Exploring the characteristics of tourism industry by analyzing consumer review contents from social media: a case study of Bamako, Mali

Sanogo Bruno\textsuperscript{a,b}, Chao Yang\textsuperscript{a}, Wenwen Tian\textsuperscript{a}, Zhong Xie\textsuperscript{a} and Yuanzheng Shao\textsuperscript{c}

\textsuperscript{a}School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China; \textsuperscript{b}Agence pour le Promotion Touristique du Mali (APTM), Hamdallaye, Bamako, Mali; \textsuperscript{c}Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, China

ABSTRACT
In this Web 2.0 era, various and massive tourist experiences and reviews presented on social networks have become important information for tourism research. In this paper, we apply social media to explore and study the tourism industry of Bamako, Mali. Over 2000 reviewers and their comments about Bamako’s hotels and restaurants from TripAdvisor and Facebook were collected. Also, we integrate official tourism statistic data and field surveying data into the online review dataset. Data mining and statistic method are used to analyze the data for purpose of exploring the characteristics about tourism industry in Bamako. And we find that: (i) Most tourists are coming to Bamako for business purpose, and they incline to choose the hotels with better service and security condition; (ii) Comments on social media would greatly affect travelers’ choice on hotels; (iii) Most travelers are satisfied about Bamako’s accommodation services.

1. Introduction
Today, social media information is regarded as an important travel information source. Various travel information contents (e.g. texts, photos, and video) are generated by billions of social media users when presenting their unique traveling experiences all over the world. People can access abundant tourism information from all kinds of social media services such as social networks (e.g. Facebook and Instagram), online review sites (e.g. TripAdvisor and Yelp), and many other social knowledge sharing sites (e.g. Wikitravel). The travel information contributed by social media users is perceived as trustworthy (Dickinger 2010), timely, and effective (O’Connor, Flanagan, and Gilbert 2010) tourism data. Tourists now are positive to share their experiences and give specific suggestions to others for hotels, restaurants, and attractions (e.g. via comments on customer service, car parking, and cleanliness) (Sotiriadis and Zyl 2013) on social networks.

Extant works of literature have illustrated that online customer reviews can be used as a major information source for researchers and practitioners. And it helps to correctly understand consumers’ preferences and demands: for example, to predict financial performance or to increase sales (Ghose and Ipeirotis 2011; Chau and Xu 2012; Clemons, Gao, and Hitt 2006; Mayzlin and Mayzlin 2006; Ye et al. 2011). Online customer reviews can empower individuals to bypass unclear and inaccurate product or service descriptions and rely directly on the first-hand experiences of other consumers to choose the products and services. Social media travel information is the very important travel information that reflects the traveler’s direct description of the travel experience. According to a presentation by TripAdvisor at the Social Media in Tourism Australia Symposium (TripAdvisor.com 2013), over 60% of travelers trend to search travel information and reference reviews from social media sites such as Facebook, TripAdvisor, Instagram, and Pinterest.

Tourism is an information-intense industry (Hannes and Klein 1999; Sheldon 1997); therefore, it is important for the tourism management department to collect tourist’s feedback to understand consumer’s behavior and travel-related information. (C. Dellarocas 2003) suggested that social media provide tourism companies with unprecedented opportunities to understand and respond to consumer preferences. Social media services like Facebook and TripAdvisor attract tremendous travelers all over the world to create and share enormous and ubiquitous travel experiences online. By analyzing the comments from the online travel review sites, e.g. TripAdvisor, hotels and other travel-related companies can understand their guest’s preference among them and their competitors, and improve customer’s travel information searching experience (Sánchez-Franco and Rondan-Cataluña 2010). In the past decade, there has been an emerging stream of studies examining the phenomenon of online consumer reviews (e.g. Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Ho-Dac, Carson, and Moore 2013). Hundreds and
thousands of community members may contribute to building online content, thereby creating the “wisdom of crowds” (Surowiecki 2005). Thus, UGC can serve as a useful source of information for enterprises that care about consumers’ demands, particularly in the hospitality industry (e.g., hotels and restaurants).

