

**How regional stereotypes are narrated in news comments: an analysis of the TikTok
platform based on the LDA Model**

Junyu Jiang,¹ Jiani Xu²

Faculty of International Media, Communication University of China

Author Note

Junyu Jiang, B.A., Communication University of China

E-mail: jainfarstrider0@gmail.com

Jiani Xu, B.A., Communication University of China

E-mail: 2747210071@qq.com

Submitted to: Communication in Post- and Neo-Authoritarian Societies (CPN)

ExOrdo submission ID: #4288

Abstract

This study examines regional stereotypes in new media comment sections, focusing on TikTok. Using the Latent Dirichlet Allocation (LDA) model, researchers analyze comments to uncover how social cognitive biases contribute to these stereotypes. They find that stereotypes often stem from collective social memories and spread rapidly on social media following news events, shaping public discourse. The research investigates the theft incident at the 2023 Midi Music Festival in Nanyang, Henan, using TikTok comment sections to identify narrative frameworks like objective reporting, irony, symbolism, and image reflection. These frameworks reflect public attitudes and the evolution of regional stereotypes in digital spaces. The study suggests that socioeconomic status, media reporting, and regional identity influence stereotype propagation. It proposes promoting positive portrayals on social media to improve regional images. This research informs strategies for fostering a harmonious online environment, advocating for technological enhancements in news platforms to diversify and positively represent geographic information dissemination.

Keywords: regional stereotype, LDA model, new media comment sections, narrative modes

How regional stereotypes are narrated in news comments: an analysis of the TikTok platform based on the LDA Model

Stereotypes and the group discrimination they engender are significant social issues in today's world. Members of the in-group tend to label those in the out-group with stereotypical tags. In recent years, racial stereotypes (Lee & Fiske, 2006), gender stereotypes (He, 2007), and occupational stereotypes (Hu & Liang, 2004) have become hot topics of research. China's vast geographical span, extremely frequent population mobility, and sharp cultural and economic development disparities have given its unique regional stereotypes a particularly important role in recent news topics. Whenever information containing references to specific regions or phrases like "people from a certain place" is disseminated on social media platforms, especially when triggered by news events or social phenomena, they can easily attract the attention of a large number of Internet users, often leading to conflicts between people from different regions.

With the emergence of interactive digital platforms, Internet audiences participate in the production of news discourse in the comment sections of news portal websites (Fuchs, 2016). Social media is showing a "decentralized" dissemination trend in all aspects, and the traditional mass media's one-to-many communication model is gradually being broken. Each netizen empowered by digital means has become a personalized communication node, and the audience can express their different views on regional stereotypes through various forms such as posting, barrage, and videos, which further exacerbates their widespread dissemination (Zhao, 2022). Previous scholars have tended to analyze the biased behavior of media reports from the media

perspective and the potential negative effects of this bias on public cognition and social operation (You, 2018 & Zhao, 2018), and have also paid attention to the regional discrimination caused by the IP address of news comments (Peng, 2019), but there has been no relevant research on the narrative logic of stereotypes, nor has there been large-scale text quantitative empirical research.

The Henan Midi Music Festival in China was held in Nanyang, Henan Province in September 2023. A major theft incident occurred in the camping area of the music festival, mainly by local villagers of Nanyang. Under the news related to this incident, the regional stereotype has triggered a huge public opinion, and the stereotype based on economic backwardness and its associated discrimination filled the comment area of the relevant news media stream. This paper takes the interactive comments in the comment area of the related TikTok news video of the incident as the text basis to explore how regional stereotypes interact in the new media comment field. Therefore, the study conducted a case study to construct an LDA topic model with related texts and revealed the narrative logic of regional discrimination through the selection of keywords. These findings provide a theoretical basis for the dissemination and resolution strategies of regional stereotypes.

Literature Review

Regional stereotypes, as a form of social stereotypes based on cognitive biases, initially stem from the construction of social memories associated with specific regions. Regions are defined as units that integrate a variety of factors—geographical, functional, social, and cultural. They are formed into distinct social structures through the interplay of politics, economy, culture,

administrative practices, and discourse (Zhou, 2019), thereby bearing the marks of human activities. Individuals often categorize human collectives based on geographical environments and oversimplify geographical portrayals, leading to specific cognitions, concepts, or expectations, and resulting in cognitive biases. Sociologist E. Goffman regards frames as the cognitive structures through which subjects "recognize and interpret the external objective world," summarizing frames as "basic elements capable of identification," and referring to the interpretive schema that people tend to use when recognizing specific things as the "primary framework," which can "render meaningless parts of a scene into something meaningful." People rely on this cognitive schema to perceive external information and generate regional impressions through media frames, that is, regions are rendered as a cognitive framework for the subject under media portrayal, and a regional image is formed. Political economists tend to analyze the political project-related regional household registration system discrimination as the source of regional stereotypes in China (Afridi et al., 2015). The Chinese Communist Party carried out market-oriented reforms in the 1970s, leading the public to accept capitalist morals (Goodman, 2014). In this process, the differences in economic development levels between regions have formed economic level-based regional discrimination.

