Study on the development trend of Yunnan coffee tourism based on LDA model

Yi Ouyang[1], and Nutteera Phakdeephirot[2*]
[1] Rattanakosin International College of Creative Entrepreneurship,
Rajamangala University of Technology Rattanakosin, Nakhon Pathom, 73170,
Thailand. Email: 1662110471104@rmutr.ac.th
[2*] Rattanakosin International College of Creative Entrepreneurship,
Rajamangala University of Technology Rattanakosin, Nakhon Pathom, 73170,
Thailand. Email: nutteera.pha@rmutr.ac.th (correspondence)

Abstract: As an emerging model of integrating local industries with tourism, Yunnan coffee tourism shows significant market potential and cultural value. This study aims to: (1) identify the main themes of Yunnan coffee tourism using the LDA model; (2) analyze the relationships and relevance among these themes; and (3) examine the opportunities and challenges for its development and sustainability. Using the LDA model, thematic analysis of tourists＇ online reviews and social media content identifies six core themes: Yunnan coffee culture, coffee tourism experiences, global coffee roasting and branding, coffee enthusiasts and baristas, coffee production and history, and coffee consumption and processing. Semantic network analysis highlights the interconnected nature of these themes, with cultural immersion and experiential elements central to shaping tourist behavior. Sentiment analysis reveals tourists＇ predominantly positive attitudes, appreciating product quality, cultural experiences, and innovative aspects, while highlighting challenges in service quality and infrastructure. This study contributes to the literature by applying advanced text-mining techniques to a niche tourism sector, offering empirical insights into development trends and tourist preferences. Practically, it provides recommendations for product innovation, branding, and sustainable management to enhance Yunnan＇s competitiveness as a coffee tourism destination. The findings underscore Yunnan coffee tourism＇s potential to drive regional economic growth and cultural dissemination, offering a model for industry-tourism integration.

Keywords: Yunnan coffee tourism, LDA model, thematic analysis, semantic network; sustainable tourism.

1. Introduction

Yunnan, as China＇s largest coffee-producing region, has witnessed significant development in its coffee industry, gradually emerging as a prominent player in the global market (Li & Sakamoto, 2021). With over 1.2 million acres of Arabica coffee plantations concentrated in Pu＇er, Baoshan, and Xishuangbanna, the region benefits from ideal high-altitude climates, fertile soils, and favorable geographical conditions. Yunnan coffee, characterized by its unique acidity and complex flavor, has gained recognition among high-end domestic and international markets, including Europe, the United States, and Japan (International Coffee Organization, 2022).

Simultaneously, Yunnan has solidified its position as a leading tourism destination in China. According to the 2022 China Tourism Yearbook, Yunnan attracted over 800 million tourists in 2021, generating nearly 1 trillion yuan in revenue, with tourism contributing over 20% to the province's GDP.

The integration of Yunnan's coffee industry with its thriving tourism sector has given rise to coffee tourism, a unique model that showcases local culture while promoting economic diversification. Coffee tourism enables visitors to experience coffee cultivation, processing, and tasting firsthand, providing opportunities for cultural immersion and sustainable local development (Pan, 2023). This emerging form of tourism not only satisfies the growing demand for personalized and experience-based travel but also holds potential for promoting regional branding and supporting rural economic revitalization.

Despite its growth potential, Yunnan coffee tourism faces several practical challenges that hinder its sustainable development. Current issues include insufficient infrastructure and service quality, a lack of innovative and diversified tourism products, and limited international visibility. Furthermore, overreliance on traditional plantation tours and basic coffee-tasting activities fails to meet tourists' evolving expectations for immersive and interactive experiences.

Existing studies on coffee tourism primarily adopt qualitative approaches, emphasizing descriptive analyses of cultural experiences, tourist satisfaction, and local economic impacts. Opoku et al. (2021) highlighted that coffee tourism enhances cultural immersion and regional branding by offering experiential activities like plantation visits and coffee tastings. Similarly, Vu et al. (2024) demonstrated that coffee tourism satisfies the growing demand for personalized and culturally immersive travel experiences, contributing to sustainable rural development.

However, quantitative, data-driven research on coffee tourism remains limited, particularly in emerging markets like China. While Casalegno et al. (2020) explored tourists' perceptions through surveys, they noted the absence of advanced analytical methods to extract deeper insights from large-scale data sources. Similarly, Yoon and Yoo (2024) emphasized the need for innovative approaches to analyze user-generated content, such as online reviews, to identify emerging themes and trends in niche tourism sectors.

Addressing these gaps, this study employs the Latent Dirichlet Allocation (LDA) model, a widely used text-mining technique, to analyze tourist reviews and social media content. The research aims to: (1) identify the main themes of Yunnan coffee tourism; (2) explore the interrelationships between these themes; and (3) evaluate the opportunities and challenges confronting Yunnan coffee tourism. By leveraging large-scale user-generated data, this study offers empirical insights into tourists' perceptions and preferences, providing actionable recommendations for enhancing tourism products, improving infrastructure, and promoting sustainable growth.

The structure of this article is as follows: Section 2 reviews the literature on coffee tourism and the LDA model to establish the theoretical foundation. Section 3 introduces the data sources, research methodology, and LDA implementation process. Section 4 presents the findings, including the core themes and their interrelationships. Section 5 discusses the practical implications, offers development suggestions, and highlights the study's contributions and limitations, while pointing toward future research directions.art of the research is about.

2. Literature review

2.1 Coffee Tourism

Coffee tourism, as an emerging form of tourism, has garnered significant global attention in recent years, particularly in coffee-producing countries where it has become a key tourism attraction and an economic driver (Vu et al., 2022). Coffee tourism offers visitors unique experiences, enabling them to understand coffee production processes from cultivation to consumption, while simultaneously contributing to local economic development (Woyesa & Kumar, 2021).

