International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-12

https://doi.org/10.70917/ijcisim-2025-0253
Article

A Study on the Communication Path and Audience Behavior Model of Popular Music Based on Multimodal Data Fusion

Yufeng Wang *

Conservatory of Music, Communication University of China, Nanjing, Jiangsu, 210000, China;
WYF19940315@163.com

Abstract: In the network era, the audience of popular music gradually occupies a dominant position in the breadth and depth of its dissemination. In this paper, we take its propagation path and audience behavior as the research entry point, integrate Bi-RNN recurrent neural network and attention mechanism, and establish a user classification model of popular music based on the integration of Bi-RNN and attention mechanism. User interest label data is introduced for the characteristics of popular music, and a deep model is used for classification, combined with the cross-entropy loss function to obtain the objective function, to build a popular music user interest classification model. The out-degree is selected as the node's low-order topological feature, combined with the cohomology value of the high-order topological feature, to form the evaluation index of the node's propagation ability (out-degree-cohomology index). By exploring the communication path and referring to the framework of Fogg's behavioral model, we set the paying behavior as the dependent variable, and selected a total of 9 independent variables from the three major categories of motivation, ability, and triggering, to construct a model of the influencing factors of the paying behavior of the popular music audience. After correlation and regression analyses, there existed a total of seven variables, namely perceived entertainment, perceived sociality, perceived cost, service quality, perceived ease of use, audition experience and content push, which showed a significant positive correlation with popular music audience behavior ($r>0, P<0.01$).

Keywords: nodal communication capability; popular music audience; out-degree-synchronicity index; influencing factors of paying behavior; Fogg's behavioral model

1. Introduction

As a cultural phenomenon, the development of popular music is closely linked to the progress of social economy, culture, art and science and technology. From performances, phonographs, records, cassettes, MP3s, radio, television to chemical network media, they are all representative media for the dissemination of popular music in each era [1-2]. Each stage of the development of popular music is related to the communication technology, and each change of the communication media provides new ways for popular music to spread [3-4]. Especially today, with the rapid development of network communication, the synergy formed by traditional communication media and network media has opened a new path for the development of popular music, and jointly constructed a new ecological environment of popular music, whose dissemination has shown drastic changes in speed, audience breadth and influence, reflecting the development trend of being more rapid, convenient, diversified and simplified [5-9].

With the support of digital technology, the dissemination form of popular music has evolved into the mode of “music content - multi-channel - audience”, personalized push music on professional music platforms, 15-30 seconds of background music and cover songs in short videos or social media, and different dissemination paths connecting to different audience groups, forming a multimodal music data [10-13]. Due to the differences in characteristics such as cultural background and age stage, audiences may have different understandings of the same music, and the behavioral paths of audience access to music are diversified, and the younger generation of users may access and understand music through multiple music platforms, short video platforms, and social platforms every day [14-17]. Therefore, the use of single-modal data for music analysis can no longer meet the current diversified forms of communication. Multimodal data fusion can quantify music lyrics text, audio melody, listener's comments and social, platform promotion video, cover image and other data and audience behavior, and construct a popular music dissemination path based on audience's interests and behaviors from the user's listening time, interactive behaviors (liking, commenting, and clicking), emotional expressions, and sharing paths.

In 2017, Herremans and Chuan [18] proposed a multimodal interactive platform for semantic music analytics that jointly analyzes music performance videos and scores, and visualizes and displays them in the platform. In 2018, Farnadi et al [19] used deep learning algorithms to extract the multimodal data features of users in social media, and performed data fusion, and in this way accurately constructed user profiles for multimodal analysis of music recommended or covered by users in social media and micro-video platforms. In 2019, Guo et al [20] combined temporal attention filters and a two-branch network structure for visual and audio-based multimodal data fusion of emotions triggered by users' viewing of different videos, and performed emotional video content analysis. 2023 In 2007, Liu et al [21] used meta-learning to create a dynamic multimodal fusion framework that dynamically matches a multimodal fusion function for each micro-video so as to simulate multimodal data associations between different micro-videos for micro-video recommendation.

In 2021, Vaswani et al [22] proposed a multimodal data fusion method based on improved attention neural networks, including data such as listening history, tags, lyrics, and acoustic content, to achieve more accurate song recommendations. In 2022, Zeng et al [23] integrated multimodal feature types of songs using multimodal feature fusion and combined it with a gradient boosting model based on decision tree constructed a popularity prediction model for popular songs. In the same year, Chen and Wu [24] used multi-scale convolutional kernel convolutional neural network and long and short-term memory network models to fuse multimodal signal data of children listening to different music genres, and determined the effects of different music genres on children's emotions and intelligence. 2023, Cui et al [25] fused multimodal data such as music text, image, video, and emotion through a multimodal aggregator, and and fused user comments, interactions, emotional tendencies and other data to form a multimodal knowledge graph network for optimizing the recommendation effect of music platforms. 2024, Tan et al [26] analyzed user preferences and published content characteristics in new media platforms through multimodal data fusion technology to obtain user profiles, personalize recommended information for platform users, and improve user service experience. 2025, Chen et al [27] proposed a playlist-song relationship decomposition model, which effectively determines the user's emotional expression in playlists or songs by recognizing the emotional similarity of music playlists and songs based on multimodal data. Hao et al [28] used feature fusion technology to fuse multimodal emotional data with audio, visual, and physiological parameters, and introduced a bi-directional long and short-term memory network to capture the multi modal data, introduced a bi-directional long short-term memory network to capture the temporal association of multi-modal data, and combined with adaptive sampling technology to realize real-time music emotion recognition.

