Tourism Critiques: Practice and Theory

Are there any differences in the tourists' perceived destination image between travel e-commerce platforms and social media platforms?

Ya-Fei Liu, Yu-Bo Zhu, Hou-Han Wu and Fangxuan (Sam) Li
College of International Tourism and Public Administration, Hainan University, Haikou, China

Received 2 May 2024
Revised 3 August 2024
Accepted 6 August 2024

Abstract

Purpose – This study aims to explore the differences in the tourists' perceived destination image on travel e-commerce platforms (e.g. Ctrip and Fliggy) and social media platforms (e.g. Xiaohongshu and Weibo).

Keywords Tourists' perceived destination image, User-generated content, LDA model, Travel e-commerce platform, Social media platform

Paper type Research paper

1. Introduction

Tourism destination image reflects the overall consumer impressions of the tourist destination (Kotler, 2002). It is a crucial factor in destination development and tourist's travel decision (Li *et al.*, 2021). Previous studies have extensively explored the definition, construction and measurements of tourism destination images (Gartner, 1993; Joo *et al.*, 2023; Woosnam *et al.*, 2020). With the advancement of information technology, online platforms have become key channels for tourists to search for information, post comments and share travel experiences (Gaffar *et al.*, 2022). User-generated content (UGC) on travel e-commerce platforms and social media platforms is considered a reliable and trustworthy way to evaluate destination images, as consumers trust these channels in their decision-making process (Kaur *et al.*, 2016; Mao *et al.*, 2021).

Differences in tourists' perceived destination images derived from various online platforms can arise due to the geographical location of users or platform-specific biases. For instance, different social media platforms are preferred in different countries, such as Mafengwo in China and Tripadvisor in other countries, leading to different representation of the same destination across cultures (Lee and Park, 2023). Guo *et al.* (2021) found that platform-specific biases can cause variations in destination images, showing that Finland's

The research was partially supported by the Hainan Provincial Philosophy and Social Science Planning (Base) Program (HNSK(JD)22-5), and the Project of Higher Education Scientific Research of Hainan Province (No. Hnky2024-5).

Tourism Critiques: Practice and Theory
Emerald Publishing Limited
2633-1225
DOI 10.1108/TRC-05-2024-0018

destination image differs across five travel e-commerce platforms: Ctrip, Qyer, Mafengwo, Tuniu and Qunaer. Thus, understanding the multiple representations of a single on the internet is essential (Choi *et al.*, 2007).

In addition to travel e-commerce platforms, general social media such as Twitter, Instagram and Weibo are widely used as data sources to represent tourism destination images (Arefieva *et al.*, 2021; Molinillo *et al.*, 2018). Critical differences exist between travel e-commerce and social media platforms. Travel e-commerce platforms, such as TripAdvisor and Ctrip, primarily focus on facilitating online travel deals (Hua *et al.*, 2015). By contrast, general social media platforms like Twitter, Instagram and Weibo emphasize online communication and interaction (Akhtar *et al.*, 2023; Liu *et al.*, 2021). Relying solely on one type of platform to form a representation of a tourism destination image can introduce significant cognitive biases, leading to irrational destination development strategies and unwise travel decisions. Therefore, it is crucial to consider the limitations and biases associated with each platform when evaluating tourism destination images.

A research gap exists regarding the differences in the tourism destination images represented by UGC on travel e-commerce platforms and social media platforms. Further investigation is needed to understand and identify these differences. To fill this gap, this study explored the difference in tourism destination images between these two platforms through UGC analysis. The main contribution of this study is the use of both an inductive approach (the LDA model and co-occurrence analysis with community detection) and a deductive approach (the word frequency analysis with manual classification) to identify the difference between tourism destination images reflected by UGC on travel e-commerce platforms and social media platforms. The specific research objectives are to:

- retrieve UGC of a specific tourism destination from both travel e-commerce platform and social media;
- construct tourism destination image frameworks from UGC on different platforms using the LDA model, co-occurrence analysis with community detection and word frequency analysis with manual classification;
- identify the differences between the tourism destination image frameworks.

2. Literature review and hypothesis development

2.1 Tourists' perceived destination image

Tourists' perceived destination image reflects the overall consumer impressions of the tourist destination (Kotler, 2002). Gartner (1993)'s theory provides a foundational framework for understanding this image, proposing that it comprises three key components: cognitive, affective and conative images. Cognitive images pertain to how tourists evaluate a destination based on their initial visit or subsequent visits (Joo *et al.*, 2023). Affective images relate to tourists' emotional responses and feelings toward the destination (Lee and Jeong, 2023). Conative images are concerned with future behaviors and intentions, which are influenced by the cognitive and affective components (Gartner and Witt, 1994; Woosnam *et al.*, 2020). The overall perceived destination image includes not only tourists' beliefs, ideas and impressions but also their thoughts and emotions about the destination's attributes (Stylos *et al.*, 2017; Chaulagain *et al.*, 2019). It encompasses both functional and psychological characteristics (Tung *et al.*, 2021), and can explain individuals' reactions and behaviors toward destinations (Afshardoost and Eshaghi, 2020).

Many researchers have worked on the tourism destination's perceived image through attributed-based or holistic approaches (Kock *et al.*, 2016; Liu *et al.*, 2021). For example,

Lee and Park (2023) expanded on Beerli and Martin's theory by categorizing destination image attributes into nine dimensions:

(1) natural resources;
(2) general infrastructure;
(3) tourist infrastructure;
(4) tourist leisure and recreation;
(5) culture, history and arts;
(6) political and economic factors;
(7) natural environment;
(8) social environment;
(9) atmosphere of the place.

This framework represents a significant advancement in the field and was used in this study using a deductive approach to construct the tourism destination image.

2.2 Tourist' perceived destination image and user-generated content

Web 2.0 technologies have enabled tourists to create and share opinions on various online platforms (Mak, 2017). Users frequently upload and share blogs, text, photos and videos, expressing their "feelings" about a tourism destination (Hunter, 2016; Taecharungroj and Mathayomchan, 2021). This UGC has become a key driver of visitor behavior, destination choice, e-word of mouth and ultimately, the reputation and success of a destination (Liu *et al.*, 2021). The online perceived image of a destination is more dynamic than its traditional counterpart, reflecting collective beliefs, knowledge, ideas, feelings and overall impressions (Hunter, 2013). Online representations through text and photographs complement and enhance traditional images found in brochures, guidebooks and surveys (Alarcon-Urbistondo *et al.*, 2023).

Advanced technological methods such as natural language processing (NLP) techniques, artificial intelligence (AI) framework and the latent Dirichlet allocation (LDA) model are commonly used to study online destination images (Lin *et al.*, 2021; Kaur *et al.*, 2016). For example, NLP techniques have been used to explore the impact of safety on online perceived destination image (Marine-Roig and Huertas, 2020), while AI frameworks assist in identifying tourism images from uploading photos (Wang *et al.*, 2020). Li *et al.* (2022) developed a multi-level visual system to analyze cognitive themes and emotional experiences of various destination images. The LDA model is particularly significant for extracting topics from UGC to measure destination image and e-reputation (Dong *et al.*, 2023; Ali *et al.*, 2021). Overall, UGC combined with advanced technological methods, such as the LDA model, provides an inductive approach to studying destination images, allowing for comprehensive and inclusive evaluation of a destination's attributes (Alarcon-Urbistondo *et al.*, 2023).

2.3 The difference between travel e-commerce platforms and social media platforms

Travel e-commerce platforms and social media platforms are two primary data sources for destination image research (Li *et al.*, 2023). On social media platforms, tourism-related photos and texts are used to investigate the perceived image of tourism destinations (Arabadzhyan *et al.*, 2021; Narangajavana *et al.*, 2017). Molinillo *et al.* (2018) compared UGC from various social media platforms such as Instagram, Facebook, YouTube and Twitter to explore differences in tourism's perceived image. By contrast, UGC from travel

e-commerce platforms, primarily online review, is used to establish perceived destination image and measure consumer satisfaction (Guo *et al.*, 2017; Hou *et al.*, 2019; Liu *et al.*, 2018; Zhang *et al.*, 2016).

Despite the importance of these platforms, the difference between travel e-commerce platforms and social media platforms remain unexplored. Studies often focus on one type of platform, either travel e-commerce or social media, without comparing the two (Guo *et al.*, 2021; Liu *et al.*, 2021; Molinillo *et al.*, 2018). This lack of comparative research limits our understanding of the similarities and differences between these platforms. Xiang *et al.* (2017) highlighted that relying on a single data source restricts the generalizability and contribution of social media analytics. Their study demonstrated significant variations in linguistic characteristics, semantic features, sentiment, rating and usefulness among online reviews on TripAdvisor, Expedia and Yelp.

