NETWORK-BASED TOURISM ATTRACTION ANALYSIS USING MULTI-SOURCE GEOTAGGED DATA

ZHOU Xiaolin

PhD

the Hong Kong Polytechnic University

2023

The Hong Kong Polytechnic University

Department of Land Surveying & Geo-Informatics

NETWORK-BASED TOURISM ATTRACTION ANALYSIS USING MULTI-SOURCE GEOTAGGED DATA

ZHOU Xiaolin

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________________ (Signed)

____________ZHOU Xiaolin____________(Name of student)

Acknowledgement

Like the research topic-tourism in this study, persons are interconnected in the world. No man is an island. This research project bears on imprint of many people. I would like to express my sincere thanks to those who walk beside me, help me along to the completion of the dissertation.

My special gratitude would go first to my chief supervisor, Professor Wenzhong, Shi for his continuous support and guidance of my PhD study and research. The formulation of the research questions and methodology cannot be divorced from his expertise. His insightful feedback and illuminating instruction are the beacon for me to polish the research idea continuously. Furthermore, he also sets an example of a mild, wise, and successful person for me.

A particular acknowledgement is owed to Dr. Anshu Zhang, Prof. Brain King, Dr. Weng Si Lei, Dr. Zhaoyu Chen and Dr. Zhewei Liu for their wonderful collaboration in various interesting research projects. Meanwhile, I would also like to relish the opportunity to thanks the administrative staff in the department of Land Surveying and Geo-informatics for their daily support.

Last but not least, I am indebted to my parents for their wise counsel and sympathetic ear through the past six year. Also, l could not have completed this dissertation without the supports of my friends and colleges in my group, providing stimulating discussions together with kind distractions to rest me.

Abstract

Tourism is a networked industry. Resources allocated to destination marketing and management should go beyond single and primary attractions. Tourism resources are significantly connected to each other based on tourist interest. Age of Big Data provides new opportunities for investigate tourist behavior and experience, especially with geotagged data tracking tourists' visitation pattern. Using geotagged information as an indication of tourist movement patterns between attractions, this study connects attractions based on tourists' common motivations. Meanwhile, tourist movements can report complex routes that reveal an attraction network of the attractiveness propagation. Information network among tourism attractions can then be abstracted and the following analysis is conducted to facilitate destination planning.

In this study, ranking analysis on attraction network is firstly conducted. Targeted at the feature of attraction network (i.e., distance is accounted as cost for tourists), we propose AttractionRank as a reasonable and fairer way to assess tourism attraction's attractiveness. AttractionRank is an extension of the PageRank algorithm by eliminating the distance-decay effect scrupulously in the attractiveness propagation process. The main rationale is to calibrate the distance-decay effect function and its effect scale via trip distribution estimation and then exclude the effect in the network linkage weight when applying weighted PageRank. Experimental results shows that AttractionRank outperforms other

two schemes and the entire ranking results are also statistically significant in data interpretation in practical sense. The results also demonstrate the potential value of AttractionRank for tourism planning in Hong Kong and other regions.

Clustering analysis on that attraction information network is then explored. The study describes the clustering effects by sorting the attractions into four clusters in Hong Kong case. A new framework was used to reveal the characteristics of these intra-cluster attractions from three dimensions: theme, visit volume, and level of importance by attractiveness propagation rank. The framework offers a theoretical contribution to the literature on clustering dimensions by providing a feasible experimentation method, and, from a practical perspective, it can guide destination marketing strategies for attraction networking. This destination-wide clustering effect analysis facilitate destination planning from a broader perspective than analysis that narrowly focus on certain attractions.

As a complex networked industry, attraction nodes in tourism system should not be concentrated on only. Tourist concerns and needs are also interconnected in VacationScape. This study finally draws upon Gunn's formative concept of organic destination image and applies it in a contemporary regional context so that a comprehensive apprehension about the attitudes of entire tourists can be revealed. This study uses a prominent Chinese social media platform - the Red – to evaluate tourist/user generated content and explores the destination image of the "9 + 2" Greater Bay Area cities in southern China (9 mainland cites plus

Macau and Hong Kong SARs). Four clusters of designative and prescriptive image are proposed. The proposed insights can benefit destination planning at local and regional levels by showing the merits of mobilizing tourism resources across cities.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

"Tourism has always been a networked industry and the usual description of tourism as a fragmented and geographically dispersed industry belies a pervasive set of business and personal relationships between companies and managers in businesses such as national tourism offices, hotels, attractions, transport, tours, travel agents and restaurants."

—— (Scott, Baggio, and Cooper, 2008)

1.1. Tourism: networked industry

The importance of tourism is beyond question as "an activity essential to the life of nations because of its direct effects on the social, cultural, educational, and economic sectors of national societies, and on their international relations" (World Tourism Organization, 1980). On the one hand, it reflects workers' rights of rest and travel in tourism origin area. On the other hand, it provides significant employment and livelihood opportunities in destinations over the world. It has become one of the world's largest economic sectors bearing on the welfare of different stakeholders. These stakeholders contain but not limit to travelers, governments, tourism enterprises, tourism employees, host communities, and tourism related organizations. (World Tourism Organization, 2005a) They interact with each other and compose a complex network in the tourism industry. Masion (2003) discussed some very visible impacts (i.e., socio-cultural, economic, and environmental) of tourism to destinations. He stated that all the impacts have positive and negative facets to the destination depending on 'the value position of the observer', which can be considered as the value of tourism to different stakeholders. With the development of tourism and the complication of stakeholders' interactions, it becomes difficult to balance the concerns of all parties. The problem then comes down to the question how to match travelers' demand to destinations' supply with quality assurance and sustainability.

Attractions are never isolate to each other in networked tourism industry. A tourist attraction itself was viewed as a system (MacCannell, 1976 p.41). On this basis, Leiper (1990) put forward a definition that "A tourist attraction is a system comprising three elements: a tourist or human element, a nucleus or central element, and a marker or informative element. A tourist attraction comes into existence when the three elements are connected." The marker is the information that links the tourist to a destination or an attraction and stimulates tourists' motivation to visit. It is basically controlled by DMOs in their efforts of tourism promotion. A nucleus (Gunn, 1972) is considered as the central element (any feature of a place) in the system as primary motivators for travel or a zone of closure (i.e., ancillary services) associated with the attraction (Page, 2015). Along with the different significance, these nuclei are assessed in a hierarchy: a primary nucleus attracts tourists as reflecting their core motivator of visiting a destination, while lower level of nuclei (secondary and tertiary) are of relatively less interest among visitors. Secondary nuclei may be those known attributes before travelling, but not yet significantly contribute to an itinerary. Tertiary nuclei are the incidental ones that tourists discover on-site (Leiper, 1990). As such, tourist attractions can be primary motivators for travel while they can also be an extra reward to one's travel when tourist may visit en route to a destination or within destination without any pre-determined decision as spontaneous visit (Lubbe, 2005). In addition, travelers would take long distance from one attraction to another stimulated by interest practically. Attractions are connected in consideration of tourist interests, especially in the destination with multi-attractions as satisfying the various needs of tourists.

Scaling from tourist attraction to destination level, interconnected concerns and tourist needs are also rewarded. Gunn (1972) proposed a 'Vacationscape' concept where a thorough understanding of destination resources at both micro and macro levels are encouraged. It emphasizes a comprehensive apprehension about the attitudes of individuals as entire persons. Meanwhile, it offers prospectively better resource utilization and the provision of meaningful visitor experiences through effective planning and development. That is, interconnected needs of tourists should be satisfied by a holonomic tourism system in the context with compatible attractions in tourism development. In line with 'Vacationscape', a large amount of tourism research was conducted to understand tourists' experience, perceptions and needs holistically in the destination: Wearing and

Foley (2017) investigated tourists experiences from a feminist perspective and reminded planners of gazing at diverse voices to achieve a balanced destination development. Virabhakul and Huang (2018) explored the serial mediating effects of service experience on tourist intensions and provided a holistic view for planners. Whereafter, the emerging development of Information and Communication Technology (ICT) provides rich set of data to explore travelers' interactions with the destination: Edwards and Griffin (2013) used GPS tracking data to investigate tourists' spatial behavior around destinations and helped improve tourist experience by providing better wayfinding service. Kim and Fesenmaier (2015) studied the relationship between travelers' real-time emotion and environment via their electrodermal activity so that service quality can be improved relevantly. Spatio-temporal sentiment changes in tourist flows were also explored with geotagged social media data (Jiang, et al., 2021). In that vein, the Great Time of Big Data is providing much more opportunity for sustainable tourism research fashionably.

1.2. Geographic concern in tourism studies

Geographic variables can provide meaningful clues to conclude tourists' travel patterns (Smith, 1985). They are key indicators to predict the demand based on the distance from the source market (McKercher & Lew, 2003). Geographic proximity was referred to as a key consideration in travel cost, and a barrier for trip in most cases. There are various factors that can be used to reflect the travel patterns, including distance decay, existence of competing opportunities, tourist's behavioral patterns, tourists' mental image and motivation when searching the information about the destination (Murphy & Keller, 1990). Distance decay theory suggests that the request for any good or service decreases as distance increases (Eldridge & Jones, 1991). It is considered a golden rule to explore the tourist movement patterns (McKercher, Chan, & Lam, 2008). It has been widely explored in tourism geography as to discover the transformation of market segments (McKercher, 2008a), effects of distance on the tourist profile and their perceptions and emotions on the destination (Bao & McKercher, 2008; Nyaupane & Graefe, 2008; Tjørve, Flognfeldt, & Tjørve, 2013; Joo et al., 2017; Schuckert & Wu, 2021).

Tourist studies related to travel distance, especially the comparisons between short-haul and long-haul tourists in their participated activities in a destination, have been widely explored in urban area (Nyaupane & Graefe, 2008; Digun-Aweto, Fawole, & Saayman, 2019). Such studies have also expanded to the field of area conservation and ecotourism development (Digun-Aweto, Fawole, & Saayman, 2019; Olaniyi, Ogunjemite, Akindele, & Sogbohossou, 2020). Compared to the mainstream studies on inter-destination tourist movement patterns using the distance decay, only a few studies focused on intra-destination movement. Smallwood, Beckley, and Moore (2012) used the protected areas in coastal marine parks to explore the factors influencing the tourists travelling from one site to another as well as some comparison of site attractiveness to a certain tourist segment despite the long travel distance. Jin et al. (2018) explored the temporal heterogeneity of tourist movement in intra-urban tourist mobility through network analysis and highlighted the role of time in tourist movements other than distance effects. In fact, distance decay curve can also be used to explore the intra-destination travel patterns. For instance, the distant relationship between accommodation hubs with attractions influences the tourist flows significantly (Paulino, Prats, & Whalley, 2020).

The reason why a tourist does not visit a destination is a combination of various deterministic variables (McKercher, 2008b). Travel distance is only one of the considerations in tourist decision making, but plays a critical role, combined with the factors like destination attractiveness, competitiveness, and other pull factors (Hooper, 2015; Lee et al., 2012; Sun & Lin, 2019; McKercher & Mak, 2019). Distance can be referred as attractiveness of destination according to the relative proximity to the tourism source (Pearce, 1989). That is, when two places are similar, tourists tend to travel to the one that is closer to their origins. However, if tourists have strong motivations or needs, the number of destination choices would become smaller. Then these potential destinations are chosen based on their relative distance rather than absolute distance (Hooper, 2015). Mou et al. (2020) found that both distance decay and attractions' attractiveness determine the spatial pattern of tourist flows. Therefore, it is possible to factor out the effect of distance in order to reveal hidden potential attractive sites. The finding of this study would be useful for tourism planning since the effect of distance can be mitigated by taking the benefits of modern transportation.

A growing number of studies have employed innovative analytical tools to examine the far-reaching consequences of tourist movements. Movement patterns are then used to frame the tourist movements in an explanatory manner, typically categorizing such movements as inter-destination and intra-destination (Raun, Ahas, & Tiru, 2016; Zheng, Huang, & Li, 2017). Lew and McKercher (2006) identified three types of intra-destination tourist movement patterns according to itinerary: point-to-point patterns, circular patterns, and complex patterns. However, the investigation method was not limited to attractions but involved a mix of functional places, including accommodations and other non-attraction spots. Thus, the method remains a means of describing physical movement patterns but likely has little power to predict connections between the attractions.

Traditional means of collecting data for analysis (e.g., interviews, online trip diaries) restrict researchers from analyzing textual information in a subjective way because of small sample sizes and researcher bias (e.g., Smallwood et al., 2012; Leung et al., 2012). Such means are also criticized for neither capturing travel behavior comprehensively nor accurately reflecting travel patterns (Vu, Li, Law, & Ye, 2015). These limitations can be overcome using a relatively accurate algorithm design. Xia et al. (2010) attempted to use an expectation-maximization algorithm to identify market segmentation based on the dominant movement patterns of tourists, using the case of a package tour to Philip Island Nature Park on the south coast of Australia. However, it was a methodological study that aimed to verify the applicability of this analytic tool for future practice; the sample size of the study was small, and the study was limited to one natural site. This limitation of single-site case studies is commonly found in tourist movement research (e.g., Chancellor, 2012; Li, Xie, Gao, & Law, 2021).

Global Positioning System (GPS) tracking technology is commonly referred to in the literature as an information system and a valuable diagnostic tool for tracking tourist movements in an accurate way, especially considering its implications for combining better wayfinding (Edwards & Griffin, 2013) and crowd control at a single attraction based on tracked intra-destination movements (Zheng et al., 2017). Correspondingly, a GIS displays tourist movements obtained via GPS technology on a map (Lü et al., 2019). Although a few studies have attempted to examine inter-destination movements (Hallo, Manning, Valliere, & Budruk, 2005), data collected through GPS cannot be used to measure or distinguish destinations on a large scale but can only indicate preference on a micro scale (Raun et al., 2016). GPS and other equivalent positioning technologies are limited to small destination areas or single attractions (e.g., McKercher & Lau, 2008; Zheng et al., 2017) and lack applicability to large-scale exploration, what to speak of to the city-level exploration required to truly understand tourist movement patterns.

At the individual level, McKercher and Lau (2008) discovered a full set of independent individual travelers' spatial-temporal movement patterns in Hong Kong, focusing on identifying the attributes (i.e., human, physical, trip, and time factors) affecting such patterns. Similarly, Gu et al. (2021) explored the spatiotemporal behavior of individual tourists in a particular wine region using a tracking app. Despite the narrow focus on individual tourists' differences, Gu et al.'s study offers a broader picture of tourist movement through big data analytics. In consideration of the inadequacy of previous tourist movement research at the micro level, the present study eliminated such practical constraints for a wider exploration of tourist movements between the attractions in a destination through their self-reported itineraries (e.g., geotagging) on location-based social networks.

Using geotagged information from social media is a novel approach compared to the traditional approach to analyzing tourist flows (Jiang et al., 2021). Vu et al. (2015) examined geotagged Flickr information for Hong Kong inbound tourists to identify attractions of interest, tourists' travel behaviors, travel routes, and travel times and further introduced a framework for effectively extracting geographical information. Chua, Servillo, Marcheggiani, and Moere (2016) analyzed the spatial, temporal, and demographic features of tourism flows in tourist attractions in Cilento, Italy using geotagged Twitter data. Similarly, Jiang et al. (2021) explored tourist flows in Beijing using Sina Weibo data. Again, these studies focused more on tourist behaviors and sentiments at the personal level than on destination attraction planning and management.

Studies using Instagram content have covered a wide range of disciplines, including linguistics in examining the role of semantics in labeling activities for photo sharing (Lin, 2010) and sociology in identifying social interactions between users (Hegde, 2016). Yu and Sun (2019) examined the role of Instagram in promoting the gastronomy destination Macao by exploring relevant user posted content. A review of current studies employing geotagged information reveals that few have made use of geotagged information on Instagram for exploring the attraction networks and effects of clustering indicated by tourist movement patterns. This study will fill the research gap comprising the ignored aspects of tourist geotagged information in such clustering effects from the destination perspective.

1.3. Information network analysis

Objects in most real systems (e.g., interpersonal communication, internet, biological network) tend to interact with each other, which are then modelled as information network. Rich structure and semantic information contained in the nodes and linkage of information network has attracted wide attentions to be studied. Nowadays, information network analysis is a trending topic in data mining for different disciplines. Researchers attempt to develop different approaches to mine hidden patterns from these interconnected linkages. For example, social network analysis and hypertext web mining are two well-known application scenarios. Particularly, Sun et al. (2009) proposed heterogeneous information network distinguished from homogeneous information network. The objects and linkages in homogeneous information network are usually the same type, while the objects and linkages in heterogeneous information network can be modelled as different types. Multi-typed objects and linkages in the network carry richer semantic information. It presents more opportunities and challenges for researchers to retrieve patterns by developing advanced mining and analysis algorithms.

Generally, information network analysis tasks contain similarity measure, clustering, classification, link prediction, ranking, and recommendation. (Shi et al., 2017) As the most basic task, similarity measure can conduct through either object feature comparison (e.g., cosine similarity) or linkage structure exploration. For example, Jeh and Widom (2003) constructed personalized-view PageRank algorithm to measure the probability from one object to another, which reveals the similarity between two objects. Following similarity measure, clustering is a task to group similar objects together. Then, each cluster has unique properties compared with the rest of the network. Modularity (Blondel et al., 2008) is a famous method to divide network as different clusters. To predict label of different classes, collective classification (Sen et al., 2008) is a common and effective network classification method, which captures correlations between objects, such as homophily, influence and confounding. Link prediction could be considered as a binary classification task. In information network analysis, ranking is a significant data mining task. For example, PageRank (Page, Brin, Motwani, & Winograd, 1999) decides the importance of a node by its linkages and their quality. HITS (Kleinberg, 1999) ranks objects by assigning pages as authority and hubs. Recommendation is the task to present appetizing products to consumers, which contains the task to measure object similarity as well as the interest of consumers. Social relationships among consumers from social network have been widely used in recommender systems.

As a complex interconnected system, different types of objects in tourism are linked with each other with different types of links. Evidently, information network in tourism tends to be heterogeneous. Therefore, traditional information network analysis for homogeneous network may not fit in tourism. Back to the basics, the most primary thing is to construct a suitable information network for research. With regard to advanced information network analysis methods for heterogeneous network, they also have received wide attention. For example, PathSim (Sun et al., 2011) has been proposed to measure objects' similarity via structural paths at meta level so that different linkages in the network are all taken into consideration. Combining link and attribute similarities has also been proposed to increase the clustering accuracy. (Yang et al., 2009) Complyed with developing information network for tourism research questions, novel analysis approaches should be also proposed.

1.4. Research gaps and objects

As is stated above, tourism is a complex interconnected network, where geographic element is significantly valued. Detailed study on constructing information network and following analysis for tourism research revealing the relationship between tourists and tourism attractions is necessary. Tourism attractions are interconnected via tourist interest. It is still rare to explore relationships between the attractions through network analysis for better marketing resource allocation and prioritization. Interconnected tourist needs are always considered in modern tourism study, but it is difficult to be quantified. Meanwhile, geographic proximity is a common factor in travel decision, revealing tourist needs. It is rare to consider quantifying the effect of geographic proximity as well as its effect scale in tourism study. It is no longer reasonable to evaluate attraction's performance in isolation. Ranking analysis task can be conducted within constructed information network so that the attractiveness of tourism attractions can be assessed. Similarly, clustering analysis task can help increase the competitiveness of a destination by balancing development of attraction bundles. Finally, attractions are connected, and destinations are connected in a regional view. Most destination image studies have neglected

Gunn's founding principles and have focused exclusively on single attractions or destinations.

To sum up, the study is objected to construct information network for tourism attractions and develop novel approaches to mine valuable information or tourism marketing. Particularly, the research is to evaluate the attractiveness of tourism attractions with consideration of attraction relationships, to model the effect of geographic proximity and its effect scale in tourism research, to discover intra-attraction features of tourism attractions in a rational way, beyond simplistic analyses based on geographic proximity or visitor volume data and to interpret organic destination image in regional context.

1.5. Structure of this study

The structure of this thesis has been arranged as follows. Chapter 1 is intended to give a comprehensive introduction including interconnected characteristics of tourism industry, development of geographic related and information network research. Research gaps and objectives are then narrowed down. Chapter 2 presents multi-source geotagged data from social media providing clue to tourist movement and behavior. Information network for tourism attractions then could be constructed. In chapter 3, ranking analysis within information network is conducted to assess attractions' attractiveness. A new algorithm - AttractionRank is proposed as a reasonable and fairer way to assess tourism attraction's attractiveness by eliminating the distance-decay effect scrupulously in the attractiveness propagation process. Chapter 4 presents the clustering analysis, which conducts destination-wide analyses of attraction clustering, and explores the relationship between attractions in different hierarchies to facilitate destination planning from a broader perspective than analyses that narrowly focus on certain attractions. Beyond tourism attractions themselves, tourist concerns and needs are also interconnected, so Chapter 5 conduct organic destination image analysis in a contemporary regional context so that a comprehensive apprehension about the attitudes of entire tourists can be revealed. In the end, concluding remarks are presented in Chapter 6.

Chapter 2. Multi-source geotagged data

Informed participation of all stakeholders is strongly encouraged in sustainable tourism development (World Tourism Organization, 2005b). Online communication plays a key role in disseminating information and shapes the tourism field into an intense informational pattern (Narangajavana, Fiol, Tena, Artola, & Garcia, 2017). With the tremendous information on social media, it allows for strategic data mining in understanding users' travel reflections, which further inform the marketing strategies for DMOs and those potential users' preferences for products and services. At the same time, social media itself also affects the beliefs, values, attitudes, intentions, and behaviors of people (Adi, Wihuda, & Adawiyah, 2017; Lai & To, 2015; Moradi, Yahya, Mohamed, & Raisian, 2017). With the recent big data technology development, this platform provides new possibilities to tourism research. Specifically, geotagged photos, posts or online check-in data on social media provides clue to tourist movement, which is unprecedentedly intensive to explore tourist behavior in a data-driven way.

