

Combining Social Network Analysis to Explore Rural Textile Consumer Preferences and Communication Paths

Jiru Ren

Department of Economics and Management, Taiyuan Institute of Technology, Taiyuan 030008, Shanxi, China

Corresponding author: Jiru Ren (e-mail: renjr2025@163.com)

ABSTRACT In rural areas, textile purchase decisions are strongly influenced by neighborhood reputation, cultural identity, and regional aesthetics and exhibit significant social communication characteristics. Traditional market research methods struggle to capture the dynamic evolution of preferences in rural textile consumption scenarios, and existing social network models often lack the multidimensional integrated modeling capabilities. This paper proposed a multidimensional modeling method that integrates social network analysis, natural language processing and a graph neural network. On the basis of the user behavior logs of rural e-commerce platforms, a weighted social network (3,685 nodes, 19,243 edges) was constructed, and seven consumer groups with geographical affiliation characteristics were identified by combining the improved Bidirectional Encoder Representations from Transformers (BERT) model and the Louvain algorithm. The GAT model was used to integrate historical purchase and sentiment features to achieve preference trend prediction at the community level, and the improved SIR model was introduced to simulate the negative word-of-mouth communication mechanism by integrating node influence and trust parameters. The experimental results show that the model constructed in this paper performs well in terms of preference classification, community division and communication path prediction. The preference emotion classification accuracy is 92.4%, and the modularity index is 0.43. This study provides a theoretical basis and practical support for the modeling of rural market consumption preferences and precise intervention strategies based on community structure.

INDEX TERMS rural textile market, social network analysis, consumer preference evolution, graph neural network, propagation path modeling

I. INTRODUCTION

WITH the rapid development of rural e-commerce and the increasing popularity of social media, the shopping behavior of rural consumers is increasingly influenced by online interactions and word-of-mouth communication [1-3]. Traditional market research methods struggle to capture

consumption preference patterns amidst this dynamic evolution [4-6], whereas user behavior modeling based on social networks provides a new perspective for understanding the consumption psychology of rural groups [7,8]. In the rural market, textiles are daily necessities and carry certain cultural functions. Their consumption behavior shows strong regional and community dependence. In recent years, the rapid development of rural e-commerce has provided a new context for research on textile consumption [9,10]. However, in the field of rural textile consumption, there is still a lack of systematic research on integrating user sentiment, community structure, and information dissemination paths. On the one hand, existing text classification models have

insufficient semantic understanding when dealing with rural review data with regional and colloquial characteristics [9,10]; on the other hand, most social network analyses ignore the impact of geographical proximity and the trust chain structure on consumer preferences [11,12]. In recent years, social network analysis technology has been widely used in consumer behavior modeling [13-15], but it still has certain defects. Sattar et al. [16] proposed a scalable distributed Louvain algorithm to improve the efficiency of community detection in large-scale graphs. Wang et al. [17] applied the improved Louvain algorithm in the division of healthcare service areas and verified its effectiveness in spatially constrained scenarios. Rostami et al. [18] evaluated the applicability of the Louvain algorithm in the medical field through a systematic review. In addition, pretrained language models such as Bidirectional Encoder Representations from Transformers (BERT) show good generalization ability in sentiment analysis tasks, but their computational cost is high, and they are prone to overfitting on small datasets. Mao et al. [19] conducted an empirical study on the bias of

the BERT model and noted its limitations in prompt-based sentiment analysis. Kit et al. [20] explored the application of the BERT model without fine-tuning to reduce resource consumption. Deepa et al. [21] evaluated the performance of the BERT model in specific sentiment analysis tasks, whereas models such as the Graph Neural Networks (GNNs) and Graph Attention Networks (GATs) performed well in predicting node-level preference trends but still had limitations. Zhao et al. [22] proposed a method for predicting agricultural product consumption behavior on the basis of GNNs, but did not fully optimize it by combining the trust chain structure and time decay effect in rural scenarios. Jain et al. [23] used GNNs to simulate the information diffusion mechanism, providing a reference for propagation path modeling, but did not consider differences in user influence and changes in trust, limiting its application effect in word-of-mouth control. Chen et al. [24] expanded the application boundaries of deep learning models in sentiment analysis, especially in fine-grained sentiment recognition. However, GNNs have not yet been combined with social network structures, making it challenging to capture the interactive impact between users. Razaque et al. [25] reported that the SIR (Susceptible-Infected-Recovered) model needs to

address more complex dependencies when simulating social network communication, resulting in reduced computational efficiency.

This paper proposes a multidimensional modeling method that integrates social network analysis, natural language processing, and GNNs, aiming to reveal the evolution of rural textile consumer preferences and their manifestations in localized communication paths. The research focuses on two core issues that are closely related to precision marketing and policy guidance in rural markets [26-28]: consumer group identification and information dissemination simulation. First, an undirected weighted social network graph is constructed on the basis of user behavior logs, and an improved Louvain algorithm is used to identify consumer groups with geographical affiliation. Second, an improved BERT model is used to extract sentiment polarity from comment texts and merge it with purchase behavior to generate community-level preference scores. A GAT model was subsequently constructed to predict the future preference trends of each community in the future, and the improved SIR (Susceptible-Infected-Recovered) model was subsequently introduced to simulate the information propagation path. The key node intervention experiment was conducted by combining node influence and user trust. The experiments indicate that the proposed method is superior to the mainstream model in terms of preference classification, trend prediction and propagation path modeling. The research results provide a new technical path for modeling consumption preferences in rural markets and provide a theoretical basis and practical support for formulating precise intervention strategies on the basis of community structure.

