song, and genre all change over time. In terms of the complexity of processing audio data, there is a big semantic gap when extracting advanced attributes, such as "type, mood, instrument and theme" from audio [21]. Nick Collins may be one of the first to use music data to study the influence of music. He studied the content-based classification of Synthpop songs on a small hand-annotated data set of 364 songs [22]. Later, he used the partial matching (PPM) variable-order Markov model for prediction experiments, but the data set used was also relatively small (248 tracks) [23]. Shalit et al. [24] used the theme modeling method for the first time to study the influence of music. Specifically, they used the dynamic topic model [25] and the document influence model [26]. The time series was extended to traditional topic modeling, allowing topics to evolve. With the recent widespread popularity of deep learning-based methods, Morton and Kim applied deep learning to content-based music impact recognition for the first time [1]. They use a deep belief network to extract features from the audio's spectral representation, although they treat influence recognition as a multi-label classification problem with only ten classes (affecting artists) in total. Xue, Wenzhe uses the DIM (Document Impact Model) to explore the topic modeling method of musician influence. At this stage, he applies k-means, which is not guaranteed to find the optimal global clustering in terms of loss reduction. He also attempted to use the song audio trained by the Siamese Convolutional Neural Network to conduct related research on the influence of musicians. The main limitation of this method is that the limited time scale will lead to loss of information, which the model cannot explain [27].

The music market has a large amount of data, and it is more suitable to adopt macro-model heuristic algorithms. DMRGT combines the advantages of macromodels, heuristic algorithms, and audio data analysis. Figure 2 is a comparison diagram of the efficiency of the DMRGT algorithm and other model algorithms. In order to improve accuracy, each musician will consider the influence of his genre, the influence of the times, and the relative influence of his portfolio. Compared with micronetworks, this model can process a large amount of data; compared with traditional PageRank, it has improved accuracy; and compared with neural networks and other models, it is more interpretable. There is no doubt that the DMRGT model has its advantages and advantages.

Type	Typical algorithm	Disadvantage	Comparison with DMRGT
Micro model	Independent Cascade Model (IC) and Linear Threshold (LT)	Time-consuming, not suitable for large-scale music market evaluation	Fast index calculation, suitable for large-scale music market evaluation
macro model	epidemic model	The accuracy is low, ignoring the topological characteristics of social networks. Unable to accurately describe the individual differences to be considered in complex evolution	Considering individual differences and considering the topological characteristics of social networks, the accuracy is high
Optimized macro model	Research with topological network and diffusion mechanism	Musicians' behavioral mechanism data is not yet complete, and the model cannot be applied to the music market temporarily	The data set used is universally searchable online
Greedy Algorithm	CELF, UBLF, mixed greed...	Algorithm complexity is high, execution time is long, and efficiency is poor	Low algorithm complexity and short execution time
Heuristic algorithm	Traditional pagerank	Low accuracy, prefer old nodes	High accuracy, adding time decay function, with preference for new nodes
Audio data analysis	DIM, artificial neural network	DIM is not easy to find a better solution, artificial neural network data cannot be explained	Strong interpretability, relatively good calculation results

Figure 2. Algorithm comparison table.

3. Methodology

3.1. Symbol Descriptions

Table 1 gives the major symbol discription for indicators calculated through the whole work procedure.

Table 1. Major symbol description.

Symbols	Description
MI	Music Influence
DMI	Dynamic Music Influence
MR	Music Ranking Value in Subnetwork
GMR	Music Ranking Value in Global Network
f_i	Number of followers for musician i
$Nor(\zeta)$	Min-max Normalization
$\text{Dist}_m(x, x')$	Gaussian distance
$\text{Music\_similarity}(x, x')$	The similarity of x and y music characteristics
$\text{Col\_music\_similarity}$	Similarity of music features within the group
$\text{Cols\_music\_similarity}$	Music feature similarity between groups
hpc	HP Filter Cycle Term

Datasets

The data source used in this article was obtained from the 2021 Mathematical Contestin Modeling: MCM PROBLEM D: The Influence of Music (available at https://www.comap.com/undergraduate/contests/mcm/contests/2021/problems/, accessed on 2 September 2021). The dataset information is described in Table 2. It contains a total of 42,770 rows of influence data, 98,340 rows of music data, and 5854 rows of artist data. Details of these data are shown in the following chart.

Table 2. Dataset description [28].

	Influence_Data	Full_Music_Data	Data_by_Artist	Data_by_Years
Number_of_Data	**42,770**	**98,340**	**5854**	**100**
	incluencer_id		A unique identification number given to the person listed as the influencer (string of digits).	
	influencer_name		The name of the influencing artist as given by the follower or industry experts (string).	
	influencer_main_genre		The genre that best describes the bulk of the music produced by the influencing artist (if available) (string).	
	influencer_active_start		The decade that the influencing artist began their music career (integer).	
Features	follower_id		A unique identification number given to the artist listed as the follower (string of digits).	
	follower_name		The name of the artist following an influencing artist (string).	
	follower_main_genre		The genre that best describes the bulk of the music produced by the following artist (if available) (string).	
	follower_active_start		The decade that the following artist began their music career (integer).	
	artist_name		The artist who performed the track (array).	
	artist_id		The same unique identification number given in the influence_data.csv file (string of digits).	

3.2. Influence Evaluation Model

This section first analyzes the feasibility of using the PageRank algorithm to build an individual influence model and the MQRT algorithm to evaluate the collective influence.

3.2.1. Individual Influence of PageRank Evaluation

The experimental data set used influence data, consisting of 42,770 records of influencers and followers among related musicians, and provided metadata, such as the music genre of predecessors, popular time, music genre and follow time of followers.

It is noteworthy that all genres in this article consider the time of popularity, i.e., a genre is one with the same genre name and popular time. Schools with the same name but different times of popularity are regarded as other schools.