This research focuses on the tourism attractions in the first place and then scales to destination level sticking to the idea that tourism is a networked system. For the city-wide destination study, Hong Kong is selected as the context and Instagram posts on the tourism attractions in Hong Kong are utilized for constructing attraction network based on tourists' interest. Guangdong-Hong

Kong-Macao Greater Bay Area is the context for regional tourism research and "the Red" platform is social media data source.

2.1. **Hong Kong and Instagram**

As a world-famous destination city, Hong Kong has long been the top tourist destination in Asia. Tourism is a major part of Hong Kong's economy, contributing approximately 4.5 percent of the GDP in 2018. Affected by local social incidents, tourist arrivals in 2019 recorded a year-on-year decline of 14.2% to 55.91 million, and tourism expenditure associated with inbound tourism also decreased by 22.7% to HKD 256.22 billion (USD 33.06 billion) according to the Hong Kong Tourism Board (HKTB) (2020). To maintain Hong Kong's status as a world-class premier tourism destination, the government is attempting to enhance the city's appeal in a balanced, healthy, and sustainable way (Hansen, 2021). Hong Kong has distinct tourist resources, including iconic urban areas, a unique cultural heritage, and a flourishing ecological environment. According to the HKTB, there are 204 attractions in three categories, natural, cultural, and urban (Figure 1). As Hong Kong offers such a wide range of attractions, it is an appropriate context for applying analytics to investigate the attractiveness of attractions in different categories to better understand tourists' preferences. This is especially relevant in attraction networking analysis for marketing segmentation and prioritizing tourist resources for effective destination management.

To investigate how geotagged social media information can be used to measure the attractiveness of a destination effectively, researchers referred to Instagram data. As of Quarter 3 in 2020, the platform has over one billion monthly active users, with females constituting 51% of the user base. A total of 29.6% of users are aged 18–24, and 33.1% are 25–34 (Iqbal, 2021). Similarly, according to the HKTB (2020), the mean age of tourists by major market area in 2018 was 36.3, and 57% of such tourists were female. Such similar demographic characteristics further affirm the appropriateness of using Instagram data in our investigation to improve the reliability of the destination attractiveness measurements. In addition, the large pool of social media users coupled with their massive sharing of data firmly indicates the strength and power of social media for the dissemination of information; this offers great opportunities for collecting and processing social media data for various purposes.

Tourist movement information can be tracked through the spatial information they self-reported on Instagram, presented as Points of Interest. Each attraction can match a Point of Interest with its own profile page that exhibits all the posts that geotagged it on the platform. A total of 606,640 Instagram posts from 202 attractions spanning four years (from December 15, 2014, to December 14, 2018) were collected in December 2018 ("Hung Shing Temple at Kau Sai Chau, Sai Kung" and "Seven Sisters Temple" were excluded due to unavailable Points of Interest on Instagram). Each post includes user ID, username, publication time, location name, and content. As shown in Figure 2, fewer than 15 sites had over 10,000 visits and tourists. Almost half of the attractions received fewer than 100 visits and tourists, showing a thick-tailed distribution. This phenomenon may help to identify tourist hotspots from the perspective of visitor flows on Instagram.

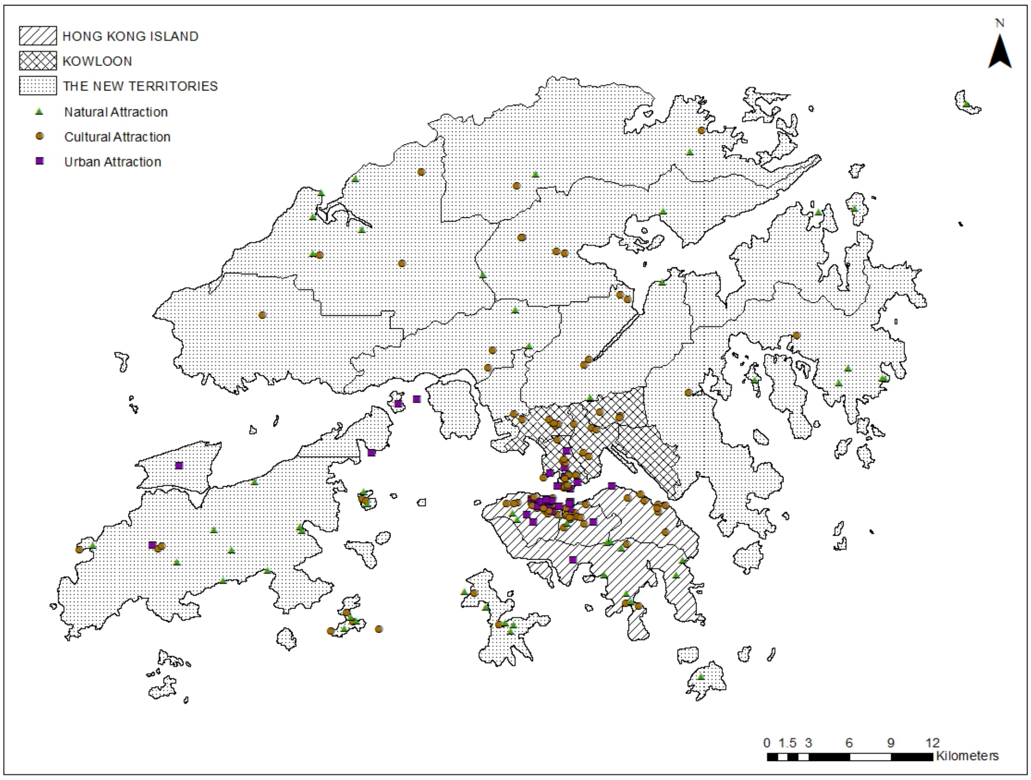


Figure 1. Attraction Distribution in Hong Kong

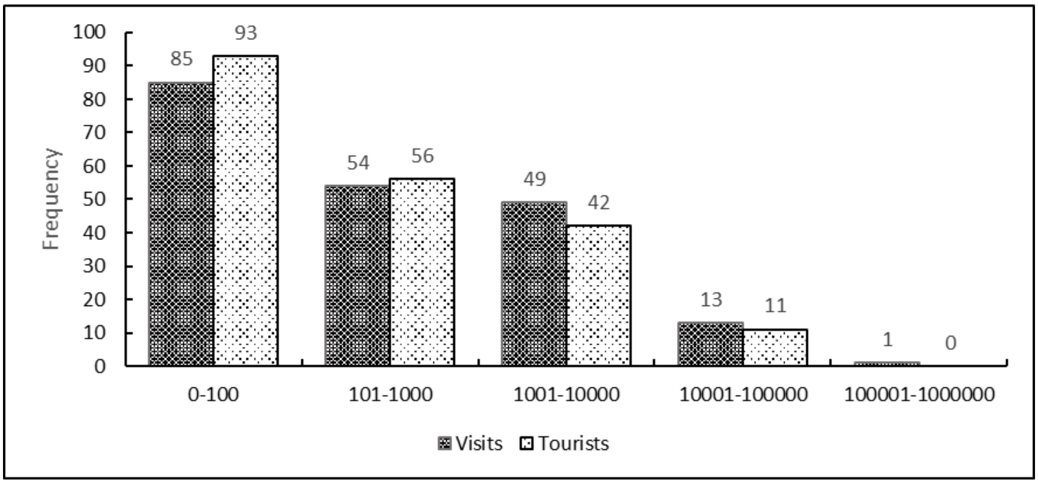


Figure 2. Frequency Distribution of Visits and Tourists per Attraction

2.2. Guangdong-Hong Kong-Macao Greater Bay Area and the Red

The geographic proximity of the 9+2 cities facilitates domestic tourist movements between Guangdong Province in mainland China and Hong Kong and Macao SARs (Figure 3). The GBA is deemed a suitable study location, particularly during the COVID-19 pandemic. When the virus was first identified in December 2019, border closures stemmed the spread. There was minima travel activity in Mainland China through early 2020, until the pandemic was largely brought under control in May. With gradual reopening of mainland tourism destinations, Chinese domestic tourists were the first to be permitted within the GBA (Yogonet, 2020). At the time of writing people may move freely and without quarantine around the nine cities within Guangdong (Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen or Zhaoqing), including those crossing the border from Macao to the GBA (since September 2020) (Teledifusao de Macao, 2020). If more optimistic forecasts materialize, tourism activities within the GBA (except Hong Kong) are expected to resume more rapidly than elsewhere within China. During the 4th quarter in 2020, about 56.7% of tourist arrivals in Macao came from the nine cities in Guangdong, indicative of the resilience of movements within the city network. This share marks a considerable increase on what was reported during the equivalent period in 2019 (53.3%) (DSEC, 2021).

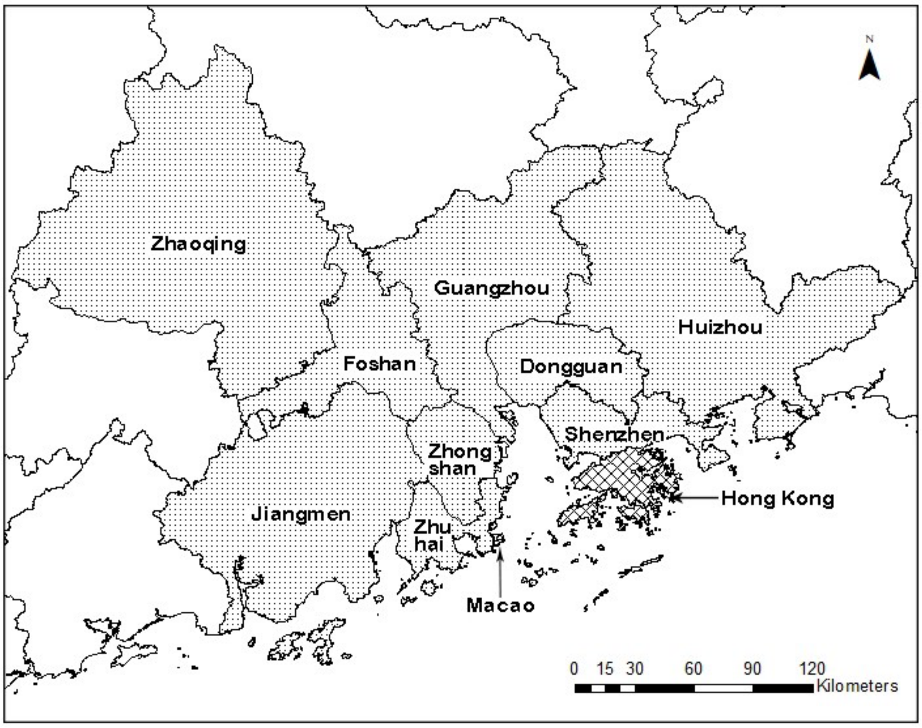


Figure 3. Map of the Greater Bay Area

The Red is a popular application (App) for social sharing (generating around 80% of its total earnings) and e-commerce (about 20%) in mainland China with over 300 million registered users in July 2019. Searchable and available topics on the platform include tourism, education and fashion (Xiaohongshu, 2021). It exemplifies the emergent platforms which can be used to explore tourism and destination image related contexts. The Red allows users to post their feelings in video or photo forms, together with descriptive texts. The popular Hashtag function summarizes the key topics about the postings and allows other viewers to search easily for the information. As social media usage in China has escalated, intending tourists have used social media platforms to seek out destination information and to share their travel experiences, particularly amongst millennials. The Red has been a fast-emerging social media platform, and it is timely to examine how its users contribute to the formation of organic destination image.

To acquire posts for each GBA city they inputted the phrase combining “city” and “travel”. This might involve searching for a keyword like “Macao Travel” (Python in Chinese characters as “澳门旅游”). All of the acquired posts were subsequently sorted based on user provided “likes”, “stars” and “comments”. In the current study, the researchers extracted the top 1,000 posts (around 400 words on average per posting) for each GBA city. The maximum number is 1,000 in light of Application Programming Interface (API) restrictions. In total, 10,957 posts were collected from the Red in May 2020. Each post includes a series of information such as “post ID”, “keyword”, “user nickname”, “publish time”, “like number”, “comment number”, “top comment”, etc. Table 1 shows examples of the data for purposes of text mining and analysis.

Table 1. Data Examples Extracted from the Red

Post ID	**User ID**	**Keyword**	**Title**	**Content**
592d5a* **	592eaf***	Dongguan Tourism (东莞旅游)	A wonderful place for an evening in Dongguan (东莞夜晚消遣好去处…)	Xiaba House, a creative art house in Dongguan… (下坝坊，一个东莞比较有创意的艺术坊…)
5d8866* **	5ad2d8***	Foshan Tourism (佛山旅游)	Another "Jiuzhai Valley" in Dongguan (佛山的 "九寨沟")	Tourism recommendation around Foshan: Another "Jiuzhai Valley" in Guangdong province, the environment is… (佛山周边旅游推荐广东的九寨沟，环境…)
5db6ac* **	5c8513***	Macao Tourism (澳门旅游)	Know more about tourism accommodation in Macao (澳门旅游住宿知多D)	Today is another sunny weekday. I would like to share some information about a five-star hotel in Macao with you… (又是一个晴朗的工作日，今天和大家分享的是澳门5星酒店…)
…	…	…	…	…

Chapter 3. AttractionRank: extended PageRank for assessing tourism attraction’s attractiveness

3.1. Introduction

The primary purpose of tourism attractions is to attract existing and potential visitors to make a journey to or within a destination. Therefore, the assessment of attraction’s attractiveness is critical for tourism planning, management, and marketing. A possible measurement is to evaluate each attraction’s performance independently like via on-site passenger flow. In practice, attractions do not stand in isolation, but connect with other attractions and collectively build the entire destination attractiveness within the attraction system. It would be more reasonable to investigate the attraction relationships rather than a single node information when valuing the attractions’ performance. Tourist movements from one attraction to another implicitly reveal a shift of motivation. Thus, every attraction contributes to the entire attractiveness propagation. With the growing application of Location Based Social Networks (LBSNs), there is an increasing opportunity to collect tourist movement when tourists geotag locations they visited. Attractions are then connected based on tourist movement in a sense: the edge is the movement from one attraction to another and its weight represents the movement frequency. It is worth noting that connections between attractions should not be treated equally yet. A movement from a well-known attraction is significantly different from a movement from a less popular one. Otherwise, an attraction that is attractive to niche market may be artificially ranked high if it has many movements only from a few specific attractions.

The mathematics of PageRank (Page, Brin, Motwani, & Winograd, 1999) sheds light on a ranking method considering the global information recursively instead of local individual information in a network. The importance of a node is decided by its links as well as the quality of these links. Thus, Weighted PageRank (Xing, & Ghorbani, 2004) could be adapted to the attraction network. However, a critical issue observed is that the majority of tourism studies highlighted the distance-decay impact on destination choice and focused on the pre-trip stage in travel decision making process (Sun, Fong, & Law, 2018). That is, when constructing the attraction network from tourist movements (i.e., behavior after decision making), the distance-decay effect is embedded in the network linkage weight. The direct Weighted PageRank result is correspondingly affected by geographic proximity. If a tourist travels long distance, the travel cost to visit an attraction increased correspondingly, thus it reveals the high level of attractiveness of this attraction. It is also common that tourists may visit secondary or tertiary attractions on site due to geographic proximity to primary attractions within a destination (Richards, 2002). Therefore, ranking result of those densely distributed attractions may also be inflated. By eliminating such a geographic effect, attractions can be ranked in a fairer way so that Destination Marketing Organizations (DMOs) may master attractions more accurately and better plan the route from an even more objective way. Another noteworthy issue is the effect scale of distance-decay effect in the attraction network, which is usually measured by a power-law or exponential functions (Barthélemy, 2011). The choice of the function and its real impacts should be handled carefully. An inappropriate or poorly calibrated function would cause poor ranking result.

3.2. PageRank algorithm and its extensions

PageRank (Page, et al., 1999) is a graph-based ranking algorithm for measuring and scoring the relative importance of each webpage within the hyperlink set. A linkage from a page to another is regarded as a vote for authority. Both the quantity and quality of such a vote are taken into consideration when calculating the page's score: if Page A was pointed to by many pages or some pages with high PageRank themselves, it would be well worth a browse with a high PageRank score (Brin & Page, 1998). By generating the ranking result through the complete graph via an iterative process, the algorithm is therefore more reliable than those focusing on little set. More specifically, PageRank score of a webpage means the probability that a user will visit it by randomly clicking on links:

$$PR(P_i) = \frac{1-d}{N} + d \times \sum_{P_j \epsilon In(P_i)} \frac{PR(P_j)}{L(P_j)} \qquad (1)$$

where p_i, p_j are webpages for evaluation, $PR(p_i)$ is the PageRank score of p_i,

d is a damping factor, N is the total number of webpages, $In(p_i)$ represents the set of webpages that link into p_i, and $L(p_j)$ means the number of outbound webpages from p_j. The formula reflects that the PageRank score of p_i depends on the PageRank score of each p_j linking into p_i. The probability distribution is assumed divided evenly at the beginning of computation. Following the link structure in the network, PageRank scores will be adjusted to equilibrium after enough iterations. Seeing the possibility of sinks (webpages having no outbound links), the equation introduces the damping factor d (0<d<1) to represent the probability that the random surfer could continue clicking on links (i.e., the random surfer does not arrive at sink webpages.) Hence, $\frac{1-d}{N}$ describes the situation that the surfer picks up any webpages randomly and starts surfing again when arriving at a sink.

Subsequently, PageRank was extended for a weighted network (Xing, & Ghorbani, 2004) by taking the importance of a link itself into consideration and proved to identify more relevant pages in a given query than standard PageRank. It was then applied in geographic scenarios to identify important places where people are likely to gather. Jiang (2009) employed the weighted PageRank (WPR) to the connectivity graph of a city and found that the result well correlated with human movement rates. Chin and Wen (2015) introduced GPR to identify the spatial concentration of human movement in a spatial network by further including geographic considerations in WPR. Both geographic proximity and location attractiveness (i.e., incoming links of a location) were included in the link weight. The major geographic proximity concern was the distance-decay effect, which was captured by the inverse distance in their model. AttractRank (Xie et. al, 2021) also employed the inverse distance as the basic probability accessing a district from another and integrated taxies' origin-destination data with WPR when calculating the city's district attraction ranking.

Like the studies mentioned above, distance factor has been included while applying PageRank algorithm in urban, economic, or cultural geography. AttractRank employed the inverse distance (power-law function) to describe the distance-decay effect directly. GPR mentioned the sensitivity of choosing distance-decay functions. They conducted a posterior correlation analysis between human movement concentration (e.g., the density of population and automobile flow) and rank result and found out the most appropriate distance-decay function and its control factor in different case studies. The algorithm was then proved to identify the spatial concentration of human movement more effectively than other traditional network metrics. However, it is difficult to collect ground truth data for such statistical test when assessing the tourism attractiveness. Our proposed method includes a priori workflow to determine the appropriate distance-decay function as well as its effect scale. Furthermore, relationship pattern inside the attraction network reveals the extent of human interests or motivations to the attractions. Motivations are argued to be correlated to the travel cost (e.g., geographic proximity: Tourists may go to secondary or tertiary attractions on the count of adjacency.) That is to say, the formation of this real-world network has embedded geographic influence. Therefore, our extended PageRank algorithm eliminates the distance-decay effect rather than take it as the linkage weight in order to relieve the possible inflation in attractiveness evaluation based on WPR.

3.3. Extending PageRank for assessing tourism attraction's attractiveness

3.3.1. Attraction network

People share their geotagged media content in LBSN so that their movement can be tracked throughout their posts. A tourist trajectory T_i among different attractions can then be represented as $T_i = \{AP_{i1}, AP_{i2}, AP_{i3}, \dots, AP_{in}\}$, where AP_{in} denotes the n^{th} attraction Point of Interest (POI) geotagged (i.e., visited) by the i^{th} LBSN user. According to the sequence in the trajectory, directed links can be formed between adjacent attractions indicating tourists' interest shifting. As a result, Attraction A_i may have multiple links to or from Attraction A_j. By integrating all tourists' trajectories, a directed weighted network $G = (V, E)$ can be constructed among all attractions. V is the set of attractions and E is a set of ordered pairs of attraction nodes. Weights are assigned to these pairs, which reveal the number of tourists moving between corresponding attraction nodes.

3.3.2. Overview of AttractionRank

Figure 4 shows the workflow of AttractionRank. First, a table of attraction POIs is built by combining descriptive information and its geo-information from LBSN by matching attraction names. Then, tourists' trajectories among attractions are extracted and the attraction network is constructed as explained in section 3.1. On the other side, a distance matrix is built based on the coordinates of attractions, where each element is the distance between corresponding attraction nodes. Then a workflow to find out an appropriate set of distance-decay function and its effect scale which is introduced in detail in section 3.4. Finally, the calibrated distance-decay function is utilized in AttractionRank in order to evaluate the attractions' space-independent attractiveness.