The distinctions between travel e-commerce platforms and social media platforms are clear. Travel e-commerce platforms operate as internet-based business models providing online trading and communication platforms for tourists and tourism suppliers, focusing on trading and evaluation functions (Hua, 2016). By contrast, social media platforms are internet-based applications build on Web 2.0 principles, allowing the creation and exchange of UGC, emphasizing communication and sharing functions (Kaplan and Haenlein, 2010; Akhtar *et al.*, 2023). Additionally, travel e-commerce platforms restrict review posting to users who have booked and consumed a certain travel product on the platform, whereas social media platforms impose no such restrictions, allowing users to freely post reviews, feelings or thoughts about a destination or travel product (Huang, 2024; Zeng and Gerritsen, 2014). Consequently, UGC on these platforms may reflect different images of a destination. Based on these differences, this study proposes the following research question (RQ):

RQ. Are there any differences in the perceived destination image derived from UGC on different online platforms? If so, what are the differences?

3. Methodology

3.1 Data collection

Sanya, located on the southernmost tip of Hainan Island in China, is often referred to as the "Hawaii of China" because of its stunning beaches, clear turquoise waters and tropical climate. This popular tourist destination generates a large amount of UGC on both travel e-commerce platforms and social media, making it an ideal location for studying perceived destination image across different types of platforms.

To collect data, we used Octoparse (https://www.octoparse.com), a robust web scraping tool that efficiently extracts both textual and metadata from various websites (Yu *et al.*, 2020; Yu and Egger, 2021). Data was collected in two phases: May 2023 and July 2024. This two-stage approach enhances the robustness of our results by increasing the data volume and mitigating the potential time bias.

This study draws data from four platforms: Ctrip, Fliggy, Xiaohongshu and Weibo. Ctrip and Fliggy represent travel e-commerce platforms, while Xiaohongshu and Weibo are social media platforms. These platforms were selected for three reasons: their widespread use among Chinese users; their representativeness in their respective fields; their frequent use as data sources in tourism image research.

Ctrip (http://www.ctrip.com), established in 1999, is a leading Chinese travel e-commerce platform with over 700 million users by 2023 (Hou *et al.*, 2019). Fliggy (https://www.fliggy.com), also known as Fei Zhu, is an Alibaba Group-owned online travel service platform launched in 2010. We used "Sanya" as a keyword to search information in Ctrip and

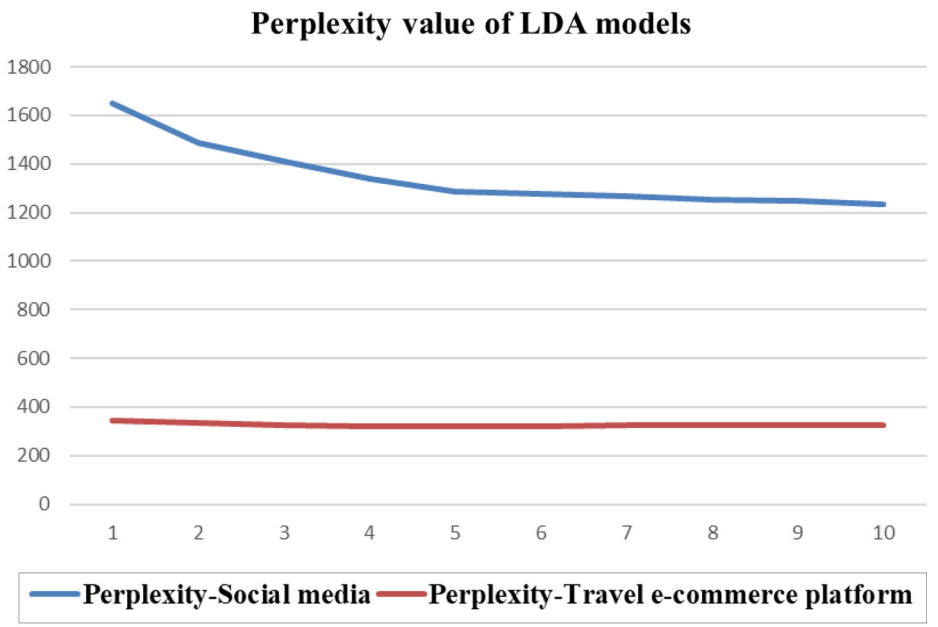
Fliggy, collecting tourism reviews for 47 hotels and 14 attractions in Sanya. This resulted in over 35,000 reviews from Ctrip and approximately 16,400 reviews from Fliggy.

Xiaohongshu (https://www.xiaohongshu.com) is a prominent social media platform in China where users share daily life experiences. Weibo, launched by Sina Corporation in 2009 and often referred to as the "Chinese Twitter," is another major social media platform. For Xiaohongshu and Weibo, we collected UGC related to "Sanya Tourism," "Sanya" and the same 47 hotels and 14 attractions. This effort yielded over 23,000 reviews from the Xiaohongshu, and more than 16,000 reviews from Weibo.

3.2 Research procedure

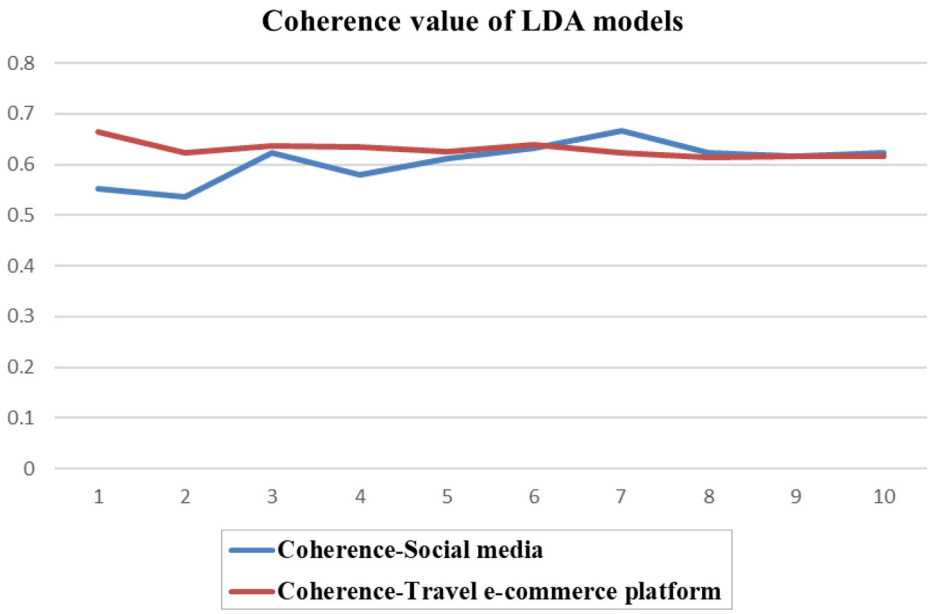
This study preformed preprocessing operations in Python such as word separation and removal of deactivated on the corpus source files to obtain the document-word matrix. We used the widely recognized Chinese stop words list developed by the Harbin Institute of Technology (Hou *et al.*, 2019). This study used three methods to extract attributes of tourist's perceived image:

(1) The LDA model. The LDA model is a widely used for theme extraction from UGC (Seyyedamiri *et al.*, 2022). We used Python to perform LDA analysis. To determine the optimal number of topics, we ran the LDA model with topic numbers ranging from 2 to 10, calculating the perplexity and coherence value of each model. Figures 1 and 2 present the perplexity value and coherence value for the LDA models applied to of UGC from both platform types. A low perplexity value indicates the model's better grasp of the underlying topic distribution in the data (Blei *et al.*, 2003), while a high



Note: The number on abscissa axis is the number of topics
Source: By authors

Figure 1. The perplexity value of LDA models



Note: The number on abscissa axis is the number of topics
Source: By authors

Figure 2. The coherence value of LDA models

coherence value shows that the words within the topics are more related, forming coherent and meaningful concepts (Röder *et al.*, 2015). The visualization results enable researchers to assess whether the topics are reasonable and relevant to the research objectives (Sakiyama *et al.*, 2020). Combining the perplexity and coherence values with the visualization results, we determined the optimal number of topics for the LDA models. For UGC on travel e-commerce platforms, an LDA model with seven topics was identified as optimal. Similarly, for UGC on social media platforms, seven topics also proved to be most appropriate number. These topics and sub-keywords collectively form the tourism destination image of Sanya from the tourist's perspective. This method follows an inductive approach to building tourism destination images.