II. ALGORITHM DESIGN

SOCIAL NETWORK CONSTRUCTION

To reveal the interactive relationship between rural textile consumers and support subsequent preference identification and propagation path modeling, this study constructed an undirected weighted social network based on the rural textile consumer behavior log. To ensure compliance with user privacy regulations, this study adhered to the requirements of China's Personal Information Protection Law and the European Union's General Data Protection Regulation (GDPR) by implementing rigorous deidentification procedures on the original user behavior logs. Direct identifiers such as full names, mobile phone numbers, and ID numbers were anonymized through one-way hash encryption to prevent traceability. Sensitive fields unrelated to geographic locations, including home addresses and IP addresses, were systematically removed to further protect user identity. In addition, textual interaction data underwent entity disambiguation processing to eliminate Personally Identifiable Information (PII), such as specific personal names, shop names,

and other identifiable entities mentioned in user comments. A multilevel data access authorization framework was also established to restrict access to deidentified datasets exclusively to core research personnel. These measures effectively severed direct identification links within the original data while preserving sufficient behavioral features critical for subsequent modeling and analysis. The deidentification process maintained data utility for social network construction, sentiment analysis, and graph neural network-based preference modeling, ensuring both regulatory compliance and analytical validity. The experimental data come from anonymized user logs provided by a large e-commerce platform, which contains the purchases, comments, likes, and forwards of 3,685 users (covering 12 administrative villages). The fields include user ID, product category (textiles are divided into three categories: cotton, synthetic fiber, and handmade fabrics), comment text, rating stars, interaction objects and timestamps.

First, the original behavior data are cleaned and preprocessed to remove outliers and duplicate records, and each interactive behavior is mapped to a specific rural textile consumer pair (u_i, u_j) . The edge weight w_{ij} between rural textile consumers is defined as the product of the interaction frequency between the two during the observation period (set to nearly 90 days) and the time decay factor, and the formula is as follows:

$$w_{ij} = \sum_{t=1}^T \frac{f_{ij}(t)}{\log(t_0 - t + 1)} \quad (1)$$

In formula (1), $f_{ij}(t)$ represents the number of interactions between users u_i and u_j on day t , t_0 is the current date, and the log function is used to control the time decay effect to ensure that recent interactions have a higher weight.

Each rural textile consumer is regarded as a node $v \in V$ in the graph. If there is at least one effective interaction between two rural textile consumers, an undirected edge $e_{ij} \in E$ is established, and the edge weight is calculated via formula (1). The final generated graph is denoted as $G = (V, E, W)$, where W is the adjacency matrix, which represents the relationship strength between nodes. The social network architecture of rural textile consumers is shown in Figure 1.

Figure 1 shows the spatial distribution of nodes corresponding to the original administrative villages with 12 labeled groups. After optimization via the Louvain algorithm, 7 consumer communities (C1–C7) are identified as the final clusters. The geographic distribution of these communities demonstrates cross-administrative village aggregation characteristics.

SENTIMENT ANALYSIS MODULE BASED ON THE BERT MODEL

This study uses Natural Language Processing (NLP) with a BERT-based sentiment analysis module to evaluate rural textile consumer reviews on e-commerce platforms. The data consisted of 3,542 valid reviews, covering categories such as cotton, synthetic fiber, and handmade fabrics. After the data were split into training (2,479), validation (708), and test sets (355), the samples were stratified by user activity and community to ensure balanced representation. Sentiment evaluation is a three-class problem, namely, positive, neutral, and negative, on the basis of manual annotation and an automated scoring system, with ratings ≥ 4 as "positive," 3 as "neutral," and ≤ 2 as "negative."

The BERT model captures contextual semantic information through a bidirectional transformer encoder structure. Suppose the input text sequence is $x = (x_1, x_2, \dots, x_T)$, where T represents the sequence length and x_t represents the t th token. After the BERT model embeds it, the following can be obtained:

$$h_t^{(0)} = W_e x_t + W_p p_t \quad (2)$$

In formula (2), $W_e \in \mathbb{R}^{d \times V}$ is the word vector matrix, V is the vocabulary size, $W_p \in \mathbb{R}^{d \times L}$ is the position encoding matrix, L is the maximum sequence length, and d is the embedding dimension. Then, the semantic features are extracted layer by layer through the multilayer transformer encoder.

The model training uses the cross-entropy loss function as the optimization target:

$$\mathcal{L} = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (3)$$

N is the batch size, y_i is the true label, and $\hat{y}_i$ is the model prediction probability.

To verify the effectiveness of this module, the proposed BERT model was compared with mainstream text-

classification models [29,30], such as TextCNN (Text Convolutional Neural Network), BiLSTM (Bidirectional Long Short-Term Memory), XGBoost+TF-IDF (Extreme Gradient Boosting + Term Frequency-Inverse Document Frequency) and Support Vector Machine (SVM) + Word2Vec. All the models were trained and tested via the same preprocessing process. For model evaluation, commonly used core indicators such as accuracy, precision, recall, F1 score (F1) and AUC value (AUC-ROC) were selected for analysis, and their calculation formulas were as follows:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

$$\text{Pre} = \frac{TP}{TP + FP}, \quad (5)$$

$$\text{Rec} = \frac{TP}{TP + FN}, \quad (6)$$

$$F1 = 2 \cdot \frac{\text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}}, \quad (7)$$

$$\text{AUC}_{\text{multi}} = \frac{1}{C} \sum_{i=1}^C \text{AUC}_i. \quad (8)$$

To further explore the preference characteristics of user comments, this study combines the TF-IDF algorithm to extract keywords from all comment texts. Specifically, first, the frequency of each word in all the documents is counted, and its IDF value is calculated. The formula is as follows:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log \left(\frac{N}{|\{d' \in D : t \in d'\}|} \right) \quad (9)$$

t represents a word, d represents a document, and N represents the total number of documents. On the basis of the TF-IDF score sorting, the top 50 keywords are extracted as high-frequency preferred words, and they are counted separately by sentiment category to form three types of vocabulary sets: positive/neutral/negative.