In Colombia, abundant coffee-growing resources and a rich coffee culture provide a solid foundation for coffee tourism. Research highlights that Colombia successfully integrates its coffee industry with tourism, creating regionally distinctive tourism brands (Duxbury & Richards, 2019). Visitors can participate in coffee-picking and production activities, sample specialty coffee beverages, and immerse themselves in the cultural allure of Colombian coffee. This form of tourism has generated significant employment opportunities, stimulated rural economies, and effectively attracted visitors, boosting local economic growth. Similarly, Brazil, the world's largest coffee-producing country, actively promotes coffee tourism. Studies show that Brazilian coffee estates attract tourists through diversified activities, including sightseeing, hands-on coffee experiences, coffee festivals, and tasting events (Dinis et al., 2021). These activities deepen visitors' understanding of coffee history and cultural significance while providing notable economic benefits for local communities. Coffee tourism thus contributes to Brazil's economic diversification and rural revitalization.

Beyond economic benefits, coffee tourism plays a crucial role in cultural dissemination and social identity. It provides visitors with opportunities to experience local traditions, production processes, and cultural narratives, fostering deeper appreciation of local heritage (Maspul & Almalki, 2023). This cultural exchange enriches tourist experiences while enhancing local residents' sense of cultural identity and pride. Through interaction with visitors, local communities can share and preserve their cultural knowledge, promoting the continuity of traditional practices (Moyle et al., 2010).

However, coffee tourism faces sustainability challenges, particularly regarding the environmental impact of coffee cultivation and tourism activities. Overdevelopment can lead to environmental degradation, including overuse of water resources and loss of biodiversity. Researchers advocate for the implementation of organic farming practices and environmental conservation measures to ensure the sustainability of coffee tourism (Akenroye et al., 2021). Many coffee estates have already adopted sustainable agricultural methods to improve product quality while minimizing ecological damage, ensuring a balance between economic and environmental objectives (Sott et al., 2020).

Coffee tourism in Yunnan is emerging as a promising tourism model. As China's primary coffee-producing region, Yunnan boasts abundant coffee resources and unique natural landscapes, offering substantial development potential (Feng et al., 2024). In recent years, various regions in Yunnan have explored coffee tourism strategies. For example, Pu'er has launched tourism routes featuring coffee plantation visits, cultural experiences, and coffee tastings, while Baoshan has developed distinctive coffee estates and tourist attractions.

However, chinese coffee tourism remains in its early stages and faces several challenges. First, the lack of product diversity limits its appeal; current coffee tourism primarily focuses on plantation visits and coffee tasting, with insufficient depth and

interactivity. Visitors increasingly demand participatory experiences, such as coffee roasting workshops and brewing tutorials, which could enhance product attractiveness. Second, inadequate promotion and low brand awareness hinder the growth of domestic coffee tourism, as many potential visitors remain unaware of the opportunities Yunnan offers. To address these issues, improving brand building and market promotion has become an urgent priority.

2.2 LDA Model in tourism research

The Latent Dirichlet Allocation (LDA) model, as an effective text-mining technique, has gained widespread application in tourism research (Lekmiti et al., 2024). LDA is particularly well-suited for processing large volumes of unstructured textual data, as it can automatically identify latent themes, offering critical insights into tourist behaviors and preferences (Kirilenko et al., 2021). This capability positions LDA as an essential tool for analyzing tourism dynamics.

The LDA model has been widely used to analyze online reviews, social media data, and travel blogs, revealing tourists' interests, motivations, and behavioral patterns. Studies show that themes such as adventure, relaxation, and cultural experiences are central to tourists' preferences, offering valuable insights for destination marketing strategies (Pomfret, 2019). Moreover, LDA has been applied to analyze hotel and attraction reviews, identifying factors that influence tourist satisfaction and decision-making (Alsayat, 2023). This highlights the importance of understanding consumer preferences in shaping tourism products. LDA has also proven valuable in exploring destination branding, as it identifies key attributes that define destination images based on online narratives and promotional content (Chen et al., 2023). LDA has emerged as a powerful tool for analyzing Chinese tourism data. Researchers have used LDA to identify common themes in tourist experiences, accommodation preferences, and perceptions of attractions, underscoring its importance in improving tourism products tailored to Chinese consumers (Huang & Chelliah). Additionally, the LDA model has facilitated the analysis of tourism trend evolution, revealing the significant impact of external factors such as the COVID-19 pandemic on tourist behavior (Wu et al., 2023). LDA also deepens the understanding of heritage tourism, emphasizing the integration of local cultural narratives into tourism products to enrich visitor experiences (Saoualih et al., 2024).

Recent advancements have further enhanced LDA's utility by integrating it with other analytical techniques, such as sentiment analysis and time-series analysis (Hashmi & Yayilgan, 2024). Combining LDA with sentiment analysis enables researchers to uncover tourists' emotional inclinations across different themes, providing comprehensive insights into satisfaction and preferences. Such integration equips tourism stakeholders with actionable data to optimize product design and marketing strategies. Despite its success, the LDA model has limitations, including high sensitivity to data quality and challenges in interpretability. Future research can explore improved parameter tuning, integration with emerging analytical techniques, and expanded applications in areas such as sustainable tourism and crisis management.

2.3 The Uniqueness of Coffee Tourism in Yunnan

Yunnan's coffee tourism is distinguished by its unique natural environment, cultural integration, diversified experiences, and strong industrial foundation (Zhang et al., 2023). These factors collectively endow Yunnan coffee tourism with distinctive appeal, making it a focal point for both domestic and international tourists.

Yunnan's exceptional natural conditions provide an ideal environment for coffee cultivation (Zhang et al., 2021). Located between the Tropic of Cancer and 15° north latitude, with altitudes ranging from 1,000 to 2,000 meters, Yunnan enjoys a microclimate characterized by fertile soil, abundant rainfall, prolonged sunlight, and significant diurnal temperature variation. These favorable conditions result in nutrient-rich coffee cherries, producing high-quality beans with unique flavors. For example, Baoshan's dry-hot river valley climate yields coffee beans with a rich, mellow taste, demonstrating Yunnan's regional diversity in coffee quality.

Cultural integration further enhances Yunnan coffee tourism's appeal. The fusion of Western coffee culture with the rich traditions of Yunnan's ethnic minorities creates a distinctive cultural landscape (Fortunel et al., 2023). Visitors not only savor locally produced coffee but also experience ethnic music, dance, and attire, fostering a rich cultural immersion. The coexistence of tea and coffee cultures adds another layer of uniqueness, allowing visitors to explore the historical and cultural nuances of these two beverages.