(2) Co-occurrence analysis with community detection. First, we preprocessed the data through tokenization and removal of stop words to create a 1,000 × 1,000 matrix where both rows and columns represent the top 1,000 most frequently occurring keywords. Each cell value in the matrix indicates the number of times the corresponding keyword pairs co-occur in the data set. The resulting co-occurrence matrix was then imported into Gephi, a clustering analysis program, where a community detection algorithm (Modularity Class) was applied to segment the data from travel e-commerce platforms and social media platforms.

(3) Word frequency analysis and manual classification. Initially, we used Python to analyze the word frequency in tourists' reviews from travel e-commerce platforms and social media platforms. Subsequently, based on the mutual destination image

framework (Beerli and Martin, 2004; Lee and Park, 2023), we manually classified the high-frequency words into destination image attributes and explored new destination image attributes. This method follows a deductive approach to build a tourism destination image.

4. Results

4.1 The result of latent Dirichlet allocation analysis

(1) Extracting attributes of tourism destination image from UGC on travel e-commerce platform using the LDA model

Table 1 presents the results obtained from applying the LDA model to UGC on travel e-commerce platforms. Figure 3 shows the visualization of these results. In Figure 3, there are seven groups of words, the words in the right represent the top 30 relevant terms in the data set. By clicking the "next topic" button, you can view the top 30 terms for each specific topic. The topics and corresponding feature words in Table 1 are derived from the most relevant terms identified in the visualization. The LDA model derived seven distinct topics from tourist reviews, primarily focused on aspects of hotel service in Sanya.

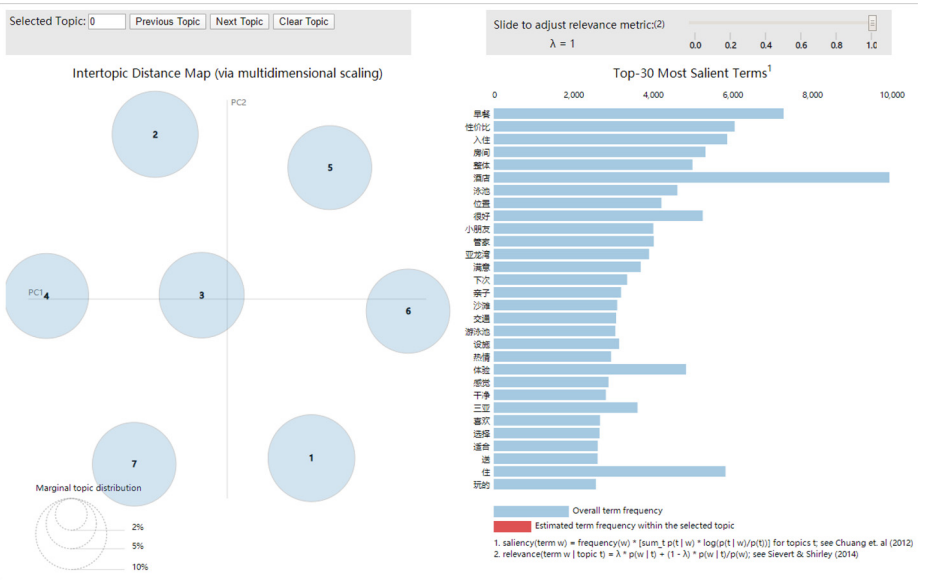
For example, the feature words for Topic 1 include breakfast, check-in, hotel, facility, environment, vacation, Marriott, Sheraton, complete, sea view room, etc. These words relate to the general perceptions of tourists to hotels. Therefore, we categorize Topic 1 as "General perceptions of hotels," and consider it a critical attribute of tourism perceived destination image of Sanya. The rest of topics are:

- Service quality. It focuses on the quality of services provided by hotels, emphasizing reliability, helpfulness and the atmosphere in guest rooms and restaurants.
- Tourist satisfaction. This topic centers on tourists' satisfaction, discussing value for money and their likelihood of revisiting.

Table 1. Topics and feature words of LDA model of UGC on travel e-commerce platform

Topics	%	Feature words
General perceptions of hotels	15.0	Breakfast, check-in, hotel, facility, environment, vacation, Marriott, Sheraton, complete, sea view room
Service quality	14.7	Overall, hotel, feeling, reliable, sea view, quiet, variety, restaurant, help, guest room
Tourist satisfaction	14.4	Value for money, like, satisfaction, next time, choice, service, revisit, thoughtful, wonderful, considerate
Comfort and cleanliness	14.2	Room, enthusiastic, very good, clean, comfortable, facilities, experience, staff, free, decoration
Convenience and accessibility	14.0	Location, beach, traffic, Haitang Bay, accommodation seaside, variety, close, convenience
General hotel experiences	13.9	Very good, swimming pool, food, nice, large, cleanliness, price, duty-free store, trip, quiet
Family-friendly activities	13.9	Children, like, playing, activity, children's playground, water world, amusement facility, patience, aquarium

Source: By authors



Source: By authors

Figure 3. The result of the LDA model on travel e-commerce platform data

- Comfort and cleanliness. It addresses the comfort and cleanliness of hotel rooms, as well as the quality of room service and facilities.
- Convenience and accessibility. This topic highlights the eases of access to key locations, such as beaches and attractions like Haitang Bay.
- General hotel experiences. It encompasses tourists' enjoyment of amenities like swimming pools, dining options and overall cleanliness.
- Family-friendly activities. This topic refers to facilities and activities suitable for families, including playgrounds, water parks and other amusement facilities.
- Extracting attributes of tourism destination image from UGC on social media using the LDA model

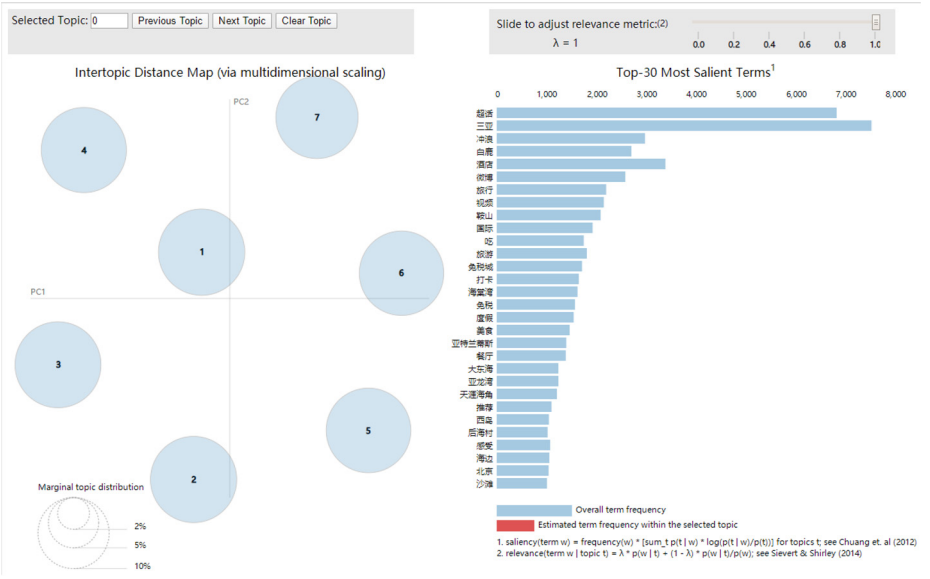
Table 2 details the results of the LDA model applied to UGC on social media platforms, while Figure 4 provides a visualization of these findings. The topics and corresponding feature words in Table 2 are derived from the most relevant terms identified in the visualization. The LDA model identified seven topics within UGC on social media platforms. For example, the feature words of Topic 1 include Sanya International Duty Free Shopping Center, influencer hotspot, beach, seascape, photography, Nanshan, scenery, sea, tropics, speed boat, etc. These feature words basically describe the scenic spots of Sanya; Thus, Topic 1 is named as "Scenic spots." The rest topics are:

- Social events and celebrity endorsements. It highlights the impact of social events and celebrity endorsements on tourism in Sanya.