Due to the lack of information gathering method, the National Tourism and Hospitality Department of Mali (Direction Nationale de l'Office Malien du Tourisme et de l'Hôtellerie, DNTH Mali) is hard to collect the effective tourism information about the tourists. For a long time, the DNTH Mali can only receive a copy of annual tourists report from the Mali customs department, e.g. where does a tourist come from, where will he/she stay, and how much time will he/she spend in Mali. In this situation, it is hard for the DNTH Mali to explore a full picture of the tourism industry, especially the hospitality industry. In this paper, we present our research by using the social media data from TripAdvisor and Facebook to explore the tourism industry of Bamako, Mali. We collect about 3000 TripAdvisor and Facebook online comments on hotels and restaurants of Bamako from more than 2000 visitors. We aim to offer an in-depth insight into Mali tourism and hospitality industry. The paper is organized as follows. First, we present the method used in our research, including sampling and text mining techniques. Second, the comments are discussed, followed by the perceptual mapping of results about hotels and restaurants. Finally, there is a discussion of limitations and implications for research and practice.

2. Relate work

Internet has changed the distribution pattern of tourism-related information and the way people plan to take their travel. The power of the tourism industry is growing fast with Web 2.0 which has facilitated and increased the use of digital devices (Giglio et al. 2019). This has allowed the tourism sector to be a big provider of information. Social media can be considered as one of the most dynamic online network tools that has been incorporating a part of social and economic domain in the real world (Xiang and Gretzel 2010; Zeng 2013).

Online reviews from social media platforms have become an important indicator of the evaluation of the travel and hospitality industry. This has great impact on the tourism by producing large amount of data about tourist behaviors, brand loyalty, etc. (Cheng et al. 2019; Misirlis and Vlachopoulou 2018). The understanding of changes in IT and consumers behavior that impact the spread and accessibility of tourism information is now important. Social media promote the synergy between online consumers and emerge as an important component of tourism domain. Many studies have explored social media; however, few have focused on the effects of social media marketing activities.

With the increasing importance of social media and online review information in the tourism operation and management, plenty of researches have studied the impact of this kind of user-generated content (UGC) in tourism industry. The spread of tourism-related information, as well as the way to plan trips, have been essentially changed over the Internet (Xiang and Gretzel 2010). On the one hand, many academic researchers have confirmed the important role social media takes in travel-related decisions. For instance, McCarthy, Stock, and Verma 2010 explored the impact of social media affecting customer preferences in the hotel industry. Tussyadiah, Park, and Fesenmaier 2011 found that UGC on social media helps audience gain information about a place by recognizing coherence in the story, and subsequently generates higher boosting in choosing the destination. On the other hand, online travel experiences and reviews offer rich information which can help the tourists with their trips, for example, planning itineraries and booking hotels. Research guided by Google showed that 84% of leisure travelers use the Internet as a planning resource. In another study conducted by (Fortis, Buhalis, and Rossides 2011), the authors found that social media have changed the way online Internet users from Russia and other former Soviet Union Republics make to their holiday plans.

In hospitality and tourism, online consumer reviews have been studied for various research issues. Previous studies using consumer review data tend to rely on a single data source and the data quality is largely anecdotal. The tourism industry is exploiting social media points of sale, based on the reputation of the destination, the opinion of consumers, etc. Social media research in tourism is still in its infancy. Although case studies focusing on the qualitative analysis of the impact of social media on tourism are needed, however, in-depth research on the influence and impact of social media on all aspects of the tourism industry and the demonstration of the economic contribution of this industry is essential. Zheng et al. (2019), using social media analytics procedure, have examined the quality of the information related to online reviews from TripAdvisor, Expedia, and Yelp, about hotels in Manhattan, New York City and discovered that huge discrepancies exist in the representation of the hotel industry on these platforms. The authors indicated that, first, these findings are based on sentiment distribution, rating and review helpfulness must be interpreted with caution; second, the text analytics tools applied to the data have their inherent limitations. More conveniently, online booking increased the number of users inclined to book hotels. Thus, online reviews and ratings by users were following this trend. Previous studies pointed out the impact of online
review and/or user rating (Wang et al. 2019). Moreover, Buted and Gillespie (2014) pointed out that the main problem in using social media is the customer’s opinions. This thought and expression are not presented leading to bad impression and unfair criticism. Levy, Duan, and Boo (2013) said “care must be done regarding their findings” because the study area was Washington D.C., i.e. Washington D.C. hotels gain relatively high guest average. However, these studies did not include the extensive geographic area.