Yunnan's coffee tourism offers diverse and immersive experiences that deepen tourists' engagement. Coffee estates allow visitors to participate in the entire coffee production process, including cultivation, picking, processing, roasting, and brewing (Wang et al., 2019). Activities such as professional coffee workshops and volunteer programs further enhance tourists' sense of participation and cultural connection.

Moreover, Yunnan's coffee tourism seamlessly integrates with its natural landscapes, including attractions like Dali's Erhai Lake and tropical rainforests. This combination enables visitors to enjoy scenic beauty while savoring high-quality coffee (Pan, 2023). Events such as coffee festivals, tasting competitions, and cultural exhibitions further amplify Yunnan's coffee tourism brand and increase its visibility.

Yunnan's strong industrial foundation and government support play a critical role in promoting coffee tourism (Qian et al., 2024). As China's largest coffee-producing region, Yunnan boasts a comprehensive coffee industry chain, including cultivation, processing, and marketing. Policies promoting agriculture-tourism integration in cities like Pu'er have successfully elevated the added value of coffee tourism, attracting large numbers of visitors and establishing Yunnan as a benchmark for integrated rural tourism.

While Yunnan coffee tourism has significant potential, challenges remain in brand development, international recognition, and sustainable growth. Addressing these issues through strategic branding, enhanced market promotion, and balanced ecological preservation will enable Yunnan coffee tourism to expand its influence in domestic and international markets, realizing its full potential.

3. Research methodology

This study employs a comprehensive text-mining approach to analyze the development trends of Yunnan coffee tourism, utilizing Latent Dirichlet Allocation (LDA) topic modeling, semantic network analysis, and sentiment analysis. LDA, a probabilistic topic modeling technique, assumes that documents are mixtures of latent topics, with each topic consisting of specific word distributions (Jelodar et al., 2019). In this study, LDA is applied to identify and extract key themes from tourist reviews and social media content, such as coffee plantation visits, cultural experiences, and product tasting, while also analyzing the relationships and relevance among these themes. Semantic network analysis is used to uncover connections between major concepts by mapping word co-occurrence and semantic similarity, visually representing how themes like "coffee production" and "local culture" interrelate (Christensen & Kenett, 2023). Additionally, sentiment analysis, based on machine

learning algorithms, classifies tourists' emotional expressions into positive, neutral, or negative categories, revealing perceptions of satisfaction, dissatisfaction, and expectations regarding Yunnan coffee tourism. This integrated methodology enables the systematic exploration of large-scale textual data, providing a quantitative and data-driven understanding of emerging trends, tourists' preferences, and the opportunities and challenges facing Yunnan coffee tourism.

4. Results

4.1 LDA model

To determine the optimal number of topics in the textual data related to Yunnan coffee tourism, this study calculated the complexity values for different topic numbers, as shown in Table 1. The validation redundancy value in Table 1 is used to assess the model's goodness-of-fit, where a lower validation redundancy value indicates a better balance between model complexity and accuracy. The results demonstrate that when the number of topics is set to 13, the validation redundancy value is relatively low, suggesting that the model achieves optimal performance under this topic number.

Table 1 Complexity Values for Different Numbers of Topics

Number of Topics	Validation Redundancy	Number of Topics	Validation Redundancy	Number of Topics	Validation Redundancy
1	230.1	6	236.7	11	270.2
2	250.2	7	233.2	12	268.5
3	266.3	8	230.5	13	255.4
4	252.5	9	250.4	14	266.3
5	243.8	10	243.9	15	273.8

Based on the above analysis, after determining the optimal number of topics as 13, the LDA model extracted the key themes and associated keywords of Yunnan coffee tourism, as shown in Table 2. Theme 1 encompasses content related to Yunnan's coffee industry and regional culture, highlighting the distinctive characteristics of Yunnan coffee culture. Theme 2 focuses on tourists' coffee experiences in cafes and urban tourism settings. Theme 3 involves topics related to global coffee roasting processes and branding. Theme 4 emphasizes activities and regional elements associated with coffee enthusiasts and baristas. Theme 5 addresses the history of coffee production and aspects related to its founders. Finally, Theme 6 concentrates on coffee consumption, processing, and production origins. Collectively, these themes reflect the multifaceted nature of Yunnan coffee tourism from diverse perspectives.

Table 2 Themes and Key Terms

Theme ID	Keywords	Label
Theme 1	Coffee, estate, culture, Yunnan, Pu'er, development, industry	Yunnan Coffee Culture
Theme 2	Café, tourism, city, premium, experience, tourists, museum	Coffee Tourism Experience
Theme 3	Roasting, branding, Korea, beans, Shanghai, cultivation, world	Global Coffee Roasting and Branding
Theme 4	China, picking, enthusiasts, Melbourne, pour-over,	Coffee Enthusiasts

|	baristas, Menglian	and Baristas
Theme 5	Founders, history, Rome, Mangshi, Kunming, Seoul, production	Coffee Production and History
Theme 6	Starbucks, Chongqing, countryside, cold brew, processing, consumption, origin	Coffee Consumption and Processing

4.2 Semantic Analysis

Figure 1 presents a visual distribution of the identified themes, mapped onto a two-dimensional plane using the Multidimensional Scaling (MDS) method to illustrate the relationships among themes. In the figure, each node represents a theme, and the distance between nodes indicates their semantic similarity or relevance. Themes that are closer together exhibit stronger semantic relationships. For example, Theme 1 (Yunnan Coffee Culture) and Theme 2 (Coffee Tourism Experience) are positioned closely, suggesting a semantic connection where Yunnan coffee culture serves as a foundational element of tourists' coffee tourism experiences.
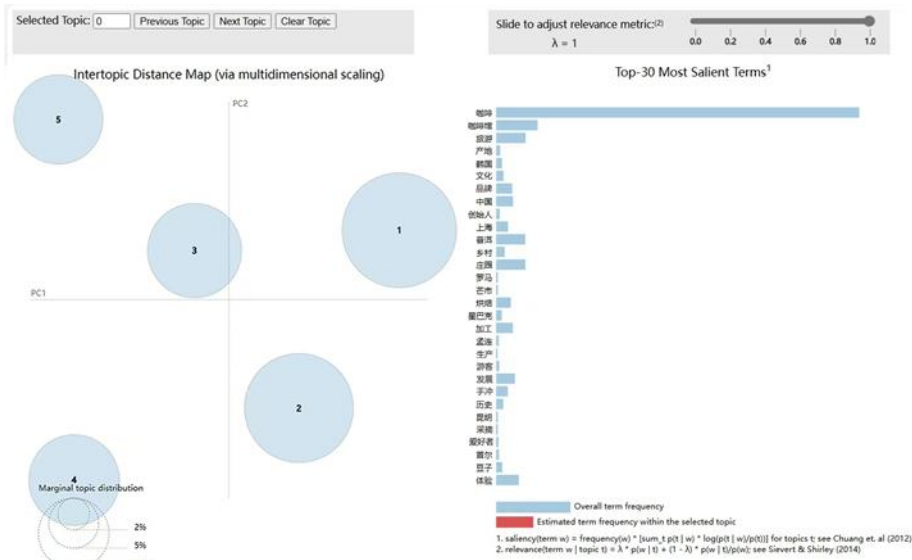