LOUVAIN ALGORITHM IDENTIFIES CONSUMER GROUPS

To identify user groups with similar interaction patterns and consumption preferences from the social network structure, this study uses the Louvain algorithm to perform multilevel clustering on the undirected weighted graph $G = (V, E, W)$ constructed in Section 2.1. The graph contains 3,685 nodes (users) and an edge weight matrix calculated based on behavioral data to reflect the interaction intensity between users. The overall architecture of the Louvain algorithm is shown in

The Louvain algorithm maximizes modularity through iterative local optimization. The specific steps are as follows: First, each node is iteratively moved to the community of one of its neighbors. The corresponding modularity gain (ΔQ) is then calculated. If $\Delta Q > 0$, the community affiliation of the node is updated; then, the nodes in the same community are merged into supernodes, a new weighted graph is constructed, and the above process is repeated until

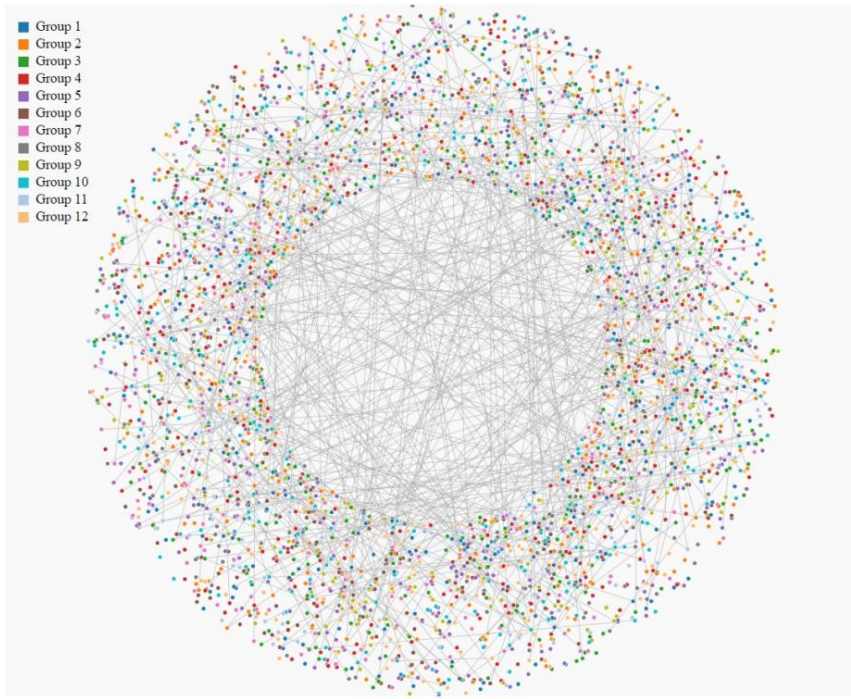


FIGURE 1. Rural textile consumer social network architecture

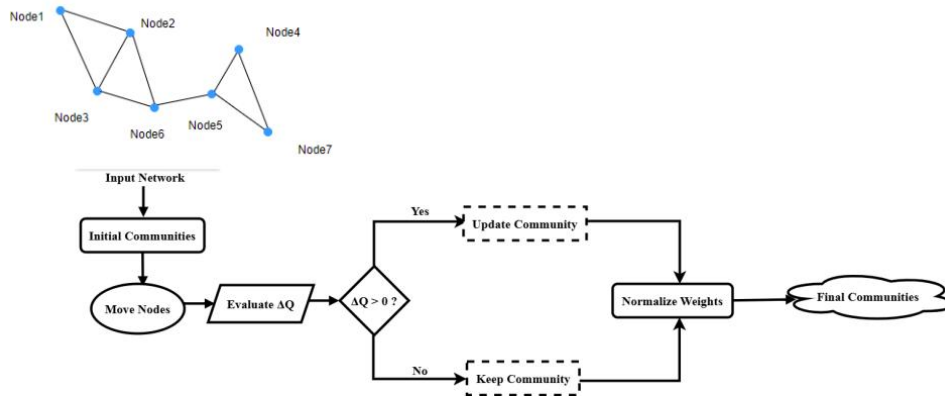


FIGURE 2. Overall architecture of the Louvain algorithm

the modularity cannot be further improved. After each node movement, the change in modularity can be expressed as:

$$\Delta Q = \frac{1}{2m} \left[\sum_j w_{ij} - \sigma_{in}(c_i) \cdot \frac{\sum_j w_{ij}}{m} \right] \quad (10)$$

In formula (10), $\sigma_{in}(c_i)$ represents the total edge weight within the current community c_i , and $\sum_j w_{ij}$ represents the total connection strength of node i .