China's society, with its significantly different regional economic developments, has seen an enormous population flow between regions. Between 2000 and 2018, China's rural population decreased by 244 million people, with an average annual reduction of over 30 million people; at the same time, the scale of rural labor also decreased accordingly, with a total reduction of

60.7207 million rural labor forces between 2000 and 2010 (Zhejiang Provincial Academy of Social Sciences, 2010). The huge inter-regional flow has provided a material basis for the exchange of regional stereotypes, while the unique development of information and communication technology has widely spread and deeply constructed regional stereotypes. Looking at the stereotypes of regions emerging from social life and the internet, most of them are labeling economically underdeveloped areas, with roots traceable back to the reform and opening up, when a large number of people from the central regions flowed out to seek work opportunities, and people from different regional cultures came into contact with each other, and outsiders were easily excluded by the locals. Sheng Yanli (Sheng, 2013) proposed the formation mechanism of Henan's regional stigma, which first involves the interaction between the outflowing people from Henan and the stigmatizers, followed by the stigmatizers' dissatisfaction with the people from Henan, then labeling the people from Henan with negative tags, spreading their views to society, and the media's reports on some negative events in Henan have intensified the prejudice, thus plunging Henan into regional stigma.

The spread of regional stereotypes initially originated from the logic of information transmission in oral narratives (Wu, 2017), where the cognitive differences and disordered expressive abilities shown in individual oral expressions have turned many words into rumors, prejudices, and jokes, which is a biased discourse structure. China's unique horizontal social structure (Fei, 2008) limits the mutual communication and trust establishment between different regions, and regional differences easily become simplified and generalized symbols in people's

minds (You, 2019), forming social stereotypes. In the era of mass communication, the media participated in the construction of social facts, and the long-term existence of regional differences formed social stereotypes, gradually solidifying into a social norm, and generating strong social pressure, and the media would follow this pressure, thus showing regional biases in reporting (You, 2019). The emergence of the internet has triggered a large amount of information dissemination. In this context, individuals actively participate in the classification and interpretation of information about other regions, and simplify information based on personal experience. The inherent differences between individuals lead to cognitive diversity. At the same time, due to the inherent limitations of the media in accurately and comprehensively disseminating information about other regions, the reliance on simplified personal inferences inevitably leads to cognitive biases, thereby gradually strengthening regional stereotypes.

However, previous researchers have been more inclined to analyze the formation and evolution of regional stereotypes from the perspective of news reporting frames, focusing on how media reports will produce regional discrimination through media reports (Ren, 2023; Yu, 2023; Zhao, 2022), and the solutions proposed are also more inclined to change the frames of the gatekeepers (Huang, 2024).

The development of social media platforms has endowed individuals with the role symbols of gradually integrating into the role of prosumers (Fuchs, 2019). The audience participates in the narrative and dissemination construction of regional symbols, gradually forming a dynamic development process of "aggregation—resonance—confrontation—

diffusion—fading." An individual's feelings, opinions, and dissatisfaction about a place can cause extensive resonance among people in different times and spaces through the network, stimulating resonance at the cognitive level, and the increasing number of individual experiences provide guidance for group experience. This homogenized individual experience, repeatedly presented through network dissemination, continuously conveys, and strengthens the same viewpoint and impression, making the cognitive resonance of netizens about a place ultimately rise to collective cognition in a subtle way, to a certain extent, shaping the netizens' attitude and view towards that place, and also affecting the netizens' cognitive expectations for that place in the future (Ren, 2023).

In today's media platformization of news stream media, the characteristics of depersonalization and high anonymity, combined with the individual's need for group belonging and identity, lead to the spread of regional stereotypes in the field of news comment areas. At this time, the narrowing and strengthening of opinions occur simultaneously, and individuals with minority opinions are often in a state of collective silence, making regional stereotypes often become more isolated and exclusive (Liu, 2015). The necessity of how regional stereotypes will be narrated in the news stream media comment areas is gradually revealed.

Field Theory and LDA Model Overview

The Latent Dirichlet Allocation (LDA) model is a semantic model used to represent topic concepts. In the past, researchers have proposed several important semantic representation models, among which the TF-IDF (Term Frequency-Inverse Document Frequency) model, LSI

(Latent Semantic Index) model, and pLSI (Probabilistic Latent Semantic Index) model are quite representative. The traditional text information discourse extraction method TDT uses K-means clustering, hierarchical agglomerative clustering, etc., to extract topics from texts, but the implementation is not strong, and the extraction effect is not good.

The basic unit of analysis based on the LDA topic modeling model is the document. Its basic idea is to express each document as a multi-distribution of a series of topics, and each topic is expressed as a multi-distribution of all the words in the vocabulary dictionary. According to the document generation model and known data, the topic structure of the document is obtained by reverse probability derivation.

Currently, topic modeling of short text data mostly adopts topic model methods, that is, by reducing the dimension of the document, the document space is represented as a topic space, and the generation process of the document is simulated. Among them, the most widely used topic model is the hierarchical Bayesian model based on the bag-of-words theory—Latent Dirichlet Allocation (LDA) supervised topic model.

Due to the LDA model's good mathematical foundation and flexible scalability, it has been widely used in the fields of text mining and information processing. Among them, the multi-level background topic model constructed by IVAN has been applied to the field of sentiment analysis, and at the same time, on social media, scholars such as Daniel and Zhang Chenyi have used the Labeled-LDA model to apply the topic tags of podcasts to topic modeling. However, it is worth noting that the LDA algorithm is most used to analyze well-structured text

documents, such as newspapers and academic journal articles, which have been reviewed, edited, and proofread in terms of grammar and spelling. In contrast, text data from social media faces many challenges, including but not limited to the word limit in social media comment areas, huge cultural differences, and the widespread use of ironic contexts; in addition, abbreviations, symbols, and deliberately truncated words in the text will also block the generation of text topics (Lei, 2012). A recent development direction of topic models is to involve humans to iteratively improve the topics generated by automatic topic modeling algorithms (Chuang et al., 2015; Hu, Boyd-Graber, Satinoff, and Smith, 2014), which may improve the interpretability of the discovered topics. Although the LDA model is very popular in the automation of text analysis tasks, there are relatively few researchers in news and mass communication who use this method to answer research questions (Lei et al., 2012).