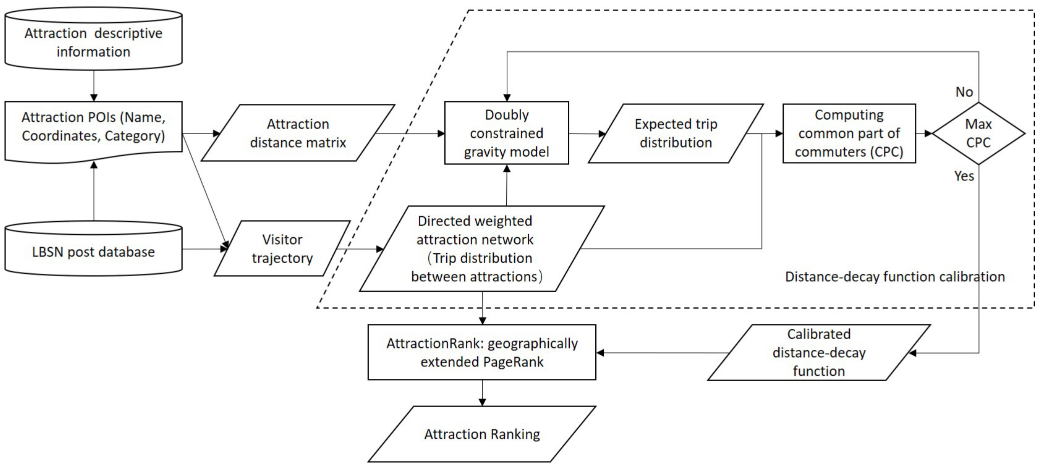


Figure 4. Workflow of AttractionRank

3.3.3. AttractionRank algorithm

In order to rank attractions in a reasonable way, we propose a geographically extended PageRank algorithm named AttractionRank. As is explained in section 3.2.1., the attraction network is a weighted directed graph. Like hyperlinks between webpages, some attractions could be sink nodes (i.e., attractions without outbound links). AttractionRank should enable an edge weight factor and preserve the damping factor based on PageRank algorithm. It is formulated as follows:

$$AttractionR(i) = \frac{1-d}{N} + d \times \sum_{j \epsilon In(i)} AttractionR(j) \times \frac{GW(j,i)}{\sum_{k \epsilon Out(j)} GW(j,k)} \quad (2)$$

where $AttractionR(i)$ represents the AttractionRank value of attraction i; d is the damping factor; N is the number of attractions in the network; $In(i)$ represents the set of attractions that link into/ destinate at attraction i; $Out(j)$ means the set of attractions that come out/ originate from attraction j; and $GW(i,j)$ is a geographically calibrated linkage weight from attraction i to j. The major geographic consideration is the travel cost between two attractions. As distance-decay effect tells, the strength of spatial interactions between two locations decays with the increasing travel cost. A link from one attraction to another indicates the corresponding tourist's interest shift, where the embedding motivation is always affected by geographic proximity. That is to say, the internal edge weight in the attraction network (i.e., the frequency of moving from one attraction to another) is influenced by geographic proximity as distance-decay. Ranking attractions in a normalized way is a process to eliminate the influence of travel cost, so $GW(i,j)$ is formulated as:

$$GW(i,j) = \frac{W_{ij}}{f(d_{ij})} \tag{3}$$

where W_{ij} is the edge weight of the link from attraction i to j in attraction network, and $f(d_{ij})$ is the function describing the distance-decay and its effect scale. Travel cost is often modeled as a power-law or exponential function, and the relative performance of the two forms in gravity model depends on the dataset used (Barthélemy, 2011). As a result, both forms are considered in AttractionRank algorithm and their effect scale is adjusted by the parameter β:

$$f(d_{ij}) = \exp(-\beta d_{ij}) \tag{4}$$

Or

$$f(d_{ij}) = d_{ij}{}^{-\beta} \tag{5}$$

Meanwhile the distance-decay effect can be utilized in gravity model to estimate the tourist movement among the attractions. By being calibrated with the real trip distributions, $f(d_{ij})$ can then be confirmed in given scenario. The process is detailed in the next section.

3.3.4. Distance-decay function calibration

Gravity law is a frequently used trip distribution law for estimating population movements. It assumes that the number of trips decays with the increasing distance. Meanwhile, the attraction network explained in section 3.1 can be considered as the observed trip distributions between attractions. By calibrating the gravity model with the observed model, the most appropriated distance-decay function with the corresponding effect scale for the given case can then be determined. According to Lenormand et. al. (2016), the doubly constrained gravity model outperforms singly constrained models (i.e., production constrained model and attraction constrained model) and unconstrained model when materializing the people commuting. The doubly constrained gravity model is thus employed in AttractionRank. By assigning O_i as all the trips originated from (generated by) attraction i, and D_j as all the trips destinated to (attracted by) attraction j, K_i and K_j as two balancing factors of the constraints. The estimated trips $\tilde{T}_{ij}$ from attraction i to attraction j can be modeled as:

$$\begin{cases} \tilde{T}_{ij} = K_i K_j f\left(d_{ij}\right) \\ \sum_{j=1}^{n} \tilde{T}_{ij} = O_i, \sum_{i=1}^{n} \tilde{T}_{ij} = D_j \end{cases} \tag{6}$$

Common part of commuters (CPC) (Lenormand et al. 2012) is then utilized to represent the goodness of fit. It is a similarity index varying between 0 and 1 describing the degree how commuting flows are correctly reproduced by the model on average. Its formula is as follows:

$$GPC\left(T, \tilde{T}\right) = \frac{\sum_{i,j=1}^{n} \min\left(T_{ij}, \tilde{T}_{ij}\right)}{\sum_{i=1}^{n} \sum_{j=1}^{n} T_{ij}} \tag{7}$$

where T_{ij} is the observed trips (i.e., the edge weight of the linkage) from attraction i to attraction j. Doubly constraint models usually use Iterative Proportional Fitting (IPF) procedure for calibration (Lenormand et al. 2016). As a least square fitting procedure of a two-way contingency table (Deming, & Stephan, 1940), IPF is sensitive to extreme values. Therefore, the existence of sink attractions with frequent zero T_{ij} as extreme value in the network would significantly affect the fitting result, which affects the calibration result of distance-decay function successively. To alleviate this problem, AttractionRank calibrate the gravity model by using median polish fitting, which is robust to extreme values (Kunut, 2010).

3.4. Experimental results

3.4.1. Distance-decay calibration results

Figure 5 shows the CPC obtained between the ground truth trip distribution and the estimated one from doubly constrained models with different parameter β values in exponential and power distance decay function (equations 4 and 5) for the study case. Both functions share the same CPC as 0.4783 when β is set as 0 which means that no distance decay effect is included in the estimated trip distribution. CPC values get larger when β is set as non-zero values. It also demonstrates the significance of the geographical proximity when estimating trip distribution. The model with exponential distance decay effect obtains the largest CPC when β is 1.40625. Power distance decay function with its effect scale at 3.25 gives a better result whose CPC is 0.60862. In addition, CPC result here is also a reasonable value according to previous empirical CPC result. (Lenormand, Bassolas, & Ramasco, 2016). Therefore, a geographically calibrated link weight (equation 3) is then valid for this case as: $W_{ij} \times {d_{ij}}^{3.25}$.

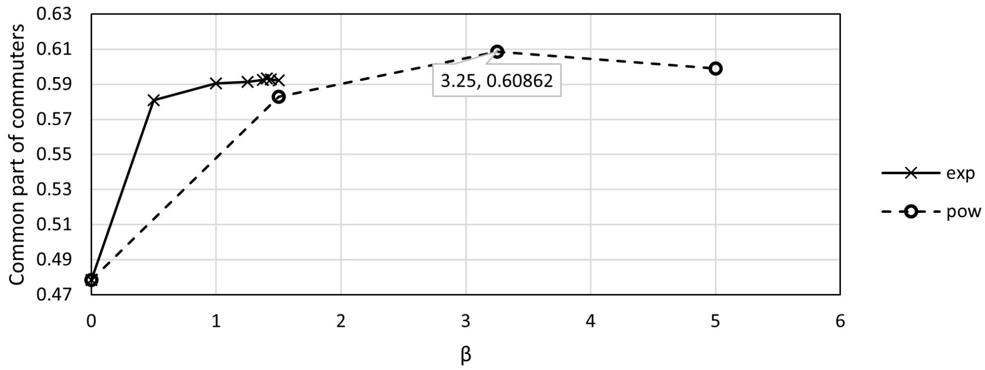


Figure 5. Common part of commuters according to the doubly constrained model: the crosses represent the exponential distance decay function; the circles represent the power distance decay function.

3.4.2. Ranking results and evaluation

202 attractions are finally ranked by our proposed AttractionRank. AttractionRank is expected to reveal hidden attractive tourism sites by taking attractiveness propagation into consideration as well as factoring out the effect of distance. The AttractionRank result is then compared with both the rank results based on passenger flow and WPR. The passenger flow on an attraction could reflect its popularity and attractiveness. Tourist hotspots can be identified by the visitor arrivals. Since some attractions have more repeat tourists, results based on number of visitors and visits would be different. We use a simple index positively correlated with both the number of visits and visitors to conveniently describe the relative popularity of an attraction: PFI = log (number of visits * number of

visitors). A rank of attractions is then generated by this index. In addition, WPR (without distance decay effect removal) is also applied to the attraction network for ranking.

To evaluate the performance of the proposed methods in revealing the attractiveness of attractions, we use the attractions in 'top-10-experiences' guide from a local perspective as listed on Hong Kong Tourism Board (HKTB) as more attractive sites (as set X below). Mann-Whitney U test is then conducted on three ranking results respectively. Table 2 shows the U test result to examine whether the rank of attractions in set X significantly differs from those not in X according to different ranking schemes. All P values are less than 0.05, which means that all three ranking methods can significantly distinguish those more attractive sites. Median ranks of those attractions not in set X are 106 in the same way based on three ranking methods. Median rank of attractions in set X according to AttractionRank is 14 among totally 202 attractions, which is smaller than 19 and 21 according to WPR and PFI respectively. Hodges-Lehmann difference ($HL\Delta$) is the median of all possible differences between an attraction's rank in set X and an attraction's rank not in set X. AttractionRank improves attractions' rank in set X by $HL\Delta = 68$, which is also larger than that of ranking results based on PFI and WPR. In this sense, all three ranking methods can effectively reflect those more attractive attractions while AttractionRank performs the best.

Figure 6 compares the AttractionRank result with FPI and WPR result of the studied 202 attractions in three distinct categories (i.e., natural, urban, and cultural attractions). Points above line Y=X means that the according AttractionRank result is higher than ranking results based on PFI or WPR. Line Y = X-20 and Y = X+20 both give a range of 20 ranking differences. In general, for the upper AttractionRanked attractions (even before top 150), their ranking result based on PFI and WPR is close. By factoring out the distance-decay effect, if attractions are promoted by AttractionRank, outstanding ranking differences are observed (larger than 20); while for those depressed attractions under AttractionRank scheme, the ranking differences are moderate (less than 20). This is especially evident for top 50 AttractionRanked attractions. The tail part of AttractionRanked sites is rated almost the same as the result based on WPR. At the same time, passenger flow on these sites is also low (PFI ranks behind 130). That is, distance-decay effect is no longer distinct if the attraction is less popular. As for attraction category, current efforts on attraction promotion mainly allocated to those natural and cultural attractions. It is reasonable that these types of attractions tend to be primary attractions, and some are far away from urban area and other unknown attractions. It really depends on tourists' intense interest to visit the sites rather than occasionally dropping by. Therefore, the values of these attractions are worth mentioning.

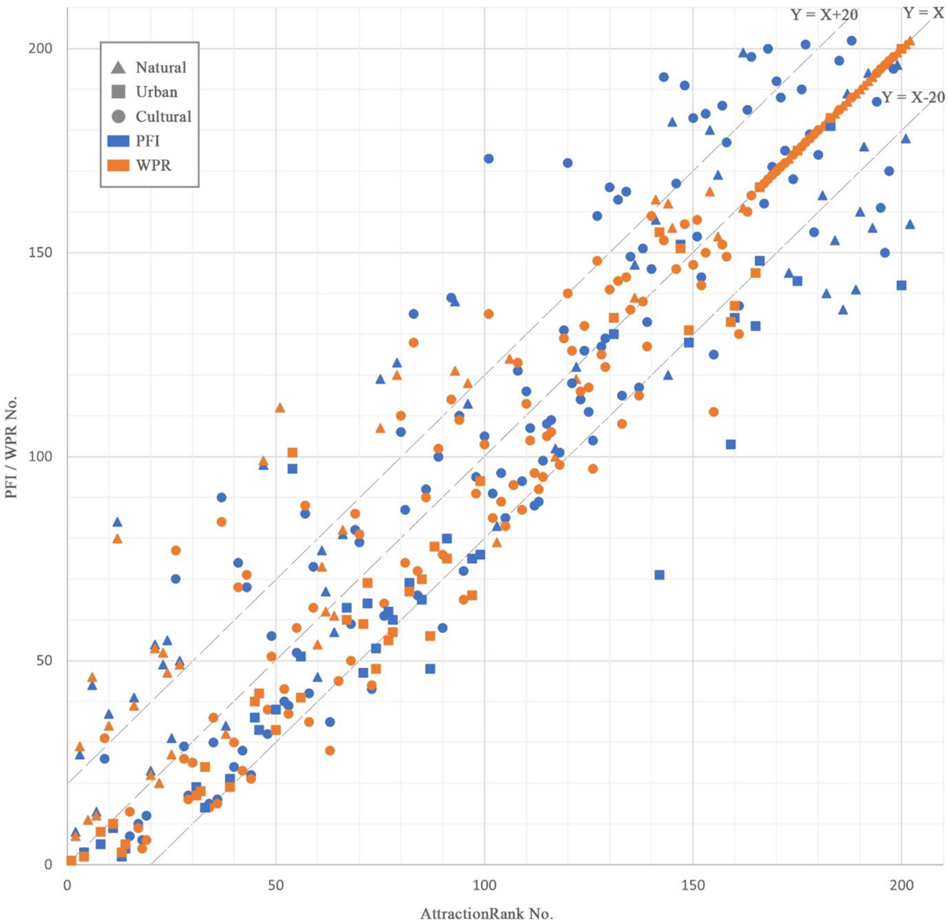


Figure 6. AttractionRank Versus PFI and WPR of 202 Attractions by Category

Table 2. Mann Whitney U Test Result for significant rank difference of attractions in set X or not based on three different methods

	Group (N)	Median Rank	U	P	$HL\Delta$
PFI	In X (13)	21	490	0.0001	64
	Not in X (189)	106			
WPR	In X (13)	19	537	0.0004	62
	Not in X (189)	106			
AttractionRank	In X (13)	14	481	0.0001	68
	Not in X (189)	106			

To provide a more specific discussion on the ranking results, the top 20 attractions based on three different ranking methods are shown in Table 3. Hong Kong International airport is ranked No.1 among all attractions whatever the ranking scheme is. This is reasonable due to its exceptional status as the international passenger hub and gateway of the destination Hong Kong.

Table ***3.*** Top 20 Attractions based on Passenger Flow Index (PFI), weighted PageRank (WPR), and Attraction*Rank (Attraction Category: (1)-Natural attractions (2)-Cultural attractions (3)-Urban attractions)*

No.	**Attraction Name (Category)**		
	PFI	**WPR**	**AttractionRank**
1	Hong Kong International Airport (3)	Hong Kong International Airport (3)	Hong Kong International Airport (3)
2	Hong Kong Disneyland (3)	Hong Kong Convention and Exhibition Centre (3)	Cheung Chau (1)
3	Hong Kong Convention and Exhibition Centre (3)	Hong Kong Disneyland (3)	Tai O (1)
4	Lan Kwai Fong (3)	PMQ (2)	Hong Kong Convention and Exhibition Centre (3)
5	Ocean Park Hong Kong (3)	Lan Kwai Fong (3)	Dragons Back (1)
6	PMQ (2)	Tai Kwun (2)	Tap Mun (1)
7	The Peninsula (2)	Cheung Chau (1)	Stanley (1)
8	Cheung Chau (1)	Ocean Park Hong Kong (3)	Ocean Park Hong Kong (3)
9	Victoria Park (3)	Hong Kong Cultural Centre (2)	Po Lin Monastery (2)
10	Hong Kong Cultural Centre (2)	Victoria Park (3)	Lantau Peak (1)
11	Dragons Back (1)	Dragons Back (1)	Victoria Park (3)
12	Tai Kwun (2)	Stanley (1)	Tung Ping Chau (1)
13	Stanley (1)	The Peninsula (2)	Hong Kong Disneyland (3)
14	Avenue of Stars (3)	West Kowloon Cultural District (2)	Lan Kwai Fong (3)
15	West Kowloon Cultural District (2)	Man Mo Temple (2)	The Peninsula (2)

16	Man Mo Temple (2)	Clock Tower (2)	Sharp Island (1)
17	Clock Tower (2)	Hong Kong Observation Wheel (3)	Hong Kong Cultural Centre (2)
18	Temple Street Night Market (3)	Temple Street Night Market (3)	PMQ (2)
19	Hong Kong Observation Wheel (3)	SoHo (3)	Tai Kwun (2)
20	Lion Rock (1)	Lion Rock (1)	Mui Wo (1)

Extended PageRank (i.e., WPR and AttractionRank) evaluates the attractiveness in the view of attraction network rather than the single site performance as PFI. Its ranking results reveal attraction's importance in terms of attractiveness propagation in the network. Attractions like Hong Kong Convention and Exhibition Center, PMQ, Hong Kong Cultural Center, Tai Kwun are ranked much higher according to WPR than PFI. These attractions share the following features: they are artistic oriented attractions and located in urban areas of the city, thus can cover a wide range of public interest, and well connected to other sites in the area. Distance-decay effect is embedded in WPR's processed attractiveness propagation. The chances are that tourists drop by adjacent mass-market sites in the urban context thus these sites get more links, and their ranking may be overestimated by WPR. Their attractiveness was cooled down but remained at high levels after eliminating distance-decay effect in the presented AttractionRank.

By contrast, the ranking of attractions meant to attract a relative niche tourist market drops in the WPR results compared to PFI. Theme parks like Hong Kong Disney Land and Ocean Park Hong Kong are mostly favored by family tourists with children; the Peninsula are popular among luxury tourists. Such attractions possess more distinctive features attracting special interest tourists and receiving frequent visits than those for the ordinary people. Their AttractionRank results drop again no matter whether they are in urban area no not. It is reasonable that special interest tourists are more determined to visit adjacent secondary and tertiary attractions rather than wander to other remote sites. It is also worth noting that the rank of attractions with common interest like Victoria Park is similar with result of PFI and WPR.

The PFI ranks of popular natural attractions like Cheung Chau, Dragons Back and Stanley are consistent with the WPR results. Hiking is a popular activity in Hong Kong and these relevant attractions are famous for both local and non-local tourists. Their attractiveness may be underestimated since they are far away from most city area attractions and tourists have to take long distance to visit. AttractionRank lifts their ranks in the result by factoring out the distance-decay effect. As a result, the ranks of other natural attractions like Tai O, Tap Mun, Lantau Peak, etc. are also promoted under AttractionRank scheme.

3.5. Discussions

In this chapter, we propose AttractionRank, a more reasonable and fairer metrics to assess tourism attractiveness beyond the travel cost. Compared with other metrics like passenger flow on site, the AttractionRanked results reveal the connections between attractions rather than considering the attraction performance separately. Furthermore, it is a space-independent measure by factoring out the distance-decay effect on the spatial interaction between attractions with a calibration process. This process employs the doubly constrained gravity model to estimate trip distribution and involves statistical tests to ensure the significance of the pattern. As concluded from the case study of Hong Kong by using Instagram data, AttractionRank performed better than the alternative schemes when compared with the official top recommended attraction list from HKTB.

Attractions whose attractiveness is promoted by AttractionRank are more likely to be primary attractions aiming at public interest instead of a niche market. Values of attractions that tourists are willing to take long distance to visit are further highlighted; while the attractiveness of attractions that are densely distributed in the area and proximate to others is not inflated. The promotion of some attractions presupposes the depression of others. Thus, the attractiveness of some urban attractions is not that high under the AttractionRank scheme. However, the depression is more moderate than the promotion and those popular urban attractions are not under-evaluated. These results suggest that the proposed method could recognize hidden attractive attractions and keep dominant attractive ones. DMOs can utilize the results directly in further destination marketing and tourism route planning. For example, more convenient transportation can be assigned to those hidden attractive attractions. With the growth of LBSNs, this method is testable and applicable for other destinations. Future study can also consider utilizing multiple LBSN data sources to alleviate the bias of user-shared geotagged information on Instagram.

Chapter 4. Destination attraction clustering: telling the story of tourist movements with geotagged information

4.1. Introduction

The competitiveness of a destination is not the result of the attractiveness of a single attraction but of the balanced development of a bundle of attractions that together satisfy the needs of visitors across various tourism markets. However, efforts to market destinations and attractions simultaneously tend to narrowly focus on single sites, based simply on the popularity of such sites among tourists as measured by visitor volume. However, there are hidden treasures that are favored by special interest tourists and receive frequent visits. These attractions are thus easily ignored by mass tourism marketing campaigns. Methodologies for marketing grouped attractions are lacking; attraction clustering marketing strategies concentrate on grouping attractions that are physically close to one another for convenience and to stimulate tourist-intensive consumption (Gu, Zhang, Huang, Zheng, & Chen, 2021; Taylor, McRae-Williams, & Lowe, 2007). This does not reflect what happens in practice. People still tend to travel long distances between attractions simply because they are driven by strong interest.

Tourism clustering has been studied as an approach for combining similar (and/or potential) tourism destinations, products, and services in cooperation to better realize the competitive advantages of each (Kol'veková et al., 2019). Such clustering is derived from the cluster theory by Porter (1990), who originally focused on business cooperation between companies. Applied to tourism, such clusters refer to complex groups of various tourism elements such as tourism operators, infrastructure, activities, products, and services (Da Cunha & Da Cunha, 2005). Recent studies have concentrated on identifying the indicators of tourism clusters or assessing clustering potential (Laing & Lewis, 2017). Clustering can be used to describe a group of tourism attractions; however, the implications of clustering are often focused on business solutions for realizing competitive strategic advantages. There are no criteria for how a network of attractions is established or for how such clustering works. It is expected that this study will fill gaps in the research by exploring relationships between the attractions through network analysis and by further clustering the attractions into categories for better marketing resource allocation and prioritization.