In this paper, we first design a popular music user classification model based on Bi-RNN and attention mechanism, the idea and main framework of the popular music user interest classification model, focusing on explaining the calculation process of the attention probability distribution and final features. Secondly, it discusses the necessity of low-order topological features and high-order topological features in comprehensive node propagation assessment, and proposes node propagation ability assessment indexes. On this basis, independent variables and dependent variables are selected to construct a model of influencing factors of popular music audience's payment behavior and analyze the operational definitions of each variable. The operational performance of the user interest classification model is examined again by comparing the hit rate performance with that of similar algorithms. Based on the user interest classification model and the node propagation ability assessment method, the propagation path of popular music in online communities is visualized. Finally, the correlation analysis of independent variables and dependent variables of popular music audience's payment behavior is conducted, and a linear equation model is constructed to carry out regression analysis among the variables to test the effectiveness of the propagation path of the influencing factors.

2. User Interest Recognition and Communication Evaluation of Popular Music

2.1. User interest classification model based on Bi-RNN and attention mechanism

LSTM and GRU are actually a special kind of recurrent neural network of RNN, and RNN actually contains LSTM and GRU.The model in this chapter combines the user fusion label information with the recurrent neural network and the attention mechanism to classify the fused popular music features into user interest categories.The popular music user classification model based on the fusion of Bi-RNN and the attention mechanism The structure is shown in Fig. 1.

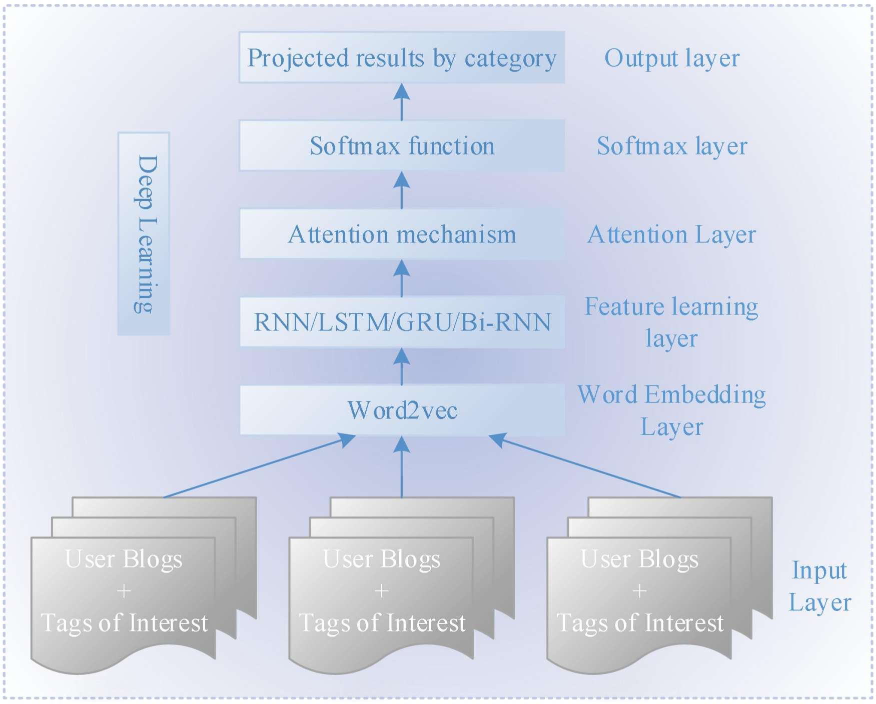

Figure 1. The structure of the popular music user classification model.

The objective function of the blog post classification model based on the fusion of Bi-RNN and attention mechanism is obtained by using softmax as the output layer normalization calculation and combining the cross-entropy loss function is shown in equation (1). The structure of the popular music user interest classification model based on the fusion of Bi-RNN and attention mechanism designed in this chapter is shown in Fig. 2.

$$sig\,\mathrm{mod}(X) = \frac{1}{1+e^{-X}} \tag{1}$$

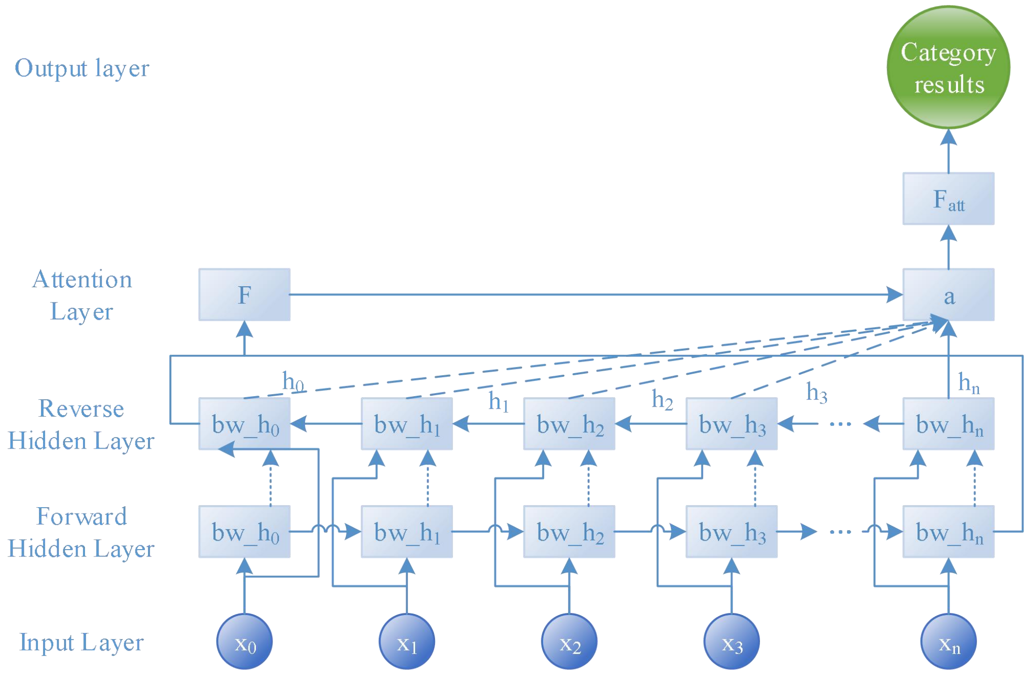

Figure 2. The structure of the popular music user interest classification model.