In the hospitality management process, consumer surveys, especially guest comment cards, have been widely used to measure hotel guest’s satisfaction (Pizam 1982). With the help of data mining technologies such as opinion mining and sentiment analysis, hospitality suppliers can better understand customers’ demands and responses based on their online reviews on social networks. Due to the unstructured nature of social media information, measuring the guest’s satisfaction is a challenging task. The text analytics method plays an important role in various types of market intelligence applications (Pang and Lee 2008). Data mining methods such as natural language processing (NLP) information extraction (IE), and artificial intelligence (AI) have been widely applied to the sentimental analysis and opinion extraction from social media online reviews. For example, Li, Ye, and Law (2013) used Latent Dirichlet Allocation (LDA) method to analyze the key factors for consumers’ satisfaction through the 42,886 online reviews of 774 star-rated hotels in Beijing. Levy, Duan, and Boo (2013) carried out a content analysis about the customers’ complaints to low-grade hotel from online review websites to understand the customers’ dissatisfaction. (Sparks and Browning 2011) revealed that hotel customers’ preference for booking is determined by price and feedback from online reviews.

The massive social media information does provide new means for tourism department and destination marketing organization (DMO) to better understand and respond tourists’ behaviors and requests. This abundant crowd-sourced traveling data offer opportunities to analyze and make statistical inferences about consumer’s behavior in hospitality industry, which can guide the tourism departments and DMOs to make corresponding policies, and implement business models. This research adopts online hospitality reviews from TripAdvisor and Facebook, and we developed a clear understanding of tourism in Bamako, Mali.

3. Study area

Republic of Mali (Mali) is a landlocked country in West Africa with 1,241,238 km². Tourism industry is an important part of the Mali’s economy. Since the northern Mali conflict in 2012, the tourism industry of Mali was deeply ruined and is still in a recovering process. Foreign travelers come to Mali with great expectation to visit the World Heritage Sites Timbuktu, Djenné and Bandiagara. The civil war put a halt to all existing projects that were initiated to open up and develop these sites, and visitor levels to these previously crowded tourism attractions were in a recession.

As Figure 1 shows, Bamako, the capital city of Mali, has a population of about 19 million which distributes in 10 administrative regions and the District of Bamako. Bamako possesses most of Mali’s tourism infrastructures and resources: the biggest international airport in Mali; 87.9% of Malian travel agencies; 55.6% hotels and 71.23% restaurants; 76.57% of the total jobs in tourism industry. According to data from the 2016 Mali inbound passengers collected by the Mali Customs Department, 98% of passengers first arrived in Bamako when they entered Mali, and 38.38% of them traveled overnight in Bamako. Research on Bamako tourism will also reflect the tourism trends of the entire country. Therefore, it is very necessary to study Bamako’s hotel and service industry and service conditions from the social media comments generated by tourists.

4. Data resource and data processing

We collected online reviews on Bamako hotels and restaurants between January 2006 and April 2017 from TripAdvisor and Facebook. Due to the low popularity of Internet facility, only 27.11% and 24.13% of hotels and restaurants were registered on the booking system or/and on the social media platform, respectively. In order to guarantee the data quality, all online reviews on these hotels and hotels are required to reach more than 5 times. From TripAdvisor website, we collected 1194 and 1484 pieces of reviews from 31 hotels and 52 restaurants, respectively. While on Facebook, we found 16 hotels and 32 restaurants which totaled 76 and 263 reviews, respectively. The reviewers’ dataset includes each reviewer’s attribute data (user ID and nationality information), and reviewer satisfaction data extracted from the review text using the natural language processing (NLP) method.