Scholars have tried to track the movements of tourists with Geographical Information System (GIS) software (McKercher & Lau, 2008) or by using selected samples of self-reported geotagged information (Smallwood, Beckley, & Moore, 2012). Social networks have been used as cost-effective marketing tools with high returns, and this method is growing in prominence in global tourism (Trihas, Perakakis, Venitourakis, Mastorakis, & Kopanakis, 2013). Although a growing number of studies use geotagged information shared by users to track user movements, these studies have mainly focused on exploring and predicting tourist movement patterns at the individual level (e.g., Jiang et al., 2021). There is still a lack of studies on the synthetic analysis of geotagged information from location-based social networks that explore attraction clustering effects at the destination level. Attractions as geographical Points of Interest may be clustered based on their geographic proximity (which affects travel cost) by employing algorithms like K-means or DBScan (Ester, Kriegel, Sander, & Xu, 1996; MacQueen, 1967). They may also be categorized by their functions or by theme (e.g., natural, cultural, or urban sites). It is apparent that it is far from sufficient to cluster attractions based on such single attributes. However, few studies have assessed attraction clustering to highlight the interrelationships between attractions using different dimensions. This study will advance the use of this method by further exploring the connection between attractions and the effects of clustering them for destination marketing planning.

To that end, we aimed to explore and synthesize the effects of clustering tourism attractions and to furthermore discover intra-attraction features of tourism attractions in a rational way, beyond simplistic analyses based on geographic proximity or visitor volume data. Using the social media platform Instagram (a photo and video sharing social network that allows users to geotag their posts), in the context of Hong Kong, this study applied big data analytics to geotagged information shared by tourists to track their movement patterns between attractions and further explored the following research questions: 1) How does geotagged information in social media reveal tourists movements? 2) What stories does the constructed attraction network reveal after analyzing the effects of clustering? 3) How can analysis of the effects of attraction clustering be used for accurate targeting in (joint) marketing efforts?

4.2. Methodology

This study constructed an attraction network based on collected Instagram posts and then conducted community detection on the network to obtain attraction clusters.

Subsequently, the attractiveness of the attractions was assessed based on two criteria: 1) a simple numeric measure of attraction appeal (hereafter AppealI) achieved by adjusting node weight, and 2) a graph-based attraction ranking. Together with their themes (HKTB's natural, cultural, and urban categories), a theoretical model of attraction clusters was built to describe their specific characteristics from three dimensions. Figure 7 shows the flowchart of the methodology.

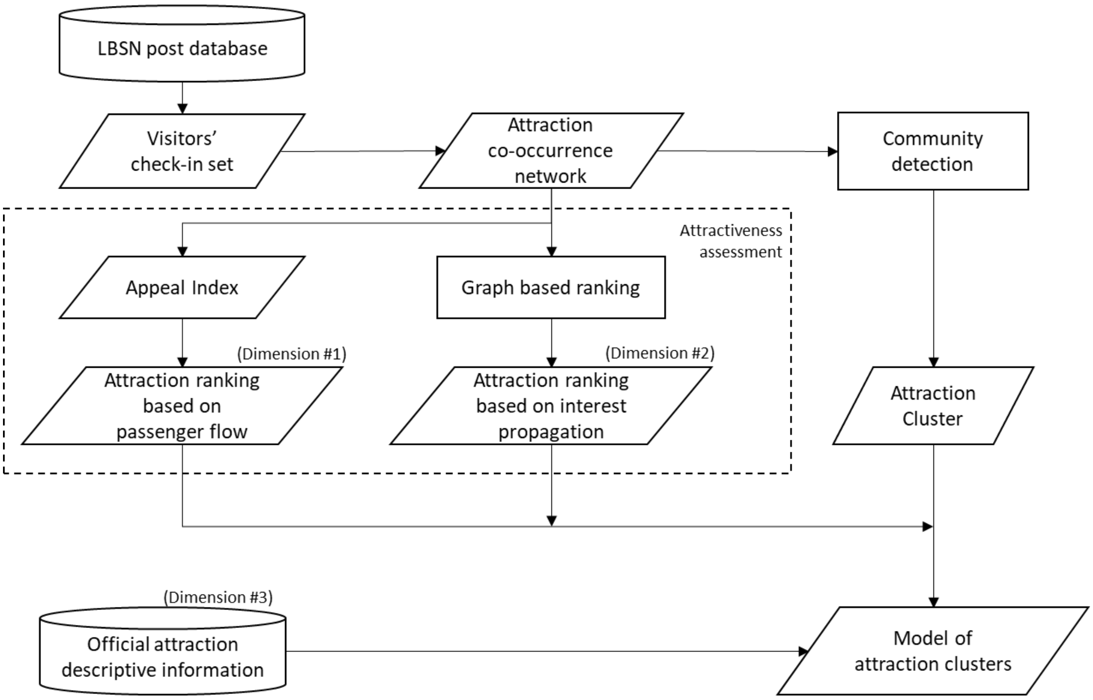


Figure 7. Overview of the Methodology

Connected attractions/ Attraction co-occurrence network: All elements (tourist, site, and marker) in an attraction system are necessarily connected (McKercher et al., 2006). Tourists' motivation to visit an attraction is always stimulated by some information (i.e., a marker) controlled by destination marketing organizations (DMOs) in their efforts to promote tourism (MacCannell, 1976, p. 41). Two attractions should be connected if both could attract the same tourist to visit them, such as through a shared element that captures the tourist's interest or stimulates their motivation. Ramires et al. (2018) used motivation to categorize tourist clusters. In the present study, the more the same tourists have visited two attractions, the stronger the relationship between the two attractions is indicated, and the connection edge between two attraction nodes reflects a process of shifting interests. The relationship pattern of the entire attraction network can then indicate the level of interest propagation and levels of motivation to visit different attractions.

The tourism movement of an Instagram user T can be tracked throughout all posts and represented as $T = \{P_1, P_2, \dots, P_n\}$, where P_n denotes the *n*th attraction the user geotagged (visited). A user could visit an attraction multiple times, so T could include repeated attraction visits. After removing duplicates in *T*, each user has an exclusive attraction check-in set. Finally, an undirected weighted network G = (V, E) is constructed, where V is the set of vertices in the network representing attractions, and E is the set of edges between vertices. The edge weight indicates the number of same Instagram users who geotagged both attractions. Meanwhile, the node weight is exactly the number of visitors to the attraction. The constructed attraction network in this study had 202 nodes and 7,812 edges.

Attraction clustering/ Community detection: To reiterate the network construction process in this study, the network originates from the visitation pattern indicating tourists' motivation. Tourists' motivation for an attraction is a comprehensive consideration of themes, travel costs, and other factors (Pearce, 1993). It is reasonable to infer that clustering based on the relationships in the attraction network could reveal tourist aggregation behaviors and common motivations to visit certain groups of attractions with comprehensive consideration of different factors.

Community detection is a method of identifying organizations inside a network as well as their functions (Lancichinetti & Fortunato, 2009). Each community has unique properties compared with the rest of the network. Fortunato and Hric (2016) recently reviewed algorithms for community detection in networks. Methods based on statistical inference, such as the degree-corrected stochastic block model (Karrer & Newman, 2011), require the number of communities, q, as preliminary knowledge. They are not suitable for this study because the attraction network is a real network whose q is unknown. Methods based on optimization, such as Louvain, or dynamics, such as Infomap, fit our purposes better. Emmons, Kobourov, Gallant, and Borner (2016) examined and compared the quality of four widely used network clustering algorithms at different scales. The Louvain algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), which is based on modularity optimization, surpassed others (including Infomap, a widely used dynamics-based method) in small-sized networks. Specifically, we employed the Louvain method to detect clusters in our 202-node attraction network. Modularity Q, the core issue of the algorithm, measures the relative density of the intra-cluster connections rather than the inter-cluster connections. It is defined as (Blondel et al., 2008):

$$Q = \frac{1}{2m}\sum_{vw}[A_{vw} - \frac{k_v k_w}{2m}]\delta(c_v, c_w) \tag{8}$$

where m is the sum of all edge weights in the network; A_{vw} is the weight of the edge between nodes v and w; k_v is the sum of edge weights that connect to node v; c_v is the community node v belongs to; and $\delta(c_v, c_w)$ is the Kronecker delta function: it is 1 if $c_v = c_w$. To detect the community, Louvain optimizes modularity in two steps: 1) each node is assigned to its own community and then moved to a neighboring community if it leads to the largest modularity increase; and 2) a new network is built from nodes in communities found in Step 1. Step 1 is then subsequently conducted on the new network. This process is repeated iteratively. In this study context, an attraction could eventually find its own community according to tourist interests or motivations.

Appeal Index—*AppealI*: The visitor flow of an attraction could reflect its appeal to tourists, and tourist hotspots can therefore be identified via node weight (the number of tourists at an attraction). However, some attractions will, distinctively, have more regular tourists. The results of designating an attraction as a tourist hotspot would be different if only the total number of visits is considered. For instance, the number of tourists at "Cheung Chau" was 10,925, which is fewer than at "Hong Kong Cultural Centre" (11,036), but the number of visits to "Cheung Chau" was 18,915, which is more than to "Hong Kong Cultural Centre" (15,745). Therefore, the appeal index *AppealI* was introduced as a convenient rule of thumb to describe the relative appeal degree of an attraction, which is positively correlated with both the number of tourists and visits. This is defined as follows:

$$AppealI = \log\ (number\ of\ tourists * number\ of\ visits) \qquad (9)$$

Graph based ranking: Attractions could be ranked by their passenger flows where

AppealI is a potential reference. High visitation is never the sole criterion when promoting tourism development. Attractiveness is defined as "the relative importance of individual (attraction or destination) benefits and the perceived ability of the destination to deliver these individual benefits" (Mayo & Jarvis, 1981, p. 201). The attractiveness of an attraction could be propagated to its connected attraction(s) in the network. An attraction with a higher volume of connections to other attractions should be more noteworthy in marketing. Attractiveness may be better assessed by comprehensively considering both the interior topology between attractions together with their passenger flows.

Graph-based ranking algorithms are committed to deciding the importance of each node in a graph (network) by computing global information recursively rather than local individual node information. In this way, the ranks of attractions emphasize their mutual connections rather than their singular appeal like *AppealI.* Algorithms such as Google's PageRank (Page, Brin, Motwani, & Winograd, 1999) and TextRank (Mihalcea & Tarau, 2004) have been successfully used in previous studies to provide Web page rankings and identify keywords. As an analogy of TextRank as an undirected weighted graph, the rank score of an attraction would be high if the score sum of its connected nodes was high. The weight score of nodes in an attraction network can be defined as follows (Mihalcea & Tarau, 2004):

$$WS(a_i) = (1-d) + d * \sum_{a_j \epsilon In(a_i)} \frac{w_{ij}}{\sum_{a_k \epsilon Out(a_j)} w_{jk}} WS(a_j) \qquad (10)$$

where $WS(a_i)$ is the weight score of attraction a_i, which is initially assigned as 1, and d is the damping coefficient. $In(a_i)$ and $Out(a_i)$ are the attractions that co-occur with attraction a_i, and w_{ij} is the weight of the edge between a_i and a_j. The final weight is obtained after iteration until convergence. The larger the weight score, the higher the attraction's ranking.

To conclude this section, this study innovatively emphasized the relationships between attractions and demonstrated a synthetic network analysis of connected attractions. Together with the clustering results, the attractiveness of attractions was assessed from two dimensions. An integrated analysis in a theoretical model is presented subsequently.

4.3. Results and discussion

All the attractions were sorted into four clusters by performing community detection on the attraction network. This section discusses the characteristics of the clustered attractions from the dimensions of themes and diverse attractiveness.

4.3.1. Category distribution

Figure 8 shows the distribution of natural, urban, and cultural attractions among the clusters. Cluster 1 has the most natural attractions—almost 70% of its attractions are displayed in green. Approximately 75% of the attractions in Cluster 2 are culturally related. Cluster 3 is dominated by urban attractions, at approximately 60%. Both cultural and urban attractions are common in Cluster 4.

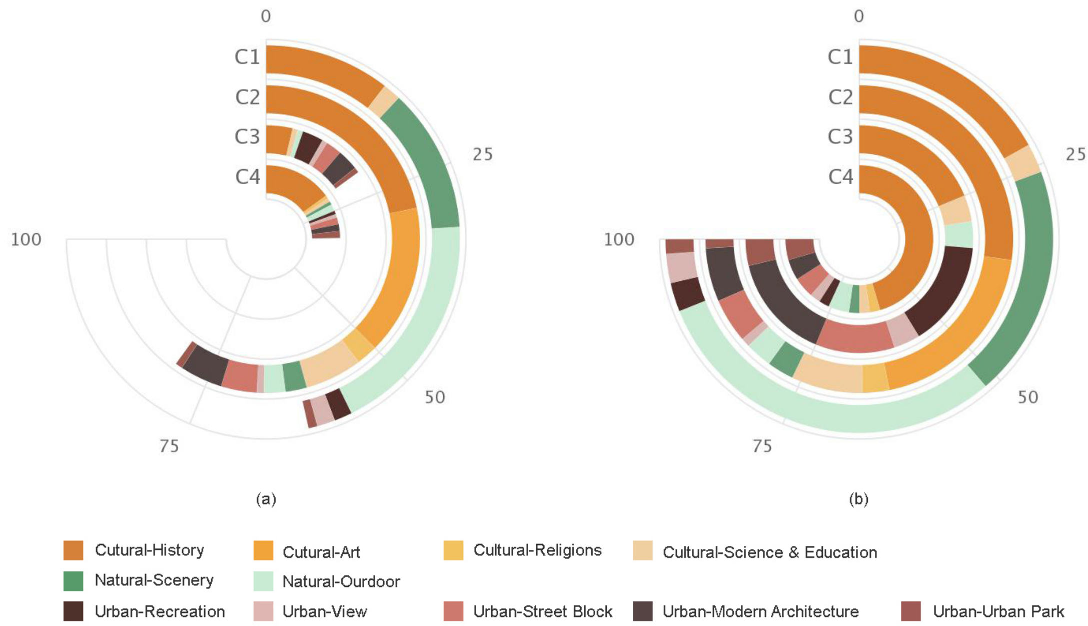


Figure 8. Attraction Clusters According to Attraction Categories by (a) Frequency and (b) Percentage

4.3.2. Attractiveness assessment

To compare the characteristics of clustered attractions, their attractiveness to tourists was also analyzed and assessed from two dimensions: the visit volume of an attraction (i.e., *AppealI)* and the graph-based between-attractions attractiveness propagation ranking. Various phenomena can be observed in the ranking list (Table 4). An attraction may receive fewer visits because it is closely correlated to another popular site and then show significant improvement in rank when a graph-based ranking is conducted. For instance, Victoria Park achieved the highest graph-based ranking, even though it ranks only ninth in visit volume. This means that Victoria Park obtains its high status through attractiveness propagation. It is understandable considering that Victoria Park is located in the central area of Hong Kong, which is full of activities involving shopping malls, cultural sites, and cross-strait ferry services traveling from Kowloon to Hong Kong Island. This multifunctional attraction can satisfy the diverse needs of tourists. A contradictory observation can be made regarding purposeful attractions such as the theme parks Hong Kong Disneyland and Ocean Park. They received a sound visit volume rank (2nd and 5th, respectively), but their relations with other attractions are relatively weak (21st and 9th, respectively). They are more likely to attract a niche tourist market than Victoria Park.

The graph-based ranking results reveal the importance of an attraction in terms of attractiveness propagation. As an overview, the top attractions have distinct features in common. They are open spaces located in central areas of the city (e.g., Victoria Park, Lan Kwai Fong, and West Kowloon Cultural District). Most importantly, these attractions are not meant to attract a niche tourist market. It can also indicate why Ocean Park is more attractive than Disneyland; Disneyland, with its cartoon images, attracts mostly family tourists with children, whereas Ocean Park, an aquarium with a wide range of species, can better serve audiences of various ages. Comparatively, attractions such as Lion Rock are favored by hikers, the Hong Kong Heritage Museum by heritage lovers, and St. John's Cathedral by Catholics; such attractions possess more distinctive features than those for the general public and rank lower, even though ordinary people may also visit them. In addition, these lower-ranked attractions are relatively small in coverage or in less competitive locations that are remote from the city center or main transport. It can be concluded that the fewer the distinctive features or the less the specific audience orientation of an attraction, the higher it ranks.

Table 4. Top 30 Attractions Based on Graph-based Ranking and Numeric Passenger Flow Measurements

Attraction	Graph-based		*AppealI*		#Visitors		#Visits	
	Rank	Weight	Rank	Value	Rank	No.	Rank	No.
Victoria Park	1	0.003179	9	8.2561	11	10510	9	17159
Hong Kong International Airport	2	0.002645	1	9.9021	1	74584	1	107024
Hong Kong Convention and Exhibition Centre	3	0.002383	3	9.2467	3	29666	3	59486
West Kowloon Cultural District	4	0.00227	15	7.7820	18	5765	14	10501
The Peninsula	5	0.002254	7	8.3326	7	12099	8	17778
Stanley	6	0.001972	13	7.9401	13	7688	13	11331
Lan Kwai Fong	7	0.001968	4	8.6671	4	18334	5	25341
Hong Kong Cultural Centre	8	0.001721	10	8.2310	8	11036	11	15745
Ocean Park Hong Kong	9	0.001614	5	8.6340	5	16461	4	26155
The Hong Kong Museum of History	10	0.001594	32	6.8896	32	2324	34	3337
University of Hong Kong	11	0.001545	22	7.3601	22	3706	21	6183
Temple Street Night Market	12	0.001523	18	7.6131	17	5789	18	7088
Tai O	13	0.001395	27	7.1111	29	2877	24	4489
SoHo	14	0.001372	21	7.3956	20	4390	23	5664
The Blue House Cluster	15	0.001323	59	5.9688	58	861	61	1081
Hong Kong City Hall	16	0.001288	28	7.0922	27	2922	28	4232
Hong Kong Observation Wheel	17	0.00124	19	7.5129	19	5253	20	6201
Tsing Ma Bridge	18	0.001239	36	6.6588	35	1982	40	2300
Peng Chau	19	0.001233	34	6.8625	39	1680	27	4337
Lion Rock	20	0.001226	20	7.4006	21	4358	22	5772
Hong Kong Disneyland	21	0.001221	2	9.3907	2	39065	2	62938
Mui Wo	22	0.001211	23	7.2153	30	2638	19	6223
Western Market	23	0.001158	87	5.0673	83	313	88	373
Tap Mun	24	0.001157	44	6.3737	44	1249	45	1893
Sharp Island	25	0.001151	41	6.4988	40	1479	42	2132
PMQ	26	0.001129	6	8.4443	6	12455	6	22335
The Hong Kong Heritage Museum	27	0.001118	56	6.1442	56	983	56	1418
Po Lin Monastery	28	0.001104	26	7.1405	24	3273	29	4222
Po Toi Island	29	0.001082	49	6.2680	52	1028	46	1803
St John’s Cathedral	30	0.001076	61	5.9334	59	841	62	1020

4.3.3. Attraction cluster models

We synthesized three dimensions so that each cluster obtained a distinctive feature that differentiates it from the others. The attraction clusters can be visualized into a matrix, sorted by the dimensions of attraction category and visit volume (*AppealI*). Based on the UNESCO definition, urban areas are characterized by cultural and natural values and attributes that cover a broader geographical setting (UNESCO, 2011). Therefore, the urban category lies between the cultural and the natural. The importance of attractiveness propagation is illustrated by the size of the circles in Figure 9. The bigger the circle, the more important the attraction in the cluster. To provide a more specific analysis of the proposed model, the top 30 attractions according to graph-based ranking are shown in Table 5 below.