Owing to the significant differences in the interaction frequency between users in the original data, some highly active users may dominate the community assignments, resulting in an excessive local aggregation effect, affecting the stability and interpretability of the overall community detection.

To address the issue of community partitioning dominated by highly active users, this study proposes a relative density-

based edge weight normalization method. Specifically, the raw edge weights w_{ij} are normalized through the following formulation:

$$w_{ij}^{\text{normalized}} = \frac{w_{ij}}{\sum_{k \in \mathcal{N}(i)} w_{ik}} \quad (11)$$

In formula (11), $\mathcal{N}(i)$ denotes the adjacent node set of node i . This approach transforms each edge weight into the relative interaction intensity between node i and its neighbor j , thereby mitigating the disproportionate influence of absolute weight differences on community partitioning. After normalization, the sum of adjacent edge weights for all nodes equals 1, effectively balancing perturbations caused by locally dominant high-activity users.

To comprehensively evaluate the quality and consistency of the Louvain algorithm's segmentation results in rural consumer social networks, this paper uses the following five

core evaluation indicators: modularity Q , normalized mutual information (NMI), adjusted Rand index (ARI), purity, and the homogeneity score. The calculation formulas are as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[w_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (12)$$

In formula (12), w_{ij} represents the edge weight between nodes i and j ; c_i represents the community to which the node belongs; and $\delta(c_i, c_j)$ is the Kronecker function, which takes a value of 1 when $c_i = c_j$ and 0 otherwise. The value range of this indicator is usually $[-0.5, 1]$. The larger the value is, the clearer the community structure and the tighter the internal connection.

$$\text{NMI}(X, Y) = \frac{\text{MI}(X, Y)}{\sqrt{H(X)H(Y)}} \quad (13)$$

In formula (13), $\text{MI}(X, Y)$ is the mutual information, which reflects the correlation between two clustering results; $H(X)$ and $H(Y)$ are the information entropies of the clustering results X and Y , respectively.

$$\text{ARI} = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}} \quad (14)$$

In formula (14), n_{ij} represents the number of samples that belong to both the i th class in clustering result X and the j th class in clustering result Y ; a_i and b_j are the numbers of samples of the corresponding classes; and n is the total number of samples. The ARI value range is $[-1, 1]$. The closer the value is to 1, the more consistent the two clustering results are.

$$\text{Purity} = \frac{1}{n} \sum_{k=1}^K \max_l |C_k \cap L_l| \quad (15)$$

In formula (15), C_k represents the k th community, L_l represents the l th true category, n is the total number of samples, and K is the total number of communities. The closer the indicator value is to 1, the purer the community classification is.

$$h = 1 - \frac{H(C|L)}{H(C)} \quad (16)$$

$H(C|L)$ is the conditional entropy of the community distribution given the true category, and $H(C)$ is the entropy of the community itself. The homogeneity score ranges from $[0, 1]$, and the higher the value is, the more consistent the community division is with the true category distribution.

During the experiment, the results of the improved Louvain algorithm in this paper were compared with those of the following three mainstream community detection algorithms: the Label Propagation Algorithm (LPA), Infomap, and spectral clustering+K-means. The performance comparison of each detection algorithm is shown in Table 1.

As shown in Table 1, the Louvain algorithm maximizes modularity without requiring preset community numbers. It

has medium time complexity and high stability and is ideal for undirected weighted graphs. The LPA algorithm is the fastest but has low stability because of its random label propagation. Infomap, based on information compression, is better for directed graphs or flow data but is less effective for undirected weighted graphs. Spectral clustering with K-means offers high stability but requires predefined community numbers and has high computational complexity, limiting its adaptability to large-scale rural social networks.

PREDICTING PREFERENCE TRENDS VIA THE GAT MODEL

To model the preference evolution of rural textile consumer groups under the social network structure, this study constructed a preference trend prediction model for community nodes based

on GAT. The model uses the seven user communities identified by the Louvain algorithm as the basic prediction unit and uses their historical behavior and social relationship information to predict the changes in the preference intensity of each community for a specific textile category in the future.

The input of this model consists of three parts: the first part is the node feature matrix $X \in \mathbb{R}^{N \times d}$, where $N = 7$ represents the number of communities, and $d = 5$ represents the feature dimensions extracted from each community, including the average purchase frequency, positive comment ratio, negative comment ratio, interactive activity index, and TF-IDF keyword similarity score of the community users in the past three months; the second part is the adjacency matrix $A \in \{0, 1\}^{N \times N}$, which represents the connection relationship between communities. If there is at least one highly interactive user pair between two communities, they are considered adjacent. The third part is the target variable $Y \in \mathbb{R}^N$, which represents the preference score of the community for a certain type of textile (such as cotton cloth) in the next stage (such as the next month). The value range is $[0, 1]$ and is calculated by weighted sentiment polarity and purchase volume:

$$y_i = \alpha \cdot \frac{\sum_{j \in C_i} s_j}{|C_i|} + (1 - \alpha) \cdot \frac{\text{purchase}_i}{\max(\text{purchase})} \quad (17)$$

C_i is the set of users in the i -th community, s_j is the sentiment score of user j (normalized to $[0, 1]$), purchase_i is the total purchase volume in the community, and α is the sentiment weight coefficient, which is used to balance the contribution of sentiment polarity and purchasing behavior in preference modeling. The adjustment process begins by setting an initial value of 0.5, indicating equal influence of sentiment and behavior. Next, the purchase volume data are standardized to eliminate differences in scale. Finally, grid search optimization is conducted through fivefold cross-validation, evaluating the model performance under different α values to select the optimal parameter that ensures a balanced integration of sentiment and behavioral features.