The PageRank algorithm is a web page ranking algorithm proposed by Google founded by Brin and Page in 1998. When the PageRank algorithm is introduced into the influence of music artists, the link relationship between pages changes. It becomes

the relationship between the influence and following of music artists, as shown in the following formula:

$$MR(i) = d \sum_{j \to i} \frac{MR(j)}{L(j)} + \frac{1-d}{|G_M|} \tag{1}$$

Among them, $M.R.(i)$ represents the music influence value of the musician i; $j \to i$ represents the music junior j following the music senior i; $L(j)$ is the number of predecessors that j, the music junior, follows; G_M is the collection of musicians; $|\ |$ represents the number of elements in the group; and d is the drag coefficient, usually 0.85.

The conditional probability of picking any predecessors to learn from is $1/L(i)$. The Formula (1) shows that when the music junior learns from predecessor j and seeks the next learning object, the probability of learning from J's predecessors is d. That is to say, all the predecessors of the predecessor j have the same probability of being selected, if any, and they all influence the next musician with the possibility $\frac{1}{L(i)}$.

However, in music, musicians do not choose to follow objects with the same probability to learn; their choices are influenced by factors such as their predecessors' popularity and the similarity of their style and genre. The ancestors that the musicians follow have a certain degree of similarity with the genre or musical genre characteristics of the musicians themselves. Therefore, whether or not the subsequent followers of the musician are affected or in what way is highly related to the length of time or the essence of the influence of other predecessors. For example, the musical innovation of many brilliant artists in the history of music development will have an essential guiding role in the direction of music development for a long time in the future. The higher the essential predecessors' achievements, the more they will be recognized, and their followers will increase accordingly. There are also outstanding followers. Influential musicians will also contribute to the prosperity and wealth of a specific music genre.

Based on the mutual enhancement relationship between the outstanding performance of musicians' followers and the highly accomplished musicians' artistic genres, this paper proposes an improved PageRank algorithm MRGT combined with the performance and popularity of emerging musicians in recent years. Furthermore, this paper investigates musicians' influences regarding music genres in recent years and the changes in musicians' influence over time. Finally, specific experiments verify that the MRGT algorithm is able to effectively quantify musicians' performance and determine the preferences of musicians.

3.2.2. Collective Influence of MQRT Evaluation

The basic idea behind MRGT is that if a musician influences more recent musicians, then the musician themself is likely to be highly influential; emerging musicians who are influenced by popular genres receive more PR. Therefore, for every musician i, the PR they receive depends on the time interval j of all the musicians they influence and the genre's performance in recent years. The specific formula can be expressed as follows:

$$MR(i) = d \sum_{j \to i} MR(j) \times \omega(i,j) + \frac{1-d}{|G_M|} \tag{2}$$

$\omega(i,j)$ is the probability of jumping from j to i.

The quality and impact of a genre are not fixed; they change with the musicians' grade. Some genres become more influential as the musicians within them become increasingly talented. On the contrary, some genres become less effective due to subsequent musicians' slow development and lag. Therefore, it is necessary to treat each genre of different periods separately. For example, "R&B, 2000–2010" and "R&B, 2010–2020" are treated as separate genres. Note that in this article, all genres refer to the genres or musicians considered in that year, unless otherwise noted.

Since there is a specific correlation between most musicians and influencers when a musician considers choosing one of the influencers of a musician he is currently studying,

new musicians from the high-impact genres of recent years are likely to be the first choice. Therefore, the weighting factor among musicians can be set as follows:

$$\omega(i,j) = \ln(f_i)[LS(g_i)]^{\alpha}[t(i,j)]^{(1-\alpha)} \tag{3}$$

where $L.S._{g_i}$ is the influence of the genre to which the musician i belongs in recent years, and $t(i,j)$ represents the time interval when the musician j is influenced by the musician i.

$$t(i,j) = e^{-\sigma(T_j - T_i)} \tag{4}$$

Here, σ is the time decay factor and T_i is time.

The influence of a genre is mainly determined by the musicians' musical development, and the better the product, the greater the power of the genre. Thus, the score of the genre $G.S.(g_k)$ can be expressed as follows:

$$GS(g_k) = \frac{1}{|G_M(g_k)|} \sum_{k \in G_M(g_k)} \text{GMR}(k) \tag{5}$$

where $GS(g_k)$ is the influence score of genre g_k and $G_p(g_k)$ is the collection of musician IDs that follow in genre V_K.

To accurately characterize the dynamic nature of the influence of a genre, we consider each period of a genre individually and consider its performance in recent years when assessing its influence. If the genre G_i belongs to year y and the genre name is N_{g_i}, then the genre can be expressed as (N_{g_i}, y), and the N of other years belongs to another periodical, such as (N_{g_i}, y) in year $Y - 1$, which can be expressed as $(N_{g_i}, y - 1)$. Genre G, in the most recent T_g year publication set and g_i, can be expressed as follows:

$$\text{pnid}(g_i) = \left\{ (N_{g_i}, y), (N_{g_i}, y - 1), \cdots, (N_{g_i}, y - t_g) \right\} \tag{6}$$

For example, the collections of music genres "R&B, 2000–2010" over the past 30 years are "R&B, 1970–1980", "R&B, 1980–1990" and "R&B, 1990–2000". The performance of the genre in recent decades $L.S.(g_i)$ can be expressed as follows:

$$LS(g_i) = \frac{1}{t_g + 1} \sum_{g_k \in pnid(g_i)} GS(g_k) \tag{7}$$

Based on the above analysis, the MRGT algorithm provides musicians' influence based on genre influence. All musicians and genres or musicians are first initialized, the score of musicians is set as $\frac{1}{G_M}$, the score of genres is set as $\frac{1}{G_G}$, and $\frac{1}{G_M}$ is the number of musicians. Moreover, V_V is the number of genres. The influence rankings composed of some of the higher-ranked musicians are presented in Figure 3.
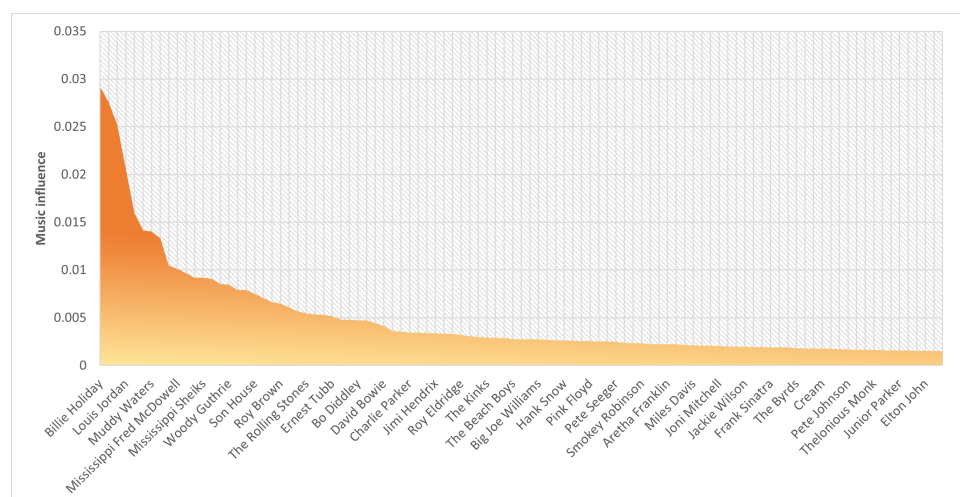
Figure 3. Top musicians' influence ranking.

3.3. Music Feature Similarity Evaluation Model

Music similarity evaluation often uses waveform evaluation directly, most of which is complex and challenging to directly apply to mass analysis. This paper proposes a fast music similarity evaluation method based on Gaussian distance and seven music features that are easy to collect. Moreover, it can obtain the average and maximum differences among music and music collection, making it possible to evaluate the similarity between musicians and genres quickly.

This section first introduces the music features used, describes its processing flow, and provides the calculation method of music similarity. There are seven ways to express the characteristics of a music melody in Appendix A. We analyze the similarity between music based on this.

We believe that the influence of each musical feature on musical similarity should be similar. In order to make the influence of different music characteristics on similarity approximately equal, we standardized the data.

For any discrete variable Z, the min-max normalization formula is as follows:

$$Nor(Z) = \frac{Z - min(Z)}{max(Z) - min(Z)} \tag{8}$$

Then, Gaussian distance was used to batch process discrete data and floating-point data.

As we used Gaussian distance to measure the similarity between music, the absolute space cannot directly reflect the difference between music for certain music features. For example, in the loudness of logarithmic distribution, there is a significant difference between 0 and -10; one has no sound, and the other has sound, while the loudness difference between -100 and -110 is relatively small as they are far away from the origin. Therefore, we used the quantile adjustment method to adjust the data further.

The function that is symmetric along the radial direction was used to map finite-dimensional data to a high-dimensional space. It is usually defined as a monotonic function of the Euclidean distance from any point x to a specific center point x' in space.

$$Dist_m\left(x, x'\right) = e^{-\frac{\|xx'\|^2}{2\sigma^2}} \tag{9}$$

x' is the center of the kernel function, and $\|x - x'\|^2$ is the Euclidean distance (L2 norm) of vector x and vector x'. As the distance between the two vectors increases, the Gaussian kernel function decreases monotonically. σ is an exogenous parameter, which can be adjusted artificially. The effective range of the Gaussian kernel function, the larger the value, and the larger the local influence range of the Gaussian kernel function. At the

same time, the selected σ cannot be too small. Otherwise, it is easy to overfit in the classification task.

We hope that the more similar the music features, the larger the value; therefore, we used the following formula to calculate the similarity of music features:

In order to achieve the purpose of the more similar features, the greater the value; therefore, we use the following formula to calculate the similarity of music features:

$$\text{Music\_similarity}(x,y) = \begin{cases} 1/\text{Dist}_m(x,y) & \text{Dist}_m(x,y) \neq 0 \\ +\infty & \text{Dist}_m(x,y) = 0 \end{cases} \quad (10)$$

$$\text{Col\_music\_similarity} = \begin{cases} \dfrac{2\times \sum\limits_{i,j\in Col, i\neq j} \text{Music\_similarity}(i,j)}{N\times(N-1)} & N \geq 2 \\ \text{Undefined} & N \leq 1 \end{cases} \quad (11)$$

As shown in Figure 4, in order to evaluate whether the Gaussian kernel function can accurately classify the genre, we produced hierarchical clustering diagrams of each genre and its main characteristics. In the production process, we used full_music_data .csv to extract the characteristics of each song and influence_data.csv to mark each song according to the composer's genre. Then, we calculated the mean value of each genre song feature, standardized each feature, and calibrated the seven regular distribution intervals with integer numbers according to the Three Sigma Guidelines; the range is $[-3,3]$, which is shown in the figure. In this diagram, the darker the color, the more distinctive the characteristic of the genre.

It can be clearly seen that each genre has a higher difference in specific song characteristics than other genres. For example, the children's genre has more tracks and longer durations, but it is not suitable for dancing. It can be said that the difference between genres in different song characteristics is the most significant difference between genres. We can also see the similarity between the various genres from the hierarchical clustering tree in the figure. The more similar the genres are on the top branch.

We also found that artists within genres are more similar to artists between genres, as measured by this musical similarity measurement model. This is the same as our cognition; that is to say, the music similarity measurement method is effective.
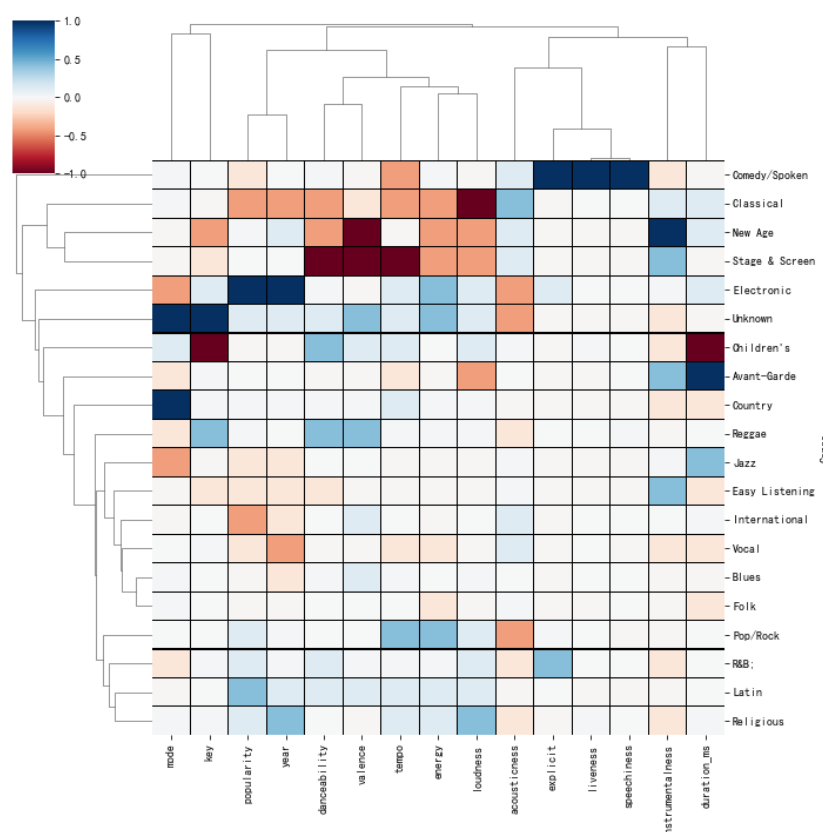
Figure 4. Genre characteristics.