Table 5. Attraction Cluster Distribution of Top 30 Attractions by Graph-based Ranking

Cluster	Attractions	Graph-based Rank	*AppealI* Rank	Category
Cluster 1	Victoria Park	1	9	Urban
	Stanley	6	13	Nature
	Tai O	13	27	Nature
	Peng Chau	19	34	Nature
	Lion Rock	20	20	Nature
	Mui Wo	22	23	Nature
	Tap Mun	24	44	Nature
	Sharp Island	25	41	Nature
	Po Toi Island	29	49	Nature
Cluster 2	Hong Kong Convention and Exhibition Centre	3	3	Urban
	West Kowloon Cultural District	4	15	Culture
	Hong Kong Cultural Centre	8	10	Culture
	The Hong Kong Museum of History	10	32	Culture
	University of Hong Kong	11	22	Culture
	The Blue House Cluster	15	59	Culture
	Hong Kong City Hall	16	28	Culture
	PMQ	26	6	Culture
	The Hong Kong Heritage Museum	27	56	Culture
	St John's Cathedral	30	61	Culture
Cluster 3	Hong Kong International Airport	2	1	Urban
	The Peninsula	5	7	Culture
	Lan Kwai Fong	7	4	Urban
	Ocean Park Hong Kong	9	5	Urban
	Hong Kong Observation Wheel	17	19	Urban
	Tsing Ma Bridge	18	36	Urban
	Hong Kong Disneyland	21	2	Urban
Cluster 4	Temple Street Night Market	12	18	Urban
	SoHo	14	21	Urban
	Western Market	23	87	Culture

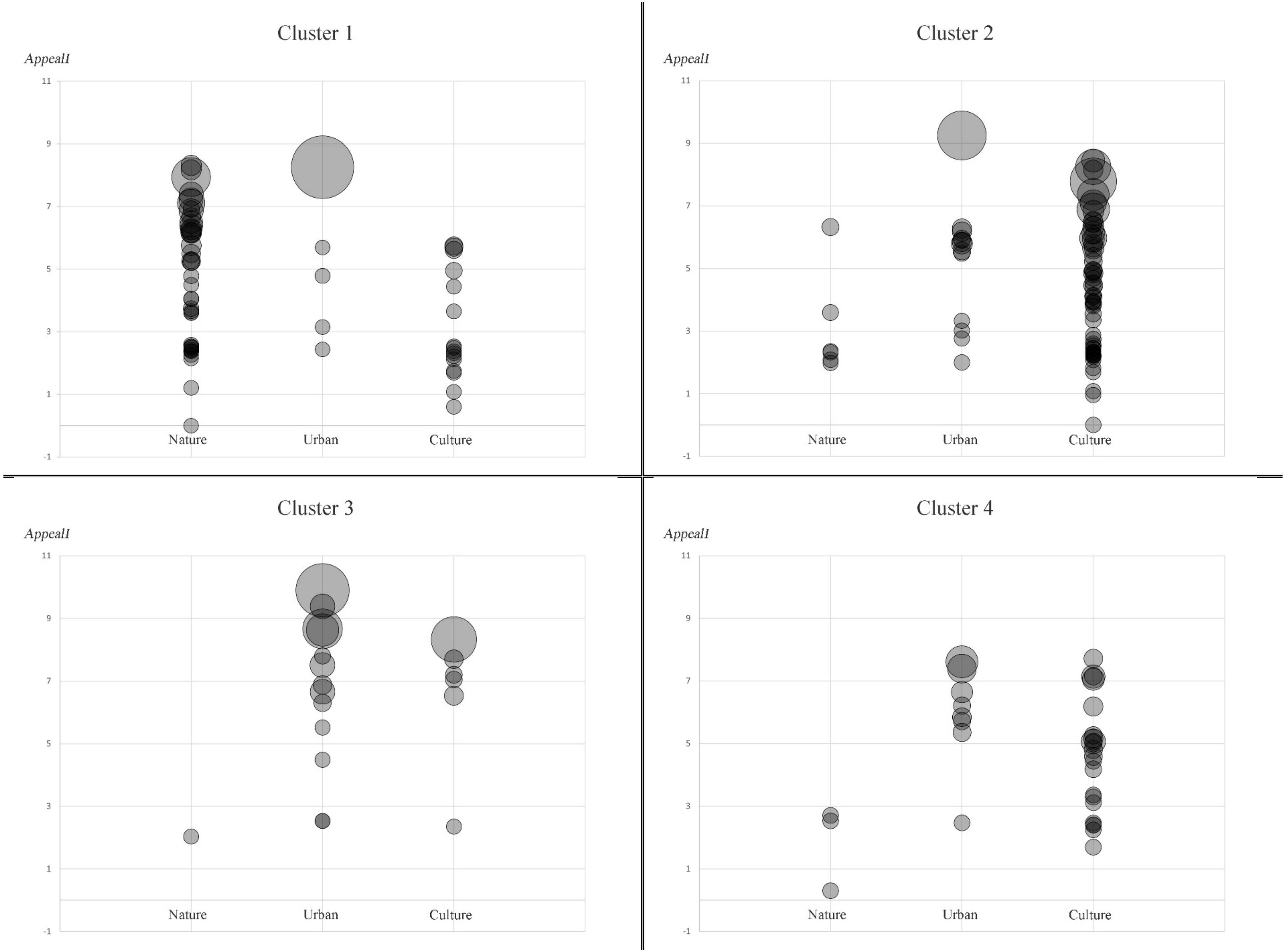


Figure 9. Model of Attractiveness of Sites in Clusters

4.3.3.1. Cluster 1—Nature-dominated attractions

Of the top 30 attractions shown in Table 5, nine attractions in Cluster 1 are natural sites, such as Stanley, Tai O, and Peng Chau. The graph-based ranking results for this cluster are relatively low, with only Victoria Park and Stanley included, ranking 1st and 6th respectively, and half of the attractions ranked lower than 20.

4.3.3.2. Cluster 2—Culture-dominated attractions

Of the top 30 attractions, only one—the Hong Kong Convention and Exhibition Centre (HKCEC)—belongs to the urban category, and the rest belong to the culture category. These cultural attractions are enclosed buildings and structures, including the Western Kowloon Cultural District, museums, and cultural or creative art-related venues. In contrast to Cluster 1, Cluster 2 has the highest number of leading attractions in the top 10 in the graph-based ranking results. Still, the top-ranking attraction in this cluster is urban, the HKCEC. The *AppealI* of attractions in this cluster is relatively higher than that in Cluster 1, but tourists' visit volume is considered fair between all attractions (Figure 10).

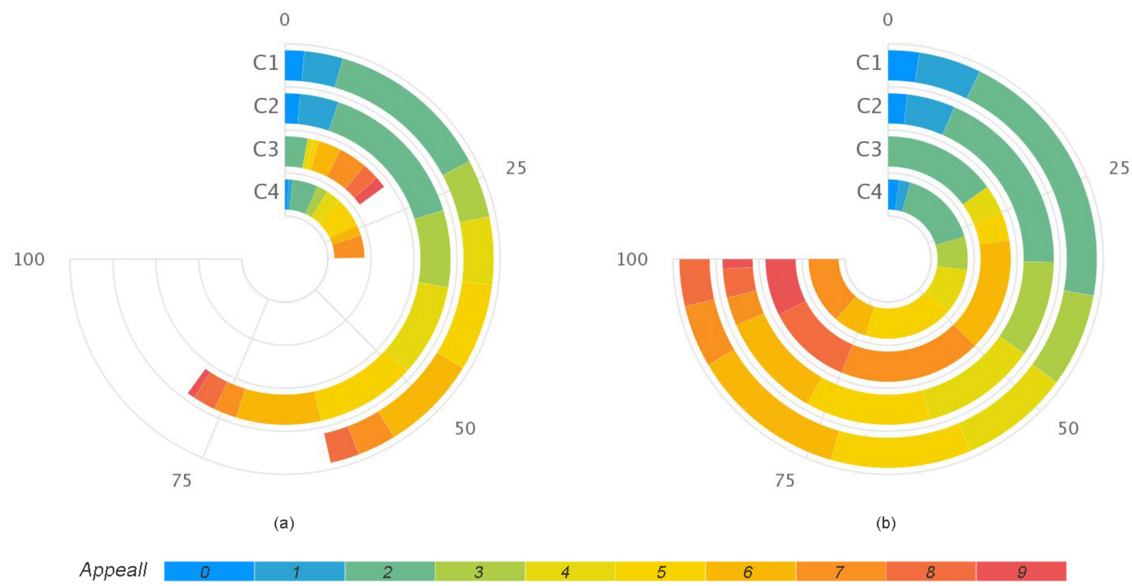


Figure 10. *AppealI* Distribution of Attraction Clusters Ranging from Lowest to Highest by (a) Frequency and (b) Percentage.

4.3.3.3. Cluster 3—Urban-dominated attractions

Cluster 3 has the lowest number of attractions but the highest *AppealI* (more than 75%). Among the top 30 attractions, seven belong to this cluster, six of which are urban, including the public facilities Hong Kong International Airport and Tsing Ma Bridge, the amusement parks Ocean Park and Disneyland, and the entertainment site Lan Kwai Fong. However, only one attraction in the cluster is a cultural site—The Peninsula.

Wall (1997) argued for classifying tourist attractions into "points, lines and areas" (p. 242), encouraging network thinking that involves connecting specific attributes of the resource, tourist behavior, and spatial distributions. Destination support services and accommodation and transportation facilities are crucial elements that enable a comfortable environment for tourists to move around in. A high standard and adequacy of supporting services can also stimulate tourist arrivals and can even be used to predict destination attractiveness (Crouch & Ritchie, 1999). It is worth mentioning that Hong Kong International Airport receives the highest rank based on both *AppealI* and attractiveness propagation, even though it is only a public support facility. Vengesayi, Mavondo, and Reisinger (2009) affirmed the role of supporting services and facilities in enhancing destination attractiveness independent of attractions. Highlighting the importance of these attractions in this cluster, four are within the top 10, which accounts for the highest proportion among the four clusters.

4.3.3.4. Cluster 4—Cultural-urban attractions

There are a few natural sites in this cluster, which is mostly dominated by cultural attractions related to history and a mix of other urban attractions. The *AppealI* in this cluster is at a medium level, between Level 2 and Level 7 (there are 10 levels, from 0–9). This means that these attractions generally received lower visit volumes. Among the top 30 attractions, only four are in Cluster 4, where two attractions are urban (i.e., Temple Street Night Market and SoHo), and the other two are cultural (i.e., Western Market and Po Lin Monastery). The graph-based ranking results of these four attractions are merely fair.

4.3.4. Urban attractions as primary attractions

The density of the attraction distribution in the matrix reveals that most attractions are located in Cluster 2, which is dominated by cultural sites, followed by Cluster 1 with its mostly natural sites, Cluster 4 with a mix of cultural and urban sites, and lastly by Cluster 3, which is dominated by urban attractions. In general, all clusters are led by one or a few urban attractions; this pattern can be detected from the largest circle in each cluster; all such circles are in the urban category, even though the phenomenon is not that obvious in Cluster 4. These urban attractions act as cores that connect all the attractions. They are considered the primary attractions that both receive a large volume of tourist visits and are important in connecting the others. The other attractions are considered secondary or even tertiary attractions. In the case of Hong Kong, the top attractions can be regarded as primary attractions based on the graph-based ranking results, which also indicate the highest level of attractiveness. For example, the primary attraction Victoria Park leads Cluster 1, assembling the most natural sites. Cluster 2 is led by the HKCEC, which assembles most cultural sites. Although both Victoria Park and the HKCEC are urban attractions, they assemble attractions with features similar to their unique characteristics. Victoria Park is an open area that is more likely to gather natural sites with features showing open spaces. Similarly, the HKCEC is a venue organizing commercial and cultural activities, so culturally-themed attractions are more likely to link to the HKCEC than to Victoria Park. Although Cluster 3 has relatively fewer attractions, its top attractions are almost all urban sites. This conforms to the phenomenon that urban attractions play an important role in each cluster, just as they do in Cluster 3. The strong characteristics of urban-based features are more obvious in Cluster 3. Comparatively, Cluster 4 is not outstanding, with both urban and cultural attractions demonstrating some attractiveness to tourists; the number of cultural attractions is slightly more than the number of urban attractions, whereas urban attractions reveal higher importance and appeal than the cultural ones.

4.4. Summary

4.4.1. Theoretical contributions

This study added to the literature by clustering attractions based on tourist movement rather than using geographic proximity or the same categorization of the sites (e.g., cultural or natural sites). Previous models that are frequently cited in the literature always categorized sites according to single dimensions, such as motivation (e.g., McKercher & Du Cros, 2002). This study manipulated a series of dimensions (i.e., theme, visit volume, and attractiveness propagation) as attraction characteristics mined from tourist movements and consolidated these in a proposed model with four clusters. Clusters corresponded to the specific tourist segments according to their features, i.e., natural, cultural, and urban and one cluster with a mix of cultural and urban attractions. Furthermore, this study extended the existing literature on tourist movement, geotagged information on social media, and big data analytics to explore intra-attraction relationships at the destination level. In addition, the proposed model can be used as a tool for DMOs to re-assess their marketing strategies in a post-COVID-19 tourism market. Tourists may move differently (e.g., to avoid overcrowding), and this model would be useful for attraction clustering based on these new movement patterns. In that vein, potential future research includes a longitudinal study that compares Hong Kong's clusters as described in this study with its post-COVID-19 clusters.

Methodically, the dimensions introduced in the proposed model can be applied in other contexts, not merely in Hong Kong. The algorithm, which has been systematically introduced, can be followed to combine the aspects of visit volume, attraction categories, and importance of attraction location in attraction network analysis. Meanwhile, *AppealI* and its impact on attractiveness propagation are two newly proposed methods to assess attractiveness. The testable model can also contribute to the evaluation of the attractiveness of sites through rankings and can further explain destination attractiveness. This model can be replicated, using the same rationale, in other destinations when there is attraction-level tourist arrival information and a clear categorization of attractions. By adopting the introduced analytical procedures, a new or similar model can be concluded for other destinations to further adjust their marketing orientations. Furthermore, this study demonstrates the possibility of applying big data analytics in tracking tourist movement patterns based on self-reported positioning information posted on social media platforms. Adopting the same method to conduct comparative studies on different social media platforms may further generalize the model.

4.4.2. Practical implications

The effects of assembling different clusters indicate the attractiveness of attractions at different levels and categories. The categorization of attractions into clusters can provide implications for better marketing orientation. That is, tourists more interested in natural attractions tend to visit the attractions in Cluster 1. The marketing implications for DMOs include the need to bundle the attractions according to their clusters and the need to invest more concentrated marketing efforts to strengthen markers between the attractions, a key role played by the urban attractions that lead each cluster. For instance, more advertisements can be posted at Victoria Park that feature information on Stanley, Tai O, and Peng Chau; information can even be posted in between these interconnected attractions. Attraction management teams can launch joint marketing efforts to organize themed tours by developing routes to link these sites. Even the HKTB can play a critical coordination role by posting relevant advertisements and event information for these attractions on a common page so that when tourists explore the information of one attraction, they are recommended to other similar attractions.

Similarly, cultural tourists are more likely to visit the attractions in Cluster 2. Cluster 3 attracts most urban tourists who like landscape views and exploring open spaces, and family tourists who are most interested as the large land coverage in these attractions that allows them to freely use the space for family gatherings and entertainment. There may be a group of tourists who cannot be easily labeled as either culture lovers or nature lovers; in such cases, Cluster 4 may be suitable for them as it combines both urban and cultural sites that can accommodate the diverse needs of those who hesitate to choose any specific attraction type. Cluster categorization can help marketers to match attractions easily with those indecisive target tourists.

This study reinforced the importance and necessity of managing and marketing attraction clusters rather than individual site management. Through clustering, marketing resources can be better utilized and concentrated on the particular attractions of similar features while at the same time packaging the attractions to serve the target tourists more accurately. Meanwhile, a reasonable balance of joint efforts, combining a primary attraction with a few secondary or tertiary attractions, can help to relieve the overcrowding issue in primary attractions by guiding tourists to these secondary attractions. This can also strengthen appreciation of secondary attractions. This is a key reason attraction management through clustering is advocated for destinations. This model also conforms to the nature of the attraction system by realizing the *marker* component effectively in a practical sense to better link *a site* with *a tourist* based on the results of observing tourist movements in attraction networks.

4.4.3. Limitations and further research

A limitation of the study is that the data processing only considered geo-tagged information as subjectively shared by social media users. It may or may not reflect a full picture of all tourist movements within Hong Kong because it depends on user willingness to make such information visible to the public. Furthermore, the possibility that the dataset includes Hong Kong residents' reflections cannot be eliminated completely. However, residents did not likely comprise a large proportion; a possible solution is tracking long-term posts of suspected resident users to further analyze their identities, whenever alternative algorithms can be applied. In addition, the study did not consider tourist profiles (e.g., age range, nationality, and gender). By adding such profiles into the analysis, the model can be refined to match the clusters with more focused tourist segments as needed. Future studies should consider these shortcomings. As stated above, potential topics should also be explored through longitudinal studies comparing Hong Kong's attraction clustering pre- and post-COVID-19.

Chapter 5. Organic destination image: tourism dynamics across China's Greater Bay Area Cities

5.1. Introduction

Many studies have examined destination image and its influence on the behavioral intentions of tourists (Chen & Tsai, 2007; Afshardoost & Eshaghi, 2020; Kim & Kim, 2020; Pan et al., 2021). Compared with induced image, which is determined by the destination, organic destination image is an accumulation of (usually less biased) information sources, and the shared experiences of those who have visited previously is more determined by the prospective traveller (MacKay & Fesenmaier, 1997). Organic destination images exist in the mind of individuals and undergo an evolution when supplementary information is received through various media (Gunn, 1988). For contemporary travellers seeking both value for money and value-added experiences, regional offerings may provide a more attractive way of satisfying their various needs. Images generated through social media may provide marketers with clues about the alignment between a regional destination, traveller needs and tendencies. The current authors pursue this line by drawing on the vacationscape concept and identifying the intersecting outcomes of "users/tourists" and "social media" in shaping multi-destination organic imagery. With particular potency for the regional context, Gunn's (1997) pioneering work highlighted the need to integrate the design, development, and management of attractions within the wider destination landscape. He contended that planning should extend beyond specific sites to the clustering of multiple attractions (developing multi-functional systems). His integrated vacationscape concept is based on development zones consisting of clusters of attractions, transportation, access, focal communities, and linkage corridors (Gunn, 1988). Gunn's (1997) original idea focused on clusters within a single destination such as areas, sites, services, and transportation. The current authors extend this thinking beyond development zones into a wider regional context. They adopt an integrated regional approach to the complex relationships between tourism components, taking account of tourist behaviors, destination choices and intra-regional movements. Gunn's (1997) vacationscape represents the connections between the overall functioning of tourism and the environment and is applied here to an emerging destination comprising a plurality of distinct cities. It offers prospectively better resource utilization and the provision of meaningful visitor experiences through effective planning and development. Destination image is viewed as a manifestation of the vacationscape construct.

Destination Management Organizations (DMOs) promote an image by drawing upon actual or desired perceptions of their offerings and disseminating support messages via various communication channels. Organic destination image emerges from traveller experiences along with various (less biased) sources of information (e.g., reports, travel writers' stories and word of mouth), distinct from marketing images emanating from destination service suppliers (Gunn, 1997; Gartner, 1993). With the that comes from accumulated impressions including those of actual visitors, organic destination imagery is an important driver of destination choice (Bilynets et al., 2021; Gartner, 1993). Tourists may be regarded as primary agents of image formation in their capacity as end users, (Nowacki & Niezgoda, 2020). Their articulations of what they perceive as real destination image is not easily managed or manipulated by DMOs and/or by other interested sponsors. Nevertheless, recent scholarly attention to organic destination image has been modest, despite the changing landscape of destination marketing. Yang (2016), for example, documented tourist awareness of their role as co-creators of destination image. The need for better insights is compounded by a gap between media projections of destination image and local perceptions (Williams-Burnett & Fallon, 2017). Social media (e.g., Wang et al., 2015) and tourists photos (Kim and Stepchenkova 2015) are potential source of organic destination image formation with the current study giving further consideration to the former.

Use of social media as a communication channel in global tourism has grown as awareness of its cost-effectiveness has increased (Trihas et al., 2013). An intensification of information flows within tourism has been widely documented (Narangajavana et al., 2017). Access to so-called big data is allowing more thorough analyses of organic image. The large-scale data available through social media platforms is enabling the mining of data which can explore and explain traveller experiences and reflections. Hence social media can be used when examining organic destination image, including how it informs destination management on a regional scale. The current study uses a social media platform to examine the preferences and reflections of tourists in forming organic destination image.

The context for the current study is Guangdong-Hong Kong-Macao Greater Bay Area (hereafter called GBA). The GBA contributes to China's national economic agenda by promoting regional development and strengthening connections between neighbouring cities and regions (Greater Bay Area, 2020). A principal aim of regional designation is to stimulate domestic tourism in southern China. The GBA city cluster consists of nine municipalities in China's Guangdong province (Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen and Zhaoqing) and two Special Administrative Regions of China (Hong Kong and Macao SARs) – the so-called "9+2 cities". Barriers are a potential impediment to integration - mainland Chinese need a visa to travel to the two SARs. The Culture and Tourism Development Plan for GBA was initiated by China's Ministry of Culture and Tourism and aims to build "… a Bay Area for Leisure … including developing more distinctive Greater Bay Area tourism products and itineraries, jointly promoting Greater Bay Area multi-destination tourism, driving the tourism development of cruise and yacht, strengthening the co-operation of tourism market regulation of the Greater Bay Area as well as nurturing of tourism talents." (Hong Kong SAR, 2021; The Department of Culture and Tourism of Guangdong Province, 2021, pp. 11-13).

Most destination image studies have neglected Gunn's founding principles and have focussed exclusively on single attractions or destinations (c.f., Lojo et al., 2020; Tamajón & Valiente, 2017). A notable recent exception was Lalicic et al (2021). Few scholars to date have investigated regional destination image encompassing multiple proximate cities. Noting the timeliness of researching the current regional context, the authors pursue novel insights about Gunn's original destination idea. They deploy an innovative big data analytical technique to examine the fundamental premise of organic image in a dynamic contemporary context. The researchers capitalise on currently accessible Chinese data, believing that a GBA investigation offers potential insights into how regional destination image is being shaped by social media user content. This study assembles the destination image for each GBA city, with a view to interpreting the presence of the region's organic destination image on one of China's fastest growing social media platforms, Xiaohongshu or Little Red Book (hereafter called the Red). The Red offers researchers timely data to examine the shaping of organic destination imagery with its escalating active usership, wide functionalities and dissemination capacity to a mass audience. The current study draws upon relevant literature and interactive content. Instead of emphasising traditional travel blogs or commentary contents in platforms like TripAdvisor where the information flows in a single direction, the authors focus on user-generated content. To understand the building of destination image, the authors consider the relationship between content on social media platforms and its interpretation. By dissecting the Red's depiction of organic destination image, the study provides destination planning insights on GBA, both local and regional, and offers suggestions for DMOs. It further explores how Gunn's vacationscape concept (1997) applies in practice in a regional context.

5.2. Methodology

Most previous investigations on the destination image of the GBA have deployed conventional methods such as questionnaire-based surveys (e.g., Kirillova et al., 2020). However, the advent if the contemporary digital marketing era has accelerated the urgency of deploying innovative methods, based on data originating from social media platforms (Molinillo et al., 2017). Examples include the sentic computing approach (Alaei et al., 2019; Zhang et al., 2020). Tourism and hospitality researchers have made increasing use of big data analytics (Mariani, 2020). Drawing upon suitable theoretical foundations, it is worth exploring the fast growing and innovative analytics that can facilitate data interpretation, and propose prospective implications for practitioners (Mazanec, 2020). Motivated by the opportunities to deploy innovative methods for an examination of online organic shaping of destination image, this study uses computer-aided lexical analysis of social media content collected from the Red, thereby allowing judgments about content, while avoiding the potential human errors that hamper manual data processing (Lai & To, 2015).