In the study of news comment areas, the comment area of news portal websites is visualized in the form of a "floor construction" mechanism, which displays the original post and its responses in the form of dialogue "blocks" (Qian, 2011). Users can reply to specific comments by quoting them, and as more netizens join the discussion, these dialogue "blocks" expand accordingly. Personal opinions can directly transform into social expressions; at the same time, by further mining the power of the masses in the forum, personal opinions can also aggregate into social expressions. Because the tree-branch structure formed by the comment area of TikTok only displays comments on the main post and some dialogue replies, other replies to user comments and dialogues are folded up. Comments on the main post can become a separate

interactive submodule. The comment area can be extended by making the main post a submodule. Specifically, replying directly to popular comments has become a shortcut to increase exposure, which can attract more users to participate. At the same time, folding the reply dialogue content can filter out many low-quality dialogue replies, while using algorithms to expose some high-quality replies, considering the needs of browsing and driving interaction.

Methods

samples

According to official data, as of April 29, 2024, the total number of TikTok users in the China region has exceeded 800 million, with daily active users exceeding 700 million, making it the most popular short video production and dissemination platform in the China region. On the TikTok platform, users can continuously interact with other users and video content through comments, likes, and shares. In the comment area of videos related to regional news events, users can directly reply to other users' comments under the comment, forming a secondary comment area. At the same time, secondary comments may trigger replies from many other users. In this one-to-one or one-to-many communication pattern, users gradually become disseminators and producers of content (Guo, 2018), accelerating the construction and spread of regional stereotypes.

On October 8, 2023, the Midi Music Festival was held in Nanyang, Henan. At the scene, many spectators' belongings were stolen by local villagers. This incident quickly sparked heated discussions on the TikTok after being released by self-media bloggers and official news media. As the news event fermented, many "regional attacks" on Nanyang and even the entire Henan

were seen in the comment area of the event video. Therefore, the researcher first searched for "Midi Music Festival" "Henan Nanyang" in the search bar of TikTok in the China region. Since the news event has been more than half a year old, and the Chinese government has actively repaired the city's image during this period by releasing promotional videos to cover negative information and limiting the flow of related videos to control public opinion, the current video content is mainly cultural tourism propaganda, and the discussion on the image of Henan is relatively small. Therefore, the researcher refined the topic search keywords to "The Midi Music Festival in Nanyang," "Nanyang Grab Event" through sample observation and discourse practice. As of June 1, 2024, the researcher selected samples in the popular video list of TikTok's keywords, focusing on the content of Henan regional stereotypes presentation. After collecting and organizing the search results, excluding the original repeated short video materials and ineffective short video content, 22 effective short video samples were manually selected.

Short videos contain a large amount of information. To reduce subsequent text processing and topic model errors, two researchers watched each video one by one, and marked the video data from two dimensions of emotional tendency and reporting angle, using more than 2 descriptive words and short sentences to reduce the error and impact caused by the degree of understanding on data. Finally, considering the purpose of this study, the researcher classified the samples into four topic categories: event narration, image restoration, public relations response, and image reflection.

To explore the user interaction caused by the video content, the researcher selected the top 25 most popular comments in the comment area of each of the 22 selected videos, and the top 15 secondary comments with more likes, totaling 9570 comments. The researcher further screened the comments according to the following criteria:

- a. Delete meaningless content (e.g., blank, punctuation, etc.).
- b. Delete content that does not involve emotional expression and regional image.
- c. Delete content that is not complete and coherent.
- d. Delete content that only has emoticons.
- e. Delete content that is completely unrelated to the news event being studied.

These disordered and large amounts of comment texts are unstructured data, which are difficult to directly mine and analyze. In response to the complexity of comment text data, this study has adopted a series of data preprocessing methods. First, in order to improve the efficiency and accuracy of text processing, the researchers used regular expressions in Python to remove "noise data" such as emoticons and special symbols in the text, and manually deleted TikTok topic tags and other texts unrelated to emotional attitudes. After the above screening, a total of 4953 comment texts were selected as analysis samples for the coding phase.

In the text analysis using the LDA topic model, words are the basic elements for constructing the model. Given the characteristics of Chinese text, the researchers selected the Jieba word segmentation tool for data word segmentation processing. In response to the problems of repeated words, personal pronouns, and colloquial expressions in the word

segmentation results, the researchers adopted the method of stop word filtering. Using Python to read each text data loop after word segmentation processing, mainly based on Baidu's stop word list for stop word matching and deletion and supplemented other stop words according to the characteristics of the text to ensure the accuracy of the word segmentation results.

LDA Model Topic Analysis

In this text analysis process, we employ Latent Dirichlet Allocation to model the context of the text and extract latent topics within a three-layer hierarchical Bayesian network model. By analyzing these comments, we anticipate that the application of unsupervised learning in text analysis will mitigate coding errors caused by supervisor bias.