Figure 1 Visualized Distribution of Topics

Figure 2 illustrates the semantic network analysis, where nodes represent words and edges represent the semantic relationships between these words. The thickness or color intensity of the edges indicates the strength of the semantic relationships. From the figure, it is evident that the node "coffee" is strongly connected to nodes such as "estate," "culture," and "cultivation," indicating that these concepts frequently co-occur in the textual data. This highlights the close connection between coffee cultivation and Yunnan's coffee culture within the context of coffee tourism. Additionally, the node "café" is connected to nodes such as "tourism," "city," and "experience," reflecting the importance of cafés as key locations for coffee experiences within urban tourism settings. These semantic relationships provide valuable insights into the thematic structure of Yunnan coffee tourism, emphasizing both cultural and experiential dimensions.
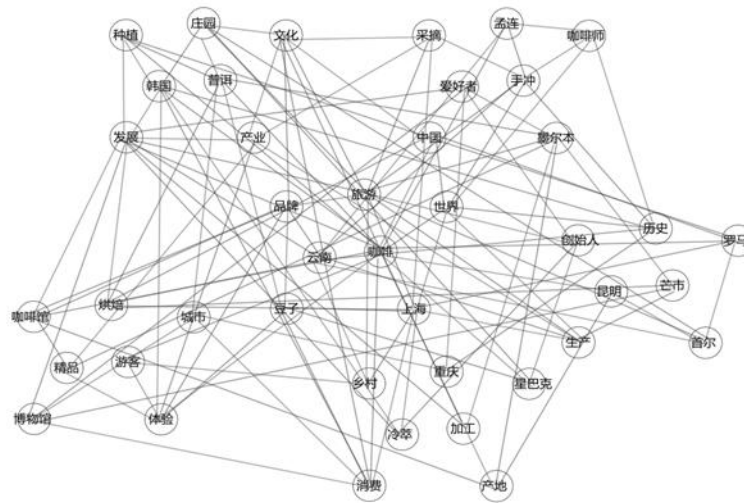
Figure 2 Semantic network analysis

Figure 3 presents a semantic network analysis incorporating clustering, where semantically related words are grouped using clustering algorithms to form distinct clusters. These clusters and their internal word connections further reveal the thematic structure of Yunnan coffee tourism. For instance, one cluster may focus on "coffee production," encompassing words such as "picking," "roasting," and "processing," which are closely related to the production stages of Yunnan coffee. Another cluster may center on "tourism experience," including words like "café," "tourists," and "experience," reflecting the experiential dimension of coffee tourism. The relationships between clusters and the connections within each group provide a clearer understanding of the semantic hierarchy and thematic associations within the textual data. This clustering analysis enables the identification of prominent themes and their underlying structures, offering deeper insights into the development trends and key focal areas of Yunnan coffee tourism.
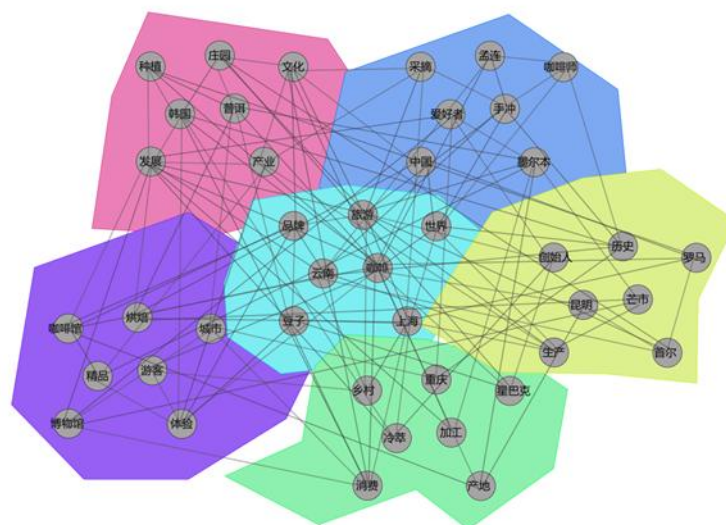
Figure 3 Semantic Network Analysis (with Clustering)