TABLE 1. Performance comparison of different detection algorithms

Detection Method	Optimization Objective	Requires Predefined Community Count	Time Complexity	Result Stability	Suitable Network Types
Louvain	Modularity maximization	No	Medium	High	Undirected weighted graph
LPA	Label propagation mechanism	No	Low	Low	Undirected/Directed graph
Infomap	Minimize description code length	No	Medium	Moderate	Directed graph, flow data
Spectral Clustering+K-means	Graph Laplacian matrix decomposition	Yes	High	High	Undirected graph

The model employs a standard two-layer GAT structure, in which each layer incorporates multihead attention mechanisms. The first layer uses 4 attention heads to encode node features into 64-dimensional representations, whereas the second layer aggregates information through 1 attention head to output a 1-dimensional preference score, thereby completing the final prediction task. The activation function uses the ELU function, and the last layer does not perform nonlinear transformation and directly outputs the predicted value. The loss function is defined as the Mean Square Error (MSE):

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (18)$$

y_i is the true preference score, and $\hat{y}_i$ is the model prediction value.

IMPROVED SIR PROPAGATION PATH MODELING

To simulate the spread of negative word-of-mouth in a social network, this study extends the classic SIR model by incorporating node influence and user trust parameters. The model defines three node states: susceptible (S), infected (I) (i.e., spreading negative information), and recovered (R). The modified state transition equations are as follows:

Infected State Equation:

$$\frac{dI_i}{dt} = \beta \cdot \varphi(v_i) \cdot \sum_{j \in N(i)} \theta_{ij} \cdot S_j \cdot I_j - \gamma \cdot I_i \quad (19)$$

In formula (19), $\varphi(v_i)$ represents the influence weight of node v_i , normalized on the basis of eigenvector centrality. $N(i)$ is the set of neighbors of node i . θ_{ij} is the trust parameter between nodes i and j , which is calculated as follows:

$$\theta_{ij} = \sigma \left(\lambda_1 \cdot \frac{w_{ij}}{\max(w)} + \lambda_2 \cdot \cos(s_i, s_j) \right) \quad (20)$$

In formula (20), $\sigma(\cdot)$ is the sigmoid function. λ_1 and λ_2 are weight parameters such that $\lambda_1 + \lambda_2 = 1$. w_{ij} is the historical interaction frequency between nodes i and j , and $\max(w)$ is the maximum interaction frequency. $\cos(s_i, s_j)$ is the cosine similarity between the sentiments of nodes i and j .

Recovered State Equation:

$$\frac{dR_i}{dt} = \gamma \cdot I_i \quad (21)$$

In formula (21), γ is the recovery rate.

The topological propagation efficiency measures the effectiveness of the information spread, which is defined as the ratio of the final infected node proportion to the average path length:

$$\eta = \frac{|I_\infty|/N}{L_G} \times 100\% \quad (22)$$

In formula (22), I_∞ is the total number of infected nodes at the end of the simulation, N is the total number of nodes in the network. L_G is the average path length of the network.

The initial infection sources are chosen from the top 5% of nodes with the highest influence, i.e., $\phi_{vi} \geq 0.72$. The propagation rate is $\beta = 0.65$, and the recovery rate is $\gamma = 0.25$, optimized through grid search.

III. EXPERIMENTS AND VERIFICATION

EXPERIMENTAL DESIGN

To verify the effectiveness of the constructed model in the rural textile consumption scenario, the experiment was based on real user behavior data and focused on preference identification, community division, trend prediction, spatial correlation analysis, and information dissemination coverage evaluation. All the experiments used a unified dataset and standardized processes to ensure the repeatability of the results and the reliability of the conclusions. The dataset of the rural textile consumer social network is shown in Table 2.

As shown in Table 2, the social network comprises 3,685 user nodes across 12 villages and 19,243 undirected weighted edges based on interactions such as likes, comments, and purchases. On average, each user interacts with 5.2 others. The network has a weighted clustering coefficient of 0.35, indicating some local clustering, and an average path length of 3.8, reflecting "small-world" characteristics. The top 5% node influence centrality threshold is set at 0.72, providing a basis for identifying key communication nodes and enabling targeted intervention to control negative word-of-mouth spread.

EVALUATION OF CONSUMER PREFERENCE CLASSIFICATION ACCURACY

The BERT model constructed in this paper is compared with the TextCNN, BiLSTM, XGBoost+TF-IDF, and SVM+Word2Vec models. All the models were trained on a 7:2:1 train/validation/test split, employing stratified sampling and temporal consistency principles to ensure data distribution alignment across partitions. After the model

TABLE 2. The dataset of the rural textile consumer social network

Metric	Value	Description
Total Number of Users (Nodes)	3,685	Rural e-commerce platform users across 12 administrative villages with geographic labels
Total Number of Edges	19,243	Undirected weighted edges based on likes, comments, shares, and purchase behaviors
Average Degree	5.20	Average number of interactions per user, reflecting rural users' social activity level
Weighted Clustering Coefficient	0.35	Measures local cohesion of interactions; higher values indicate stronger community ties
Average Path Length	3.80	Average shortest path between any two users, demonstrating "small-world" characteristics
Proportion of Largest Connected Component (LCC Size %)	82.60%	Proportion of nodes in the largest connected subgraph, indicating strong overall network connectivity
Node Influence Centrality Threshold (Top 5%)	≥ 0.72	Threshold for high-influence nodes (normalized) used as initial infection sources in the improved SIR model

training is complete, the accuracy, F1, precision, recall, and AUC-ROC of each category are calculated to evaluate the semantic understanding and classification capabilities of each model for texts with regional and colloquial characteristics. The results are shown in Table 3.