Table 2. Topics and feature words of LDA model of UGC on social media

Topics	%	Feature words
Scenic spots	14.7	Sanya International Duty Free Shopping Center, influencer hotspot, beach, seascape, photography, Nanshan, scenery, sea, tropics, speed boat
Social events and celebrity endorsements	14.7	Celebrity, Weibo, experience, activities, enjoy, friends, tourists, consumption, entertainment, charm
Hotel recommendations and experiences	14.6	Hotel, Haitang Bay, recommendation, accommodation, service, new, fun, reservation, sea, free
Shared experiences and tips	14.3	Weibo super topic, Weibo, video, duty-free, vacation, travel tip, Wuzhizhou Island, life, record, shopping festival
Photo-worthy locations	14.2	Yalong Bay, sea, sharing, low light, forest park, beautiful, scenery, filter, weather, darkness
Relaxation spots	14.1	Sand beach, island, play, sky, choice, sunshine, suggestion, routine, comfortable, accommodation
Local foods and special activities	13.4	Surfing, eating, delicious, wake surfing, Zaopi Vinegar, swimming pool, taste, address, hot pot, dining

Source: By authors



Source: By authors

Figure 4. The result of the LDA model on social media data

- Hotel recommendations and experiences. The topic covers recommendations and experiences related to hotel accommodations in Sanya, particularly in well-known hotel clusters in Haitang Bay.
- Shared experiences and tips. It emphasizes vacation activities, shopping and notable sites like Wuzhizhou Island, as shared by tourists on social media platforms.
- Photo-worthy locations. This topic includes recommendations for scenic spots and photo worthy locations in Sanya, such as Yalong Bay, and consider weather conditions.
- Relaxation spots. It highlights popular relaxation spots in Sanya, including beaches, islands and comfortable accommodations.
- Local foods and special activities. This topic focuses on tourists' recommendations of local foods and special sports activities.

4.2 The result of co-occurrence analysis with community detection

The co-occurrence analysis method, enhanced by community detection algorithms, provides a more in-depth understanding of keyword relationships within UGC. This approach goes beyond simple frequency analysis and LDA model, offering insights into the relational dynamics of terms within the data set (Liu *et al.*, 2019; Zhang *et al.*, 2019). By analyzing how keywords co-occur, we can identify key themes and trends that may otherwise remain hidden. The community detection algorithm clusters keywords into distinct groups based on their co-occurrence matrix, revealing the underlying structure and thematic composition of the data. This method complements the LDA results by capturing the context and interrelationships between words in a more nuanced way:

(1) Extracting attributes of tourism destination image from UGC on travel e-commerce platforms using co-occurrence analysis with community detection.

Table 3 presents the results from the co-occurrence analysis conducted on UGC from travel e-commerce platforms. Figure 5 illustrates the co-occurrence matrix-based topic model. The topics and corresponding feature words in Table 3 are summarized based on the most relevant terms identified from the co-occurrence matrix diagram in Figure 5. The analysis, visualized using Gephi, categorizes the travel e-commerce platforms data into seven topics:

- natural and photogenic aspects of Sanya;
- tourists' experiences with hotels;
- quality of service provided by hotel staff;
- cleanliness, facilities and overall comfort of guest rooms;
- suitability for family-friendly activities;
- tourists' experiences with food and dining; and
- general satisfaction of tourists.

(2) Extracting attributes of tourism destination image from UGC on social media using co-occurrence analysis with community detection.

Table 4 presents the findings from the co-occurrence analysis conducted on UGC on social media platforms. Figure 6 illustrates the result of the co-occurrence matrix-based topic model. The topics and corresponding feature words in Table 4 are summarized based on the

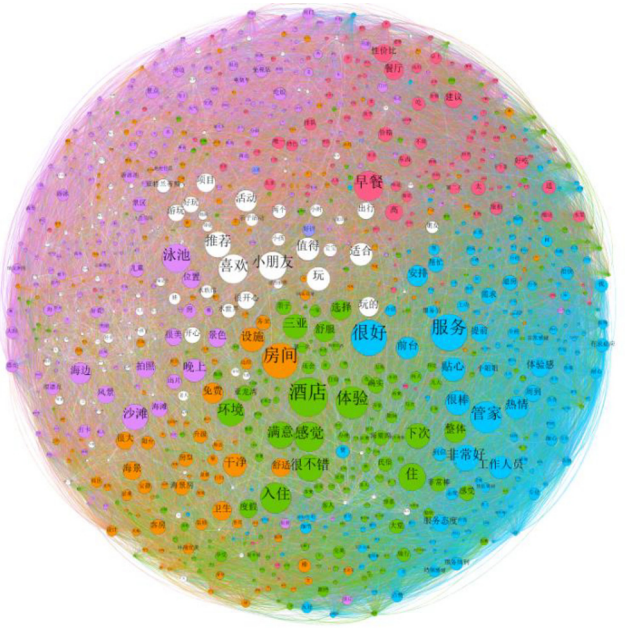
Table 3. Topics and feature words of co-occurrence matrix diagram on travel e-commerce platform

Topics	%	Feature words
Natural and photogenic aspects of Sanya	22.5	Pool, beach, location, scenery, photography, seaside, evening, landscape, beautiful, duty-free store, children, coast, room, photogenic, tourist area, influencer hotspot, transportation, very beautiful, ocean, sea
Tourists' experiences with hotels	20.0	Hotel, check-in, experience, environment, accommodation, Sanya, next time, satisfaction, very good, feeling, choice, will return, overall, vacation, family, comfortable, will come again, Yalong Bay, B&B, Haitang Bay
Quality of service provided by hotel staff	17.9	Service, very good, butler, enthusiastic, considerate, excellent, front desk, staff, arrangement, great, in advance, help, attentive, young lady, service attitude, sense of experience, check-out, needs, thorough, introduction
Cleanliness, facilities, and overall comfort of rooms	14.5	Facilities, clean, hygiene, upgrade, sea view, free, comfortable, spacious, complete, sea view room, balcony, guest room, tidy, room type, decor, cleaning, quiet, design, returned
Suitability for family-friendly activities	12.8	Children, like, recommend, play, suitable, worth it, fun, playing, activities, very happy, activity, happy, playing, aquarium, Atlantis, outing, baby, water park, exciting, cute
Tourists' experiences with food and dining	11.9	Breakfast, value for money, restaurant, high, delivery, queue, delicious, price, suggestion, travel photography, eat, fruit, taste, gift, variety, too much, things, a lot, no need, package
General satisfaction of tourists	0.4	Good review, five stars, money, spend

Source: By authors

most relevant terms identified from the co-occurrence matrix diagram in Figure 6. The analysis were visualized using Gephi, categorizing the social media data into 11 topics:

(1) tourists' experiences and recommendations related to various activities and attractions;
(2) services and infrastructure available to tourists;
(3) natural beauty of Sanya;
(4) tourists' recommendations and experiences with local cuisine;
(5) tourists' recommendations of hotels;
(6) cultural and historical attractions;
(7) recommendations for photogenic locations and activities;
(8) media coverage and news related to Sanya tourism;
(9) theme parks in Sanya;
(10) aspects of daily life and personal documentation; and
(11) air transportation facilities.



Source: By authors

Figure 5. The co-occurrence matrix diagram of travel e-commerce platform

4.3 The result of word frequency analysis

In the word frequency analysis, we examined 36,439 words from UGC on travel e-commerce platforms and 67,084 words from UGC on social media platforms. Words appearing 300 times or more were selected to shape the perceived destination image of Sanya (Jiang *et al.*, 2021). These high-frequency words were classified into various dimensions and attributes using the perceived tourism image framework proposed by Beerli and Martin (2004) and Lee and Park (2023). Then, we calculated the weight of each dimension according to the following Format 1.

Format 1:

$$Weight_a = \frac{\sum_1^n Frequency\ of\ Word_i}{\sum_1^m Frequency\ of\ Word_j}$$

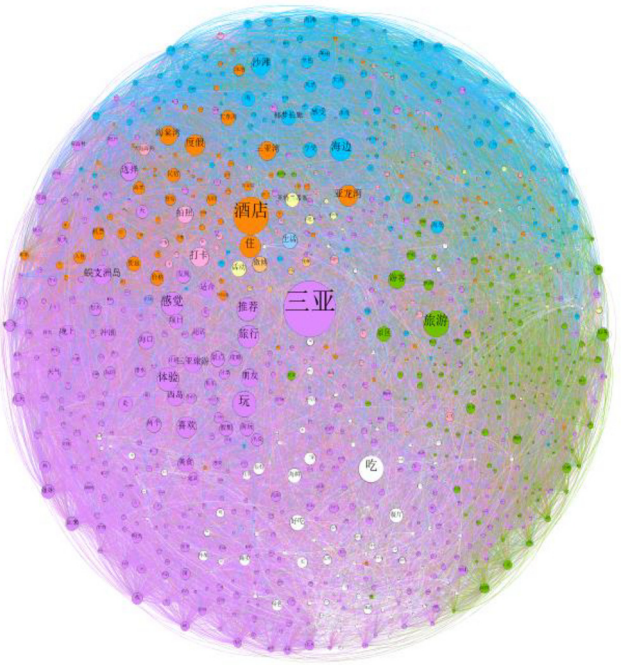
In Format 1, *weight* refers to the weight of attribute *a*, *i* represents the high-frequency words that belong to attribute *a*, and *n* represents total number of words in attribute *a*. The *Frequency of Word i* is the number of times word *i* appears in the data set. *j* represents the high-frequency words on travel e-commerce platform and social media platform, and the *Frequency of Word j* is the number of times word *j* appears in the data set. *m* represents total number of frequency words used to construct the tourism perceived destination image. For the travel e-commerce platforms, *m* is 294, and for the social media platforms, *m* is 201.