4.3 Sentiment Analysis

Figure 4 presents the distribution of public sentiment toward Yunnan coffee tourism, categorized into 11 sentiment types, including positive, negative, neutral, expectation, and disappointment. The horizontal axis represents the sentiment categories, while the vertical axis indicates the sentiment scores. Words with high positive sentiment scores, such as "premium," "experience," and "innovation," reflect tourists' recognition of Yunnan coffee's quality, immersive tourism experiences, and innovative elements. Conversely, negative sentiment words like "processing," "consumption," and "challenges" suggest potential issues in these areas that require improvement. Neutral sentiment words, such as "estate" and "Yunnan," predominantly serve as objective descriptors. Expectation-related words, including "potential" and "growth," indicate public confidence in the future development of Yunnan coffee tourism. On the other hand, disappointment-related terms, such as "backward" and "misunderstanding," highlight certain negative perceptions or dissatisfaction among the public. Through sentiment analysis, this study provides a comprehensive understanding of public attitudes toward Yunnan coffee tourism, identifying both strengths and areas for improvement. These insights can serve as a valuable reference for stakeholders to enhance tourism products and services, address existing challenges, and align development efforts with tourists' expectations.
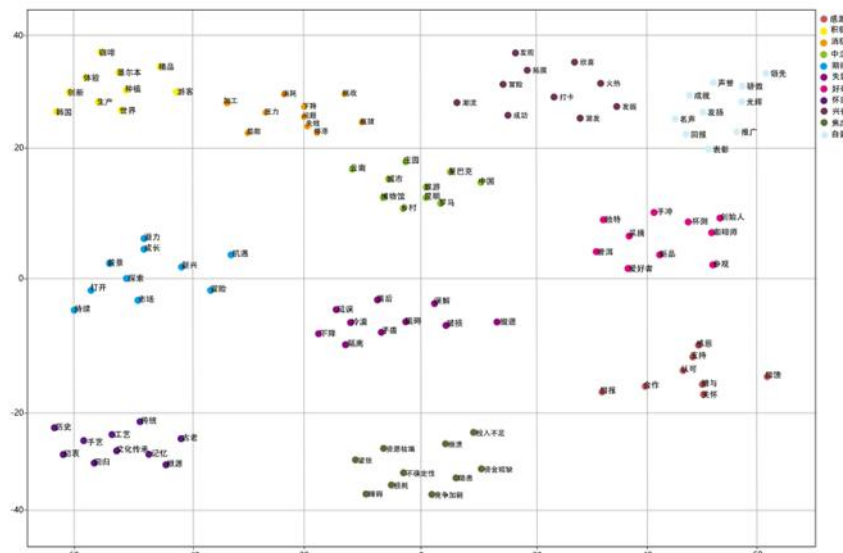

Figure 4 Public Sentiment Towards Yunnan Coffee Tourism

5. Discussion and Conclusions

This study employed the LDA topic model, semantic network analysis, and sentiment analysis to analyze textual data from tourist reviews and social media platforms, extracting the main themes of Yunnan coffee tourism and identifying the relationships between these themes. The findings reveal six key themes: Yunnan coffee culture, coffee tourism experiences, global coffee roasting and branding, coffee enthusiasts and baristas, coffee production and history, and coffee consumption and processing. These themes highlight the diversity, appeal, and evolving trends of Yunnan coffee tourism. Semantic network analysis further illustrates the strong interconnections between themes, such as the links between cultural immersion, coffee production processes, and experiential tourism. Sentiment analysis reveals overall positive perceptions of Yunnan coffee tourism, while also

identifying key areas requiring improvement, such as infrastructure, service quality, and product innovation.

5.1 Theoretical implications

This study makes significant contributions to the theoretical understanding of niche tourism development by adopting a data-driven, quantitative approach. Unlike previous studies that rely primarily on qualitative methods and descriptive cultural analysis (Ferreira & Ferreira, 2020; Lee et al., 2021), this study utilizes the LDA topic model and semantic network analysis to systematically identify themes and trends in Yunnan coffee tourism. This approach enhances the methodological rigor of tourism research and provides a replicable framework for analyzing large-scale textual data in other niche tourism contexts.

The study advances experience economy theory (Pine & Gilmore, 1999) by demonstrating the importance of immersive cultural experiences in driving tourist engagement and satisfaction. The strong association between Yunnan coffee culture and coffee tourism experiences highlights how integrating cultural heritage into tourism enhances the perceived value and appeal of niche destinations. Furthermore, the study contributes to sustainable tourism development theory by identifying opportunities and challenges in balancing economic benefits with ecological preservation. Compared with prior research, this study offers novel insights into tourists' evolving preferences and uncovers actionable themes that align with emerging trends in experiential and sustainable tourism.

5.2 Practical implications

The findings of this study provide valuable guidance for various stakeholder groups, including local governments, tourism operators, coffee producers, and policymakers.

The study highlights the need for infrastructure development, including improved transportation, accommodations, and tourism facilities, to meet rising tourist expectations. Governments should also promote policies that encourage sustainable coffee production and eco-friendly tourism practices.

Tourism Operators can design innovative, immersive experiences to cater to tourists' growing demand for cultural engagement. Examples include professional coffee workshops (roasting and brewing), guided plantation tours, and activities integrating local ethnic culture into coffee tourism.

By leveraging Yunnan's coffee culture and international reputation, producers can collaborate with tourism stakeholders to enhance product branding and develop integrated tourism packages, combining coffee products with experiential activities.

The study underscores the importance of balancing development and ecological sustainability. Policies promoting resource conservation, environmental protection, and local employment should be prioritized to ensure long-term growth.

The insights from sentiment analysis further highlight areas for targeted improvements, such as addressing service quality issues and enhancing marketing efforts to raise the global visibility of Yunnan coffee tourism.

5.3 Limitations and future research

While this study provides significant findings, it has certain limitations. First, the reliance on online reviews and social media data may introduce sample bias, as these platforms may not fully represent all tourist demographics or experiences. Future research can incorporate survey questionnaires and field interviews to enrich data sources and enhance representativeness.

Second, the LDA model, while effective in topic extraction, has limitations in terms of topic interpretability and parameter sensitivity. Future studies can explore

advanced topic modeling techniques, such as BERTopic or deep learning-based methods, to improve analytical accuracy and depth.

Lastly, while this study identifies the opportunities and challenges of Yunnan coffee tourism, future research could further examine its practical applications through empirical collaborations with tourism enterprises. Such partnerships could facilitate experimental studies to test the effectiveness of specific interventions, such as innovative product offerings and sustainable practices.

5.4 Conclusion

This study provides a systematic and data-driven analysis of Yunnan coffee tourism, identifying six key themes and revealing their interconnections through the LDA model and semantic network analysis. By uncovering tourists' perceptions and sentiment trends, the study offers theoretical contributions to niche tourism development and practical insights for stakeholders, including local governments, tourism operators, and policymakers. Yunnan coffee tourism holds significant potential to drive regional economic growth, enhance cultural dissemination, and promote sustainable development.

To unlock this potential, efforts should focus on enriching cultural experiences, improving infrastructure and service quality, fostering product innovation, and ensuring environmental sustainability. By addressing these priorities, Yunnan coffee tourism can further enhance its competitiveness in domestic and international markets, positioning itself as a successful model for integrating local industries with tourism.