In Fig. 2, F denotes the sum of the final hidden layer state values in the respective independent directions in the Bi-RNN, which is called the final state of the Bi-RNN, a denotes the attentional probability distribution of the hidden layer unit states for the final state at some moments, where the component a_n denotes the attentional probability of the Bi-RNN state h_n for the final state at the moment n , h_n is obtained by summing the states in their respective independent directions at that moment, and F_{att} denotes the attention-weighted popular music feature vector.

Models based on the attention mechanism generally contain two parts of the computational process, one is the computational process about the probability distribution of attention, and the other is the computational process of the final features based on the attention distribution. In this model, the attention probability of the output data under n moments for the final state is calculated as in Eqs. (2)-(3):

$$a_n = \frac{\exp\left(h_n'\right)}{\sum_{i=1}^{N} \exp\left(h_i'\right)} \tag{2}$$

$$h_n' = h_n^t UF \tag{3}$$

The formula utilizes the Softmax function as the calculation of the probability distribution of attention, where N denotes the number of input sequence elements. u is the weight matrix, F denotes the sum of the final hidden layer state values of the respective independent directions in the Bi-RNN, and h_n denotes the summation of the hidden layer state values of the bi-directional at the moment of n , which is calculated as shown in Eq. (3). In this model, based on the final feature F_{att} of the attention distribution, the computation process is expressed as equation (4):

$$F_{att} = \sum_{i=1}^{N} a_n h_n \tag{4}$$

Where N denotes the number of input sequence elements, a_n denotes the attentional probability of the output data for the final state at the moment of n , and h_n denotes the sum of the hidden layer states under two independent directions at the moment of n . After obtaining the text feature vector F_{att} based on the attention mechanism, the probability distribution of the user's interest category is computed through the softmax function of the output layer, and the computation process is expressed as Eqs.

(5)-(6):

$$y_n = softmax\left(F'_{att}\right) = \frac{\exp\left(F'_{att(i)}\right)}{\sum_{j=1}^{T} \exp\left(F'_{att(j)}\right)} \tag{5}$$

$$F'_{att} = VF_{att} \tag{6}$$

where T is the number of category labels and V denotes the weight matrix of the model output layer. $F'_{att(i)}$ denotes the i th component value in the vector F'_{att}, and the vector length is equal to the number of categorized categories. After classification by softmax function, the probability distribution y of user categories based on the attention mechanism can be obtained, and the cross-entropy loss is sought with the real category distribution Y, which is expressed as equation (7):

$$E(Y, y) = -Y \log(y) \tag{7}$$

where Y denotes the probability distribution of the true category and y denotes the probability distribution of the user's interest category predicted by the model.

2.2. Node dissemination capacity assessment

The higher-order topological features extracted in this paper characterize the strength and frequency of interaction behaviors among multiple nodes through the persistence of various higher-order topological structures, as a way of reflecting the impact of complex contagion caused by frequent node interactions on their propagation capabilities. However, in the process of information dissemination, not all nodes have frequent interactions. Therefore, in order to evaluate the propagation ability of all nodes, in addition to considering the higher-order topological features of nodes, this paper also needs to combine other features for evaluation.

The opposite of high-order topological features are low-order topological features, which quantitatively characterize the relationship between pairs of nodes in the form of proximity and distance. Since simple contagion is one of the most basic propagation mechanisms in information dissemination and it corresponds to the propagation behavior between pairs of nodes, all nodes have low-order topological features. Because of this, most of the network topology-based methods for evaluating the propagation capability of nodes are based on low-order topology. However, low-order topological features cannot describe the interaction behavior of multiple nodes, nor can they characterize the impact of the complex contagion mechanism behind them on node propagation capability. In contrast, higher-order topological features are able to characterize this effect, but they are unable to describe the most fundamental relationships between pairs of nodes in an information dissemination network and cannot characterize the effect of simple contagion.

Since both simple and complex contagion exist in information dissemination, both the relationship of pairs of nodes and the behavior of multi-node interactions are important to assess when evaluating node dissemination capabilities. The former can reflect the influence of nodes on nodes, and the latter can reflect the influence of nodes as part of a group on other nodes and other groups. Therefore, it is necessary to combine low-order topological features with high-order topological features to realize the assessment of comprehensive node propagation capability. For example, Fig. 3 illustrates a simple information dissemination network and the dissemination ability of some of its nodes, setting the node out-degree as its low-order feature and the number of triangular structures near the nodes as its higher-order feature, then it can be seen that, when considering only the low-order feature, node *A* and node *B* have the same dissemination ability, failing to make a distinction through the triangular structure, which may lead to complex contagion; and, when considering only the higher-order feature, node When considering only the high-order features, node *A* and node *C* have the same propagation ability, failing to differentiate by the number of nodes directly connected to them, and thus ignoring the influence of simple contagion; only when combining low-order features and high-order features can we consider the influence of simple and complex contagion at the same time, and differentiate the propagation ability of node *A*, node *B* and node *C*.