Then, we processed the data into three datasets: hotel comments dataset, restaurant comments dataset, and reviewers’ dataset. The hotel and restaurant review datasets include attribute information (location, rating, price level, etc.) for each hotel and restaurant, and the comment information. Beside that, we added security information for the hotel’s dataset. Safety is an important reference for tourists when they choose hotels, but it is usually difficult to learn from online reviews. We developed a hotel safety dataset by conducting field surveys of safety indexes for each hotel. The investigations were done by the authors after making the criterion with advisors from DNTH and from the United Nations Multidimensional Integrated Stabilization Mission in Mali (MINUSMA).
4.1 Extract tourism information from social media comments

After the 2012 northern Mali conflict, some hotels and restaurants have been closed because of lacking customers (e.g. Hotel Royal, Hotel Le Refuge) or requisitioned as administration headquarters (e.g. case of hotel Azalai Nord Sud which is now the headquarter of the European Union Training Mission (EUTM)). We used the hotel registration data from DNTH as the reference data, and eventually kept 31 hotels and 52 restaurants as our research objects. Finally, we gathered 1270 hotel reviews and 1747 restaurant reviews from 1250 hotel reviewers and 1152 restaurant reviewers, respectively. Also, due to the diversity in the language of the reviewers, all the reviewers’ comments were translated into English for the convenience of data processing.

We analyzed tourism information in social media reviews from both user information and commentary content. The user information includes the user’s attribute and activity. We can extract data from user’s profile information such as user name (ID), gender information, nationality information and age information. Besides, we can also tell when a user visit or revisit a place from his/her active information. The user name (ID) is usually unique in a social media platform, so we can locate all the information about a user.

4.2 Field investigation about hotel security condition

The DNTH produces the annual statistical reports on the tourism industry of Mali every year. The administrative information on the hotels their official ranking were obtained from those reports. Moreover, the investigations about hotels’ security level were made in conjunction with the DNTH.

In Mali, safety and security conditions would largely affect tourist’s choices about hotels. Mali has gone through a series of terrorist attacks on certain tourism infrastructures. For instance, in March 2015, a masked gunman opened fire at La Terrasse restaurant; then in November, Radisson Blu Hotel was raided by Islamist militants who killed 20 people. As a result, many hotel owners have equipped safety and security facilities (e.g. X-Ray, metal detectors or explosives) and hired security guards for their property.

To better understand how important the security conditions would affect tourist’s choice on hotels, we took a fieldwork to investigate main Mali hotels’ security facilities. The evaluation criteria are based on the existence of checkpoint on each tourism infrastructure. These checkpoints must have metal detectors (the mirror to control cars, x-rays to scan objects, hand scanners or magnetic gantries) and qualified security agents provided by the government (National Guard or Police) or security companies. Many luxury hotels are equipped with security check-points, where tourist would pass a metal detector while the luggage should go through x-ray scanner. However, not all the hotels were well equipped with security facilities. In this case, we rank the safety and security conditions based on their full extent of security equipment: we give 1 score to each safety and security equipment (i.e. mirror = 1, X-ray = 1, metal detector door = 1); the presence of security agents from government has 1 as score (national guard or police) and the existence of private companies’ agents also scores 1. As Table 1 shows, we categorized hotels

Figure 1. Location and administration division of Bamako.
in Bamako into 5 security levels: high level concerns hotels whose security score is 5; very good level is about hotels that gain 4 scores; average level is available to hotels that obtain 3 for security score; poor level is for hotels with 2 security scores and very poor level is for the hotels with 1 as security score.

4.3 Sentiment analysis on reviewer’s comments

In the tourism industry, sentiment and opinion analysis allow service providers to identify the advantages and disadvantages of their facilities and services, for example, hotel room cleanliness, the staff at a tourist spot, or the service at a restaurant. The sentimental analysis result can provide useful feedback information and can help to spot a problem and get down to the solution immediately; on the other side, from user’s comments, we can read the user’s experiences about a place. Mining sentimental information from these online reviewer’s comments can help to understand user’s opinions and attitudes about the hotels and restaurants. Then, it can help to figure out tourists’ preferences and flavors.