The dimensions and attributes identified in the analysis are summarized in Tables 5 and 6, which list the representative high-frequency words associated with each attribute. The perceived destination image of Sanya, as derived from UGC on travel e-commerce platform,

Table 4. Topics and feature words of co-occurrence matrix diagram on social media

Topics	%	Feature words
Tourists' experiences and recommendations related to various activities and attractions	39.1	Sanya, travel, play, super topic, recommendation, cuisine, experience, surfing, Wuzhizhou Island, feeling, Sanya tourism, like, West Island, holiday, Haikou, project, strategy, May Day, friends, night
Services and infrastructure available to tourists	16.3	Tourism, tourist, Sanya International Duty Free Shopping Complex, scenic spot, yacht, consumption, duty-free, service, leisure, Luhuitou, Spring Festival, new, period, high, cultural tourism, free, in advance, shopping, transportation
Natural beauty of Sanya	14.0	Beach, seaside, ocean, sea, Coconut Dream Corridor, feeling, island, enjoy, world, ocean, tropical, beautiful, tour, coast, scenery, sunset, romantic, sunshine, wonderful, forest park
Tourists' recommendations and experiences with local cuisine	10.7	Eat, seafood, delicious, restaurant, family, Zaopei vinegar, buy, coconut chicken, duty-free store, taste, very good, specialty, environment, this restaurant, coconut, hotpot, food, indeed, Qingbuliang, night
Tourists' recommendations of hotels	9.3	Hotel, vacation, Yalong Bay, Haitang Bay, stay, Sanya Bay, ticket, B&B, Dadonghai, price, sea view, swimming pool, cheap, Phoenix Island, check-in, accommodation, expensive, East, center, family
Cultural and historical attractions	3.2	Culture, Tianya Haijiao, Nanshan, freedom, maritime, cultural tourism, walk, matter, tourist area, South Sea, Tianya, tradition, rainforest, Guanyin, nearby, area, sport, review, Yanoda, history
Recommendations for photogenic locations and activities	2.2	Influencer hotspot, photograph, wear, beautiful, picture, internet celebrity, webpage, one photo, photogenic, clothes, dress, details, click, match, celebrity, swimsuit, lottery, immersive
Media coverage and news related to Sanya tourism	1.8	Weibo, video, release, reporter, livelihood, news, say, post, live broadcast, traveler, media, text, headline, moments, resort, daily, Sina, Hainan Daily
Theme park	1.7	Atlantis, activity, Water World, theme, The Lost Chambers Aquarium, water, show, amusement park, series, display, pink, content, promotion
Aspects of daily life and personal documentation	1.4	Life, feeling, two adults and one child, official, ceremony, record, festival, one adult and one child, ideal, launch, parade, camping, full
Air transportation facilities	0.3	Airport, Phoenix International Airport

Source: By authors



Source: By authors

Figure 6. The co-occurrence matrix diagram of social media

consists of eight dimensions and 20 attributes. Meanwhile, UGC on social media platforms reflects nine dimensions and 21 attributes.

(1) Extracting attributes of tourism destination image from UGC on travel e-commerce platforms with word frequency analysis

From the word frequency analysis of UGC on travel e-commerce platforms, it was found that the tourist infrastructure (37.1%) and the atmosphere of the place (27.5%) are the two most influential dimensions shaping Sanya's perceived tourism image. Within the atmosphere of the place dimension, attributes such as "pleasant" and "satisfactory" are highly emphasized. In the tourist infrastructure dimension, "hotels" stands out as a critical attribute. Other significant dimensions include service (13.0%), natural environment (7.2%) and tourist leisure and recreation (6.3%). Although dimensions like social environment (3.9%), general infrastructure (3.2%) and price level (1.8%) are associated with fewer high-frequency words, they still contribute to the overall tourism perceived image of Sanya.

Notably, the study identified some new attributes that were not present in the previously established destination image frameworks. These include duty-free stores, ocean, special sports and theme parks, which underscore Sanya's unique advantages as a tourism destination.

(2) Extracting attributes of tourism destination image from UGC on social media platforms with word frequency analysis

Table 5. Dimensions and attributes of the perceived destination image of Sanya derived from UGC on travel e-commerce platforms

Dimensions	Attributes	Keywords
Tourist infrastructure (37.1%)	Hotel (84.0%)	Hotel (30,565), room (16,488), check-in (11,991), butler (11,471), experience (7,454), environment (6,920), facility (6,029), breakfast (5,382), reception (3,397), check-out (1,642)
	Restaurant (7.5%)	Restaurant (2,091), eat (1,952), delicious (1,282), taste (891), buffet (684), dinner (508), food (479), cake (432), take-out (398), Chinese restaurant (304)
	Attraction (6.7%)	Yalong Bay (1,797), Haitang Bay (1,539), attraction (1,180), Wuzhizhou Island (961), tourist spot (906), park (773), instructor (771), West Island (763), river (472), ticket (369)
	Duty-free store (1.8%)	Duty-free store (1,494), shopping (1,433)
Atmosphere of the place (27.5%)	Pleasant (56.5%)	Good (12,866), like (4,912), very good (4,006), wonderful (3,490), interesting (2,798), happy (2,452), comfortable (2,294), pleasant (1,590), exciting (1,121), nice (958)
	Satisfactory (28.2%)	Next time (5,451), satisfactory (4,440), recommendation (3,979), suitable (3,417), still come (2,986), revisit (2,456), top (545), first-rate (483), first-choice (463), deserving (335)
	Family-oriented (14.2%)	Child (7,406), parents and children (2,322), infant (1,228), family (789), parents-child activity (739), elder (594), children's activity (453)
	Unique (1.1%)	Special (678), sole (615)
Service (13.0%)	Service (100%)	Service (23,009), enthusiastic (5,280), thoughtful (5,171), service attitude (1,906), thorough (1,788), attitude (1,370), active (1,344), patient (1,327), attentive (1,098), grant whatever is requested (1,024)
Natural environment (7.2%)	Overall environment (54.3%)	Overall (2,761), scenery (2,303), photo (2,141), picturesque (1,266), sight (748), vision (723), beautiful environment (690), weather (579), hot (491)
	Ocean (45.7%)	Beach (3,872), seascape (2,382), seaside (2,136), sea (2,025), sand (532)
Tourist leisure and recreation (6.3%)	*Theme Park* (57.4%)	Amusement facility (2,613), aquarium (1,514), Atlantis (1,360), water world (1,210), white whale (607), mermaid (557), marine organism (535), rafting (475), children's playground (459), performance (418)
	Special sport (35.2%)	Swimming pool (6,563), swimming (687), infinity pool (601), diving (394), beach comb (356)
	Nightlife (7.4%)	Night (2,000)

(*continued*)

Table 5. Continued

Dimensions	Attributes	Keywords
Social environment (3.9%)	Cleanness (58.8%)	Clean (5,438), sanitation (3,123), neat (1,223)
	Friendly resident (31.0%)	Helpful (2,350), help (2,152), friendly (653)
	Overcrowded (10.2%)	Line-up (1,384), crowded (319)
General infrastructure (3.2%)	Transportation (85.0%)	Transportation (1,108), electro mobile (911), close (909), car (633), convenience (619), distance (607), airport (574), walking 9531), sightseeing vehicle (469), taxi (347)
	Infrastructure (15.0%)	Facility (436), hardware (433), well-equipped (399), public lavatory (333)
Price level (1.8%)	Price level (100%)	Free (2,454), cost-effective (2,450), price (1,181), expensive (497), cheap (374), charge (357)