5.2.1. Data cleansing and preprocessing

The authors combined "title" and "content" into a "full script" to best accommodate subsequent text mining. They then eliminated HTML tags and special symbols in "full script". Some irrelevant posts were also removed to enhance the validity of the data set and to avoid data availability bias (Ruths & Pfeffer, 2014), This was mainly because: 1) the content was advertising oriented. For instance, posts were used by a studio for wedding photography promotion while tagging the keyword "tourism" to be eye-catching; and 2) the city name mentioned in the post was not the subject of discussion. For example, in some posts our studied cities served as a point of departure rather than a destination. Another special case is that "Zhongshan" was covered by the Chinese name of the Sun Yat-sen Mausoleum ("Zhong Shan Ling"). As a result, posts acquired by searching "Zhongshan Tourism" included some posts introducing the Sun Yat-sen Mausoleum — a famous attraction in Nanjing. A total of 8,295 posts remained after data cleaning (Table 6).

Table 6. Data Volume of Each City/Region in the GBA

City Name	Post Amount	Average Number of Words per Post
Hong Kong	974	523
Macao	962	519
Huizhou	949	375
Guangzhou	937	515

Shenzhen	897	413
Zhuhai	799	452
Jiangmen	721	367
Foshan	662	389
Dongguan	571	374
Zhaoqing	543	404
Zhongshan	280	378

Since all content on the Red is in Chinese characters, the researchers deployed an open-sourced Python package jieba (GitHub, 2020). This is commonly used for Chinese word segmentation in big data analytics (Liu et al., 2020; Xiao et al., 2018). The following sequential steps highlight the data processing that was undertaken prior to analysis.

(1) Custom dictionary construction: Since jieba is a general word segmentation tool, most of the proper nouns for the study area are not included in the lexicon. The researchers proceeded to create a customized dictionary to improve the accuracy of the segmentation results, particularly concerning tourism related terms. The dictionary included names of top 50 tourism hotspots, central business districts, popular restaurants and those offering local cuisine by taking the references from Meituan.com (a Chinese leading e-commerce platform for catering, traveling and other entertainment and lifestyle services, covering 2,800 cities across China with 460 million users and 6.3 million merchants) (Meituan, 2020).

(2) Stop-word corpus customization: the work of extracting keywords is based on word frequencies. Certain words (such as "a", "the", "on", "in" in English) are common in all contexts, though have little lexical content. These words are referred to as stop words and would be ignored in further natural language analysis. For the purposes of the current study, three general Chinese stop-word lists provided by Baidu.com, Sichuan University and Harbin Institute of Technology were combined to filter the posts.

(3) Chinese word segmentation: by employing the custom dictionary and stop-words corpus, the Jieba package cut the sentences into text segmentations for purposes of text mining. Each post *N* could be represented as $N = \{t_1, t_2, \dots, t_i\}$, where t_i represents the *i*th segmented term in the post. As a result, each city's term pool *C* would then be represented as $C = \{N_1(t_{11}, t_{12}, \dots, t_{1i}), N_2(t_{21}, t_{22}, \dots, t_{2i}), \dots, N_j(t_{j1}, t_{j2}, \dots, t_{ji})\}$, where N_j denotes the *j*th virtual post of the given city.

5.2.2. Graph-based text mining

The researchers undertook Graph-based text mining and analysis on the term that was described above - pool C. We deployed TextRank in this study, a graph-based ranking model for unsupervised keyword extraction (Mihalcea & Taray, 2004). It has been used recently in tourism analytics and other fields (Chen et al., 2020;

Spruit & Ferati, 2020). An undirected weighted graph G was constructed to represent relationships between the terms: a node is a segmented term t_{ji}; an edge between two nodes would be added if two terms co-occurred in a specific window in the paragraph, and the weight of the edge is the co-occurrence frequency. For the purposes of the current study the specific window size was set at 5. This means that each term would find its nearest 5 terms on both left and right sides to count co-occurrence. The final step was the construction of a graph based on the co-occurrence pattern.

TextRank was then deployed for term graph G. As an algorithm derived from Google's PageRank (Brin & Page, 1998), the significance of each node is represented by its weight, and it is defined as follows:

$$WS(t_i) = (1-d) + d * \sum_{t_j \epsilon In(t_i)} \frac{w_{ij}}{\sum_{t_k \epsilon Out(t_j)} w_{jk}} WS(t_j)$$

where $WS(t_i)$ is the weight of a term t_i, which is assigned as 1 initially. d is the damping coefficient. $In(t_i)\ and\ Out(t_i)$ means the terms which co-occur with term t_i, w_{ij} is the weight of edge between t_i and t_j. The final weight was obtained after the iteration until convergence. The higher the weights, the more importance that is indicated by the terms. We also combined candidate keywords (top-weighted terms) with one another to discover possible proper nouns and then updated the dictionary to remedy the probable incomprehension of our custom dictionary. The experiment was then rebooted after updating the dictionary.

Ultimately the researchers extracted the 30 highest weighted terms. These were the most commonly and popularly used terms or phrases by users of the Red and had strong connections with other phrases.

5.2.3. Data analysis

The 30 terms offer a starting point for qualitative exploration. Two rounds of data analysis were undertaken with codes and pattern coding (Miles & Huberman, 1994). First round data were analyzed based on the top 5 key terms (top weighting score) for each city inside GBA. The top 5 key terms are compared with the results of network analysis, score of co-occurrences between paired terms. 2 illustrates a keyword term graph of one city – Macao. The graph indicates that the higher the edge weight, the higher the co-occurrence between the two terms. This indicates that when users discuss topic A (referring to the term), it may correlate with the discussion on topic B. The second round of data analysis was based on the results from the first. The researchers iterate the comparison between the first-round results and the factual tourism information of each city as well as to detect the intertwined relationship between 11 cities. The 9+2 cities are subsequently categorized into four clusters with distinctive features.

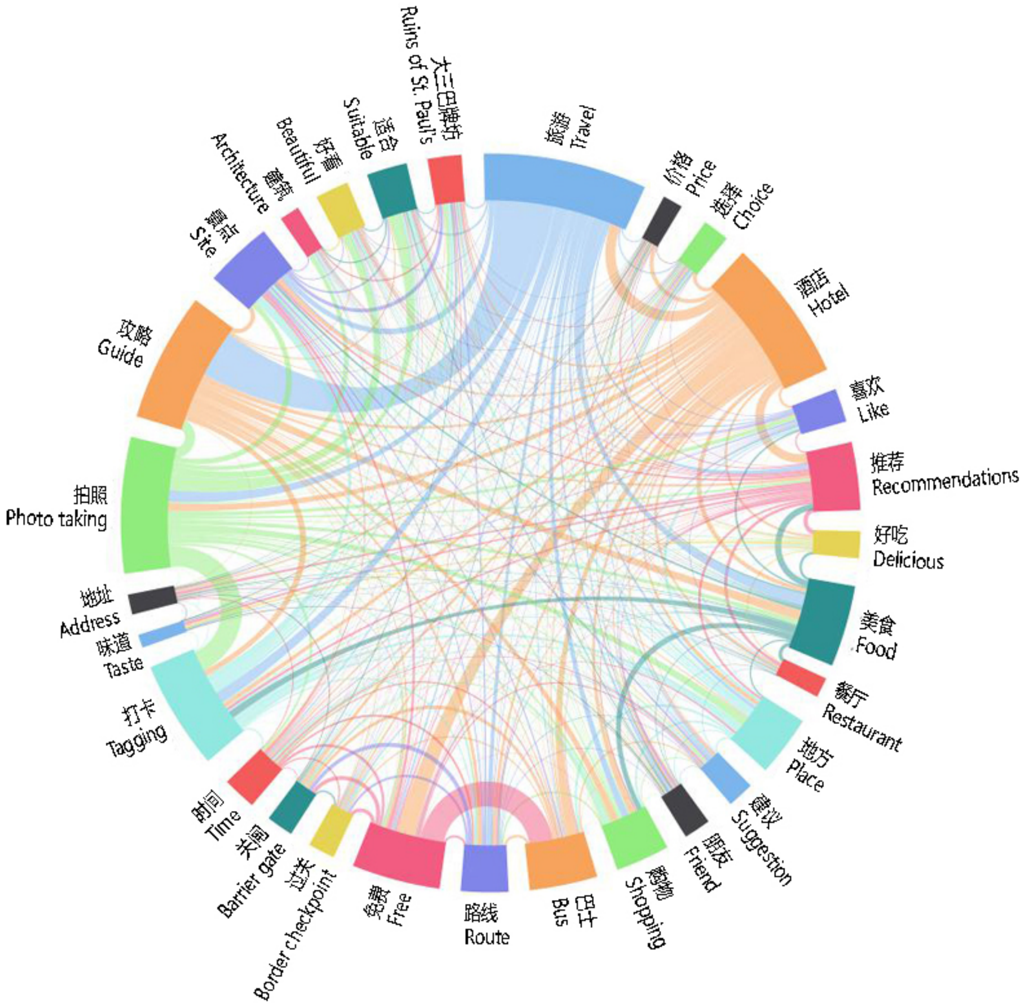
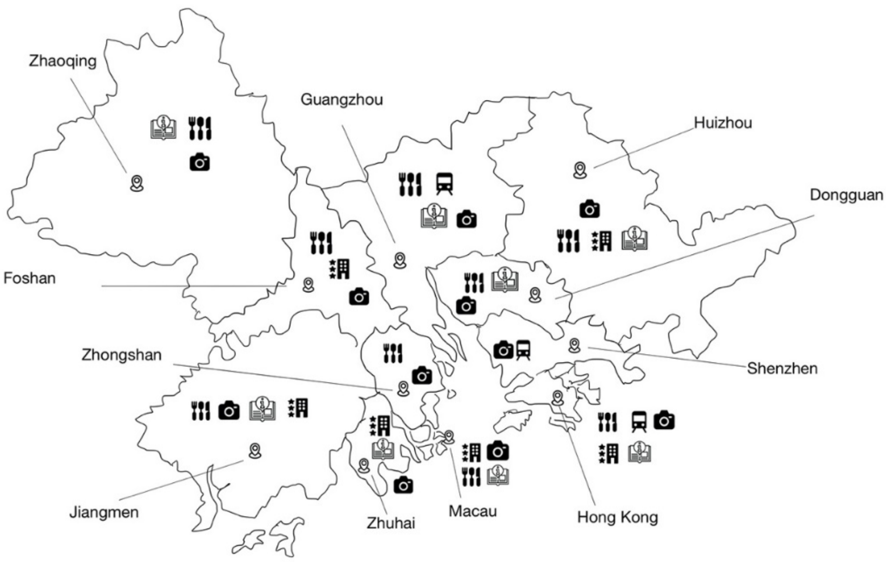


Figure 11. Keyword Term Graph for Macao

5.3. Findings and discussion

The information in Table 7 is based on the TextRank and network analysis. TextRank reveals popular topics about a destination that are discussed and of concern (referring to the keywords), and network analysis correlates the discussion topics into a meaningful pattern (e.g., "photo taking" correlates with "tagging"). Based on initial impressions text rank offers a good picture of popular discussion about GBA city destinations. For instance, "hotel", "travel" and "photo taking" are the top ranked keywords – these three terms were mentioned frequently by users in relation to Macao. They reveal that UGC discussions focus extensively on hotels in Macao (Table 7). However, a very different discussion pattern is revealed when the data about each destination is scrutinized using network analysis, including for Macao. In the latter case, the two pairs of keywords “free and shuttle bus” and “hotel and free” show a relatively high co-occurrence intensity in the data set (see Figure 11). The reason may be that Macao’s integrated resorts provide free shuttle bus services around the city. When users discuss “free”, they are mainly referring to the provision of shuttle bus services by hotels and resorts. This illustrates that social media users are aware of such service provision in the applicable destination—Macao—and engage in widespread dissemination through social media. UGC on social media enables us to detect the popular destination related discussion topics amongst users. User online postings, sharing and commenting contributes to the generation of organic destination image for cities.

The deployment of text mining allows a consolidated summary of the identified discussion topics among the GBA cities. Consistent with Marine-Roig’s (2019) destination image component model and Lojo et al.’s (2020) study, a destination image is the interplay of designative, appraisive and prescriptive perceptions that combine both the physical characteristics of tourism resources and tourists’ subjective emotional responses to the physical environment. The manifestation of these image components varies by destination (Lalicic, 2021). In the current study, the 11 cities are summarized into clusters based on the results of the TextRank and network analysis. It is revealed that GBA destination images are principally designative and prescriptive, namely designative images in the case of gastronomy, hotel, and the metro network, and prescriptive in the case of recommendation. Each cluster separately indicates an agglomeration of images shared across the cities with different themes. Collectively they form the GBA regional destination image (organic imagery). The most frequent discussion topics among users for each city (with the highest ranking and the highest degree of co-occurrence in pairs) are labelled with the corresponding icons in Figure 12.



Legend of the Map

City	Photo taking	Gastronomy
Hotel	Recommendation	Metro

Figure 12. Organic Destination Image of GBA of China

Table 7. TextRank and Network Analysis Results

City	TextRank			Network Analysis		
	Rank	Term	Weight	Term 1	Term 2	Co-occurrence
Macao	1	Hotel	0.012168893	Free	Bus	672
	2	Travel	0.007054372	Photo taking	Tagging	576
	3	Photo taking	0.007034504	Hotel	Free	544
	4	Gastronomy	0.005520648	Photo taking	Suitable	352
	5	Recommendation	0.005426136	Travel	Guide	347
Hong Kong	1	Recommendation	0.005354382	Photo taking	Tagging	429
	2	Photo taking	0.005347604	Metro	Exit	331
	3	Hotel	0.005346537	Tsim Sha Tsui	Exit	227
	4	HKD	0.005231943	Gastronomy	Guide	225
	5	Tsim Sha Tsui	0.004629847	Central	Exit	225
Zhaoqing	1	Travel	0.007821093	Zhaoqing city	Duanzhou district	340
	2	Recommendation	0.006074165	Travel	Shops Exploration	300
	3	Photo taking	0.004866643	Address	Zhaoqing city	295
	4	Like	0.004699158	Address	Duanzhou district	252
	5	Scenic Area	0.004456439	Travel	Gastronomy	204
Dongguan	1	Photo taking	0.009037603	Photo taking	Place	283
	2	Travel	0.008132425	Photo taking	Suitable	238
	3	Place	0.006040104	Dongguan City	Address	222
	4	Like	0.004532138	Photo taking	Tagging	184
	5	Recommendation	0.004449509	Gastronomy	Shops Exploration	180
Jiangmen	1	Photo-taking	0.00942649	Shops Exploration	Gastronomy	533
	2	Travel	0.008874113	Gastronomy	Guangzhou	389
	3	Hotel	0.007632563	Photo taking	Tagging	347
	4	Shops exploration	0.005820214	Travel	Gastronomy	338
	5	Recommendation	0.004999585	Travel	Shops Exploration	337
Guangzhou	1	Photo taking	0.010162756	Metro	Line	1192
	2	Recommendation	0.007609708	Metro	Transportation	842
	3	Travel	0.007184418	Photo taking	Suitable	578
	4	Metro	0.006633133	Metro	Exit	578
	5	Gastronomy	0.005263489	Photo taking	Tagging	563
Zhuhai	1	Photo taking	0.009654043	Photo taking	Beautiful	401
	2	Travel	0.008470515	Photo taking	Travel	366
	3	Hotel	0.007508436	Photo taking	Suitable	340
	4	Recommendation	0.005989538	Photo taking	Place	339
	5	Tagging	0.005212067	Photo taking	Tagging	330
Zhongshan	1	Photo taking	0.010470211	Japan	Scenic Area	495
	2	Travel	0.009792348	Zhongshan City	Address	192
	3	Japan	0.007627564	Photo taking	Suitable	117
	4	Scenic Area	0.007542745	Travel	Guangzhou	117
	5	Zhongshan City	0.005121801	Travel	Gastronomy	112

Color Coding of Text Rank and Keys

Gastronomy	Gastronomy Cluster	Recommendation	Recommendation Cluster
Hotel	Hotel Cluster	Metro	Metro Cluster
Photo taking	Most frequent search terms		

Table 8. TextRank and Network Analysis Results (Continued)

City	TextRank			Network Analysis		
	Rank	Term	Weight	Term 1	Term 2	Co-occurrence
Huizhou	1	Hotel	0.011085484	Gastronomy	Shops Exploration	672
	2	Travel	0.009272453	Hotel	Vacation	512
	3	Photo taking	0.008516587	Travel	Guide	379
	4	Recommendation	0.005813913	Travel	Shops Exploration	364
	5	Homestay	0.004381108	Photo taking	Suitable	330
Shenzhen	1	Photo taking	0.019976822	Photo taking	Sacred Place	3265
	2	Site	0.007162088	Photo taking	Site	1288
	3	Tagging	0.006789062	Photo taking	Free	1186
	4	Travel	0.006730604	Site	Tagging	1113
	5	Sacred Place	0.005075805	Photo taking	Tagging	1074
Foshan	1	Travel	0.010895184	Address	Foshan City	408
	2	Photo taking	0.009994806	Photo taking	Suitable	371
	3	Hotel	0.007250367	Travel	Guangzhou	319
	4	Guangzhou	0.006463466	Travel	Gastronomy	308
	5	Gastronomy	0.005407499	Photo taking	Guangzhou	284

Color Coding of Text Rank and Keys

Gastronomy	Gastronomy Cluster	Recommendation	Recommendation Cluster
Hotel	Hotel Cluster	Metro	Metro Cluster
Photo taking	Most frequent search terms		

Note: Only the top five results are shown with consideration of the length of the display.

5.3.1. Designative image of gastronomy

With the exception of Shenzhen and Zhuhai, a keyword shared among the GBA cities is “gastronomy”. Gastronomy appears as a high-frequency search term (ranked 4th and 5th respectively) in the case of Macao and Guangzhou in UGC on the Red based on TextRank. However, the term does not have high co-concurrence with other discussion topics. This suggests that users of the Red prefer to share their perceived destination image about topics other than gastronomy. Nonetheless, gastronomy has a relatively high ranking in searches, indicative that users are interested in browsing information about the topic. It is notable that UNESCO designated Macao as a Creative City of Gastronomy under the Creative Cities Network in 2017 (United Nations Educational, Scientific, and Cultural Organization, 2020). The Macao government has devoted considerable effort to promoting a gastronomy related city image (Macao Government Tourism Office, 2020). However, gastronomy is not a strong attribute in Macao’s destination image on the Red. Guangzhou exhibits a similar pattern, even though the world famous Cantonese cuisine originates from the English name for the city (Canton). When gastronomy is encountered as a keyword, it has a high co-occurrence in the remaining seven mainland cities (Shenzhen and Zhuhai report no significant results on this image and are exceptions with) with “shops exploration” (探店), “guide” (攻略) and “travel” (旅游). These findings reveal an interesting phenomenon - that a city may capitalize on the potential brand appeal of city of gastronomy designation under the UNESCO Creative Cities

Network. If DMOs adopt well-crafted strategic marketing strategies, they can attract visitors using this important tourism resource. Neither Macao nor Guangzhou have capitalized on this important tourism resource on the basis of the data presented. Some potential lessons may be learnt from Foshan's Shunde district – the latter is also listed as a Creative City of Gastronomy under the UNESCO Network and the study findings indicate that it has successfully developed a gastronomy linked destination image. These findings provide detailed implications to DMOs of the cities where efforts to promote gastronomy have been ineffective and demand serious review and attention before the cities can benefit from membership of the UNESCO Creative Cities Network (Lai et al., 2018).

5.3.2. Designative image of hotel

"Hotel" is a key term appearing frequently across most of the cities in the top 5 text rank, except in the case of Guangzhou, Shenzhen, Zhaoqing, Dongguan and Zhongshan (Table 7). It is well understood that tourists make extensive use of social media platforms to search for accommodation (Banerjee & Chua, 2016). In the city of Zhongshan there is little evidence of searching for accommodation and the possible reasons for this are: 1) the destination has fewer attractions within its geographical limits prompting tourists to make only single day trips. Tourists to Zhongshan typically visit a single tourist spot for photo-taking — Zhongshan City Studio (中山影视城) (a Japanese style film studio); social media users rarely search or post hotel or homestay-related information; 2) Zhongshan is adjacent to Macao and Zhuhai, where the keyword “hotel” is a popular search term on the relevant posts (ranked 1st and 3rd in Macao and Zhuhai respectively). Social media users who are planning a day trip to Zhongshan may opt to search for accommodation in either Zhuhai or Macau; and 3) the researchers also identified a lack of international chain hotels in Zhongshan, whereas they are widespread in Macau and Zhuhai. Given the geographic proximity of the three cities, tourists opt to stay in one city where international chain hotels are their choice and roam around the city cluster. A similar pattern is evident for Dongguan which is located between Huizhou and Shenzhen. The latter two cities offer tourists a wide variety of international chain hotels.

Three city groupings are evident across the hotel clusters with the first consisting of Zhongshan, Macau and Zhuhai (3 cities). The second group comprises Huizhou, Dongguan, and Shenzhen (3 cities). The third group consists of Hong Kong, Jiangmen, Foshan, Zhaoqing and Guangzhou (5 cities). A travel visa is required to cross the border into Hong Kong, because of its status as a special administrative region, whereas no such restriction exists amongst the various cities in Mainland China; thus, Hong Kong is “stand-alone”, despite its proximity to the neighboring cities. Locational convenience makes it easy for tourists to travel around the four other proximate cities (Jiangmen, Foshan, Zhaoqing and Guangzhou). The key term “hotel” ranks much higher for Jiangmen and Foshan and much lower for Zhaoqing (below 5) and for Guangzhou (below 10). In seeking to interpret these findings, the researchers triangulated the data by identifying the number of international chain hotels in these cities. An interesting pattern emerged in the case of cities with more international chain hotels - “hotel” has a higher rank as a key search term. The greater the average number of international chain hotels per square kilometer, the higher the appearance of the keyword “hotel” in TextRank. There are many and frequent advertising-related postings about international chain hotels on the Red, thereby offering a prospective explanation for tourist “hotel” related search behaviors. Such postings undoubtedly influence the search behaviors of social media users and presumably their clicking through advertising messages. Cities that house more international chain hotels appear to capture more “hotel” searches.