In this research, we used natural language processing methods for reviewer’s polarity detection, i.e. analyzing positive and negative emotions in customer reviews to determine customer’s sentiment orientation on hotels and restaurants. We applied an open sourced NLP toolkit (NLTK) under Python platform. NLKT has a sentiment analysis lexicon library and pre-trained sentiment analysis models that can be used directly to process sentences in hotel and restaurant reviews and to obtain sentiment analysis results. For instance, a customer from TripAdvisor wrote down his comments on Radisson Blu Hotel, Bamako: “the room in deep need for maintenance, broken shower dead bulbs and the room very dark. the staff untrained and rude.”. We use NLKT VADER component, which is designed by Hutto and Gilbert (2014), to analyze these sentences, and we got the analyzing result as follows: “negative emotion score: 0.366, neutral emotion score: 0.634, positive emotion score: 0.0”. Then, we can learn the guest keep a negative opinion on the Radisson Blu Hotel. The opinion information on review data can help us understand the quality of service in the Bamako hospitality industry and the differences in understanding of services among different traveler groups.

Finally, we combine the online review data, hotel registration data from DNTH, and the field investigation security data together, then, we reorganized Bamako’s accommodation information as follows: (1) hotel’s information (ID, hotel’s name, location, star rating, styles, ranking, overnight price, Business customer numbers, leisure customer numbers, owners’ nationalities, safe and security level) as presented in Table 1; (2) restaurant’s information (ID, restaurant’s name, location, cuisine, price, number of reviews, and owner’s nationalities) as shown in Table 2; (3) reviewers’ opinions on hotels and restaurants.

Table 1. Reviews on hotels of Bamako.

| ID   | Hotels’ Name           | Stars | Styles   | Ranking | Overnight price | Business | Leisure | Reviews | Owners Nationalities | Safe & Security Level |
|------|------------------------|-------|----------|---------|-----------------|----------|---------|---------|----------------------|----------------------|
| H001 | Auberge Djamilla       | none  | Budget   | 4.5     | 19              | 3        | 15      | 17      | Swiss                | 1                    |
| H002 | Azalai Hôtel Salam     | 5      | Luxury   | 3.5     | 161             | 82       | 20      | 108     | Malian               | 5                    |
| H003 | Azalai Grand Hotel     | 4      | Mid-range| 3.5     | 108             | 56       | 18      | 77      | Malian               | 5                    |
| H004 | Azalai Hôtel Dunia     | 2      | Budget   | 3       | 70              | 0        | 2       | 6       | Malian               | 5                    |
| H005 | Bada Lodge             | none   | unspecified | 5      | 123             | 1        | 0       | 13      | French               | 3                    |

Table 2. Reviews on restaurants of Bamako.

| ID   | Restaurant             | Cuisine                  | Price | Meals        | Reviews | Owners nationalities |
|------|------------------------|--------------------------|-------|--------------|---------|----------------------|
| R001 | African Grill          | African Cuisine          | $$$   | Lunch, Dinner | 16      | Malian               |
| R002 | Annapuma Restaurant    | Indian Cuisine           | $$$   | Dinner       | 11      | Indian               |
| R003 | Appaloosa Restaurant   | American Cuisine         | $$$   | unspecified  | 132     | American             |
| R004 | Badala Restaurant      | European Cuisine         | $$$   | unspecified  | 30      | French               |
| R005 | Badalodge Restaurant   | European Cuisine         | $$$   | Breakfast, Dinner, Brunch | 50 | French |

Table 3. Summary of emotional opinions of reviews on trip advisor and Facebook (1270 hotels and 1747 restaurants).

| Item                      | TripAdvisor | Facebook |
|---------------------------|-------------|----------|
|                           | N           | Percent  | N           | Percent  |
| Reviews hotels            | 1194        | 94.01    | 76          | 5.99     |
| Reviews restaurants       | 1484        | 84.94    | 263         | 15.05    |
| Pos. opinions hotels      | 1103        | 86.85    | 53          | 4.18     |
| Negative opinions hotels  | 91          | 7.16     | 23          | 1.81     |
| Positive opinions restaurants | 1203   | 68.86    | 205         | 11.73    |
| Negative opinions restaurants | 281   | 16.08    | 58          | 3.32     |