Notes: Attributes in italics refer to those attributes that are not from the mutual destination image framework but are newly identified by text mining analytics in this study. The numbers in the bracket behind dimensions and attributes represent their weights, and the numbers in the bracket behind keywords represent their frequencies
Source: By authors

Table 6 Dimensions and attributes of the perceived destination image of Sanya derived from UGC on social media platforms

Dimensions	Attributes	Keywords
Tourist infrastructure (37.6%)	Attraction (41.1%)	Yalong Bay (2,342), Haitang Bay (1,956), Sanya Bay (1,487), Wuzhizhou Island (1,410), West Island (1,281), attraction (1,178), Houhai (1,096), Dadonghai (994), Coconut Dream Corridor (943), Clear Water Bay (923)
	Hotel (26.4%)	Hotel (8,988), B&B (1,059), accommodation (544), sea-view room (385), Edition (378), villa (362), Marriott (351), Mandarin (330)
	Duty-free store (13.6%)	Duty-free (1,781), promotion (1,364), shopping (684), duty-free store (559), brand (388), off-island (377), product (317)
	Special food (11.0%)	Seafood (1,611), Zaopi Vinegar (685), coconut chicken (580), coconut (436), Qingbuliang (396)
	Restaurant (7.9%)	Delicious (1,549), restaurant (1,249), taste (556), hot pot (434), dining (347)
Recommendation (18.4%)	Recommendation (100%)	Recommendation (2,002), influencer hotspot (1,173), travel tip (1,104), discovery (732), link (601), inquiry (530), wearing (517), advice (511), sharing (470), sunscreen (321)
Natural environment (13.7%)	*Ocean* (52.4%)	Sand beach (1,499), seaside (1,459), sea (1,024), island (912), seascape (867), ocean (768), beach (601), seawater (428), sea breeze (353), wave (300)
	Overall environment (47.6%)	City (1,210), weather (885), hot (785), tropics (710), scenery (558), sunset (554), rain (515), sunshine (513), summer (469), nature (360)
Tourist leisure and recreation (9.8%)	*Special activity* (41.0%)	Surfing (1,475), yacht (1,127), diving (883), swimming pool (822), swimming (611), beach comb (310)
	Theme park (38.7%)	Atlantis (1,656), amusement facility (1,111), water world (630), Eternal Love (444), aquarium (349)
	Nightlife (10.5%)	Night (895), night market (537)
	Recreation (9.8%)	Relaxation (872), entertainment (465)
Atmosphere of the place (9.3%)	Pleasant (75.4%)	Like (1370), enjoyable (779), happy (572), free (564), good (542), nice (514), exciting (421), comfortable (340), attractive (309)
	Family-oriented (17.8%)	Family (693), parents and children (422), mother (408), child (397), infant (392)
	Unique (4.1%)	Special (533)
	Satisfactory (2.8%)	Deserving (360)

(*continued*)

Table 6 Continued

Dimensions	Attributes	Keywords
General infrastructure (4.9%)	Transportation (100%)	Air ticket (1,287), walking (737), car rental (721), airport (617), trip (568), transportation (478), flight (456), plane (427), center (410), Phoenix Airport (385)
Price level (3.0%)	Price level (100%)	Consumption (1,057), price (995), cheap (739), free (530), expensive (519)
Culture, history and art (2.6%)	Culture (77.1%)	Culture (1,028), cultural tourism (1,014), East (453), civilization (304)
	Celebrity effect (22.9%)	Wu Lei (418), internet celebrity (415)
Service (0.6%)	Service (100%)	Service (901)

Note: Attributes in italics refer to those attributes that are not from the mutual destination image framework but are newly identified by text mining analytics in this study. The numbers in the bracket behind dimensions and attributes represent their weights, and the numbers in the bracket behind keywords represent their frequencies.
Source: By authors

The perceived destination image of Sanya derived from UGC on social media platforms consists of nine dimensions and 21 attributes. The weight of each dimension is distributed as follows: 37.6% for tourist infrastructure, 18.4% for recommendation, 13.7% for natural environment, 9.8% for tourist leisure and recreation, 4.9% for general infrastructure, 3.0% for price level, 2.6% for culture, history and art, 0.6% for service. Some novel attributes emerged from the analysis of social media platforms, including recommendations, satisfactory, duty-free stores, special food, celebrity effect, ocean, theme parks and special activities. These attributes reflect the distinct nature of UGC on social media platforms, where users often share personal experiences, provide recommendations or seek advice about specific destinations.

These findings highlight the varied dimensions that shape Sanya's perceived image and reveal how different platforms can offer unique insights into what tourists value in their travel experiences.

Table 7 provides a comprehensive overview of the data sets and the results obtained from the three methods.

5. Discussion and conclusion

Tourists' perceived destination image is crucial in destination management. Previous studies have emphasized the construction of destination images and their impacts on tourism behavior and destination development (Lee and Jeong, 2023; Molinillo *et al.*, 2018). The importance of destination image in marketing strategy and tourism choice is well documented (Afshardoost and Eshaghi, 2020; Hallmann *et al.*, 2015). However, accurately measuring the destination image remains a challenge for both scholars and destination managers (Beerli and Martin, 2004; Bui *et al.*, 2022). With the development of information technology, leveraging UGC and machine learning has emerged as a comprehensive approach to measure destination images accurately (Buhalis and Law, 2008).

5.1 Main findings

Our study identified several key differences in the tourism destination image between travel e-commerce platforms and social media:

- Travel e-commerce platforms provide a comprehensive and logistical view of tourism experiences, particularly focusing on hotel accommodations, service quality and facility amenities. These platforms emphasize practical elements like booking

Table 7 Summary of data sets and results of data analysis

Type of platforms	Name of platforms	Amount of review data	Number of dimensions in word frequency analysis	Number of attributes in word frequency analysis	Number of topics in LDA analysis	Number of topics in co-occurrence matrix-based analysis
Travel e-commerce platform	Ctrip Fliggy	35,346 16,391	8	20	7	7
Social media	Xiaohongshu Weibo	23,492 16,044	9	21	7	11

Source: By authors

procedures, check-in and check-out processes and detailed service reviews. By contrast, social media platforms highlight overall enjoyment, distinctive experiences and personal recommendation.

- The tourism destination image of Sanya on travel e-commerce platforms is centered around hotels. By contrast, the image derived from social media is richer, covering a broader range of attributes.
- Both platforms highlight the importance of tourism infrastructure, indicating that travelers prioritize the completeness and diversity of infrastructure during their travels. This suggests that destinations should focus on developing robust tourism infrastructure.
- Social media platforms offer a more comprehensive portrayal of cultural and environmental aspects than travel e-commerce platforms. Users on social media frequently highlight natural beauty, cultural attractions and social activities, reflecting a greater appreciation of these elements.
- Our study identified new attributes such as duty-free stores, oceans, special sports, theme parks and recommendations. While the first four attributes are emerging on both platforms, recommendation are unique to social media.

These differences in destination images can be attributed to the distinct nature of travel e-commerce and social media platforms. The primary function of the tourism platform is the transaction of tourism product, with user reviews influencing trading decisions (Hua, 2016; Zhong *et al.*, 2023). Only those who have booked tourism products are allowed to post reviews, which focuses the UGC on e-commerce platforms being more targeted and functional. By contrast, social media platforms are centered around communication, allowing users to freely post notes and comments (Akhtar *et al.*, 2023; Zeng and Gerritsen, 2014). This results in more diverse UGC on social media, encompassing a wider range of the tourism perceived destination image.

Additionally, our findings indicate that recommendations are a significant attribute in UGC on social media. UGC on these platforms may play different roles in the decision-making process of tourists. Information on social media may contributes primarily to the travel demand perception and information search stages (Leung *et al.*, 2013), while information on travel e-commerce platforms is more relevant during the travel preparation and satisfaction evaluation stage (Xiang *et al.*, 2017). These differences reflect deeper platform biases.

Specifically, the openness and interactivity of social media make it a primary channel for tourists to share personal experiences and opinions, leading to highly personalized and subjective information. By contrast, UGC on travel e-commerce platforms tends to be more objective, reflecting users' feedback on specific tourism products, though it may exhibit some selection bias. Moreover, social media UGC often includes rich emotional expressions and personal stories, whereas e-commerce UGC is inclined towards functional information. These biases impact how tourists perceive destinations and make decisions.

In summary, studying the differences in tourism destination images between these platforms not only highlights the characteristics and biases of UGC but also provides valuable insights into the tourist decision-making process. These findings can help guide tourism marketing strategies and platform optimization.