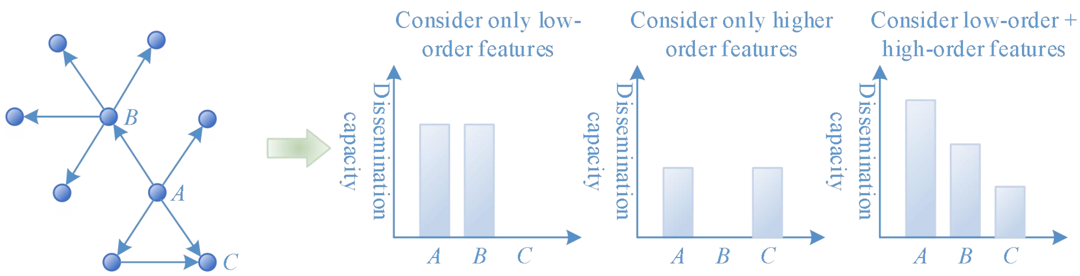

Figure 3. Information dissemination network and the spread capability of some nodes.

Currently, the vast majority of studies based on network topology to evaluate node propagation ability are based on low-order topology, and the extracted node features are all low-order topological features. For the selection of this feature, there have been several classical approaches. For example, degree centrality, H-index and kernel are considered for local topology, and tight centrality and meso-centrality are considered for global topology. The former attaches importance to the role of neighborhood information in the evaluation of node propagation capability, and considers neighborhood information sufficient to characterize most of the node features with lower complexity; the latter considers the role played by nodes in the whole network comprehensively with higher complexity. In the information dissemination network, the out-degree of a node indicates the number of times it has been forwarded, and the use of out-degree can most directly characterize a node's dissemination ability, and the time complexity of obtaining out-degree is also low. Therefore, in this paper, the out-degree is chosen as a low-order topological feature of a node.

Combining the cohomology value that characterizes the high-order topological features of a node and the out-degree that characterizes the low-order topological features of a node, the node propagation ability assessment index can be constructed: out-degree-cohomology index (OPI). The out-degree-homotopy index of node v is calculated as Eqs. (8)-(9):

$$OPI(v) = \frac{od(v) - \min[od(V)]}{\max[od(V)] - \min[od(V)]} \cdot PV(v) \tag{8}$$

$$PV(v) = Sigmoid\left[pv_1^D(v) - \alpha\right] + Sigmoid\left[pv_1^{VR}(v) - \alpha\right] + \beta \tag{9}$$

Where: od is the out-degree of the node; PV is the homotopy value of the node, which combines the homotopy features pv_1^D and pv_1^{VR} computed by the 2 filter functions, respectively; α and β are adjustable parameters, which are both set to 0 in this paper. Considering that the range of values that can be taken by the node out-degree and the cohomology value are usually very different, the maximum-minimum normalization is used for the node out-degree and the nonlinear function Sigmoid is used for the cohomology value of the node in order to adjust the relative sizes of the two to the appropriate intervals.

3. Model of Influencing Factors of Popular Music Audience Payment Behavior

In this paper, by exploring the influencing factors of popular music audience's paying behavior and inverting the popular music audience's behavior, we set the dependent and independent variables in this chapter to construct a model of influencing factors of popular music audience's paying behavior.

3.1. Theoretical Modeling

This study takes Fogg's behavioral model as the theoretical framework, and based on a systematic review of relevant literature in the academic world, combined with the first-hand data obtained from in-depth interviews, and taking into full consideration of the characteristics of the content production, user consumption mode, and product operation mechanism of the knowledge payment industry, this study constructs a theoretical model of user behavior suitable for knowledge payment scenarios. The theoretical model constructed in this study contains nine core variables, including three variables based on the motivation element of Fogg's behavioral model, including perceived usefulness, perceived entertainment, and perceived sociability; three variables based on the ability element, including perceived cost, service quality, and perceived ease of use; and three variables based on the trigger element, including auditioning experience, social influence, and content push. Paying behavior, as a dependent variable, is directly influenced by the first nine independent variables. The theoretical model of influencing factors of popular music paying behavior is shown in Figure 4.

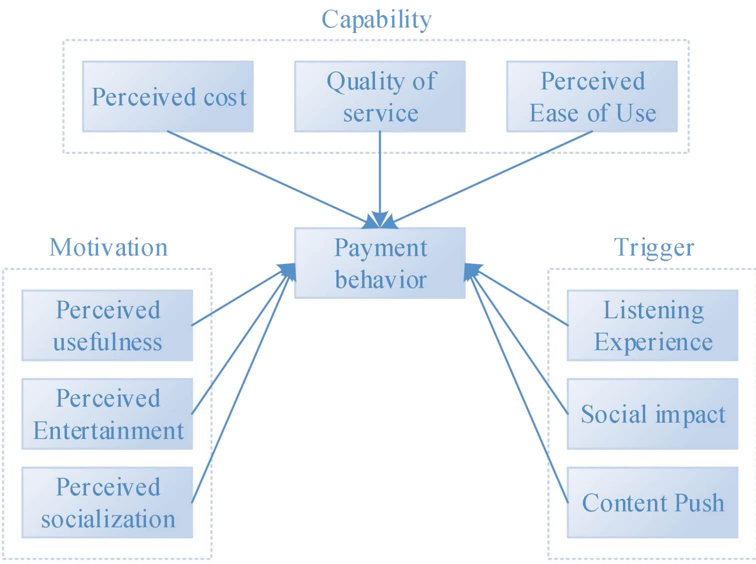

Figure 4. Theoretical model of influencing factors of payment behavior.