4. Experiment

4.1. Musician Influence Experiment

To verify the algorithm's correctness, this paper randomly selected the ranking of musicians of popular music acts for verification. We set the 50 most excellent pop musicians in history chosen by "Rolling Stone" magazine as the reference standard after consulting various authoritative data. Due to the difference in selection years, this article excluded 13 famous musicians in Rolling Stone magazine but not in the MRGT data set. Figure 5 gives the experimental results as a bubble graph in which the size and color of the bubble changed with the extent of the influence; the greater the power, the larger the bubble.

It is worth noting that the part above the bubble graph's regression line represents the group of musicians in the MRGT model who ranked and surpassed "Rolling Stone". Most of the orders that reached the "Rolling Stone" scale are emerging musicians. The active age is mainly in the late 1960s and the 1970s and 1980s. These emerging musicians have less influence than older musicians due to the decline factor, so that their power declines less, and they catch up with the old musicians who initially ranked closely. Moreover, if our paper removes the decline factor, the ranking will be further consistent with the order of "Rolling Stone". For the convenience of explanation, here, we directly eliminate the musicians who have surpassed the regression line, that is, the eight emerging musicians, such as Madonna. The model is consistent with the ranking of "Rolling Stone". The fit was further improved, reaching 48.3%.
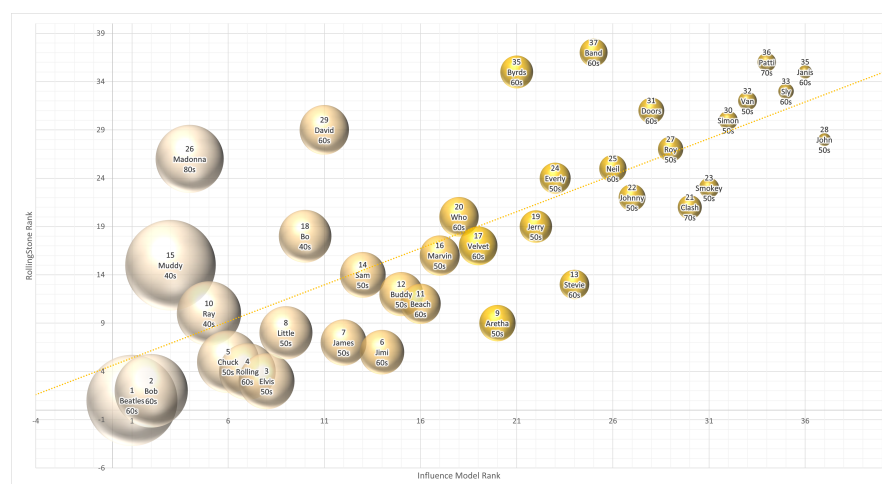
Figure 5. Bubble chart of Rolling Stone and MRGT ranking of year and musician.

This further experiment proves that the model in this paper is a good model for emerging musicians. Simultaneously, for better musicians, the time decay model does not significantly impact their rankings. The preference for new musicians does not negate the historical influence of the earlier musicians. For example, Muddy in the 1940s, Ray and others are still ranked highly. This reflects that the model retains the characteristics of objective and fairness based on specific preference settings.

In the traditional sense, the so-called ranking must be based on a unified index system. Many are based on an orchestra's annual budget, number of tours, broadcast audiences, the specifications and number of musicians, and the number of followers for the current order of orchestras. For example, CNN in the United States has selected twenty musicians as candidates worldwide to commemorate the broadcast of the art and culture program "icon" and set the world's five significant musicians by voting. This article considers that the voting process is very long, which also leads to a lengthy ranking process. Using votes as an indicator to rank musicians will also be affected by unfavorable factors, such as scouring the rankings and canvassing voters. The most significant advantage of the MRGT algorithm is that it automatically sorts musicians based on various indicators, such as time and genre. A computer algorithm that avoids unfavorable factors, such as brushing and canvassing votes and the cranking speed, is rapid.

4.2. Music Similarity Experiment

We hope to analyze the development trend of the genre through the similarity of the genre and try to prove our conjecture through historical moments. We adopted HP filtering for our research. Our article first obtains the similarity index between the song pairs in the data set. The first step is to preprocess the data. Then, randomly select 100 groups to calculate the similarity between the three groups. The first group selects a specific type of sample to calculate the average value, reflecting its degree of looseness. The data is taken from inside and outside the genre to reflect its degree of outlier in the second group. The third group is the average of the first two samples, reflecting a new branch of genre. By applying the HP filter to the three event sequences, the HP filter obtains the average period item and the trend item and then checks for outliers of the average period item. The point of change is the period of significant change. In Reference [29], HP filtering principle assumes a time series.