5.3.3. Designative image of metro networks

Transportation is an essential part of a destination zone (Gunn, 1994). Based on the current findings, the underground transportation (“metro network”), appears frequently in posts from those cities offering metro services, namely Guangzhou, Hong Kong (ranked within the top 5) and Shenzhen (ranked within the top 10). From the network analysis, “metro” is closely associated with terms “line”, “transport”, and “exit” when users post detailed itineraries with comprehensive metro service information. Compared to the diverse tourism resources in these cities (e.g., gastronomy, landmarks, shopping, accommodation), this finding illustrates that social media users prefer handy local transport information that has been shared by previous visitors, to other tourist information. Resonant with previous research, this finding highlights the utilitarian behavior of online social media users (Zheng et al., 2019). It further affirms the role of social media in disseminating information and could potentially inform destination marketers in the case of appealing information for tourists. As this finding reveals, it is strongly suggested that DMOs ensure availability on social media platforms of detailed local transportation information or a metro guide that includes exit and station information. Alternatively, they might develop smart applications (e.g., WeChat Mini Programs) incorporating metro service information and tourist information to engage with tourists who are familiar with digital information.

5.3.4. Prescriptive image of recommendation

The keyword "recommendation" (referring to a user's post which offers recommendations on various things) is ranked amongst the top for eight cities. Shenzhen, Foshan and Zhongshan are exceptions. This finding reflects that Red users are actively recommending most (eight) of the GBA cities. A similar pattern is evident in the case of other keywords such as "suggestion", "address" and "time" (Figure 11). The results reveal that the sharing of information by experienced tourists, such as addresses of places of interest, operating hours and practical tips is valuable to users of the Red who are seeking out tourist information. As previous researchers reported about social media platforms like TripAdvisor, (Banerjee & Chua, 2016), users often seek recommendations or advice prior to travel (Tham et al., 2013). The Red exhibits a similar pattern. Since GBA is a newly developed regional destination, potential tourists seem hungry for recommendations before deciding to travel. Recommendations or reviews of GBA tourism products and service offerings are essential to stimulate tourist interest and travel motivations. Such information will be imperative to project a positive destination image to potential tourists, and to satisfy those seeking recommendations. This finding provides insights to DMOs in each GBA city by providing information about tourist interests, especially those who surf online.

"Recommendations" from key opinion leaders (KOLs) or tourists who have previously travelled to the cities are important sources of information and marketing resources which can benefit from skilful online management. Previous researchers have observed that online recommendations and reviews are generally considered to be more reliable, up-to-date, and pleasant to read (organic), than straightforward promotional text that is derived from DMOs (induced) (Simeon et al., 2017). The current findings indicate potential for DMOs to adopt alternative marketing strategies by considering user generated contributions to organic destination image. Furthermore, DMOs can deploy designated personnel to monitor reviews posted by social media users and respond promptly to facilitate information flow on social media platforms.

It is notable that "recommendation" is a less popular key term than "photo taking", which appeared among the top five in all GBA cities. “Recommendation” does not co-occur with other search terms due to “recommendation” being a key search term associated with too many key terms. The data reveal no strong co-occurrence of recommendation with a specified key term. This offers high plasticity to the building and shaping of a city’s destination image. As was stated previously, the posts acquired from the Red were sorted by the researchers on the basis of sophisticated criteria containing the number of “likes”, “stars” and “comments” given by the users. It is the relevance to users, rather than the time of publication; thus, “recommendation” could pair up with many different key terms. To supplement the logic of social media algorithms, DMOs and other tourism bureaux might carefully cultivate the applicable destination image with a set of key terms that enhance the social media searchability of tourism offerings.

5.3.5. GBA—a regional destination image of photo taking

“Photo taking” emerged as the most prominent text in the data and ranked consistently within the top 5. The term “photo taking” also co-occurs frequently with others, such as “tagging”, “places” and “suitable” across all 11 cities, even in the case of Zhongshan which has the fewest overall postings. “Photo taking” is associated with a variety of recommended photogenic activities and attractions as evidenced in the posting excerpts. Most of the posts extracted from the data set are embedded with the key term “photo taking”. This contrasts with conventional tourist behaviors such as sightseeing, visiting friends and family, shopping, socializing, entertainment, and recreation (McIntosh et al., 1995). The following are some verbatim observations:

> Zhongshan is worth visiting. The building structures are colorful, which is suitable for those people who like photo taking. It is really beautiful in pictures!!! (Post on Zhongshan)
>
> The library and café inside Longxihu Park are worth visiting, especially for photo taking. (Post on Jiangmen)
>
> After renovation, Yeli Island is open to the public, and it is full of flowers planted nearby, especially when the rapeseed flowers bloom. It is the best time for photo taking, as it looks quite beautiful in photos. (Post on Zhuhai)

This finding offers thought-provoking insights for DMOs. The inclusion of a striking photo of a tourism offering on the Red could be described as being "worth a thousand words". The Red platform allows posts to be shared with photos/videos and descriptive wording and provides prospective tourists with visuals of the tourism resources thereby influencing their behavioral intentions. An appropriate photo on the Red may prompt a consequential thread of comments and posts, particularly on other Chinese social media platforms. Though not unique amongst social media platforms, it is notable that a single posting on the Red can lead to thousands of comments and post sharing. Comments and post sharing are important indicators of social media user engagement. The higher the engagement, the more effective the marketing strategies (Gretzel & Yoo, 2013). A post, a photo or a short video clip going viral can benefit DMOs; therefore, providing photo-taking spots and photo-taking opportunities during visits to GBA cities is a key to building a coherent destination image.

5.4. Summary and implications

By dissecting the organic destination image of the region, the study results provide insights about the various GBA cities that are relevant to destination planning. The authors have assembled four main clusters for the destination image of the GBA cities incorporating designative and/or prescriptive images. Though the images exhibit informational features, all designative images have a relationship to tourist activities within destinations, notably dining, living, traveling, and sightseeing. The results show that the regional destination image of GBA on the Red concentrates on designative image (Marine-Roig, 2019) with informational descriptions closely related to tourist resources and services. It is revealed that tourist inferences and attitudinal responses construct a designative destination image; while prescriptive destination image possesses the capacity to persuade others through recommendation of destinations, related tourism products or services. By way of comparison the consolidated study findings have not generated any obvious appraisive destination images. This suggests that the 9+2 cities are still at their infancy in the tourism product life cycle and offer considerable further development potential. An aggregation effect is evident since each cluster has distinct features applicable across multiple cities. The association of some cities with several clusters demonstrates the capacity of a channel such as the Red to generate diverse destination images that constitute rich territorial tourism resources.

This study has advanced exploration of the concept of organic destination image in vacationscape which has been substantially overlooked since its origins in the 1980s and has emphasized regional development within China's national policymaking agenda. The findings contribute to the literature on regional planning for tourism by shedding light on organic destination image based on patterns of social media UGC, as well as extending Gunn's vacationscape design principles in a cross-regional context. The authors have highlighted the role of social media platform users as agents of image formation and as an information source in the GBA regional context. They have adopted a holistic approach to examine destination image at a regional level, consistent with Gunn's (1977) original proposition that destination management and marketing should extend beyond single destinations. It has considered a zone where a complete destination picture extends beyond attractions and considers a collective of connected elements (e.g., transportation and services). Connected elements such as a well-developed service infrastructure provide the groundwork for delivering satisfactory visitor experiences. The authors conclude that experienced visitors who have shared their experience on the Red are contributing to the weaving of organic regional destination image. The study demonstrates the active involvement of users in the construction of "organic" destination images, distinct from the more DMO influenced induced counterparts.

DMOs are becoming more sensitive to a wide range of market interests. Attractions, transportation, services, information, and promotion are the key fundamentals of destination supply. Tourist perceptions on the supply are critical for DMOs as references for better destination design and viable marketing solutions. The results from big data analysis are particularly helpful. Scholars have noted the growing use of big data analytics for textual analysis, especially themes forming a basis for storytelling (c.f., Marine-Roig, 2021, Lalicic, 2021). Methodologically, the current study extends the application of big data analytics to destination image and adds knowledge by exploring GBA destination image on a social media platform. The researchers have analysed relevant data with a systematic deployment of text analytics. The pool of relevant data provides sufficient information to interpret the shaping of organic destination images for the 9+2 GBA cities.

The managerial insights arising from this destination image study offer implications for DMOs to formulate marketing strategies and to refine their service provisions. The Chinese Central Government has an ambitious GBA development goal with its 70 million people and high per capita incomes of GDP 23,371 USD (Greater Bay Area, 2020). The data that has become newly available via the Red can inform tourism market development, especially in light of the accelerating contribution of China's tourism market to the global economy. The study findings open potential pathways for DMOs to cultivate destination image via digital marketing strategies, particularly as tourism recovers from the pandemic. Furthermore, there is growing awareness of the merits of developing “smart” tourism destinations and the capacity of tourists to obtain information, access services and exchange opinions via social media platforms and though OTAs (Jovicic, 2019). The enhanced functionality of social media platforms empowers users to share their opinions vividly. Social media users are acting like spokespersons when they actively share their opinions about destinations. Such UGC is valuable because it represents an organic destination image based on perceptions amongst previous tourists, thereby equipping them with the capacity to affect others. DMOs in the designated cities can capitalise on the findings by considering the cluster information and developing customized marketing strategies based on destination image as reported on the Red. It has been found, for example, that Hong Kong, appears in all four clusters. The Hong Kong Tourism Board could target its strategies at social media user segments with greater interest in gastronomy, accommodation, transportation and recommendations. In contrast, cities which only appear in a single cluster such as Zhongshan (gastronomy) and Shenzhen (metro network), might diversify their destination image by increasing the visibility of additional tourism resources on social media platforms such as the Red.

5.5. Limitations

The study has focussed on examining available textual data and excluded consideration of photos. Future researchers might engage in a mining of photo sharing in the Red or other platforms. Furthermore, the researchers were unable to access demographic profiles of social media users. Ideally these should be considered when generalizing the findings to other social media platforms (Ruths & Pfeffer, 2014). Future studies might examine each cluster along with users' demographic information to generate detailed market segmentations profiles. Such a study might provide insights for DMOs and other practitioners to formulate marketing strategies targeted at contemporary tourists who use social media platforms.

Chapter 6. Conclusions

Tourism is a complex interconnect network. Tourism attractions and tourists are interconnected with each other and respectively. Tourist attractions play a critical role in representing and delivering an experience of a particular sense of place and in shaping the appeal of a destination that serves as the basis for destination development and competitive growth. This study firstly uses geotagged information as an indication of tourist movement patterns between attractions, so that attractions are connected based on tourists' common motivations. Ranking and clustering analysis are then conducted on this network, which has contributions to the tourism research where stresses that attractions do not stand in isolation but connect with other attractions and collectively build the entire destination attractiveness within the attraction system. Then the study moves focus to the connection among tourist concerns. An organic destination image in regional context is examined so that a comprehensive apprehension about the attitudes of entire tourists can be revealed.

In ranking analysis task, a new method AttractionRank is proposed to assess the attraction's attractiveness, which is critical for tourism planning, management, and marketing. This research focused on the attraction relationships rather than a single node information when valuing the attractions' performance. Furthermore, it is a space-independent method. By eliminating the influence of distance-decay effect, the resulted rank of attractions could reveal tourists' interest more explicitly, recognize hidden attractive attractions, and keep dominant attractive ones. DMOs can utilize the results directly in further destination marketing and tourism route planning. For example, more convenient transportation can be assigned to those hidden attractive attractions. With the growth of LBSNs, this method is testable and applicable for other destinations. Future study can also consider utilizing multiple LBSN data sources to alleviate the bias of user-shared geotagged information on Instagram.

For clustering, a new framework is used to reveal the characteristics of these intra-cluster attractions from different dimensions. An appropriate examination of attraction clustering has become extremely important in informing the marketing of a destination from a broad perspective. Urban attractions act as cores that connect all the attractions in Hong Kong. They are considered the primary attractions that both receive a large volume of tourist visits and are important in connecting the others. The other attractions are considered secondary or even tertiary attractions. The framework offers a theoretical contribution to the literature on clustering dimensions by providing a feasible experimentation method, and, from a practical perspective, it can guide destination marketing strategies for attraction networking.

Furthermore, idea of the interconnected tourism elements has been extended to a regional study context. Tourist concerns on attractions are also interconnected. A comprehensive apprehension about the attitudes of entire tourists is explored. By dissecting the organic destination image of the region, the study provides insights about the various GBA cities that are relevant to destination planning. The results show that the regional destination image of GBA on social media concentrates on designative image with informational descriptions closely related to tourist resources and services. It is revealed that tourist inferences and attitudinal responses construct a designative destination image; while prescriptive destination image possesses the capacity to persuade others through recommendation of destinations, related tourism products or services. The findings contribute to the literature on regional planning for tourism by shedding light on organic destination image based on patterns of social media UGC, as well as extending Gunn's vacationscape design principles in a cross-regional context.

In the end, innovations of this study can be summarized as: (1) It connects tourism attractions via tourist interest / movement and builds the network, (2) It ranks attractions' attractiveness in line with attractiveness propagation, (3) It ranks attractiveness by factoring out the effect of geographic proximity (calibrated) ,(4) It explores intra- attraction relationship in the network, (5) It consolidates different dimensions in a proposed model, (6) It explore advanced the concept of organic destination image in regional intra-destination research.

Reference

Adi, P. H., Wihuda, F., & Adawiyah, W. R. (2017). The role of social media browsing intention for behavioral outcomes of young consumers. *Market-Trziste, 29(1),* 39–57. DOI:10.22598/mt/2017.29.1.39

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

Agapito, D., Valle, P. O. d., & Mendes, J. d. C. (2013). The Cognitive-Affective-Conative Model of Destination Image: A Confirmatory Analysis, *Journal of Travel & Tourism Marketing*, *30*(5), 471-481.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research, 58*(2), 175-191.

Arif, A. S. M., & Du, J. T. (2018). Understanding collaborative tourism information searching to support online travel planning. *Online Information Review, 43*(3), 369–386.

Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management, 53,* 125–131.

Bao, Y. F., & McKercher, B. (2008). The Effect of Distance on Tourism in Hong Kong: A Comparison of Short Haul and Long Haul Visitors, *Asia Pacific Journal of Tourism Research, 13*(2), 101-111.

Barry, S. J. (2014). Using social media to discover public values, interests, and perceptions about cattle grazing on park lands. *Environmental Management, 53*, 454-464.

Barthélemy, M. (2011). Spatial networks. *Physics Reports, 499(1)*, 1–101.

Bilynets, I., Knezevic Cvelbar, L., & Dolnicar, S. (2021). Can publicly visible pro-environmental initiatives improve the organic environmental image of destinations? *Journal of Sustainable Tourism*. DOI: 10.1080/09669582.2021.1926469

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10). doi:10.1088/1742-5468/2008/10/P10008

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1–7), 101-117

Chancellor, C. (2012). Applying travel pattern data to destination development and marketing decisions. *Tourism Planning & Development, 9*(3), 321-332.

Chen, C. F., & Tsai, D. (2007). How destination image and evaluative factors affect behavioral intentions? *Tourism Management, 28*(4), 1115-1122.

Chen, L.-F, &, Tsai, C.-T. (2016). Data mining framework based on rough set theory to improve locationselection decisions: A case study of a restaurant chain. *Tourism Management*, 53, 197-206.

Chen, W., Xu, Z., Zheng, X., Yu, Q., & Luo, Y. (2020). Research on sentiment classification of online travel review text. *Applied Sciences, 10*(15), 5275. DOI:10.3390/app10155275

Chin, W.C.B., Wen, T.H., (2015). Geographically Modified PageRank Algorithms: Identifying the Spatial Concentration of Human Movement in a Geospatial Network. *PLOS ONE 10(10)*: e0139509.

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management, 57*, 295-310.

Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, & Yu, P. S. (2017). A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 29(1), 17–37. https://doi.org/10.1109/TKDE.2016.2598561

Crouch, G., & Ritchie, B. (1999). Tourism, competitiveness, and societal prosperity. *Journal of Business Research, 44*, 137-152.

Da Cunha, S.K., & Da Cunha, J. C. (2005). Tourism cluster competitiveness and sustainability: Proposal for a systemic model to measure the impact of tourism on local development. *Brazilian Administration Review, 2*(2), 47–62. doi: 10.1590/S1807-76922005000200005.

Deming, W. E., & Stephan. F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics, 11(4),* 427-444

Digun-Aweto, O., Fawole, O. P., & Saayman, M. (2019). The effect of distance on community participation in ecotourism and conservation at Okomu

National Park Nigeria. *GeoJournal, 84,* 1337-1351.

DSEC. (2021). Tourism statistics. Retrived from https://www.dsec.gov.mo/getAttachment/4e45ab48-81e9-42fd-a73a-d86d8eed113a/C_TUR_FR_2020_Q4.aspx (accessed 4 May 2021).

Edwards, D. & Griffin, T. (2013). Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management. *Journal of Sustainable Tourism, 21(4),* 580-595.

Eldridge, D., & Jones, J. P. (1991). Warped Space: A Geography of Distance Decay. *Professional Geographer, 43*(4), 500-511.

Emmons, S., Kobourov, S., Gallant, M., & Borner, K. (2016). Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. *PLoS ONE, 11*(7), e0159161. doi:10.1371/journal.pone.0159161

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U. M. Fayyad (eds.). *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226-231). AAAI Press

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide, Physics Reports. *Physics Reports, 659*, 1-44. DOI:10.1016/j.physrep.2016.09.002

Gaffar, V., Tjahjono, B., Abdullah, T., & Sukmayadi, V. (2021). Like, tag and share: bolstering social media marketing to improve intention to visit a nature-based tourism destination. Tourism Review, DOI: 10.1108/TR-05-2020-0215

Garay, L. (2019). Visitspain. Breaking down affective and cognitive attributes in the social media construction of the tourist destination image. *Tourism Management Perspectives, 32*, 100560.

Gartner, W. C. (1993). Image Formation Process. *Journal of Travel & Tourism Marketing, 2*(2/3), 191-215.

Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management, 72,* 306–312.

GitHub. (2020). jieba. Retrieved from https://github.com/fxsjy/jieba (accessed 4

May 2021).

Greater Bay Area. (2020). Outline development plan. Retrieved from https://www.bayarea.gov.hk/en/outline/plan.html (accessed 4 May 2021).

Gretzel, U., & Yoo, K. H. (2013). Premises and promises of social media marketing in tourism. In S. McCabe (Ed.), *The Routledge Handbook of Tourism Marketing* (pp. 491–504). Taylor and Francis.

Gu, Q., Zhang, H., Huang, S. S., Zheng, F., & Chen, C. (2021). Tourists' spatiotemporal behaviors in an emerging wine region: A time-geography perspective. *Journal of Destination Marketing and Management, 19,* DOI: 10.1016/j.jdmm.2020.100513

Gunn, C. A. (1972). *Vacationscape: Designing tourist regions*. Austin, TX: University of Texas Press.

Gunn, C. A. (1988). *Tourism planning*. New York: Taylor and Francis.

Gunter, U., & Önder, I. (2020). An exploratory analysis of geotagged photos from Instagram for residents of and visitors to Vienna. *Journal of Hospitality & Tourism Research*, DOI: 10.1177/1096348020963689

Hallo, J. C., Manning, R. E., Valliere, W., & Budruk, M. (2005). *A case study comparison of visitor self-reported travel routes and GPS recorded travel routes*. Paper presented at the 2004 northeastern recreation research symposium, Newtown Square, PA.

Hansen, B. (2021, March 15). The Hong Kong Tourism Board announces strategies to drive tourism recovery and sustainable development. *Hospitalitynet*. Retrieved from https://www.hospitalitynet.org/news/4103442.html (accessed 1 May 2021).

Hong Kong Tourism Board. (2020). A Statistical Review of Hong Kong Tourism 2019. Retrieved from https://securepartnernet.hktb.com/filemanager/intranet/ir/ResearchStatistics/paper/Stat-Review/StatReview2019/Statistical%20Review%202019.pdf

Hong Kong Special Administrative Region. (2021). HKSAR Government welcomes promulgation of Culture and Tourism Development Plan for Guangdong-Hong Kong-Macao Greater Bay Area by Ministry of Culture and Tourism. Retrived from https://www.info.gov.hk/gia/general/202101/04/P2021010400745.htm?fontSize=1 (accessed 4 May 2021)

Hooper, J. (2015). A destination too far? Modelling destination accessibility and distance decay in tourism. *GeoJournal, 80*, 33-46.

Hu, Y., & Ritchie, J. R. B. (1993). Measuring Destination Attractiveness: A Contextual Approach. *Journal of Travel Research, 32*(2), 25-34.

Hu, W., & Wall, G. (2005). Environmental management, environmental image and the competitive tourist attraction. *Journal of Sustainable Tourism, 13*(6), 617–635.

Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management, 47,* 68–76.

Iqbal, M. (2021). Instagram Revenue and Usage Statistics (2021). Retrieved from https://www.businessofapps.com/data/instagram-statistics/

Jeh, G., & Widom, J. (2003). Scaling personalized web search. *Proceedings of the 12th International Conference on World Wide Web, 271–279*. https://doi.org/10.1145/775152.775191

Jiang B. (2009). Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science, 23(7)*, 823–837.