To evaluate the classification effectiveness of the topic model, the source text is divided into 1103 bags-of-words models, and the trend of topic coherence is assessed. To optimize the results of topic modeling, we use Grid Search to obtain the best model for text construction under four thematic videos, employing two parameters: the number of topics, which ranges from 1 to 14, and the learning decay parameter, which controls the learning rate, tested at values of 0.7, 0.8, and 0.9. The optimal model is determined by the log likelihood score and perplexity resulting from the combination of these two parameters.

Different types of videos contain similar narrative logic; therefore, constructing a unified topic model may lead to misleading results due to narrative differences across videos. To address this issue, tweets can be grouped based on common features such as author or publication time.

Hong and Davison (2010) proposed a method where all tweets from the same author published at different times are combined into a single document. This method mixes all topics from each user over different periods. Another study (Zhao et al., 2011) chose to combine all tweets from all users within a unit of time into a single document, mixing all topics from each time unit. These methods enable more effective topic modeling of tweets, reducing misleading results caused by different narrative logics.

The LDA training generated a list containing 23 "topics" and the probabilities of all words associated with each topic. To determine the actual meaning of these "topics," two communication researchers read all words with a probability higher than 1% in each topic and proposed labels they believed represented the topic. All topic models were merged and organized.

Topic Validation and Evaluation

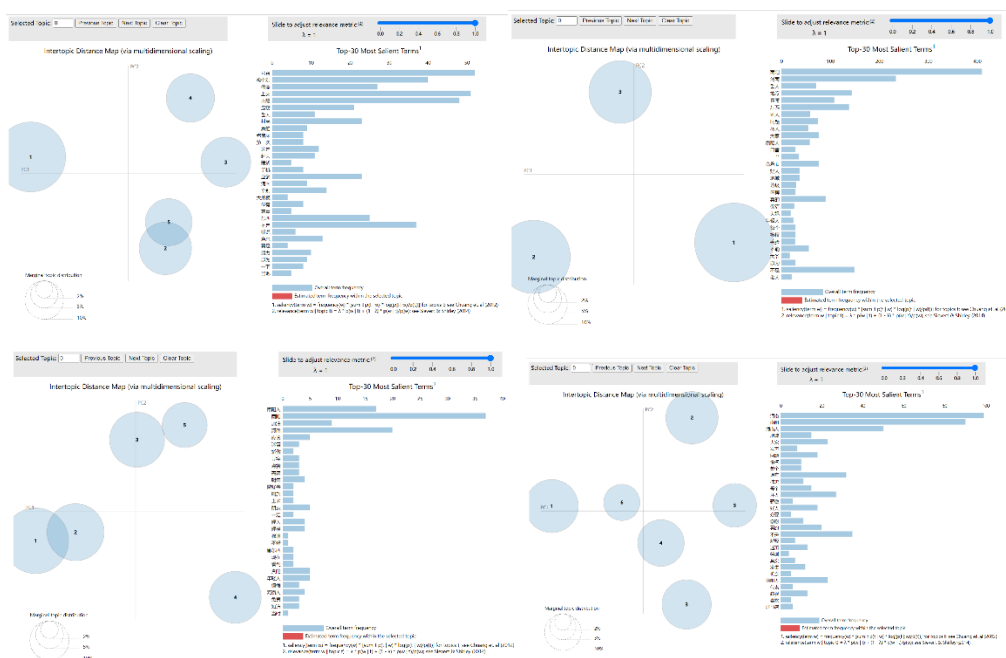
To explore the effectiveness and representativeness of the merged topics, we extracted text materials from the topics for manual evaluation. We randomly selected 100 comment samples from different video types. This sample does not represent the entire population. We did not intend for human coders to evaluate a representative sample of documents, as this is logically impossible. Specifically, approximately one-third of the files ($n = 32$) were randomly selected from the entire dataset within each set of 100 comment texts. Then, 32 files were randomly selected to represent the results based on keywords, ensuring at least eight comments were labeled as containing each of the three predetermined topics. Two new independent researchers, who had not previously seen these comments, read the documents and independently identified

the topic of each document. The coders then discussed their interpretations and re-read the documents until they reached a consensus. For each comment, the coders discussed and determined which method generated a "topic" closer to their own evaluation results. For the LDA analysis results, coders were asked to read the labels and the related word lists until three consecutive random selections did not yield unclassifiable text comments.

Conclusion

The analysis includes five topics, the number of articles corresponding to each topic, and ten characteristic words under each topic. The five topics are explained as four narrative sublines: T1, T2, T3, T4.

The tables above visualize the LDA topic model, where larger circles indicate a higher proportion of the topic in all documents, and the distance between circles represents the distance between topics. The bar chart represents the list of characteristic words that best explain the topics.



Topic 1 presents an objective narrative framework of the event, exhibiting a relatively balanced distribution of emotional tendencies. Within this framework, discussions predominantly

Table 1: Objective Narrative of Events		
TOPIC	FRAME	CHARACTERISTIC WORDS
1	To state objective events.	villagers, theft, belongings, take away, train station, go home, Nanyang, theft, music festival, large-scale, heard, know, believe, mistaken for, underpants, tent, extremely rare, smear
2	To propose recommendations for handling the event.	call for, submit, should, must, do not, commit a crime, sentence, punishment, severe penalty

involve video content recounting local events, using descriptive terms such as "villagers" (1.1.1), "theft" (1.1.2), "Nanyang" (1.1.3), and "music festival" (1.1.4) to convey basic facts. Notably, in the elucidation of the event, commentators tend to amplify the absurdity of the incident through dramatic expressions like "mistakenly thought" (1.1.5) and "underpants" (1.1.6), thereby expressing mild emotional inclinations towards the event. Additionally, the discussion not only focuses on objective facts but also on the handling of the event, using terms such as "call to action" (1.2.1) and "sentencing" (1.2.2). These terms reflect moral judgments, offering suggestions to individuals, government bodies, and social organizations involved in the event, thereby expressing a reasonable expectation for the handling of the event and a pursuit of social justice.