The HP filtering principle assumes that the time series $X = \{x_1, x_2, \cdots, x_n\}$ contains the long-term trend part and the periodic fluctuation component, then $X = XT + XC$. The trend sequence $XT = \{xt_1, xt_2, \cdots, xt_n\}$, representing the long-term trend component, and the fluctuation sequence $XC = \{xc_1, xc_2, \cdots, xc_n\}$, representing the recurring fluctu-

ation component, are obtained by HP filtering X. XT is the solution to the minimization problem [29].

$$\min\left\{\sum_{i=1}^{n}(x_i - xt_i)^2 + \lambda \sum_{i=2}^{n-1}[(xt_{i+1} - xt_i) - (xt_i - xt_{i-1})]^2\right\} \qquad (12)$$

Among them, $\sum_{i=1}^{n}(x_i - xt_i)^2$ is the sum of the fluctuation sequence's squares, representing the fluctuation degree of the fluctuation sequence or the trend sequence's tracking degree to the original series. Ref. [30] $\sum_{i=2}^{n-1}[(xt_{i+1} - xt_i) - (xt_i - xt_{i-1})]^2$ is the sum of squares of the second difference sequence of the trend sequence and represents the trend component's smoothness. $\lambda(\lambda \geq 0)$ is the penalty factor that controls the smoothness of the trend sequence and becomes the smoothness parameter. Empirically, we assume that λ is equal to 14,400. The solution of the HP filter optimization problem is as follows. By taking the different partial derivatives, we can obtain the system of equations [31] :

$$\begin{cases} \frac{\partial S}{\partial g_1} = -2(y_1 - g_1) + 2\lambda(g_3 - 2g_2 + g_1) = 0 \\ \frac{\partial S}{\partial g_2} = -2(y_2 - g_2) + 2\lambda(g_4 - 2g_3 + g_2) - 4\lambda(g_3 - 2g_2 + g_1) = 0 \\ \cdots \\ \frac{\partial S}{\partial g_{T-1}} = -2(y_{T-1} - g_{T-1}) + 2\lambda(g_{T-1} - 2g_{T-2} + g_{T-3}) - 4\lambda(g_T - 2g_{T-1} + g_{T-2}) = 0 \\ \frac{\partial S}{\partial g_T} = -2(y_T - g_T) + 2\lambda(g_T - 2g_{T-1} + g_{T-2}) = 0 \end{cases} \qquad (13)$$

The matrix form of the system is as follows:

$$\left[I + \lambda \begin{pmatrix} 1 & -2 & -2 & \cdots & 0 & 0 \\ -2 & 4+1 & -2-2 & \cdots & 0 & 0 \\ 1 & -2-2 & 1+4+1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1+4 & -2 \\ 0 & 0 & 0 & \cdots & -2 & 1 \end{pmatrix}\right] \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ \cdots \\ g_{T-1} \\ g_T \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_{T-1} \\ y_T \end{pmatrix} \qquad (14)$$

where I is the unit matrix. Then, we can obtain the main trend using the above equation [32]. Figure 6 shows the trend term of the HP filter. Figure 7 shows the cycle term of the HP filter.

Observing Figure 7, we find that the country has obvious fluctuations, so we might as well take out the trend term and cycle term of the country genre separately, such as in Figure 8. In the Figure 8, in 1941, the similarity within the country music genre dropped sharply, and a sudden change occurred. This represents the country style from pluralism to unanimity.

Looking at the history, it was found that country music did indeed undergo a great fusion during World War II [33].
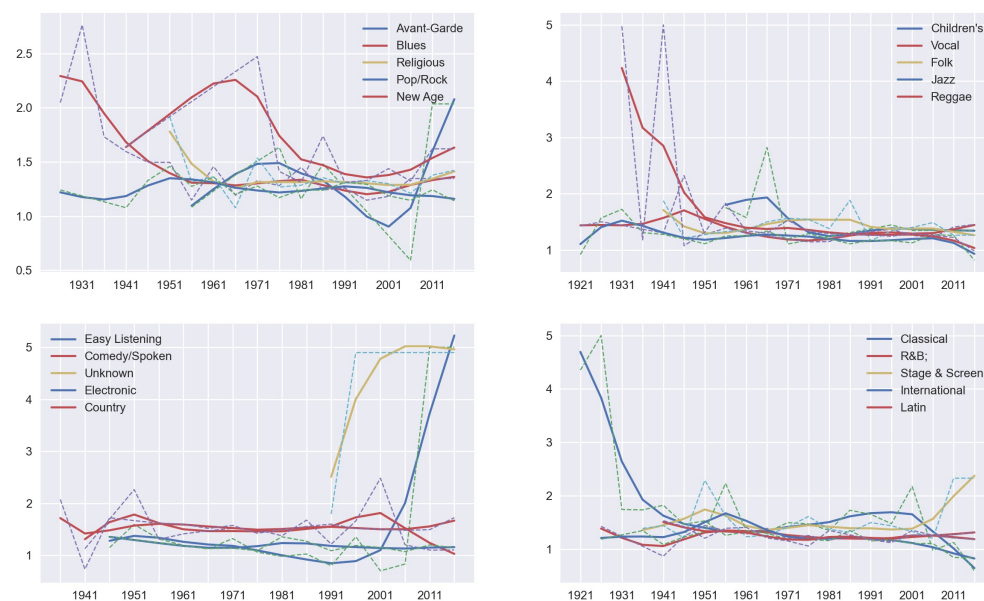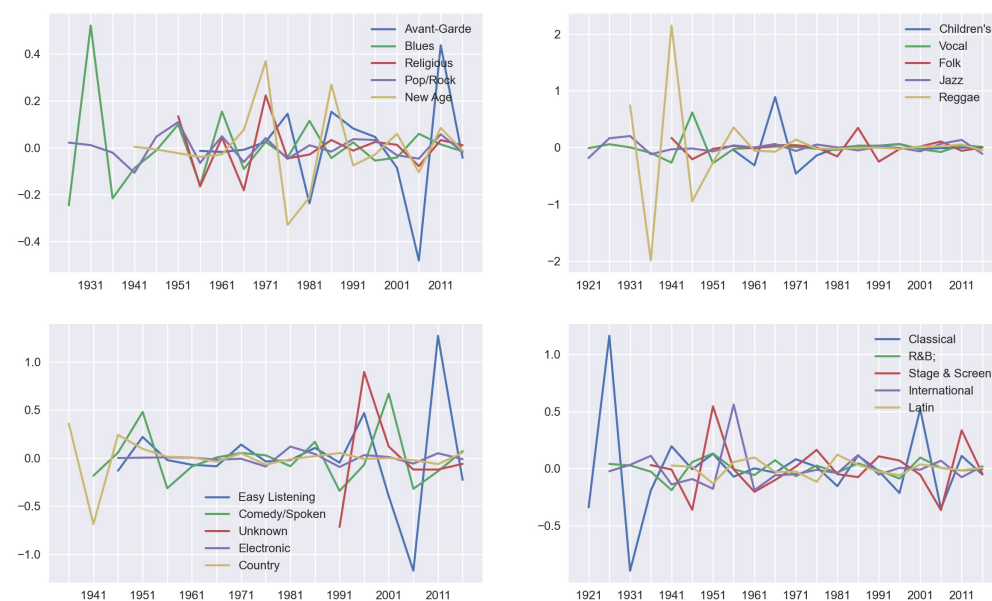
Figure 6. The trend term of HP filter.