6. New knowledge Contribution

This study generates significant new knowledge in the field of niche tourism development, particularly in the context of coffee tourism, by integrating data-driven methodologies and offering theoretical, methodological, and practical advancements.

First, the study introduces a quantitative, data-driven framework for analyzing large-scale textual data, addressing the limitations of prior studies that predominantly relied on qualitative, descriptive methods. By applying the LDA topic model, semantic network analysis, and sentiment analysis, this research establishes a replicable model for uncovering latent themes, analyzing their interconnections, and evaluating public perceptions in niche tourism contexts. This framework expands the analytical toolkit available for tourism researchers and can be applied to similar studies in other regions or industries.

Second, this study identifies six key themes in Yunnan coffee tourism—Yunnan coffee culture, coffee tourism experiences, global coffee roasting and branding, coffee enthusiasts and baristas, coffee production and history, and coffee consumption and processing. These themes offer a comprehensive conceptual understanding of the multidimensional nature of coffee tourism, demonstrating how cultural immersion, production processes, and global branding interrelate to create a unique and appealing tourism experience. This thematic structure provides a new lens through which the development of niche tourism can be understood and analyzed.

Third, the study contributes to the advancement of experience economy theory by empirically illustrating the importance of immersive cultural experiences in shaping tourist satisfaction and engagement. It demonstrates that tourists' perceptions of coffee culture and experiential activities serve as central factors driving interest in coffee tourism, highlighting the value of integrating cultural and production-related experiences into tourism products.

Furthermore, this research contributes to sustainable tourism development theory by identifying both opportunities and challenges in balancing economic growth with environmental and cultural preservation. The findings emphasize the need for innovative, sustainable tourism practices that align with tourists' expectations for ecological responsibility and cultural authenticity, offering guidance for future sustainable tourism frameworks.

Lastly, the integration of sentiment analysis in this study provides a novel approach to understanding tourists' emotional responses and preferences. By categorizing sentiment into positive, negative, and expectation-driven dimensions, this research identifies key strengths and weaknesses in Yunnan coffee tourism, offering actionable insights for stakeholders to address critical challenges and optimize tourism products and services.

In summary, this study generates new knowledge through the development of a data-driven analytical framework, the identification of a comprehensive thematic structure for coffee tourism, and its contributions to existing theoretical frameworks. These findings not only advance the academic understanding of niche tourism but also offer a practical roadmap for integrating industry and cultural tourism, promoting sustainable growth, and enhancing stakeholder decision-making.

Data Availability
The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author: 1662110471104@rmutr.ac.th

Interest Declaration
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The research did not involve any financial, personal, or other relationships with other people or organizations that could inappropriately influence (bias) our work. This submission is an honest, accurate, and transparent account of the study being reported, without any material omissions. Any roles of funding bodies have been clearly delineated; these bodies had no influence over the content of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process
During the preparation of this work the author(s) used ChatGPT in order to polish English. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Appendix
**Table 1 Complexity Values for Different Numbers of Topics**

Number of Topics	Validation Redundancy	Number of Topics	Validation Redundancy	Number of Topics	Validation Redundancy
1	230.1	6	236.7	11	270.2
2	250.2	7	233.2	12	268.5
3	266.3	8	230.5	13	255.4
4	252.5	9	250.4	14	266.3
5	243.8	10	243.9	15	273.8

(Source:generated by this research)

Table 2 Themes and Key Terms

Theme ID	Keywords	Label
Theme 1	Coffee, estate, culture, Yunnan, Pu'er, development, industry	Yunnan Coffee Culture
Theme 2	Café, tourism, city, premium, experience, tourists, museum	Coffee Tourism Experience
Theme 3	Roasting, branding, Korea, beans, Shanghai, cultivation, world	Global Coffee Roasting and Branding
Theme 4	China, picking, enthusiasts, Melbourne, pour-over, baristas, Menglian	Coffee Enthusiasts and Baristas
Theme 5	Founders, history, Rome, Mangshi, Kunming, Seoul, production	Coffee Production and History
Theme 6	Starbucks, Chongqing, countryside, cold brew, processing, consumption, origin	Coffee Consumption and Processing

(Source:generated by this research)
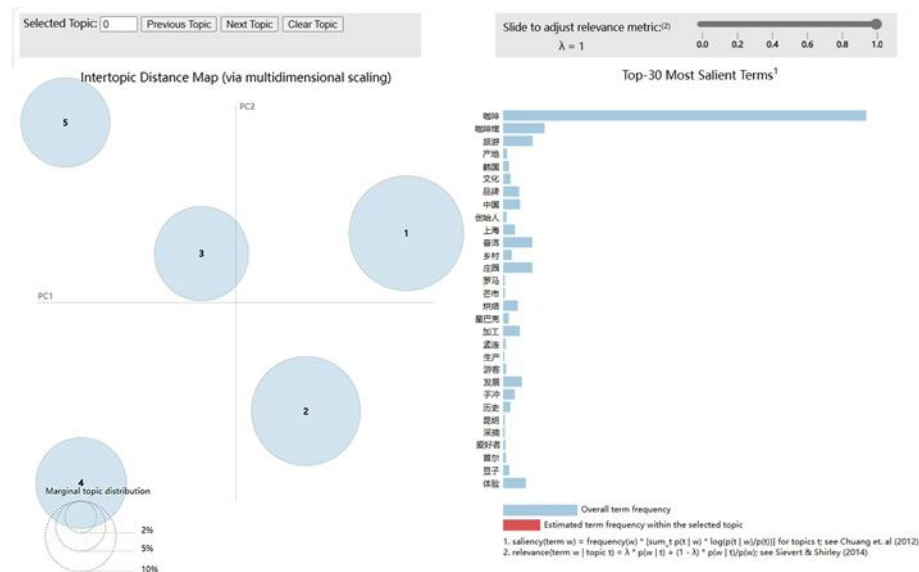
Figure 1. Visualized Distribution of Topics
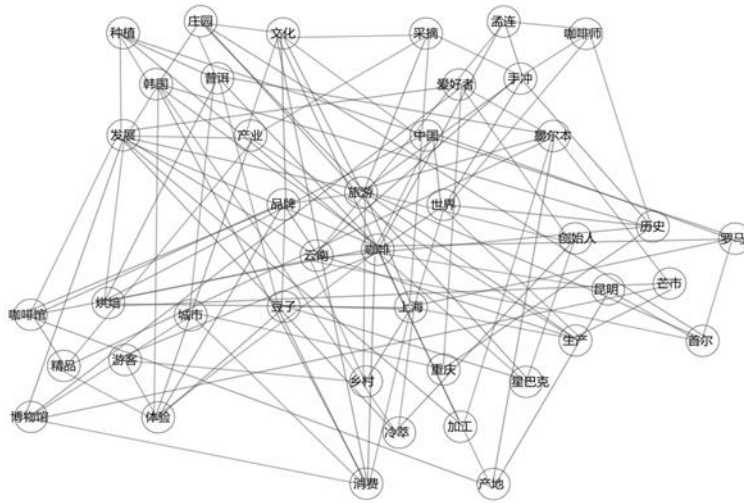(Source: generated by this research)