1.1.1 *It's probably taken by the nearby villagers.*

1.1.2 *This is definitely theft!*

1.1.3 *It's Nanyang again! The Nanyang Music Festival disaster (This is a pun; the pronunciation of "disaster" in Chinese is like "festival.").*

1.1.4 *The Midi Music Festival didn't make headlines, but the theft did.*

1.2.1 *The government should call on these villagers to collectively return the items.*

1.2.2 *Which village is this? They knowingly committed theft; they must be sentenced.*

Table 2: Irony Narrative		
TOPIC	FRAME	CHARACTERISTIC WORDS
1	Objective Statement Form	Procurement, unwanted, respect, customs, people from Henan
2	Allusion in Poetry and Literature	samples above

Topic 2 exhibits a framework of irony and rhetorical narrative, within which sarcasm and negative emotional tendencies are particularly evident. Firstly, the text mentions two obvious rhetorical threads. The first thread conceals sarcasm through seemingly objective statements. For example, using terms like "stocking up" (see 2.1.1), "nobody wants" (see 2.1.2), and "custom" (see 2.1.3), it implies a stereotype that Henan people have a tradition of stealing, and suggests that this behavior should be considered a "folk custom" and respected (Grice, 1975). Although this expression appears to be an objective description on the surface, it actually implies the absurdity or irrationality of this view through irony.

The second thread relies on historical allusions and related famous literary works from the region to convey sarcasm through the core actions of specific events. For example, some comments repeatedly mentioning the story of the Three Kingdoms where General Dian Wei of had his double halberds stolen in Nanyang, Henan (2.2). This comparison emphasizes the concept of Henan people stealing, and by referencing historical anecdotes, metaphorically suggests the absurdity and irrationality of such behavior (Grice, 1975).

Additionally, some comments analyze poems by the famous Tang Dynasty poet Liu Yuxi, focusing on the core event location, which is also part of the verse, "Nanyang." These comments alter the phrase originally expressing noble character undeterred by humble surroundings, "No matter how dilapidated the house is, my soul is beautiful enough." to "No matter how egregious the theft is, my soul is kind enough." (2.3), thereby changing the meaning and implying that Henan people take pride in stealing instead of feeling ashamed (Searle, 1979).

Within the framework of speech act theory, the ironic nature of this text can be further explained. According to speech act theory, irony is considered an indirect speech act because it conveys a meaning opposite to the literal expression through implication (Grice, 1975). Stating that Henan people have a tradition of stealing and that it should be respected suggests the absurdity of this view. Participants in the discussion will interpret this text by understanding its true intention through the most relevant cognitive effort, thereby inferring it as an expression of sarcasm or irony.

Furthermore, comparing the statement "Henan people stealing" with examples from historical anecdotes can be seen as a metaphor (Searle, 1979). This metaphor implies the concept of Henan people loving to steal through comparison, highlighting its absurdity and irrationality.

2.1.1 *They even drove cars over; people might think they are stocking up.*

2.1.2 *What do you mean by stealing? It's called picking up things nobody wants.*

2.1.3 *Those who have been to Nanyang, please follow local customs. Those who haven't, please respect the local history and customs.*

2.2 [197 AD, the second year of Jian'an] During the Three Kingdoms period in the state of Wei, Cao Cao's renowned general Dian Wei had his double halberds stolen by Che Hu'er in Wancheng, ultimately leading to his death in Wancheng (present-day Nanyang).

2.3 Nanyang has the former residences of Zhuge Liang and Confucius. Confucius would ask, what impact and harm does theft cause?

Table 3: Symbolic Descriptive Narrative Logic		
TOPIC	FRAME	CHARACTERISTIC WORDS
1	Distinguishing between events and regional symbols	good person, bad person
2	Distinguishing between Nanyang and the entirety of Henan Province	we the one hundred million people in Henan Province
3	Stating and addressing malignant incidents	Henan, Nanyang, Zhumadian, representative
4	Dialectical form	Y'all betcha, We'uns, Figured

Topic 3's symbolic descriptive framework shows a moderate emotional bias, divided into opposing regional stereotypes and emphasizing regional stereotypes.

Regarding opposition to regional stereotypes, the discussion emphasizes that although the event occurred in Nanyang, it does not imply that all residents are related to the event or share the same behavioral characteristics. Commentators shift the focus from mere geographical location or regional population to a deeper analysis of their actions and morals by describing participants in the event as "good people (3.1.1)" and "bad people." This approach aims to prevent individual character flaws or localized behavioral deviations from being overly generalized into negative impressions of the entire region. Additionally, commentators use phrases like "we the one hundred million people (3.2.1)" to highlight Nanyang's uniqueness and

individual characteristics within Henan Province, thereby opposing the spread of overall stereotypes of Henan.