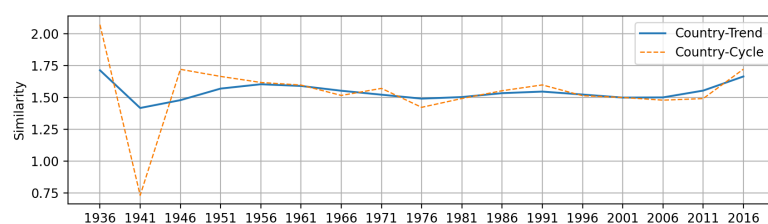**Figure 7.** The cycle term of HP filter.


Figure 8. The cycle term of HP filter.

4.3. Music Algorithm Ranking Comparison

In related works, this article found, from a principle perspective, that the micromodel is time consuming and is not suitable for large-scale music market evaluation; the macromodel cannot accurately describe the individual differences to be considered in the complex

evolution, and the accuracy is low; the greedy algorithm has high algorithm complexity and execution time and poor efficiency. The DMRGT algorithm is a fusion algorithm based on a heuristic algorithm plus audio data analysis technology. Compared with the traditional PageRank, we added the time decay function, genre weight, and audio data analysis influence factor. However, whether the new algorithm is better than the heuristic algorithm has not yet been confirmed.

Here, we determined the inferring artist influence by comparing the The Document Influence Model (DIM) algorithm with the DMRGT algorithm. The above two algorithms were used to sort the data set in this article and compare it with the Rolling Stone ranking. The Rolling Stone ranking order was taken as the abscissa and the DMRGT relative ranking order as the ordinate to draw the graph. As shown in Figure 9, the graphs are distributed around the $y = x$ axis, reflecting that the DMRGT model can largely fit the ranking method in the music market.

The Rolling Stones ranking order was taken as the abscissa and the DIM relative ranking order as the ordinate to draw the graph. The results are shown in Figure 9 and occupy the top eight places, as shown in Figure 10. It can be seen that the DMRGT ranking is positively correlated with the Rolling Stone ranking.
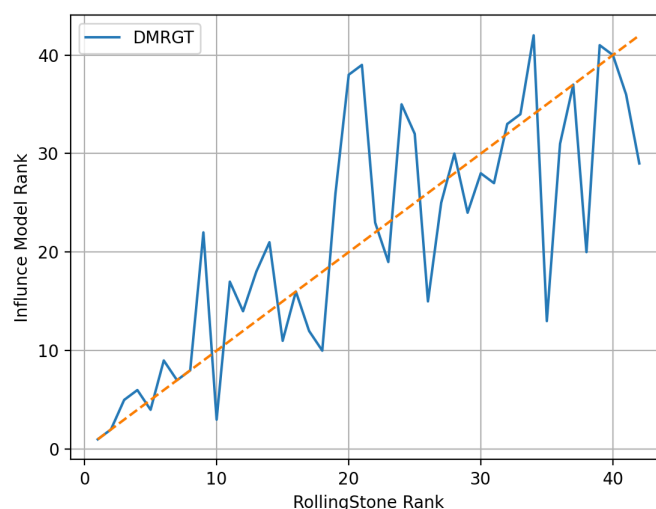
Figure 9. Comparison of DMRGT ranking and Rolling Stone ranking.
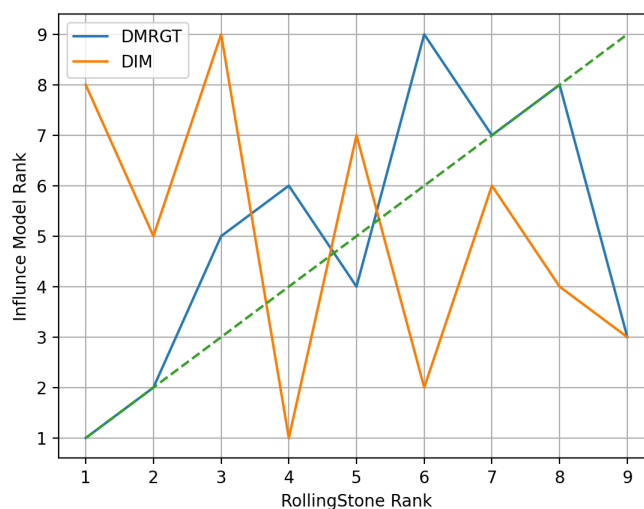
Figure 10. Ranking comparison between DMRGT algorithm based on Rolling Stone ranking and DIM algorithm.