Figure 2 Semantic network analysis
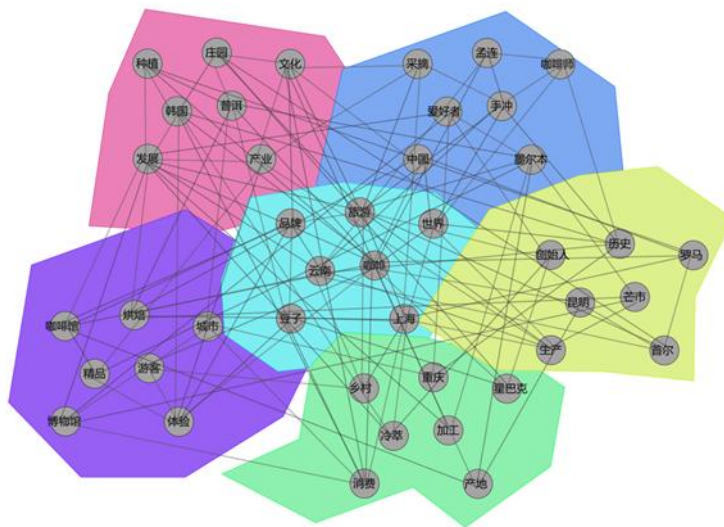(Source: generated by this research)

Figure 3 Semantic Network Analysis (with Clustering)
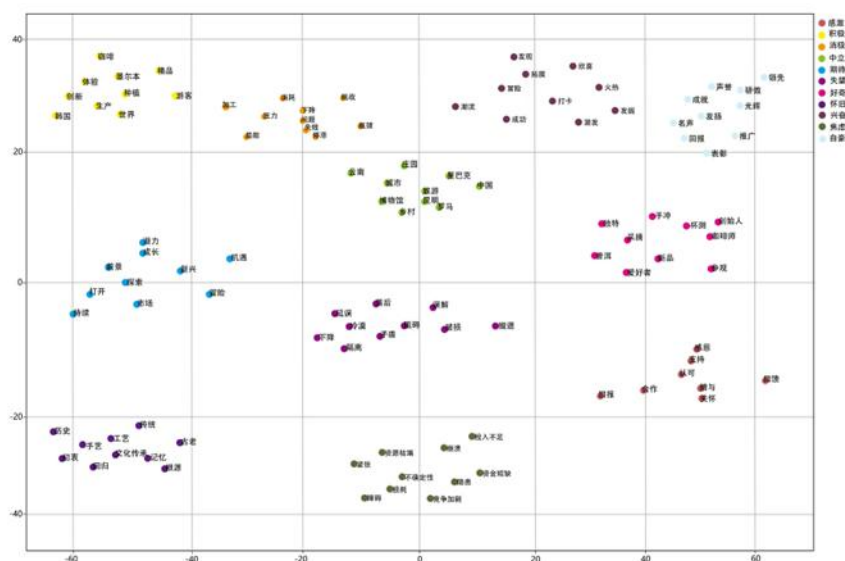(Source: generated by this research)

Figure 4 Public Sentiment Towards Yunnan Coffee Tourism
(Source: generated by this research)

Reference

Akenroye, T. O., Dora, M., Kumar, M., Elbaz, J., Kah, S., & Jebli, F. (2021). A taxonomy of barriers to the adoption of sustainable practices in the coffee farming process. *Journal of Cleaner Production, 312*, 127818. https://doi.org/10.1016/j.jclepro.2021.127818

Alsayat, A. (2023). Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca. *Neural Computing and Applications, 35*(6), 4701-4722. https://doi.org/10.1007/s00521-022-07992-x

Casalegno, C., Candelo, E., Santoro, G., & Kitchen, P. (2020). The perception of tourism in coffee-producing equatorial countries: An empirical analysis. *Psychology & Marketing, 37*(1), 154-166. https://doi.org/10.1002/mar.21291

Chen, Q., Xu, S., Liu, R., & Jiang, Q. (2023). Exploring the Discrepancy between Projected and Perceived Destination Images: A Cross-Cultural and Sustainable Analysis Using LDA Modeling. *Sustainability, 15*(12), 92-96. https://doi.org/10.3390/su15129296

Christensen, A. P., & Kenett, Y. N. (2023). Semantic network analysis (SemNA): A tutorial on preprocessing, estimating, and analyzing semantic networks. *Psychological Methods, 28*(4), 860. https://doi.org/10.1037/met0000463

Dinis, M. G., Melo, C. S., & Sousa, J. M. B. M. (2021). Coffee tourism in Portugal: an attraction case study. *International Journal of Culture, Tourism and Hospitality Research, 15*(3), 399-412. https://doi.org/10.1108/IJCTHR-08-2020-0194

Duxbury, N., & Richards, G. (2019). Towards a research agenda for creative tourism: Developments, diversity, and dynamics. *A research agenda for creative tourism*, 1-14. https://doi.org/10.4337/9781788110723.00008

Feng, T., Zhang, C., & He, J. (2024). Mellower: A Specialty Coffee Brand From China. *FUDAN*, 1-21. https://doi.org/10.12156/FUDAN.CASE202216011