On the other hand, some commentators, based on regional stereotypes, intentionally or unintentionally link this event to previous serious incidents within Henan Province (3.3.3), deepening and reinforcing stereotypes of Henan people as "low-quality" and "poor." Moreover, dialects serve as the most direct form of regional identity and cultural symbols. Due to their complexity in Chinese word segmentation, researchers manually filter and process texts using dialects through vocabulary identification, classification, content understanding, and contextual analysis. While expressing emotional tendencies, commentators cleverly combine dialect vocabulary and expressions characteristic of Henan's regional features, such as " Y'all betcha (3.4.1)," " We'uns (3.4.2)," and " Figured" integrating regional cultural identifiers with subjective emotional expressions to deepen stereotypes and perceptions of Henan's region.

3.1.1 *There are good and bad people everywhere; we should condemn the thieves.*

Regional attacks are unacceptable.

3.2.1 *Biased media are also setting the tone. Can Nanyang represent one hundred million people in Henan? Talk about Nanyang, not all of Henan!*

3.3.3 *Henan has a bad reputation, and the contributions from the Nanyang and Zhumadian (two regionals of Henan Province) area are indeed significant.*

3.4.1 *Y'all betcha*

3.4.2 *I reckon nobody wants it.*

Table 4: Narrative Logic of Reflection on Regional Stereotypes after Image Restoration		
TOPIC	FRAME	CHARACTERISTIC WORDS
1	Exemplifying Outstanding Regional Characteristics	free of charge, well done, do good deeds, friend, proud action, practical action, thankful
2	Proposing Image Restoration Suggestions	Like, contribute, promote, government, volunteers, strict enforcement, approach, announce
3	Support for Image Restoration Outcomes	Increasingly, believe, hope, go for it
4	Subjective Critique and Emotional Output Regarding the Event	Thief, stealing, crazy, criminal activity, rat droppings, disgraceful
5	Expressing Positive Emotions Toward Hometown	Pride, South-to-North Water Diversion Project
6	Aspiring to Change Public Opinion	Believe, gratitude, speak out, smear
7	Calling for Image Restoration	Effort, change, hope

Topic 4 focuses on reversing regional stereotypes by listing excellent examples from Nanyang. The article first enumerates a series of positive behaviors and examples, such as "free of charge" (4.1.1), "great!" (4.1.2), and "doing good deeds!" aiming to showcase many praiseworthy and commendable behaviors in the Nanyang area. Through these examples, commenters attempt to change the public's negative impression of Nanyang, emphasizing that there are many positive and uplifting individuals and events in the region, thereby reversing the public's stereotypes about it.

Additionally, related comments offer specific suggestions for image restoration, such as "likes" (4.2.1), "promotion" (4.2.2), "government support," and "volunteer actions." These suggestions aim to improve Nanyang's public image through positive social actions and guided public opinion, promoting a positive perception and evaluation of Nanyang (Benwell & Stokoe, 2006).

In terms of emotional expression, related comments display subjective criticism and emotional output about the event, using negative terms such as "thief" (4.3.1), "illegal activities," and "disgrace" (4.3.2). These expressions reflect the commenters' strong condemnation and negative emotions towards the theft incident and their deep concern and worry about Nanyang's damaged image (van Dijk, 1993).

Moreover, some comments actively oppose regional discrimination, expressing pride and support for their hometown. Words like "South-to-North Water Diversion" (4.4.1) reflect the appreciation and promotion of Nanyang's unique natural and cultural resources. Through these expressions, commenters try to emphasize Nanyang's uniqueness and diversity as a region, opposing the generalization of a single event into a negative impression of the entire region (Wetherell & Potter, 1992).

4.1.1 *The younger generation has good qualities. The older generation is disgraceful. This morning I saw many private cars in Nanyang offering free rides in the rain to take people back to their hotels.*

4.1.2 *No matter when, the four provinces will always be together. As a Hebei person, I believe most Henan people are great!*

4.2.1 *Thumbs up to the young people working hard to maintain the city's image.*

4.2.2 *Thanks to the editor for promoting Henan, especially my hometown, Xichuan County in Nanyang. You are welcome to visit.*

4.3.1 *Please don't let any thief go unpunished.*

4.3.2 *Nanyang needs a good cleanup, making all Henan people lose face.*

4.4.1 *As a person from Xichuan, Nanyang, the source of the South-to-North Water*

Diversion, I welcome everyone to come to Nanyang to verify. There are many simple and kind people in Nanyang. You are welcome to come to Nanyang, to Xichuan, to taste Danjiang water, and to see the seagulls.

Discussion

This article reveals how regional stereotypes and related discriminatory behaviors are narrated and constructed in new media comment sections. The study hypothesizes that the spread of regional stereotypes in comments constitutes a digital shift of symbolic violence on new media platforms. This phenomenon is characterized by the widespread portrayal of poverty and backwardness, reflecting socioeconomic disparities and unfair wealth distribution. The digital shift of these stereotypes exacerbates associated negative emotions. Furthermore, the established narrative frameworks are resistant to simple efforts of positive regional dissemination (Zhou, 2004).

In the news comment sections, regional stereotypes are narrated and constructed in a unique commentary area formed by Chinese new media and their related news activities on platforms, creating a relatively independent field. The media field is part of the broader media domain, and its development and renewal are subject to the interventions and constraints of previously established power fields such as economics and politics. The main components of news comment sections include publishers, users, and comment content. Therefore, studying the internal structure of these comment sections involves analyzing the information published, user

behavior habits in producing comments, and the comment content (Tang, 2020). This study uses the selection and optimization of the LDA model to construct discourse themes based on video topics from a large corpus of comment texts. This approach elucidates how regional stereotypes begin to be narrated, aiding in understanding how internet users' regional imaginations and the dissemination and development of these stereotypes as symbols occur in virtual communities.