The performance comparison in Table 3 shows that the BERT model outperforms the other models in rural textile review sentiment classification, achieving 92.4% accuracy, 92.2% F1, 93.0% precision, 91.5% recall, and 96.2% AUC-ROC. Traditional models such as SVM+Word2Vec lag behind, with an accuracy of 80.1%, F1 of 79.9%, and AUC-ROC of 87.8%. Deep learning models such as TextCNN and BiLSTM also perform better than traditional models do but still fall short of BERT. The TextCNN achieves an accuracy of 84.6%, and the accuracy of the BiLSTM is 86.7%. BERT's superior performance in terms of all metrics highlights its stronger ability to capture complex semantics and regional characteristics in rural reviews, making it more robust and adaptable for tasks such as consumer group identification and communication path analysis.

COMMUNITY SEGMENTATION CONSISTENCY EVALUATION

Figure 3 compares the performances of the Louvain, LPA, Infomap, and spectral clustering+K-means algorithms in segmenting the rural textile consumer social network. The horizontal axis shows evaluation metrics: modularity Q (intracommunity compactness), NMI and ARI (consistency with ground-truth labels), and purity and homogeneity score (matching between community classification and true categories). The vertical axis represents the score for each metric.

As shown in Figure 3, the Louvain algorithm outperforms the other three algorithms across all five metrics. For example, it scores 0.43 in Modularity Q, higher than LPA (0.39), Infomap (0.41), and Spectral Clustering+K-means (0.37), indicating better community structure clarity. The Louvain algorithm also leads in NMI (0.58), ARI (0.51), purity (0.76), and homogeneity (0.69), outperforming the others in each category. These results demonstrate that the Louvain algorithm more accurately identifies consumer groups with geographic ties in undirected weighted social

networks, offering high stability and interpretability.

PREFERENCE TREND PREDICTION ERROR EVALUATION

Figure 4 shows the performance comparison of the proposed GAT model with mainstream regression models such as the XGBoost regressor and linear regression in the task of predicting rural textile consumption preference trends. The evaluation indicators include the MAE, RMSE, R^2 , and Spearman rank correlation.

As shown in Figure 4, the GAT model outperformed the XGBoost regressor and linear regression in terms of all the metrics. It achieves an MAE of 0.09, which is lower than those of XGBoost (0.13) and linear regression (0.17), and an RMSE of 0.13, which is better than those of XGBoost (0.18) and linear regression (0.23). The GAT model also scored 0.92 for R^2 , surpassing XGBoost (0.86) and linear regression (0.81), and achieved a Spearman rank correlation of 0.89, which was higher than that of XGBoost (0.81) and linear regression (0.76). These results highlight the GAT model's superior adaptability and accuracy, making it well suited for modeling consumption preferences and analyzing propagation pathways in rural textile markets.

SPATIAL CORRELATION ANALYSIS OF THE PREFERENCE EVOLUTION TREND

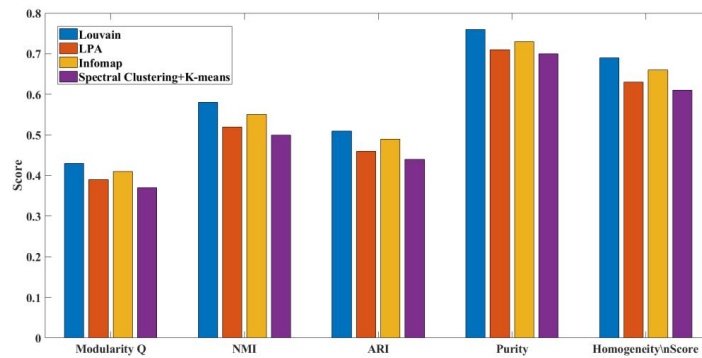
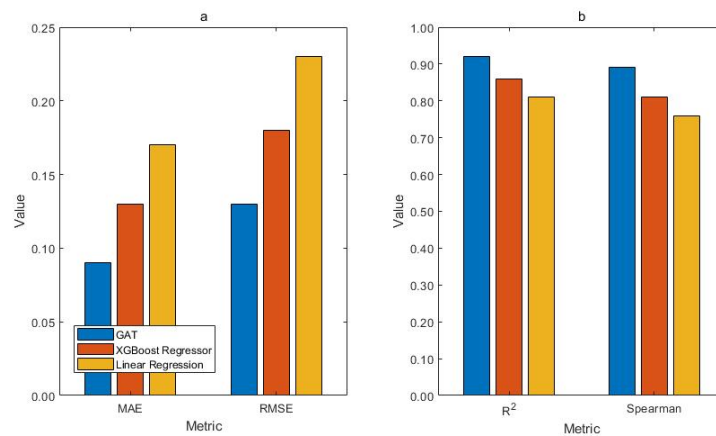
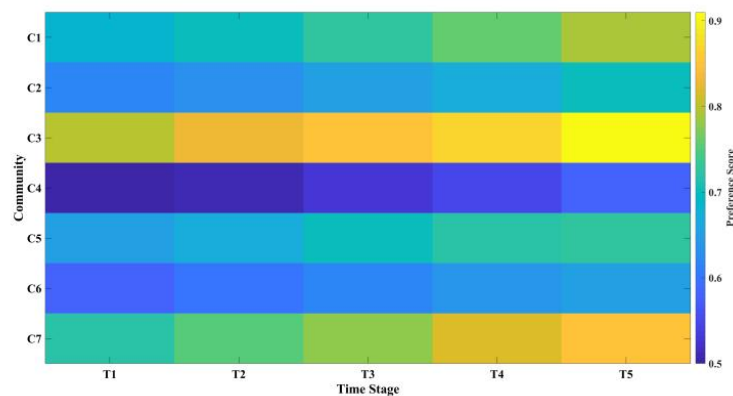
On the basis of 7 cross-administrative village consumer groups (C1–C7) generated via the Louvain algorithm, this experiment quantitatively analyzed the spatiotemporal evolution of their textile consumption preferences.