Fortunel, F., Hu, Y., & Le Duc, N. (2023). The Construction of Coffee Qualities: Geo-economics Crossroads between China and Vietnam. *TRaNS: Trans -Regional and -National Studies of Southeast Asia, 11*(2), 141-157. https://doi.org/10.1017/trn.2022.14

Hashmi, E., & Yayilgan, S. Y. (2024). A robust hybrid approach with product context-aware learning and explainable AI for sentiment analysis in Amazon user reviews. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-024-09896-5

Huang, X., & Chelliah, S. Attributes Influencing Tourist Satisfaction: Sentiment Analysis and Topic Modeling of Online Reviews. *Journal of China Tourism Research*, 1-20. https://doi.org/10.1080/19388160.2024.2440323

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169-15211. https://doi.org/10.1007/s11042-018-6894-4

Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tourism management, 83*, 104241. https://doi.org/10.1016/j.tourman.2020.104241

Lekmiti, A., Stolk, P. J., Taylor, A., Ramachandran, S., & Yap, N. K. (2024). Text mining in tourism and hospitality research: a bibliometric perspective. *Journal of hospitality and tourism technology, ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JHTT-05-2024-0284

Li, N., & Sakamoto, Y. (2021). Trends in Specialty Coffee. In L. Louzada Pereira & T. Rizzo Moreira (Eds.), *Quality Determinants In Coffee Production* (pp. 407-443). Springer International Publishing. https://doi.org/10.1007/978-3-030-54437-9_9

Maspul, K. A., & Almalki, F. A. (2023). Preserving Local Wisdom: Unaizah's Coffee Culture and Dates Farming Sustaining Cultural Heritage. *J-CEKI: Jurnal Cendekia Ilmiah, 2*(6), 639-664. https://doi.org/10.56799/jceki.v2i6.2280

Moyle, B., Croy, G., & Weiler, B. (2010). Tourism interaction on islands: the community and visitor social exchange. *International Journal of Culture, Tourism and Hospitality Research, 4*(2), 96-107. https://doi.org/10.1108/17506181011045172

Opoku, E., Wang, S., & Muñoz, K. (2021). A better brew: COVID-19 and sustainable outcomes for coffee tourism in Ali Mountain, Taiwan. *Journal of Responsible Tourism Management, 1*(1), 71-87. https://doi.org/10.47263/JRTM.01-01-06

Pan, Q. (2023). The Past, Present and Future of Coffee Tourism. *Open Journal of Business and Management, 11*(2), 688-703. https://doi.org/10.4236/ojbm.2023.112037

Pomfret, G. (2019). Conceptualising family adventure tourist motives, experiences and benefits. *Journal of Outdoor Recreation and Tourism, 28*. https://doi.org/10.1016/j.jort.2018.10.004

Qian, J., Zeng, Y., Tang, X., & Hu, X. (2024). Empowering left-behind places in Southwest China: participation in coffee value chains as place-based development. *Cambridge Journal of Regions, Economy and Society, 17*(2), 375-392.

Saoualih, A., Safaa, L., Bouhatous, A., Bidan, M., Perkumienė, D., Aleinikovas, M., Šilinskas, B., & Perkumas, A. (2024). Exploring the Tourist Experience of the Majorelle Garden Using VADER-Based Sentiment Analysis and the Latent Dirichlet Allocation Algorithm: The Case of TripAdvisor Reviews. *Sustainability, 16*(15), 63-78. https://www.mdpi.com/2071-1050/16/15/6378

Sott, M. K., Furstenau, L. B., Kipper, L. M., Giraldo, F. D., Lopez-Robles, J. R., Cobo, M. J., Zahid, A., Abbasi, Q. H., & Imran, M. A. (2020). Precision techniques and agriculture 4.0 technologies to promote sustainability in the

coffee sector: state of the art, challenges and future trends. *IEEE Access, 8*, 149854-149867. https://doi.org/10.1109/ACCESS.2020.3016325

Vu, O. T. K., Alonso, A. D., Martens, W., Ha, L. D. T., Tran, T. D., & Nguyen, T. T. (2022). Hospitality and tourism development through coffee shop experiences in a leading coffee-producing nation. *International Journal of Hospitality Management, 106*, 103300. https://doi.org/10.1016/j.ijhm.2022.103300

Vu, O. T. K., Alonso, A. D., Tran, T. D., Martens, W., Do, L., Nguyen, T. T., Atay, E., & Akbari, M. (2024). Coffee culture unravelled: exploring the coffee shop experience model in the Vietnamese context. *Tourism Recreation Research*, 1-19. https://doi.org/10.1080/02508281.2023.2295621

Wang, M.-j., Chen, L.-H., Su, P.-a., & Morrison, A. M. (2019). The right brew? An analysis of the tourism experiences in rural Taiwan's coffee estates. *Tourism Management Perspectives, 30*, 147-158. https://doi.org/10.1016/j.tmp.2019.02.009

Woyesa, T., & Kumar, S. (2021). Potential of coffee tourism for rural development in Ethiopia: a sustainable livelihood approach. *Environment, Development and Sustainability, 23*(1), 815-832. https://doi.org/10.1007/s10668-020-00610-7

Wu, B., Wang, L., & Zeng, Y.-R. (2023). Interpretable tourism demand forecasting with temporal fusion transformers amid COVID-19. *Applied Intelligence, 53*(11), 14493-14514.

Yoon, H. Y., & Yoo, S.-C. (2024). Finding tourism niche on image-based social media: Integrating computational methods. *Journal of Vacation Marketing, 30*(4), 874-889. https://doi.org/10.1177/13567667231180994

Zhang, J., Zhang, R., Li, Q., Zhang, X., & He, X. (2023). Spatial Sifferentiation and Differentiated Development Paths of Traditional Villages in Yunnan Province. *Land, 12*(9), 1663. https://doi.org/10.3390/land12091663

Zhang, S., Liu, X., Wang, X., Gao, Y., & Yang, Q. (2021). Evaluation of coffee ecological adaptability using Fuzzy, AHP, and GIS in Yunnan Province, China. *Arabian Journal of Geosciences, 14*(14), 1366. https://doi.org/10.1007/s12517-021-07795-9