3.2. Operational definitions of theoretical model variables

The variables of the model are operationally defined as follows:

(1) Motivation

1) Perceived usefulness. Users perceive paid music as more satisfying for enjoying music or improving some skill compared to free products.

2) Perceived Entertainment. The subjective evaluation of the user's pleasure and engagement in the process of using paid music.

3) Perceived socialization. Users feel the information exchange and interaction between individuals and singers, and between individuals and singers when using paid music.

(2) Ability

1) Perceived cost. When using paid music, users comprehensively assess the multiple costs to be invested, covering time costs, economic costs and energy costs.

2) Service quality. The quality of services provided by paid music, including user privacy protection, timeliness of customer service response, etc.

3) Perceived ease of use. Users' subjective evaluation of the ease of operation of paid music.

(3) Trigger

1) Trial listening experience. The user's perceived experience and evaluation when auditioning music, including but not limited to sound quality, reading content, interface design, ease of use and other aspects of the experience.

2) Social impact. The extent to which users are influenced by the suggestions, recommendations or behaviors of people around them (e.g., friends, family, colleagues) when using and paying for music.

3) Content Push. The platform will carry out moderate information push, the push content not only includes the information of the music itself and promotional activities, but also will cover the content shared by the members to be liked, commented on or retweeted by other users when the alerts.

(4) Paying behavior: users are willing to pay to listen to popular music.

4. Analysis of Popular Music's Communication Path and Audience Behavior

4.1. Performance test of user interest classification model

In this paper, the hit rate (HR) index was selected to evaluate the operational effect of the proposed popular music user interest classification model. TF-IDF and TextRank were used as the control groups, and HR@5 and HR@10 experiments were conducted on ten groups of data (numbered from G1 to 10 in sequence). The results are shown in Table 1. The average HR@5 value of the model algorithm in this paper on ten sets of data is 0.772, and the average HR@10 value is 0.879. Not only does it far exceed the other two similar algorithms, but it also remains stable above 0.700.

Table 1. The HR value performance of the algorithm.

	HR@5			HR@10		
	Textual	TextRank	TF-IDF	Textual	TextRank	TF-IDF
G1	0.872	0.580	0.576	0.985	0.629	0.587
G2	0.709	0.485	0.493	0.851	0.545	0.523
G3	0.790	0.495	0.483	0.919	0.559	0.509
G4	0.821	0.602	0.568	0.921	0.639	0.605
G5	0.772	0.483	0.508	0.873	0.545	0.509
G6	0.839	0.485	0.491	0.927	0.557	0.495
G7	0.866	0.591	0.577	0.925	0.631	0.591
G8	0.663	0.372	0.398	0.783	0.447	0.419
G9	0.619	0.405	0.436	0.757	0.473	0.459
G10	0.773	0.492	0.478	0.845	0.553	0.521
Average	0.772	0.499	0.501	0.879	0.558	0.522
Variance	0.0059	0.0046	0.0031	0.0069	0.0058	0.0051

The HR@5 values of the proposed model are shown in Fig. 5 as the parameters change, and the proposed model can better track the changes in the individual user's preferences. When α is 1.0, the model goes in to consider the user's short-term interest, and when it is 0, the model considers only the long-term interest. Also from Fig. 5, it can be seen that the HR value has a higher value at αò [0.1,0.9] than when α<0.1 or α>0.9, i.e., it is more capable of obtaining the user's real interest words when both short-term interest and long-term interest are considered. It is further inferred that when γ=0.1 has the best HR value.

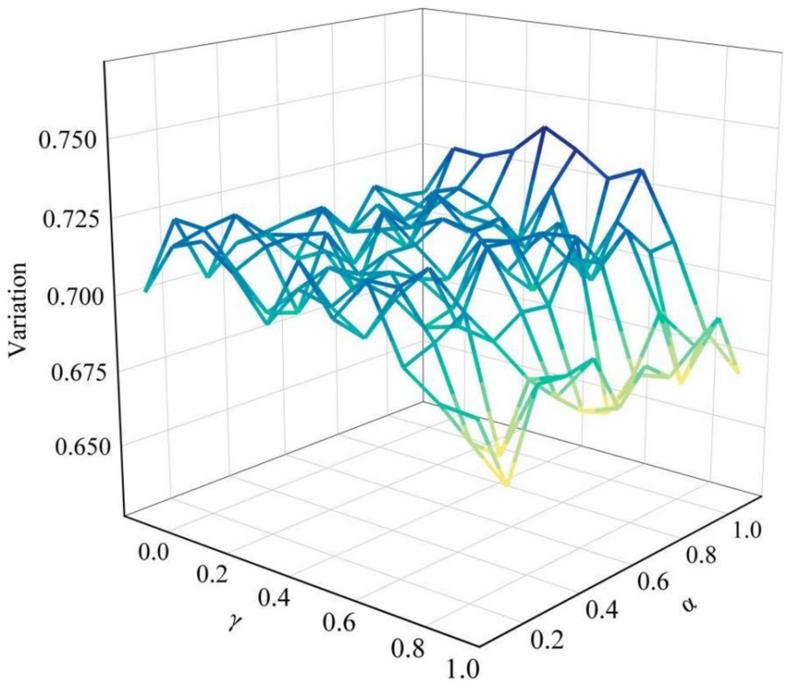

Figure 5. The HR@5 value of the model varies with the parameters.

4.2. Visualization of popular music communication networks

4.2.1. Overall communication network presentation

The node communication ability assessment index is applied to analyze in the popular music public communication network community to obtain the structural characteristics of the composition of this community. Its overall community network is shown in Figure 6, where nodes of the same color, except for the bright yellow color, indicate that they belong to the same online community.