Jiang, W., Xiong, Z., Su, Q., Long, Y., Song, X., & Sun, P. (2021). Using Geotagged Social Media Data to Explore Sentiment Changes in Tourist Flow: A Spatiotemporal Analytical Framework. *ISPRS International Journal of Geo-Information, 10(3)*, 135.

Jin, C., Cheng, J., & Xu, J. (2018). Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility. *Journal of Travel Research, 57*(6), 779-791.

Joo, D., Woosnam, K. M., Shafer, C. S., Scott, D., & An, So. (2017). Considering Tobler's first law of geography in a tourism context. *Tourism Management, 62*, 350-359.

Jovicic, D. Z. (2019). From the traditional understanding of tourism destination to the smart tourism destination. *Current Issues in Tourism, 22*(3), 276–282.

Karrer, B., & Newman, M.E.J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 83*(1), 016107.

Karna, R. K., Qiang, Y., & Karn, A. L. (2017). Facebook as destination image formation tool after Nepal earthquake: The collective essence of events in social media dialogue by travel marketers. *Transformations in Business & Economics, 16*(3C(42C)), 397–412.

Kim, H., Stepchenkova, S. (2015). Effect of tourist photographs on attitudes towards destination: Manifest and latent content. *Tourism Management, 49*, 29-41.

Kim, J. J., & Fesenmaier, D. R. (2015). Measuring Emotions in Real Time: Implications for Tourism Experience Design. *Journal of Travel Research, 54(4),* 419-429.

Kim, M., & Kim, J. (2020). Destination authenticity as a trigger of tourists' online engagement on social media. *Journal of Travel Research, 59*(7), 1238-1252.

Kirillova, K., Park, J., Zhu, M., Dioko, L. D., & Zeng, G. (2020). Developing the coopetitive destination brand for the Greater Bay Area. *Journal of Destination Marketing & Management, 17,* 100439. DOI:10.1016/j.jdmm.2020.100439

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632. https://doi.org/10.1145/324133.324140

Kladou, S., & Mavragini, E. (2015). Assessing destination image: An online marketing approach and the case of TripAdvisor. *Journal of Destination Marketing and Management, 4*(3), 187-193.

Kol'veková, G., Liptáková, E., Štrba, L., Kršák, B., Sidor, C., Cehlár, M.,

Khouri, S., & Behún, M. (2019). Regional Tourism Clustering Based on the Three Ps of the Sustainability Services Marketing Matrix: An Example of Central and Eastern European Countries. *Sustainablity, 11*, doi:10.3390/su11020400

Kuhnt, S. (2010). Breakdown concepts for contingency tables, *Metrika, 71(3),* 281 - 294.

Lai, L. S. L., & To, W. M. (2015). Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research, 16(2),* 138–152.

Lai, M. Y., Khoo-Lattimore, C., & Wang, Y. (2018). A perception gap investigation into food and cuisine image attributes for destination branding from the host perspective: The case of Australia. *Tourism Management, 69,* 579–595.

Laing, C., & Lewis, A. (2017). Exploring clustering as a destination development strategy for rural communities: The case of La Brea,Trinidad. *Journal of Destination Marketing & Management, 6*, 184-195.

Lalicic, L., Marine-Roig, E., Ferrer-Rosell, B., & Martin-Fuentes, E. (2021). Destination image analytics for tourism design: an approach through Airbnb reviews. *Annals of Tourism Research*, 86, DOI: 10.1016/j.annals.2020.103100.

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E, 80*(5), 056117. doi:10.1103/PhysRevE.80.056117

Leiper, N. (1990). Tourist attraction systems. *Annals of Tourism Research, 17(3)*, 367–384.

Lee, H., Guillet, B. D., Law, R., & Leung, R. (2012). Travel motivations and travel distance with temporal advance: A case study of Hong Kong pleasure travelers. *Journal of Destination Marketing & Management, 1*, 107-117.

Lenormand, M., Bassolas, A., & Ramasco, J.J. (2016). Systematic comparison of trip distribution laws and models, *Journal of Transport Geography*, *51*, 158-169.

Lenormand, M., Huet, S., Gargiulo, F., & Deffuant, G. (2012). A universal model of commuting networks. *PLoS One, 7*, e45985.

Leung, X. Y., Wang, F., Wu, B., Bai, B., Stahura, K. A., & Xie, Z. (2012). A social network analysis of overseas tourist movement patterns in Beijing: The impact of the Olympic Games. *International Journal of Tourism Research, 14*, 469-484.

Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research, 33*(2), 403-423.

Li, Y., Xie, J., Gao, X., & Law, A. (2021). A method of selecting potential development regions based on GPS and social network models – from the perspective of tourist behavior. *Asia Pacific Journal of Tourism Research, 26*(2), 183-199.

Lin, Y. R. (2010). *Community discovery in dynamic, rich-context social networks.* (Doctoral dissertation, Arizona State University). Available from ProQuest Dissertation and theses database (3425810).

Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., Chu, B., Zhu, H., Akinwunmi, B., Huang, J., Zhang, C. J. P., & Ming, W. K. (2020). Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. *Journal of Medical Internet Research, 22*(4), e19118. DOI:10.2196/19118

Lojo, A., Li, M., and Xu, H. (2020). Online tourism destination image: components, information sources, and incongruence. *Journal of Travel & Tourism Marketing, 37*(4), 495-509.

Lü, G., Batty, M., Strobl, J., Lin, H., Zhu, A. X., & Chen, M. (2019). Reflections and speculations on the progress in geographic information systems (GIS): A geographic perspective. *International Journal of Geographical Information Science, 33*(2), 346-367.

Lubbe, B.A. (2005). Part 4: The supply of tourism: sectors. In *Tourism Management in Southern Africa (2nd Eds)*, (pp. 101 – 112). Cape Town: Pearson Education South Africa.

Macao Government Tourism Office. (2020). Macao joins the world again to promote Sustainable Gastronomy Day fostering preservation of gastronomic culture through video production. Retrieved from https://www.gastronomy.gov.mo/#whats-on (accessed 4 May 2021).

MacCannell, D. (1976). *The tourist: A new theory of the leisure class*. London: Macmillan.

MacKay, K. J., & Fesenmaier, D. R. (1997). Pictorial element of destination in image formation. *Annals of Tourism Research, 24*(3), 537-565.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press.

Mason, P. (2003). *Tourism Impacts, Planning and Management*. Burlington MA: Butter worth-Mannheim (Elsevier).

Mayo, E. J., & Jarvis, L. P. (1981). *The psychology of leisure travel: Effective marketing and selling of travel service*. Boston, MA: CBI Publishing.

Mainka, A., Hartmann, S., & Stock, W. G. (2015). Looking for friends and followers: A global investigation of governmental social media use. *Transforming Government: People, Process and Policy, 9*(2), 237–254.

Mariani, M. (2020). Big Data and analytics in tourism and hospitality: a perspective article. *Tourism Review, 75*(1), 299-303.

Marine-Roig, E. (2019). Destination image analytics through traveller-generated content. *Sustainablity, 11*, 3392. DOI:10.3390/su11123392

Marine-Roig, E. (2021). Measuring online destination image, satisfaction, and loyalty: Evidence from Barcelona Districts. *Tourism and Hospitality, 2*, 62-78.

Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management, 4*, 162-172.

Mayo, E. J., & Jarvis, L. P. (1981). *The psychology of leisure travel: Effective marketing and selling of travel service*. Boston, MA: CBI Publishing.

Mazanec, J. A. (2020). Hidden theorizing in big data analytics: With a reference to tourism design research. *Annals of Tourism Research, 83*, DOI: 10.1016/j.annals.2020.102931

McKercher, B. (2008a). Segment transformation in urban tourism. *Tourism Management*, 29, 1215-1225.

McKercher, B. (2008b). The implicit effect of distance on tourist behavior: A comparison of short and long haul pleasure tourists to Hong Kong. *Journal of Travel & Tourism Marketing*, 25(3-4), 367-381.

McKercher, B., Chan, A., & Lam, C. (2008). The Impact of Distance on International Tourist Movements. *Journal of Travel Research, 47*(2), 208-224.

McKercher, B., & Du Cros, H. (2002). *Cultural Tourism: The Partnership Between Tourism and Cultural Heritage Management*. New York: Hayworth Hospitality Press.

McKercher, B., & Lau, G. (2008). Movement patterns of tourists within a destination. *Tourism Geographies, 10*(3), 355-374.

McKercher, B., Lew, A. A. (2003). Distance decay and the impact of effective tourism exclusion zones on international travel flows. *Journal of Travel Research, 42*, 159-165.

McKercher, B., & Mak, B. (2019). The impact of distance on international tourism demand. *Tourism Management Perspectives, 31*, 340-347.

McKercher, B., Mei, W. S., & Tse, T. M. (2006). Are Short Duration Cultural Festivals Tourist Attractions? *Journal of Sustainable Tourism, 14*(1), 55-66.

McGregor, A. (2000). Dynamic texts and tourist gaze: Death, bones and buffalo. *Annals of Tourism Research, 27*(1), 27-50.

McIntosh, R. W., Goeldner, C. R., & Ritchie, J. B. (1995). *Tourism: Principles, practices, philosophies (7th ed.)*. New York: Wiley.

Mihalcea, R., & Tarau, P. (2004). *TextRank: Bringing Order into Text*, Barcelona, Spain.

Moradi, L., Yahya, Y., Mohamed, I., & Raisian, K. (2017). A study of factors influencing online purchasing intention within E-tourism setting. *Journal of Environmental Management and Tourism, 4(20),* 882–895.

Middleton, V., & Clarke, J. (2001). *Marketing in travel and tourism*. Oxford: Butterworth-Heinemann.

Mihalcea, R., & Taray, P. (2004). *TextRank: bringing order into texts.* Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 404-411.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Calif.: Sage.

Missaoui, S., Kassem, F., Viviani, M., Agostini, A., Faiz, R., & Pasi, G. (2019). LOOKER: A mobile, personalized recommender system in the tourism domain based on social media user-generated content. *Personal and Ubiquitous Computing, 23,* 181–197.

Molinillo, S., Liébana-Cabanillas, F., & Anaya-Sánchez, R. (2017). Destination image on the DMO's platforms: Official website and social media. *Tourism & Management Studies, 13*(3), 5–14.

Mou, N., Zheng, Y., Makkonen, T., Yang, T., Tang, J., & Song, Y. (2020).

Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China. *Tourism Management, 81*, https://doi.org/10.1016/j.tourman.2020.104151

Murphy, P. E., Keller, C. P. (1990). Destination travel patterns: An examination and modeling of tourist patterns on Vancouver Island, British Columbia. *Leisure Sciences, 12*, 49-65.

Narangajavana, Y., Fiol, L. J. C., Tena, M. A. M., Artola, R. M. R., & Garcia, J. S. (2017). The influence of social media in creating expectations. An empirical study for a tourist destination. *Annals of Tourism Research, 65,* 60–70. DOI:10.1016/j.annals.2017.05.002

Ngai, E. W. T., Moon, K. K., Lam, S. S., Chin, E. S. K., & Tao, S. S. C. (2015). Social media models, technologies, and applications: An academic review and case study. *Industrial Management & Data Systems, 115*(5), 769–802.

Nowacki, M., & Niezgoda, A. (2020). Identifying unique features of the image of selected cities based on reviews by TripAdvisor portal users. *Scandinavian Journal of Hospitality and Tourism, 20*(5), 503-519.

Nyaupane, G. P., & Graefe, A. R. (2008). Travel distance: A tool for nature-based tourism market segmentation. *Journal of Travel & Tourism Marketing, 25*(3-4), 355-366.

Olaniyi, O. E., Ogunjemite, B. G., Akindele, S. O., & Sogbohossou, E. A. (2020). Temporal and distance decay analysis of land use/land cover around ecotourism hotspots: evidence from Pendjari National Park, Benin. *GeoJournal, 85*, 53-66.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web.* (Technical Report1999-66). Stanford InfoLab.

Page, S. (2015). *Tourism* management (5th ed.). Abingdon, Oxon: Routledge.

Pan, X., Rasouli, S., & Timmermans, H. (2021). Investigating tourist destination choice: Effect of destination image from social network members. *Tourism Management, 83,* 104217. DOI:10.1016/j.tourman.2020.104217

Paulino, I., Prats, L., & Whalley, P. A. (2020). Establishing Influence Areas of Attractions in Rural Destinations. *Tourism Planning & Development, 17*(6), 611-635.

Pearce, D. (1989). *Tourist development* (2nd ed.). Harlow: Longman Scientific.

Pearce, P. (1991). Analysing tourist attractions. *Journal of Tourism Studies, 2*(1), 46–55.

Pearce, P. (1993). Fundamentals of tourist motivation. In D. Pearce & R. Butler (Eds.), *Tourism Research: Critiques and Challenges* (pp. 85-105). London: Routledge and Kegan Paul.

Pocock, D., & Hudson, R. (1978). *Images of the urban environment*. London: Macmillan.

Porter, M. E. (1990). *The competitive advantage of nations*. New York: Free Press.

Ramires, A., Brandão, F., & Sousa, A. C. (2018). Motivation-based cluster analysis of international tourists visiting a WorldHeritage City: The case of Porto, Portugal. *Journal of Destination Marketing & Management, 8*, 49-60.

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management, 57*, 202-212.

Richards, G. (2002). Tourism attraction systems: Exploring cultural behavior. *Annals of Tourism Research, 29*(4), 1048–1064.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063–1064.

Schuckert, M., & Wu, J. (2021). Are neighbour tourists more sensitive to crowding? The impact of distance on the crowding-out effect in tourism. *Tourism Management, 82,* https://doi.org/10.1016/j.tourman.2020.104185

Scott, N., Baggio, R., & Cooper, C. (2008). *Network analysis and tourism: from theory to practice*. Channel View Publications.

Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., & Eliassi-Rad, T. (2008). Collective Classification in Network Data. *The AI Magazine,* 29(3), 93–106. https://doi.org/10.1609/aimag.v29i3.2157

Shao, J., Li, X., Morrison, A. M., & Wu, B. (2016). Social media micro-film marketing by Chinese destinations: The case of Shaoxing. *Tourism Management, 54,* 439–451.

Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017).

Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology, 8*(2), 220–238.

Smallwood, C. B., Beckley, L. E., Moore, S. A. (2012). An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia, *Tourism Management, 33*, 517-528.

Smith, S. L. J. (1985). U.S. vacation travel patterns: Correlates of distance decay and the willingness to travel. *Leisure Sciences, 7*(2), 151-174.

Spruit, M., & Ferati, D. (2020). Text mining business policy documents: Applied data science in finance. *International Journal of Business Intelligence Research, 11*(2), 28–46.

Sultan, M. T., Sharmin, F., Badulescu, A., Gavrilut, D., & Xue, K. (2021). Social Media-Based Content towards Image Formation: A New Approach to the Selection of Sustainable Destinations. *Sustainablity, 13*(8), DOI: 10.3390/su13084241

Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., & Wu, T. (2009). RankClus: integrating clustering with ranking for heterogeneous information network analysis. *Proceedings of the 12th International Conference on Extending Database Technology*, 565–576. https://doi.org/10.1145/1516360.1516426

Sun, Y., Han, J., Yan, X., Yu, P. S., & Wu, T. (2011). PathSim: meta path-based top-K similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment,* 4(11), 992–1003. https://doi.org/10.14778/3402707.3402736

Sun, Y,-Y., & Lin, P.-C. (2019). How far will we travel? A global distance pattern of international travel from both demand and supply perspectives. *Tourism Economics, 25*(8), 1200-1223.

Tamajón, L. G., & Valiente, G. C. (2017). Barcelona seen through the eyes of TripAdvisor: Actors, typologies and components of destination image in social media platforms. *Current Issues in Tourism, 20*(1), 33–37.

Taylor, P., McRae-Williams, P., & Lowe, J. (2007). The determinants of cluster activities in the Australian wine and tourism industries. *Tourism Economics, 13*(4), 639-656.

Teledifusao De Macao. (2020). Quarantine exemption extended to all Guangdong province. Retrieved from https://port.tdm.com.mo/c_radio/index.php?ra=nd&id=22857

Tham, A., Croy, G., & Mair, J. (2013). Social media in destination choice: Distinctive electronic word-of-mouth dimensions. *Journal of Travel & Tourism Marketing, 30*(1–2), 144–155.

The Department of Culture and Tourism of Guangdong Province. (2021). Greater Bay Area Culture and Tourism Development Plan. Retrived from http://zwgk.mct.gov.cn/zfxxgkml/ghjh/202012/P020201231518402967699.pdf (accessed 4 May 2021).

Tjørve, E., Flognfeldt, T., & Tjørve, K. M. C. (2013). The Effects of Distance and Belonging on Second-Home Markets, *Tourism Geographies, 15*(2), 268-291.

Trihas, N., Perakakis, E., Venitourakis, M., Mastorakis, G., & Kopanakis, I. (2013). Social media as a marketing tool for tourism destinations: The case of Greek municipalities. *Journal of Marketing Vistas, 3*(2), 38–48.

United Nations Educational, Scientific, and Cultural Organization. (2020). Shunde: About the creative city. Retrieved from https://en.unesco.org/creative-cities/shunde (accessed 4 May 2021).

UNESCO. (2011). Recommendation on the Historic Urban Landscape. Retrieved from http://www.historicurbanlandscape.com/themes/196/userfiles/download/2014/3/31/3ptdwdsom3eihfb.pdf

Vengesayi, S., Mavondo, F. T., & Reisinger, Y. (2009). Tourism destination attractiveness: attractions, facilities, and people as predictors. *Tourism Analysis, 14*, 621-636.

Virabhakul, V., & Huang, C.H. (2018). Effects of service experience on behavioral intentions: Serial multiple mediation model. *Journal of Hospitality Marketing & Management, 27*(8), 99-1016.

Vu, H. Q., Li, G., Law, R., & Ye, H. B. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*, 222-232.

Wall, G. (1997). Tourist attractions: points, lines and areas. *Annals of Tourism Research, 24*(1), 240-243.

Wang, D., Chan, H., & Pan, S. (2015). The impacts of mass media on organic destination image: A case study of Singapore. *Asia Pacific Journal of Tourism Research, 20*(8), 860–874.

Wang, Y., & Sparks, B. A. (2016). An eye-tracking study of tourism photo stimuli: Image characteristics and ethnicity. *Journal of Travel Research, 55*(5), 588–602.

Wearing, S.L. & Foley, C.T. (2016) Understanding the Tourist Experience of Cities, *Annals of Tourism Research, 65*, 97-107.

Williams-Burnett, N. J., & Fallon, J. (2017). What really happens in Kavos. *Journal of Place Management and Development, 10*(2), 183-195.

World Tourism Organization. (1980). Manila declaration on world tourism, *UNWTO Declarations, 1(1).* Retrieved from https://www.e-unwto.org/doi/pdf/10.18111/unwtodeclarations.1980.01.01

World Tourism Organization. (2005a). Report on the WTO survey on the implementation of the global code of ethics for tourism*. Ethics of Tourism, 2005(3)*. Retrieved from https://www.e-unwto.org/doi/epdf/10.18111/ethicsintourism.2005.3.t45n3x06777g133g

World Tourism Organization. (2005b) *Making Tourism More Sustainable - A Guide for Policy Makers.* Retrieved from https://www.e-unwto.org/doi/book/10.18111/9789284408214

Xia, J., Evans, F. H., Spilbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourists. *Tourism Management 31*, 464-469.

Xiao, Y., Li, B., & Gong, Z. (2018). Real-time identification of urban rainstorm waterlogging disasters based on Weibo big data. *Natural Hazards, 94*(2), 833–842.

Xiaohongshu. (2021). About Xiaohongshu. Retrieved from https://www.xiaohongshu.com/protocols/about (accessed 4 May 2021).

Xie, G., Zhang, R., Li, Y., Huang, L., Wang, C., Yang, H., & Liang, J. (2021). AttractRank: District Attraction Ranking Analysis Based on Taxi Big Data, *IEEE Transactions on Industrial Informatics*, *17(3)*, 1679-1688.

Xing, W. and Ghorbani, A. (2004). *Weighted PageRank algorithm.* Second Annual Conference on Communication Networks and Services Research CNSR'04, pp.305–314. 19–21 May 2004, Fredericton, NB, USA Yang, X. (2016). Tourist co-created destination image*. Journal of Travel & Tourism Marketing, 33*(4), 425-439.

Yang, T., Jin, R., Chi, Y., & Zhu, S. (2009). Combining link and content for community detection: a discriminative approach. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 927–936. https://doi.org/10.1145/1557019.1557120

Yogonet. (2020). Macau partially restarts tourist visas, tightens restrictions for Hong Kong visitors. Retrieved from https://www.yogonet.com/international/noticias/2020/08/10/54252-macau-partially-restarts-tourist-visas-tightens-restrictions-for-hong-kong-visitors (accessed 4 May 2021).

Yu, C.-E., & Sun, R. (2019). The role of Instagram in the UNESCO's creative city of gastronomy: A case study of Macau. *Tourism Management, 75,* 257–268.

Zamir, M. H. (2017). *Anatomy of a social media movement: Diffusion, sentiment, and network analysis*. (Doctoral dissertation, University of South Carolina). Available from ProQuest Dissertation and theses database (10606458).

Zhang, K., Chen, Y., & Lin, Z. (2020). Mapping destination images and behavioral patterns from user-generated photos: a computer vision approach. *Asia Pacific Journal of Tourism Research, 25*(11), 1199-1214.

Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place? *Tourism Management, 59*, 267-28

Zheng, X., Men, J., Yang, F., & Gong, X. (2019). Understanding impulse buying in mobile commerce: An investigation into hedonic and utilitarian browsing. *International Journal of Information Management, 48,* 151–160.