5. Application

We used the trend item results generated by the HP filter to integrate the MRGT influence index to obtain an MRGT model with era characteristics, which we call the DMRGT model. According to Figure 7, the cycle items were obtained by HP filtering according to Pinet similarity time series data and normalized. The cycle term is less than -2 times the standard deviation, and a mutation is considered to have occurred. We constructed a dynamic Music Influence index (ln(popularity)*MI*cycle) of each musician, and the revolutionaries were judged according to the DMI size. The transformative power of the *I* musician in time *t* is shown in the following:

$$\text{DMI} = \frac{hpc(t) * MI * \sum_{m \in [t-10,t+10]} \text{popularity}\,(i,m)}{\sum_{i} m_i} \tag{15}$$

Among them, *MI* represents music influence, and HPC is the cycling term of HP filtering. The lower the HPC value is, the more significant the sudden change in a short time. In popularity(*i*,*m*), *I* represent musicians, and *m* represents a person's average popularity in the last ten years. The DMI values of artists of different genres are shown in Figure 11.
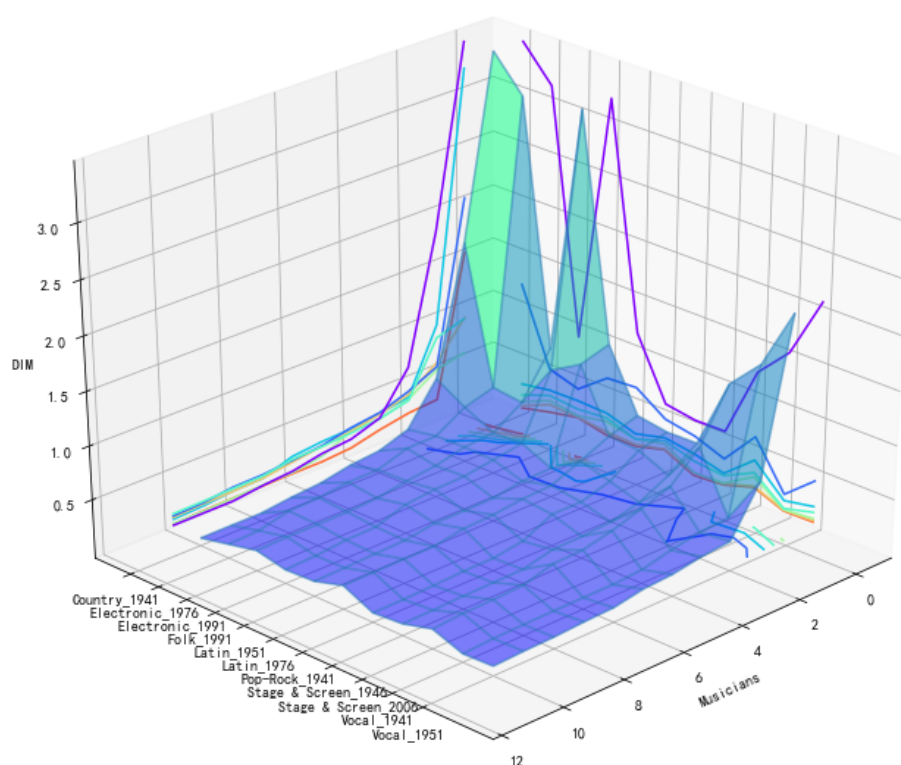

Figure 11. DMI of revolutionary artists of different genres.

Music companies can use this indicator to sort the age, genre, and musicians to identify the most suitable musicians for signing.

6. Conclusions

The Gaussian distance can be used to divide genres, and the distance within the genre after its division is significantly smaller than that outside the genre. The trend items of the music characteristics of each genre after the HP filtering process can be used as an indicator of the change in the times. In a specific period, the DMI index constructed by considering the influence of social networks, indicators of changes in the era, and the ratio of creative music to contemporary music can effectively reflect the influence of musicians.

As a macromodel heuristic algorithm, the DMRGT model is explanatory and can handle batch data. Compared with the ordinary macromodel, we do not assume that each node is similar. We have fully considered the different influences of different genres, different eras, and music created by each artist in a certain era on the node. Therefore, the accuracy of DMRGT is higher. During the comparison process, the accuracy of the influence index we constructed is even better than that of neural network algorithms.

Some people predict market changes through music attributes [34], but it is rare to predict them through influence and music attributes. Even if we discuss the influence of social networks, our model is relatively good. By constructing DMI indicators, music companies can find musicians with more investment value and receive assistance in their business decisions. The shortcoming of this article is that it builds a macronetwork that can handle massive amounts of data.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; writing—original draft preparation, H.G., Y.L., L.W., Y.J. and Z.L.; writing—review and editing, H.G. and Y.L.; validation, Y.L.; formal analysis, Y.L.; supervision, H.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: https://www.comap.com/undergraduate/contests/mcm/contests/2020/problems/, accessed on 2 September 2021.

Appendix A

Music Feature [28]:

- Danceability: A measure of how suitable a track is for dancing based on a combination of musical elements, including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable, and 1.0 is the most danceable (float).
- Energy: A measure representing the perception of intensity and activity. A value of 0.0 is least intense/energetic and 1.0 is most intense/energetic. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy (float).
- Valence: A measure describing the musical positiveness conveyed by a track. A value of 0.0 is most negative, and 1.0 is most positive. Tracks with high valence sound more positive (e.g., happy, cheerful and euphoric), while tracks with low valence sound more negative (e.g., sad, depressed and angry) (float).
- Tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, the tempo is the speed or pace of a given piece and is derived directly from the average beat duration (float).
- Loudness: The overall loudness of a track in decibels (dB). Values typically range between −60 and 0 dB. Loudness values are averaged across the entire track and are useful for comparing the relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude) (float).
- Mode: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track. Major is represented by 1, and minor is 0.
- Key: The estimated overall key of the track. Integers map to pitches using standard pitch class notation, e.g., 0 = C, 1 = C #/Db and 2 = D. If no key is detected, the value for the key is −1 (integer).