5.2 Theoretical implications

This study contributes to the literature on tourism destination image in several significant ways.

First, it pioneers the integration of multiple data sources from different online platforms, specifically travel e-commerce platforms and social media. While existing studies have recognized the benefits of using various data sources to measure destination image, they have predominantly relied on UGC from either travel e-commerce platforms (Guo *et al.*, 2021) or social media platforms (Molinillo *et al.*, 2018), but not both. By incorporating UGC from both platform types, our research fills this gap and provides a more comprehensive understanding of tourists' perceived destination image for Sanya.

Second, this study uses a unique combination of methodologies. We used three approaches:

(1) the LDA model to extract topics from UGC, representing the tourism destination image;

(2) word frequency analysis to identify high-frequency words, followed by manual classification into dimensions and attributes based on a mature tourism destination framework; and

(3) the co-occurrence analysis method, enhanced by community detection algorithms, provides a more in-depth understanding of keyword relationships within UGC (Lee and Park, 2023; Qian *et al.*, 2022; Zhong *et al.*, 2023).

This multiple approach enhances the robustness and depth of the analysis, demonstrating the value of combining quantitative and qualitative analyses.

Finally, this study identifies new attributes of Sanya's perceived tourism destination image, including duty-free stores, oceans, special sports, theme parks and recommendations. These attributes, especially those unique to social media, reflect the evolving nature of tourism destination images in the digital age. The findings further suggest that a combination of information technology and manual analysis can yield richer insights, ultimately advancing our understanding of how tourism images are constructed and perceived in the digital era.

In summary, our study makes significant theoretical contributions by broadening the scope of data sources, using a robust methodological framework and identifying new attributes of destination images, all of which enrich the existing literature on tourism destination image.

5.3 Practical implications

This study offers important practical implications by highlighting the distinct roles of social media and travel e-commerce platforms in shaping tourism destination images. It provides valuable insights for both destination managers and tourists.

For destination managers, the study suggests two key implications. First, by analyzing feedback from tourists on both platforms, managers can gain a more holistic understanding of tourists' perceptions and feelings about the destination. This comprehensive insight allows for more informed decision-making in destination management. Second, recognizing the differences between social media and travel e-commerce platforms enables managers to tailor their promotional strategies effectively. By understanding the unique characteristics and user behaviors of each platform, they can create targeted marketing campaigns that resonate with each platform's specific audience, thereby enhancing the effectiveness and impacts of their promotional efforts.

For tourists, the study underscores the importance of consulting both travel e-commerce platforms and social media platforms when making travel decisions. Utilizing information from both sources provides a more comprehensive view of the destination, helping tourists make more informed and well-rounded decisions regarding their travel plans.

Furthermore, the findings can be generalized to other contexts by applying similar analytical approaches to different destinations. An important practical implication of this study is to consider the differences between platforms when devising destination marketing strategies and tourist communication strategies. Additionally, the potential long-term impacts on destination marketing strategies and tourism behavior include more personalized and effective marketing campaigns, improving tourist satisfaction and destination management practices.

5.4 Limitations and future research direction

This study demonstrates the existence of differences in the tourism perceived image of a destination that is derived from UGC on different kinds of platforms (social media and travel e-commerce platforms). Future research may explore the effect of these differences on tourists' travel decisions and destination marketing organizations' decisions.

References

Afshardoost, M. and Eshaghi, M.S. (2020), "Destination image and tourist behavioural intentions: a meta-analysis", *Tourism Management*, Vol. 81, p. 104154.

Akhtar, M.J., Azhar, M., Khan, N.A. and Rahman, M.N. (2023), "Conceptualizing social media analytics in digital economy: an evidence from bibliometric analysis", *Journal of Digital Economy*, Vol. 2, pp. 1-15.

Alarcon-Urbistondo, P., Rojas-de-Gracia, M.M. and Casado-Molina, A. (2023), "Proposal for employing User-Generated content as a data source for measuring tourism destination image", *Journal of Hospitality and Tourism Research*, Vol. 47 No. 4, pp. 643-664.

Ali, T., Marc, B., Omar, B., Soulaimane, K. and Larbi, S. (2021), "Exploring destination's negative e-reputation using aspect based sentiment analysis approach: case of Marrakech destination on TripAdvisor", *Tourism Management Perspectives*, Vol. 40, p. 100892.

Arabadzhyan, A., Figini, P. and Vici, L. (2021), "Measuring destination image: a novel approach based on visual data mining. A methodological proposal and an application to European islands", *Journal of Destination Marketing and Management*, Vol. 20, p. 100611.

Arefieva, V., Egger, R. and Yu, J. (2021), "A machine learning approach to cluster destination image on Instagram", *Tourism Management*, Vol. 85, p. 104318.

Beerli, A. and Martin, J.D. (2004), "Factors influencing destination image", *Annals of Tourism Research*, Vol. 31 No. 3, pp. 657-681.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

Buhalis, D. and Law, R. (2008), "Progress in information technology and tourism management: 20 years on and 10 years after the internet–the state of eTourism research", *Tourism Management*, Vol. 29 No. 4, pp. 609-623.

Bui, V., Alaei, A.R., Vu, H.Q., Li, G. and Law, R. (2022), "Revisiting tourism destination image: a holistic measurement framework using big data", *Journal of Travel Research*, Vol. 61 No. 6, pp. 1287-1307.

Chaulagain, S., Wiitala, J. and Fu, X. (2019), "The impact of country image and destination image on US tourists' travel intention", *Journal of Destination Marketing and Management*, Vol. 12, pp. 1-11.

Choi, S., Lehto, X.Y. and Morrison, A.M. (2007), "Destination image representation on the web: Content analysis of Macau travel related websites", *Tourism Management*, Vol. 28 No. 1, pp. 118-129.

Dong, Y., Li, Y., Hua, H.Y. and Li, W. (2023), "Perceived tourism authenticity on social media: the consistency of ethnic destination endorsers", *Tourism Management Perspectives*, Vol. 49, p. 101176.

Gaffar, V., Tjahjono, B., Abdullah, T. and Sukmayadi, V. (2022), "Like, tag and share: bolstering social media marketing to improve intention to visit a nature-based tourism destination", *Tourism Review*, Vol. 77 No. 2, pp. 451-470.

Gartner, W.C. (1993), "Image formation process", *Journal of Travel and Tourism Marketing*, Vol. 2 No. 2-3, pp. 191-215.

Gartner, W.C. and Witt, S.F. (1994), "Tourism research in the 90s", *Annals of Tourism Research*, Vol. 21 No. 2, pp. 422-424.

Guo, Y., Barnes, S.J. and Jia, Q. (2017), "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation", *Tourism Management*, Vol. 59, pp. 467-483.

Guo, X., Pesonen, J. and Komppula, R. (2021), "Comparing online travel review platforms as destination image information agents", *Information Technology and Tourism*, Vol. 23 No. 2, pp. 159-187.

Hallmann, K., Zehrer, A. and Müller, S. (2015), "Perceived destination image: an image model for a winter sports destination and its effect on intention to revisit", *Journal of Travel Research*, Vol. 54 No. 1, pp. 94-106.

Hou, Z., Cui, F., Meng, Y., Lian, T. and Yu, C. (2019), "Opinion mining from online travel reviews: a comparative analysis of Chinese major OTAs using semantic association analysis", *Tourism Management*, Vol. 74, pp. 276-289.

Hua, N. (2016), "E-commerce performance in hospitality and tourism", *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 9, pp. 2052-2079.

Hua, N., Morosan, C. and DeFranco, A. (2015), "The other side of technology adoption: examining the relationships between E-commerce expenses and hotel performance", *International Journal of Hospitality Management*, Vol. 45, pp. 109-120.

Huang, Y.C. (2024), "The roles of social media and mutual relationships between travel attitudes and brand equity", *Asia Pacific Journal of Marketing and Logistics*, Vol. 36 No. 2, pp. 390-407.

Hunter, W.C. (2013), "China's chairman Mao: a visual analysis of Hunan province online destination image", *Tourism Management*, Vol. 34, pp. 101-111.

Hunter, W.C. (2016), "The social construction of tourism online destination image: a comparative semiotic analysis of the visual representation of Seoul", *Tourism Management*, Vol. 54, pp. 221-229.

Jiang, Q., Chan, C.S., Eichelberger, S., Ma, H. and Pikkemaat, B. (2021), "Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists", *Current Issues in Tourism*, Vol. 24 No. 17, pp. 2501-2522.