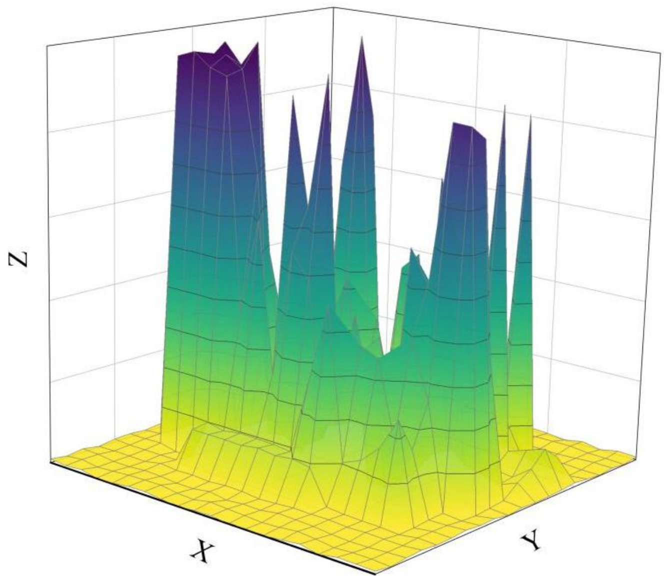

Figure 6. Community network division.

4.2.2. Reorganized communities

The overall community network with delineated structure is reorganized to obtain a total of 50 community networks (numbered 1-50), but retaining the interaction boundaries between their community networks. The new community networks are obtained by sorting them in size according to the number of community network audiences as shown in Fig. 7. Among them, the pop music community networks numbered 1-39 have an audience size between [1,100], which soars from number 40, ranging from 1000-5000.

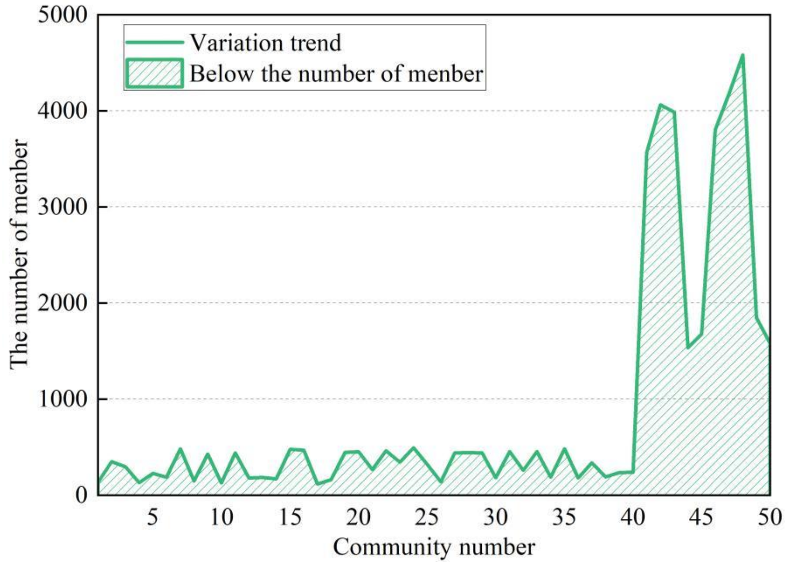

Figure 7. The number of community network members.

Each node in the reorganized popular music network is a community, and the boundaries between each community and other communities are retained and weights are recalculated to obtain the reorganized network and the corresponding number of community members in Fig. 8. The size of the nodes represents the number of members of the community, and the distance between the nodes is the size of the weights between the nodes and the other networks, which is the propagation path and propagation depth of the popular music among the community networks. Depth. When the number of community members is large, the difference in distance from other nodes is small, i.e., the weights are equal to the boundaries of other community networks.

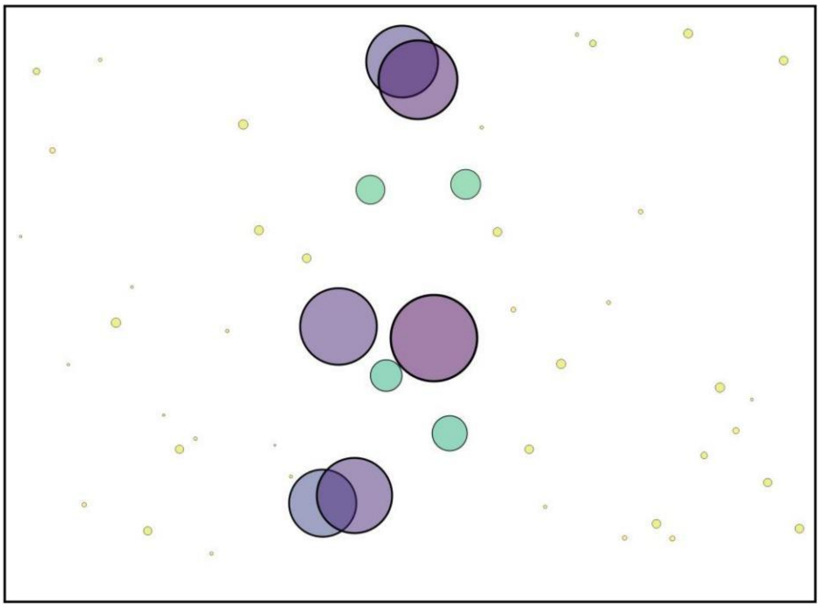

Figure 8. Reorganize the number of network and community members.