To reveal the spatial correlation and dynamic evolution trend of rural consumer preferences, this paper analyzes the spatial preference evolution of different communities (C1 to C7) in multiple time stages (T1 to T5). The results are shown in Figure 5, where the horizontal axis represents the time stage and the vertical axis represents the community name; the color depth in the figure reflects the preference score, and the preference score represents the preference intensity of different communities for spatial selection.

As shown in Figure 5, the preference score of the C3 community is always the highest, reaching 0.91 at T5, indicating that the consumption willingness of this community

TABLE 3. Performance comparison of each model in the rural textile review sentiment classification task

Model Name	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)	AUC-ROC (%)
BERT	92.4	92.2	93.0	91.5	96.2
TextCNN	84.6	84.2	85.1	83.3	91.0
BiLSTM	86.7	86.3	87.2	85.5	92.1
XGBoost + TF-IDF	82.3	82.0	83.0	81.0	89.4
SVM+Word2Vec	80.1	79.9	80.8	79.0	87.8

**FIGURE 3.** Consistency evaluation results of the community division of each algorithm**FIGURE 4.** Evaluation results of the preference trend prediction error of each model. (a) MAE and RMSE evaluation results of each model; (b) R^2 and Spearman's evaluation results of each model.**FIGURE 5.** Spatial correlation analysis results of textile consumption preferences in various communities (administrative villages)

has grown fastest. The score of the C4 community is the lowest, at only 0.50 at T1, and it has not exceeded 0.58 throughout the entire observation period, indicating that its overall acceptance is low. The remaining communities, such as C1, C5 and C7, showed a medium to high preference level and a steady upward trend, indicating that the consumption preferences of these communities gradually increased over time and that the dissemination path increased. Overall, the results of Figure 5 reveal the spatiotemporal evolution of consumption preferences and provide a visual basis for precision marketing interventions based on community structure.

IV. CONCLUSION

This paper proposes a method for modeling rural textile consumption preferences by integrating social network analysis, natural language processing, and GNNs. A weighted social network (3,685 nodes, 19,243 edges) was constructed on the basis of rural e-commerce user behavior logs, and seven geographically related consumer groups were identified via the improved Louvain algorithm. The BERT model was used for sentiment feature extraction, the GAT model was used for trend prediction, and the improved SIR model was used to simulate negative word-of-mouth communication. The experimental results show that the proposed method performs well in terms of preference sentiment classification (accuracy = 92.4%) and modularity (modularity $Q = 0.43$). However, this study has two limitations: limited data coverage (12 villages) and insufficient multimodal data fusion. Future research can focus on dynamic preference modeling, personalized intervention strategies, agricultural e-commerce applications, and the integration of multimodal data sources, including text, images, and physiological signals.