References

1. Morton, B.G.; Kim, Y.E. Acoustic features for recognizing musical artist influence. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 1117–1122.
2. Travers, J.; Milgram, S. The small world problem. *Phychol. Today* **1967**, *1*, 61–67.
3. Tang, J.; Sun, J.; Wang, C.; Yang, Z. Social Influence Analysis in Large-scale Networks. In Proceedings of the 2009 ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD'09 ed., Paris, France, 28 June–1 July 2009. [CrossRef]
4. Peng, S.; Wang, G.; Xie, D. Social influence analysis in social networking big data: Opportunities and challenges. *IEEE Netw.* **2016**, *31*, 11–17. [CrossRef]
5. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 137–146. Available online: https://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf (accessed on 2 September 2021).
6. Leskovec, J.; McGlohon, M.; Faloutsos, C.; Glance, N.; Hurst, M. Patterns of cascading behavior in large blog graphs. In Proceedings of the 2007 SIAM International Conference on Data Mining, SIAM, Minneapolis, MN, USA, 26–28 April 2007; pp. 551–556.
7. Gruhl, D.; Guha, R.; Liben-Nowell, D.; Tomkins, A. Information diffusion through blogspace. In Proceedings of the 13th International Conference on World Wide Web, New York, NY, USA, 17–20 May 2004; pp. 491–501.
8. Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **1978**, *83*, 1420–1443. [CrossRef]
9. Li, P.; Liu, K.; Li, K.; Liu, J.; Zhou, D. Estimating user influence ranking in independent cascade model. *Phys. Stat. Mech. Appl.* **2021**, *565*, 125584. [CrossRef]
10. Sun, J.; Tang, J. A survey of models and algorithms for social influence analysis. In *Social Network Data Analytics*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 177–214.
11. Li, K.; Zhang, L.; Huang, H. Social influence analysis: models, methods, and evaluation. *Engineering* **2018**, *4*, 40–46. [CrossRef]
12. Daley, D.J.; Kendall, D.G. Epidemics and rumours. *Nature* **1964**, *204*, 1118–1118. [CrossRef] [PubMed]
13. Wang, H.; Deng, L.; Xie, F.; Xu, H.; Han, J. A new rumor propagation model on SNS structure. In Proceedings of the 2012 IEEE International Conference on Granular Computing, Hangzhou, China, 22–24 October 2014; pp. 499–503.
14. Wang, Y.Q.; Yang, X.Y.; Han, Y.L.; Wang, X.A. Rumor Spreading Model with Trust Mechanism in Complex Social Networks. *Commun. Theor. Phys.* **2013**, *59*, 510–516. [CrossRef]
15. Xia, L.L.; Jiang, G.P.; Song, B.; Song, Y.R. Rumor spreading model considering hesitating mechanism in complex social networks. *Phys. Stat. Mech. Appl.* **2015**, *437*, 295–303. [CrossRef]
16. Peng, S.; Zhou, Y.; Cao, L.; Yu, S.; Niu, J.; Jia, W. Influence analysis in social networks: A survey. *J. Netw. Comput. Appl.* **2018**, *106*, 17–32. [CrossRef]
17. Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; Glance, N. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 420–429. [CrossRef]
18. Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 199–208.
19. Zhou, C.; Zhang, P.; Zang, W.; Guo, L. On the Upper Bounds of Spread for Greedy Algorithms in Social Network Influence Maximization. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 2770–2783. [CrossRef]
20. Zhang, B.; Wang, Y.; Jin, Q.; Ma, J. A pagerank-inspired heuristic scheme for influence maximization in social networks. *Int. J. Web Serv. Res. (IJWSR)* **2015**, *12*, 48–62. [CrossRef]
21. Van Den Oord, A.; Dieleman, S.; Schrauwen, B. Deep content-based music recommendation. In *Neural Information Processing Systems Conference (NIPS 2013)*; Neural Information Processing Systems Foundation (NIPS): Lake Tahoe, NV, USA, 2013; Volume 26.
22. Collins, N. Computational Analysis of Musical Influence: A Musicological Case Study Using MIR Tools. In Proceedings of the ISMIR, The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010), Utrecht, The Netherlands, 9–13 August 2010; pp. 177–182. Available online: https://pdfslide.net/documents/computational-analysis-of-musical-influence-analysis-of-musical-influence-a.html (accessed on 2 September 2021).
23. Collins, N. Influence in Early Electronic Dance Music: An Audio Content Analysis Investigation. In Proceedings of the ISMIR, The 13th International Society for Music Information Retrieval Conference, Porto, Portugal, 8–12 October 2012; pp. 1–6.
24. Shalit, U.; Weinshall, D.; Chechik, G. Modeling musical influence with topic models. In Proceedings of the International Conference on Machine Learning. PMLR, 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16–21 June 2013; pp. 244–252. Available online: http://proceedings.mlr.press/v28/shalit13.html (accessed on 2 September 2021).
25. Blei, D.M.; Lafferty, J.D. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 113–120.
26. Gerrish, S.; Blei, D.M. A language-based approach to measuring scholarly impact. In Proceedings of the ICML, 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010. Available online: http://www.cs.columbia.edu/~blei/papers/GerrishBlei2010.pdf (accessed on 2 September 2021).
27. Xue, W. Modeling Musical Influence through Data. Ph.D. Thesis, Harvard University, Cambridge, MA, USA, 2018.

28. 2021 Mathematical Contestin Modeling: MCM PROBLEM D: The Influence of Music Dataset. Available online: https://www.comap.com/undergraduate/contests/mcm/contests/2021/problems/ (accessed on 2 September 2021).
29. Du Toit, L. *Optimal HP Filtering for South Africa*; Stellenbosch University, Department of Economics, Bureau for Economic Research: Stellenbosch, South Africa, 2008.
30. Jia, C.L.; Xu, W.X.; Wang, F.T.; Wang, H.N. Track irregularity time series analysis and trend forecasting. *Discret. Dyn. Nat. Soc.* **2012**, *2012*, 387857.
31. Saha, S.K.; Ghoshal, S.P.; Kar, R.; Mandal, D. Cat swarm optimization algorithm for optimal linear phase FIR filter design. *ISA Trans.* **2013**, *52*, 781–794. [CrossRef] [PubMed]
32. Ravn, M.O.; Uhlig, H. On adjusting the HP-filter for the frequency of observations. *Rev. Econ. Stat.* **2002**, *84*, 371–376. [CrossRef]
33. Li, W.; Shi, W.H. Three Great Fusions of Country Music. *Audiov. Technol.* **1997**, 161–163. Available online: http://qikan.cqvip.com/Qikan/Article/Detail?id=683847483199710025 (accessed on 2 September 2021).
34. Maymin, P. Music and the market: Song and stock volatility. *N. Am. J. Econ. Financ.* **2012**, *23*, 70–85. [CrossRef]