4.3. Validity Analysis of Factors Influencing Payment Behavior

4.3.1. Correlation analysis

Using Pearson's correlation analysis, the degree of correlation between a total of 10 variables, namely (I11) perceived usefulness, (I12) perceived entertainment, (I13) perceived sociability, (I21) perceived cost, (I22) quality of service, (I23) perceived ease of use, (I31) auditioning experience, (I32) social influence, (I33) content push, and (D1) paying behavior, was analyzed. Whether the nature of the relationship between the variables is positive or negative is reflected by the positive or negative correlation coefficient. The definition of the degree of their correlation influence between very weak, weak, moderate, strong, and very strong depends on the fall point of the absolute value, corresponding to the fall point intervals of [0,0.2], [0.2,0.4], [0.4,0.6], [0.6,0.8], and [0.8,1.0], in that order. The correlation analysis between the variables is shown in Table 2.

Table 2. Correlation analysis among various variables.

	I11	I12	I13	I21	I22	I23	I31	I32	I33	D1
I11	1									
I12	0.224 ***	1								
I13	0.406 ***	0.519 ***	1							
I21	0.731 ***	0.563 ***	0.741 ***	1						
I22	0.504 ***	0.624 ***	0.724 ***	0.481 ***	1					
I23	0.596 ***	0.746 ***	0.355 ***	0.312 ***	0.777 ***	1				
I31	0.275 ***	0.374 ***	0.283 ***	0.645 ***	0.289 ***	0.742 ***	1			
I32	0.471 ***	0.554 ***	0.257 ***	0.598 ***	0.217 ***	0.289 ***	0.616 ***	1		
I33	0.455 ***	0.259 ***	0.436 ***	0.592 ***	0.252 ***	0.706 ***	0.236 ***	0.579 ***	1	
D1	0.733 ***	0.685 ***	0.762 ***	0.631 ***	0.434 ***	0.674 ***	0.712 ***	0.661 ***	0.556 ***	1

Note:***, **, and * represent significance levels of 1%, 5%, and 10% respectively

***,the same below.

Overall, all the independent variables showed a significant effect relationship with the dependent variables ($p<0.01$) and all the variables showed a positive correlation ($r>0.000$). The correlation coefficients between (I11) perceived usefulness, (I12) perceived entertainment, (I13) perceived sociality, (I21) perceived cost, (I22) service quality, (I23) perceived ease of use, (I31) audition experience, (I32) social influence, and (I33) content pushing, a total of 10 variables, and (D1) paying behavior were 0.733, 0.685, 0.762, 0.631, 0.674, 0.712, 0.661, and 0.556.

4.3.2. Regression analysis

From the content of the analysis in the previous subsection, it is known that there is a linear relationship between all the variables to varying degrees. In this subsection, linear regression equations are constructed for them, and the predictor variables that have an impact on the outcome variables are explored through regression analysis. The structural relationship between the variables and the results of their standardized path coefficient tests are shown in Table 3.

Table 3. The regression results of the model.

Relation	Standardized path coefficient	*S.E.*	*S.R.*	*P*
I11→D1	-0.029	0.051	-0.018	0.513
I12→D1	0.235	0.048	4.112	0.000***
I13→D1	0.272	0.051	4.598	0.000***
I21→D1	0.162	0.043	1.148	0.0024***
I22→D1	0.135	0.057	1.689	0.0079***
I23→D1	0.156	0.052	1.733	0.0054***
I31→D1	0.132	0.060	1.383	0.0073***
I32→D1	0.097	0.054	0.072	0.362
I33→D1	0.163	0.053	1.432	0.0021***

There are (I11) perceived usefulness and (I32) social influence totaling 2 independent variables on (D1) paying behavior with significance $P>0.1$, did not show statistically significant difference, i.e., the path of perceived usefulness and social influence to paying behavior is invalid.

(I12) Perceived Entertainment to (D1) Payment Behavior Path r=0.235, P=0.000***, (I13) Perceived Sociality to (D1) Payment Behavior Path r=0.272, P=0.000***, (I21) Perceived Cost to (D1) Payment Behavior Path r=0.162, P=0.0024***, and (I22) Service Quality to (D1) paid behavior path r=0.135, P=0.0079***, (I23) perceived ease of use to (D1) paid behavior path r=0.156, P=0.0054***, (I31) audition experience to (D1) paid behavior path r=0.132, P=0.0073***, (I33) content push to (D1) paid behavior path r=0.163, P=0.0021***. The effects of the above seven variables on paid behavior all reached the level of statistical significance, and thus were able to reject the original hypothesis and confirm the validity of their path to paid behavior.

5. Conclusion

In this paper, we propose a user interest classification model based on Bi-RNN and attention mechanism for interest recognition and classification of popular music users under multimodal data, and design an out-degree-homotopy index to evaluate the spreading depth of popular music in multiple online communities. Among them, the user interest classification model has a more accurate hit rate compared with similar models. In ten sets of experimental data, the average HR@5 value is 0.772 and the average HR@10 value is 0.879.

Taking audience payment behavior as the dependent variable and a total of 3 dimensions from motivation, ability, and trigger as the independent variables, we constructed a model of influencing factors of popular music audience payment behavior. All the independent variables showed significant influence relationship with the dependent variables ($P<0.01$), and all the variables showed positive correlation ($r>0.000$). Among them, perceived entertainment, perceived sociality, perceived cost, service quality, perceived ease of use, audition experience and content push, a total of seven variables are effective paths for popular music to guide audience's paying behavior (P<0.01***).

References

1. Dubois, L. E., Renard, S., & Guttentag, D. (2025). Towards monopolistic music promotion: an analysis of North American concert tours. Cultural Trends, 34(3), 431-448.
2. Bennett, A. (2017). Popular music, media and the narrativization of place. In Sonic Synergies: Music, Technology, Community, Identity (pp. 69-78). Routledge.