Joo, D., Cho, H. and Woosnam, K.M. (2023), "Anticipated emotional solidarity, emotional reasoning, and travel intention: a comparison of two destination image models", *Tourism Management Perspectives*, Vol. 46, p. 101075.

Kaplan, A.M. and Haenlein, M. (2010), "Users of the world, unite! The challenges and opportunities of social media", *Business Horizons*, Vol. 53 No. 1, pp. 59-68.

Kaur, A., Chauhan, A. and Medury, Y. (2016), "Destination image of Indian tourism destinations: an evaluation using correspondence analysis", *Asia Pacific Journal of Marketing and Logistics*, Vol. 28 No. 3, pp. 499-524.

Kock, F., Josiassen, A. and Assaf, A.G. (2016), "Advancing destination image: the destination content model", *Annals of Tourism Research*, Vol. 61, pp. 28-44.

Kotler, P. (2002), *Marketing Places*, Simon and Schuster, New York, NY.

Lee, S. and Jeong, E.L. (2023), "An integrative approach to examining the celebrity endorsement process in shaping affective destination image: a K-pop culture perspectives", *Tourism Management Perspectives*, Vol. 48, p. 101150.

Lee, J.S. and Park, S. (2023), "A cross-cultural anatomy of destination image: an application of mixed methods of UGC and survey", *Tourism Management*, Vol. 98, p. 104746.

Leung, D., Law, R., Van Hoof, H. and Buhalis, D. (2013), "Social media in tourism and hospitality: a literature review", *Journal of Travel and Tourism Marketing*, Vol. 30 Nos 1/2, pp. 3-22.

Li, Y., He, Z., Li, Y., Huang, T. and Liu, Z. (2023), "Keep it real: assessing destination image congruence and its impact on tourist experience evaluations", *Tourism Management*, Vol. 97, p. 104736.

Li, C., Cao, M., Wen, X., Zhu, H., Liu, S., Zhang, X. and Zhu, M. (2022), "MDIVis: visual analytics of multiple destination images on tourism user generated content", *Visual Informatics*, Vol. 6 No. 3, pp. 1-10.

Li, H., Lien, C.-H., Wang, S.W., Wang, T. and Dong, W. (2021), "Event and city image: the effect on revisit intention", *Tourism Review*, Vol. 76 No. 1, pp. 212-228.

Lin, M.S., Liang, Y., Xue, J.X., Pan, B. and Schroeder, A. (2021), "Destination image through social media analytics and survey method", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2219-2238.

Liu, M., Xue, J. and Liu, Y. (2021), "The mechanism leads to successful clickbait promotion in WeChat social media platforms", *Asia Pacific Journal of Marketing and Logistics*, Vol. 33 No. 9, pp. 1952-1973.

Liu, X., Schuckert, M. and Law, R. (2018), "Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews", *Tourism Management*, Vol. 66, pp. 38-46.

Liu, Y., Huang, K., Bao, J. and Chen, K. (2019), "Listen to the voices from home: an analysis of Chinese tourists' sentiments regarding Australian destinations", *Tourism Management*, Vol. 71, pp. 337-347.

Liu, M., Liu, Y., Mo, Z. and Ng, K. (2021), "Using text mining to track changes in travel destination image: the case of Macau", *Asia Pacific Journal of Marketing and Logistics*, Vol. 33 No. 2, pp. 373-395.

Mak, A.H. (2017), "Online destination image: comparing national tourism organisation's and tourists' perspectives", *Tourism Management*, Vol. 60 No. C, pp. 280-297.

Mao, Y., He, J., Morrison, A.M. and Andres Coca-Stefaniak, J. (2021), "Effects of tourism CSR on employee psychological capital in the COVID-19 crisis: from the perspective of conservation of resources theory", *Current Issues in Tourism*, Vol. 24 No. 19, pp. 2716-2734.

Marine-Roig, E. and Huertas, A. (2020), "How safety affects destination image projected through online travel reviews", *Journal of Destination Marketing and Management*, Vol. 18, p. 100469.

Molinillo, S., Liébana-Cabanillas, F., Anaya-Sánchez, R. and Buhalis, D. (2018), "DMO online platforms: Image and intention to visit", *Tourism Management*, Vol. 65, pp. 116-130.

Narangajavana, Y., Fiol, L.J.C., Tena, M.Á.M., Artola, R.M.R. and García, J.S. (2017), "The influence of social media in creating expectations. An empirical study for a tourist destination", *Annals of Tourism Research*, Vol. 65, pp. 60-70.

Qian, L., Zheng, C., Wang, J., Perez Sanchez, M.D., Parra Lopez, E. and Li, H. (2022), "Dark tourism destinations: the relationships between tourists' on-site experience, destination image and behavioural intention", *Tourism Review*, Vol. 77 No. 2, pp. 607-621.

Röder, M., Both, A. and Hinneburg, A. (2015), "Exploring the space of topic coherence measures", Proceedings of the, eighth ACM international conference on Web search and data mining, pp. 399-408.

Sakiyama, M., Fujii, N., Kokuryo, D. and Kaihara, T. (2020), "Visualization of group discussion using correspondence analysis and LDA in ideathon", *Procedia CIRP*, Vol. 88, pp. 595-599.

Seyyedamiri, N., Pour, A.H., Zaeri, E. and Nazarian, A. (2022), "Understanding destination brand love using machine learning and content analysis method", *Current Issues in Tourism*, Vol. 25 No. 9, pp. 1451-1466.

Stylos, N., Bellou, V., Andronikidis, A. and Vassiliadis, C.A. (2017), "Linking the dots among destination images, place attachment, and revisit intentions: a study among British and Russian tourists", *Tourism Management*, Vol. 60, pp. 15-29.

Taecharungroj, V. and Mathayomchan, B. (2021), "Traveller-generated destination image: analysing Flickr photos of 193 countries worldwide", *International Journal of Tourism Research*, Vol. 23 No. 3, pp. 417-441.

Tung, V., Tse, S. and Chan, D. (2021), "Host-guest relations and destination image: compensatory effects, impression management, and implications for tourism recovery", *Journal of Travel and Tourism Marketing*, Vol. 38 No. 8, pp. 833-844.

Wang, R., Luo, J. and Huang, S.S.(. (2020), "Developing an artificial intelligence framework for online destination image photos identification", *Journal of Destination Marketing and Management*, Vol. 18, p. 100512.

Woosnam, K.M., Stylidis, D. and Ivkov, M. (2020), "Explaining conative destination image through cognitive and affective destination image and emotional solidarity with residents", *Journal of Sustainable Tourism*, Vol. 28 No. 6, pp. 917-935.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017), "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism", *Tourism Management*, Vol. 58, pp. 51-65.

Yu, J. and Egger, R. (2021), "Color and engagement in touristic Instagram pictures: a machine learning approach", *Annals of Tourism Research*, Vol. 89, p. 103204.

Yu, C.E., Xie, S.Y. and Wen, J. (2020), "Coloring the destination: the role of color psychology on Instagram", *Tourism Management*, Vol. 80, p. 104110.

Zeng, B. and Gerritsen, R. (2014), "What do we know about social media in tourism? A review", *Tourism Management Perspectives*, Vol. 10, pp. 27-36.

Zhang, Z., Li, H., Meng, F. and Li, Y. (2019), "The effect of management response similarity on online hotel booking: field evidence from Expedia", *International Journal of Contemporary Hospitality Management*, Vol. 31 No. 7, pp. 2739-2758.

Zhang, Z., Zhang, Z. and Yang, Y. (2016), "The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior", *Tourism Management*, Vol. 55, pp. 15-24.

Zhong, L., Morrison, A.M., Zheng, C. and Li, X. (2023), "Destination image: a consumer-based, big data-enabled approach", *Tourism Review*, Vol. 78 No. 4, pp. 1060-1077.

About the authors

Ya-Fei Liu is a Lecturer at the College of International Tourism and Public Administration, Hainan University, China. Her research focuses on tourism big data mining and tourism firm management.

Yu-Bo Zhu is an undergraduate student at College of International Tourism and Public Administration, Hainan University, China. His research focuses on tourism destination image.

Hou-Han WU is a graduate student at College of International Tourism and Public Administration, Hainan University, China. His research focuses on tourism big data analysis.

Fangxuan (Sam) Li is a Professor from College of International Tourism and Public Administration, Hainan University. His research interests include tourist behavior, tourism marketing and travel experience. Fangxuan (Sam) Li is the corresponding author and can be contacted at: lifangxuan12345@126.com