In the news comment sections, the video titles serve as the initial catalyst for introducing topics and setting the agenda. This, in turn, lays the foundation for constructing the narrative field within the videos. Comments and interactions within the comment sections play a crucial role in inciting the initial momentum behind regional stereotypes. Notably, the censorship mechanisms adopted by Chinese news portals primarily aim to quell collective dissent against the ruling party rather than promote the dissemination of negative events about specific regions. This mechanism, to some extent, resists efforts of positive regional dissemination and reinforces regional stereotypes.

The study found that any discriminatory phenomena emerging between different regions are rooted in a collective symbolic dissatisfaction with the existing economic situation, leading to negligible symbolic violence exerted by the online community. Meanwhile, in the information field, there is a tendency to suppress more relevant and diverse discourses about a specific region, resulting in extreme expressions of cultural aphasia. On an individual level, digital nomads away from their place of origin exhibit high sensitivity to self-identity disorders. This susceptibility arises from their fear of potential isolation in social groups and gradual detachment

from actual territories. This psychological phenomenon significantly influences their non-virtual social behaviors.

Furthermore, the ironic and symbolic descriptions based on dialect significantly amplify the positive and diverse dissemination of symbols, reinforcing regional stereotypes. Commenters use emotionally charged words and descriptive expressions to further deepen the impact of regional stereotypes in the audience's minds. This narrative approach not only reinforces existing stereotypes but also adds an emotional factor, increasing the audience's acceptance of regional stereotypes. Particularly, by disguising irony with seemingly objective statements and using historical allusions and famous poems for metaphor, they suggest and emphasize negative characteristics of specific regions. This indirect speech act conveys meanings opposite to the literal expression through irony, further deepening regional stereotypes.

Notably, the real-name system for internet protocol addresses implemented by the Chinese government in the first half of 2023 provides support. This system facilitates the bidirectional interaction of geographic areas concerning topics and the geographic locations associated with commenters' internet protocol addresses. This mechanism sustains symbolic violence caused by regional stereotypes, affecting the cognition and behavior of information recipients. The technical structure of IP address announcements can also reinforce the simplified, rough, and one-sided connotations of the region in the audience's perception.

These four thematic frameworks collectively construct narrative threads around regional stereotypes in new media comment sections. Topic 1, the framework of objective event

narration, presents the basic facts and shows a lower level of low-grade emotional tendency in discussions related to the event location, Nanyang City. Topic 2 expresses views on local events through irony and rhetoric, containing evident sarcasm and negative emotions. Topic 3, the symbolic description framework, provides mechanisms for the construction and dissemination of regional stereotypes, showcasing two aspects of emotional tendencies. In Topic 4, the narrative logic of image restoration and reflection focuses on reversing and repairing regional images, revealing the possibility of correcting regional stereotypes on new media platforms, offering strategies and directions for improving and enhancing regional images.

In the framework of fact-based objective narration, public discussions often focus on the event itself rather than the formation of stereotypes. Therefore, media and content creators should provide the public with a comprehensive and balanced perspective when reporting on events with regional characteristics, thereby reducing potential biases in the reports and minimizing unfair or one-sided views of specific regions.

Within the dual framework of irony and rhetoric narration and symbolic description, commenters use dialect and historical cultural allusions to deliver sarcasm and biased views with strong emotional colors, utilizing cultural symbols to deepen regional stereotypes. To improve this, dialect should be used as a core element of local cultural dissemination, combined with lively and interesting themes to increase audience acceptance and interest in local dialect and culture. Additionally, focusing on spreading regional diversity and positive stories will

significantly balance and enrich public cognition of regions, guiding positive public opinion development.

In the narrative logic of image restoration and reflection, the public often calls for image restoration of the parties involved due to pride in their hometown or deep concern about the event. Regional governments and social organizations should actively respond to this demand by organizing and promoting positive social activities to repair and enhance regional images. Utilizing new media platforms to showcase the good image and positive aspects of the region can effectively reduce public stigmatization and promote a more comprehensive and positive understanding of the region.

This study helps us understand regional imagination in virtual communities and provides a reference for creating a more harmonious online environment. Future research can further explore reducing regional discrimination in new media through technical means and policy regulations, strengthening public education, and guiding users to form healthy online behavior habits. Additionally, refining the technical structure of news platforms will help improve the dissemination environment of regional stereotypes and promote more diverse and positive geographic information dissemination.

References

- Afridi, F., Li, S. X., & Ren, Y. (2015). Social identity and inequality: The impact of China's hukou system. *Journal of Public Economics*, 123, 17-29.
- An, Y. (2020). A history of instructional media, instructional design, and theories. *International Journal of Technology in Education*. <https://doi.org/10.13016/5f0da3d8.5f0e2b4c>
- Biot, M. A. (1956). Variational and Lagrangian methods in viscoelasticity.
- Bierstead, R. (1970). *The social order* (3rd ed., pp. 290-292). New York, NY: McGraw-Hill.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Bourdieu, P., Passeron, J. C., Nice, R., Bourdieu, P., & Bottomore, T. B. (2013). *Reproduction in education, society, and culture*. Los Angeles; London; New Delhi; Singapore; Washington, D.C.: Sage.
- Chevtava, E., & Denizci-Guillet, B. (2021). Digital nomads' lifestyles and coworkation. *Journal of Destination Marketing & Management*.
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Duan, Y., Song, X., & Chen, J. (2016). Sense of place: What it means to people. In D. Yifu, S. Xiukui, & C. Jinfeng (Eds.), *Aesthetic Studies: Spring 2016 Volume* (pp. 317-331).
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. T. (2007). Topics in semantic representation.