AUTHOR CONTRIBUTIONS

All work in this study was independently completed by Jiru Ren.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] K. Sutherland, S. Casey, and G. Crimmins, "Social media behaviors and barriers impacting women e-commerce entrepreneurs in rural Australia: A pilot study," *Community Development*, vol. 56, no. 1, pp. 39-58, 2025.
- [2] Y. Song, L. Li, S. Sindakis, S. Aggarwal, C. Chen, and S. Showkat, "Examining e-commerce adoption in farmer entrepreneurship and the role of social networks: data from China," *Journal of the Knowledge Economy*, vol. 15, no. 2, pp. 8290-8326, 2024.
- [3] T. M. Wanigapura, T. Guruge, I. V. Kuruppu, and P. C. Abeysiriwardana, "Diversified impact of electronic word-of-mouth (eWOM) on consumer communities: a developing country perspective," *Decision*, pp. 1-17, 2025.
- [4] I. Ghosal, B. Prasad, M. P. Behera, and A. Kumar, "Depicting the prototype change in rural consumer behaviour: An empirical survey on online purchase intention," *Paradigm*, vol. 25, no. 2, pp. 161-180, 2021.
- [5] I. Antosova, M. Purny, and J. Stavkova, "Changes in consumer purchasing decisions: Traditional and emerging factors in the dynamic marketing landscape over 15 years," *Marketing i Menedžment Inovacij*, vol. 14, no. 3, pp. 85-96, 2023.
- [6] Y. Ren, B. Castro Campos, Y. Peng, and T. Glauben, "Nutrition transition with accelerating urbanization? Empirical evidence from rural China," *Nutrients*, vol. 13, no. 3, pp. 921-939, 2021.
- [7] J. Yang, J. Li, and Y. Cao, "Analysis of peer effects on consumption in rural China based on social networks," *Applied Economics*, vol. 55, no. 6, pp. 617-635, 2023.
- [8] K. Osei-Frimpong, G. McLean, N. Islam, and B. A. Otoo, "What drives me there? The interplay of socio-psychological gratification and consumer values in social media brand engagement," *Journal of Business Research*, vol. 146, pp. 288-307, 2022.
- [9] H. Zhang, E. Millan, K. Money, and P. Guo, "E-commerce development, poverty reduction and income growth in rural China," *Journal of Strategy and Management*, vol. 18, no. 1, pp. 148-176, 2025.
- [10] V. R. Revathy, A. S. Pillai, and F. Daneshfar, "LyEmoBERT: Classification of lyrics' emotion and recommendation using a pre-trained model," *Procedia Computer Science*, vol. 218, pp. 1196-1208, 2023.
- [11] H. Wang and J. Leskovec, "Combining graph convolutional neural networks and label propagation," *ACM Transactions on Information Systems (TOIS)*, vol. 40, no. 4, pp. 1-27, 2021.
- [12] Z. A. Rahman, B. A. Talip, and H. Sarirah, "Exploring Customer Review of Local Agriculture Product Acceptance in Malaysia: A Concept Paper on Sentiment Mining," *International Journal on Perceptive and Cognitive Computing*, vol. 10, no. 1, pp. 29-39, 2024.
- [13] D. Modi and L. Zhao, "Social media analysis of consumer opinion on apparel supply chain transparency," *Journal of Fashion Marketing and Management: An International Journal*, vol. 25, no. 3, pp. 465-481, 2021.
- [14] L. Ren, B. Zhu, and Z. Xu, "Robust consumer preference analysis with a social network," *Information Sciences*, vol. 566, pp. 379-400, 2021.
- [15] H. Fe, "Social networks and consumer behavior: Evidence from Yelp," *Journal of Economic Behavior & Organization*, vol. 209, pp. 1-14, 2023.
- [16] N. S. Sattar and S. Arifuzzaman, "Scalable distributed Louvain algorithm for community detection in large graphs," *The Journal of Supercomputing*, vol. 78, no. 7, pp. 10275-10309, 2022.
- [17] C. Wang, F. Wang, and T. Onega, "Network optimization approach to delineating health care service areas: Spatially constrained Louvain and Leiden algorithms," *Transactions in GIS*, vol. 25, no. 2, pp. 1065-1081, 2021.
- [18] M. Rostami, M. Oussalah, K. Berahmand, and V. Farrahi, "Community detection algorithms in healthcare applications: A systematic review," *IEEE Access*, vol. 11, pp. 30247-30272, 2023.
- [19] R. Mao, Q. Liu, K. He, W. Li, and E. Cambria, "The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1743-1753, 2022.
- [20] Y. Kit and M. M. Mokji, "Sentiment analysis using pre-trained language model with no fine-tuning and less resource," *IEEE Access*, vol. 10, pp. 107056-107065, 2022.
- [21] D. Deepa and A. Tamilarasi, "Bidirectional encoder representations from transformers (BERT) language model for sentiment analysis task," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 7, pp. 1708-1721, 2021.
- [22] R. Zhao and X. Han, "Prediction Method of Fruit and Vegetable Product Consumption Behavior Based on Graph Neural Network," *Pak. J. Agri. Sci.*, vol. 61, no. 4, pp. 1235-1245, 2024.
- [23] L. Jain, R. Katarya, and S. Sachdeva, "Opinion leaders for information diffusion using graph neural network in online social networks," *ACM Transactions on the Web*, vol. 17, no. 2, pp. 1-37, 2023.
- [24] X. Chen, H. Xie, S. J. Qin, Y. Chai, X. Tao, and F. L. Wang, "Cognitive-inspired deep learning models for aspect-based sentiment analysis: A retrospective overview and bibliometric analysis," *Cognitive Computation*, vol. 16, no. 6, pp. 3518-3556, 2024.
- [25] A. Razaque, S. Rizvi, M. Almiani, and A. Al Rahayfeh, "State-of-art review of information diffusion models and their impact on social network vulnerabilities," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 1, pp. 1275-1294, 2022.
- [26] R. Palalic, V. Ramadani, S. Mariam Gilani, S. Gërguri-Rashiti, and L. P. Dana, "Social media and consumer buying behavior decision: What

- entrepreneurs should know?" *Management Decision*, vol. 59, no. 6, pp. 1249-1270, 2021.
- [27] A. S. Krishen, Y. K. Dwivedi, N. Bindu, and K. S. Kumar, "A broad overview of interactive digital marketing: A bibliometric network analysis," *Journal of Business Research*, vol. 131, pp. 183-195, 2021.
- [28] J. Son and L. S. Niehm, "Using social media to navigate changing rural markets: The case of small community retail and service businesses," *Journal of Small Business & Entrepreneurship*, vol. 33, no. 6, pp. 619-637, 2021.
- [29] W. Zhang, P. K. Chintagunta, and M. U. Kalwani, "Social media, influencers, and adoption of an eco-friendly product: Field experiment evidence from rural China," *Journal of Marketing*, vol. 85, no. 3, pp. 10-27, 2021.
- [30] J. Yadav, A. Yadav, M. Misra, N. P. Rana, and J. Zhou, "Role of social media in technology adoption for sustainable agriculture practices: Evidence from twitter analytics," *Communications of the Association for Information Systems*, vol. 52, no. 1, pp. 833-851, 2023.