3. Shin, S. (2017). Niche, ethnic and global operations: Models of production and circulation of East Asian popular music. European Journal of East Asian Studies, 16(1), 5-35.
4. Verboord, M., & Brandellero, A. (2018). The globalization of popular music, 1960-2010: A multilevel analysis of music flows. Communication Research, 45(4), 603-627.
5. Cwynar, C. (2020). Brick, Mortar, and Screen: Networked Digital Media, Popular Music, and the Reinvention of the Public Radio Station. Journal of Radio & Audio Media, 27(1), 74-92.
6. Shukri, A. A., Zainal, C. N. A. S. C., & Bakar, A. S. A. (2025). Exploring the Digital Wave: Social Media Strategies in K-Pop Influence on Fan Engagement and Artist Promotion. Pena International Journal of Media, Journalism and Mass Communication, 1(1), 36-42.
7. Terroso-Saenz, F., Soto, J., & Muñoz, A. (2023). Music mobility patterns: How songs propagate around the world through spotify. Pattern Recognition, 143, 109807.
8. Wang, S. (2018). Music, social media and public pedagogy: Indie music in the post-Cantopop epoch. Asian Education and Development Studies, 7(1), 42-52.
9. Hidayah, A. R., & Retpitasari, E. (2023). Cultural Studies and Media Ecology: Identifying Mass Culture and Cultural Proliferation through Javanese Pop Music on Social Media. Jurnal Spektrum Komunikasi, 11(3), 349-361.
10. Chen, M., & Tai, K. H. (2025). The Dissemination of Social Media and Pop Music: Influence and Trends. Metaverse Basic and Applied Research, 4, 3.
11. Zhou, C., Zhao, H., Ma, X., & Zhang, M. (2025, May). User Experience and Retention in Chinese Music Streaming: Insights from QQ Music and NetEase Cloud Music. In International Conference on Human-Computer Interaction (pp. 306-323). Cham: Springer Nature Switzerland.
12. Radovanović, B. S. (2022). TikTok and sound: Changing the ways of creating, promoting, distributing and listening to music. INSAM Journal of Contemporary Music, Art and Technology, (9), 51-73.
13. Chen, J., He, L., Liu, H., Yang, Y., & Bi, X. (2024). Background music recommendation on short video sharing platforms. Information Systems Research, 35(4), 1890-1908.
14. Hasan, G., & Scorpianti, E. (2022). The influence of usefulness, entertainment, interaction, enjoyment, and familiarity to purchase music platform subscriptions in Millennials and Gen-z communities through trust mediation. Jurnal Manajemen Dan Pemasaran Jasa, 15(2), 161-176.
15. Donze, P. L. (2017). Gender and popular culture: a comparison of promoter and listener preferences for popular music artists. Sociological perspectives, 60(2), 338-354.
16. Grebosz-Haring, K., & Weichbold, M. (2020). Contemporary art music and its audiences: Age, gender, and social class profile. Musicae Scientiae, 24(1), 60-77.
17. Yoon, K. (2018). Global imagination of K-pop: Pop music fans' lived experiences of cultural hybridity. Popular Music and Society, 41(4), 373-389.
18. Herremans, D., & Chuan, C. H. (2017, January). A multi-modal platform for semantic music analysis: visualizing audio-and score-based tension. In 2017 IEEE 11th International Conference on Semantic Computing (ICSC) (pp. 419-426). IEEE.
19. Farnadi, G., Tang, J., De Cock, M., & Moens, M. F. (2018, February). User profiling through deep multimodal fusion. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 171-179).
20. Guo, J., Song, B., Zhang, P., Ma, M., & Luo, W. (2019). Affective video content analysis based on multimodal data fusion in heterogeneous networks. Information Fusion, 51, 224-232.
21. Liu, H., Wei, Y., Liu, F., Wang, W., Nie, L., & Chua, T. S. (2023). Dynamic multimodal fusion via meta-learning towards micro-video recommendation. ACM Transactions on Information Systems, 42(2), 1-26.
22. Vaswani, K., Agrawal, Y., & Alluri, V. (2021, November). Multimodal fusion based attentive networks for sequential music recommendation. In 2021 IEEE seventh international conference on multimedia big data (BigMM) (pp. 25-32). IEEE.
23. Zeng, H., Yuan, Q., Guo, L., & Xu, S. (2022, November). Song popularity prediction model based on multi-modal feature fusion and LightGBM. In Proceedings of the 8th International Conference on Communication and Information Processing (pp. 28-32).
24. Chen, W., & Wu, G. (2022). A multimodal convolutional neural network model for the analysis of music genre on children's emotions influence intelligence. Computational Intelligence and Neuroscience, 2022(1), 5611456.
25. Cui, X., Qu, X., Li, D., Yang, Y., Li, Y., & Zhang, X. (2023). Mkgcn: multi-modal knowledge graph convolutional network for music recommender systems. Electronics, 12(12), 2688.
26. Tan, X., Wu, D., & Li, W. (2024, October). Application of Multimodal Data Fusion in Intelligent Content Recommendation System for New Media Platforms. In 2024 First International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-5). IEEE.
27. Chen, Y., Yuan, H., Ma, B., Wang, L., & Qian, Y. (2025). Beyond songs: Analyzing user sentiment through music playlists and multimodal data. ACM Transactions on Multimedia Computing, Communications and Applications, 21(3), 1-24.
28. Hao, X., Li, H., & Wen, Y. (2025). Real-time music emotion recognition based on multimodal fusion. Alexandria Engineering Journal, 116, 586-600.