- Psychological Review, 114(2), 211-244.
- Hu, Q., Pan, X., Luo, J., & Yu, Y. (2022). The effect of service robot occupational gender stereotypes on customers' willingness to use them. *Frontiers in Psychology*, 13, 985501. <https://doi.org/10.3389/fpsyg.2022.985501>
- Hrebícková, M., & Graf, S. (2014). Accuracy of national stereotypes in Central Europe: Outgroups are not better than ingroup in considering personality traits of real people. *European Journal of Personality*, 28, 60-72. <https://doi.org/10.1002/per.1904>
- Huang, L., Clarke, A., Heldsinger, N., & Tian, W. (2019). The communication role of social media in social marketing: A study of the community sustainability knowledge dissemination on LinkedIn and Twitter. *Journal of Marketing Analytics*, 7, 64-75. <https://doi.org/10.1057/s41270-019-00053-8>.
- Jackson-Brown, G. (2010). Understanding community media. *Journalism & Mass Communication Educator*, 65(2), 141-156. <https://doi.org/10.1177/1077695810365957>
- Kaban, A., & Girolami, M. (2003). On an equivalence between PLSI and LDA. In *Annual ACM Conference on Research and Development in Information Retrieval* (pp. 433-434).
- Kravtsov, Y. A., Ostrovsky, L. A., & Stepanov, N. S. (1974). Geometrical optics of inhomogeneous and nonstationary dispersive media.
- Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the Stereotype Content Model. *International Journal of Intercultural Relations*, 30(6), 751–768. <https://doi.org/10.1016/j.ijintrel.2006.06.005>

- Lippman, W. (1965). Public opinion (pp. 63-64). New York, NY: Free Press.
- Lumsden, K., & Morgan, H. M. (n.d.). Cyber-trolling as symbolic violence: Deconstructing gendered abuse online. In N. Lombard (Ed.), *The Routledge Handbook of Gender and Violence*. Routledge.
- Malek, E. (2012). Half-maximal supersymmetry from exceptional field theory. *Fortschritte der Physik*, 60(4-5).
- Marengo, E. A. (2013). A new theory of the generalized optical theorem in anisotropic media. *IEEE Transactions on Antennas and Propagation*, 61(4).
- McNay, L. (1999). Gender, habitus and the field: Pierre Bourdieu and the limits of reflexivity. *Theory, Culture & Society*, 16(1), 95-117.
- Parikka, J. (2016). Deep times of planetary trouble.
- Peng, A. Y. (2021). Amplification of regional discrimination on Chinese news portals: An affective critical discourse analysis. *Convergence*, 27(5), 1343-1359.
<https://doi.org/10.1177/1354856520977851>
- Romim, N., Ahmed, M., Talukder, H., & Islam, M. S. (2020). Hate speech detection in the Bengali language: A dataset and its baseline evaluation. *arXiv.org [cs.CL]*.
- Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *UAI '04: Proceedings of the 20th Conference on Uncertainty and Artificial Intelligence* (pp. 487-494).
- Scolari, C. A. (2012). Media ecology: Exploring the metaphor to expand the theory.

Communication Theory, 5.

Van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, 4(2), 249-

283. <https://doi.org/10.1177/0957926593004002006>

Vitale, T. (2015). Comment on Loris Caruso/4. Territorial conflicts and new forms of left-wing

political organization: from political opportunity structure to structural contexts of

opportunities. *Sociologia*, 53(3).

Urry, J., & Larsen, J. (2011). *The tourist gaze 3.0*. Sage.

Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the*

legitimation of exploitation. Columbia University Press.

Xu, L., & Wen, C. (2023). "What is home": A study of the sense of place among young digital

nomads in a mobile society. *Chinese Youth Research*, 2023(08), 70-79.

<https://doi.org/10.19633/j.cnki.11-2579/d.2023.0119>

Fu, H. (2010). The paradigm of psychological research on regional culture and regional

stereotypes. *Social Psychological Science*, 25(5), 13-17.

Jiang, Y., & Zhang, H. (2009). A preliminary discussion on cross-cultural research of regional

personality stereotypes in China. *Social Science Frontier*, 24(10), 157-160.

Zhang, H., Jiang, Y., Zhao, W., An, G., Zhang, X., Hu, Z., & Zhang, W. (2012). Theoretical

exploration and empirical research on cross-cultural psychology in Chinese regions.

Advances in Psychological Science, 20(8), 1229-1236.

Zhang, H., & Jiang, Y. (2010). An empirical study on regional stereotypes: Taking the natives of

- Gansu and Inner Mongolia as research subjects. *Journal of Xihua University: Philosophy and Social Sciences Edition*, 29(3), 107-111.
- Chen, Q. (2022). Research on stereotypes of Northeast China in rural-themed short videos [Doctoral dissertation, Changchun University of Technology].
- Wang, X. (2021). The media construction of regional image and the diachronic transmission of stereotypes of Northeast China. *Social Sciences Front*, 2021(02), 267-273.
- Zhang, L. (2016). Regional conflict, attribution of responsibility, and individual attitude from the perspective of stereotypes: An empirical study based on the youth groups of Hong Kong and Guangzhou. *Jinan Journal (Philosophy and Social Sciences Edition)*, 38(05), 59-67+130.