The results indicate that both reviewers from TripAdvisor and Facebook platforms have shown satisfaction with the restaurant and hospitality industries in Bamako. The positive sentiment of the hotel reviews was 91.03%, the negative sentiment was 8.97%; the positive mood of the restaurant reviews was 80.59%, and the negative sentiment was 19.4%. However, reviewers from Facebook are more picky, and the rating satisfaction in
5. Data analysis and findings

After all the possible features that would affect tourists’ choices on hotels and restaurants were listed, we used principal component analysis (PCA) method to analyze the most important features that would affect tourists’ options. PCA is a widely used algorithm in the social and physical sciences analysis research. In this research, we listed the features as a set of variables, and we adopted Statistical Product and Service Solutions (SPSS) software, applied PCA to examine and identify variables which relatively contribute more in tourist’s choice.

Kaiser-Meyer-Olkin (KMO) measure of sample Adequacy generally indicates whether or not the variables can be grouped into a smaller set of underlying factors. KMO varies from 0 to 1 and should be 0.60 or more to proceed. We use 0.50 for KMO score as a more lenient cut-off. If the KMO is less than 0.50, the results of the factor analysis probably will not be very useful. Some authors set this value at 0.50. For reference, Kaiser put the following values on the results: 0.00 to 0.49 unacceptable; 0.50 to 0.59 miserable; 0.60 to 0.69 mediocre; 0.70 to 0.79 middling; 0.80 to 0.89 meritorious; 0.90 to 1.00 marvellous. (Kayole and Charity 2018)

5.1. Hotels’ feature analysis

After we input the hotels’ feature data, we calculated the KMO score, which was 0.655 (higher than 0.60), while the P-value for Bartlett’s sphericity test was sig.0.000. In the PCA analysis process, we calculated seven features: Hotels’ Stars, Hotels’ Style, Hotels’ price, No. of Reviews, Positive opinions, Owners Nationalities, Security Level. After that, we extracted three components. The component 1 had two strong and positive loadings: Positive opinions (0.954) and the number of reviews (0.850) (Table 4). It described the positive opinions and number of reviews frequency about the hotels. In component 2, the security level is the strongest (0.855) followed by star (0.747). Only the style had a strong index (0.689) in component 3.

Then, we explored the relationship between the purpose of the tourists to Bamako and their choice of hotels. Through Figure 2, we noticed that except unspecified hotels, where the leisure trip takes the majority, the business trips increases as the hotel standard increases. This trend was proved by a positive and moderate correlation between hotels’ rating (star) and nights stay numbers of business-purposed trip.

We applied the independent T-test to evaluate whether the business and leisure have equal means on some variable. The results have indicated that the mean of condition 1 (business) is 25.39 and the mean of condition 2 (leisure) is 14.84, which shows the difference of 10.55. In addition, Levine’s Test for Equality of Variances that determines if the two conditions have about the same or different amounts of variability between scores was check, and the result was 0.047, which means business travelers and leisure travelers have different preferences and choices when choosing a hotel.

5.2. Restaurants’ feature analysis

For the restaurant’s dataset, the KMO and P-value for Bartlett’s sphericity test was 0.648 and sig. 0.000, respectively; We calculated seven features: Cuisine Type, Price, No. of Reviews, Restaurant Rating, Restaurant Type, Positive opinions, and Owners nationalities. And we used PCA method to extract two components; Component 1 had three strong and positive loadings: price (0.769), Cuisine (0.724) and reviews (0.722) as shown in Table 5. It described the opinions frequency of the restaurants. In Component 2, the restaurant rating was the only strong index (0.700). The results suggest that when a user chooses a restaurant, the cuisine type, price, and popularity of the restaurant will be the overall decision-making factor of the user.

Then, we explore the preferences of customers in different languages for choosing a restaurant by studying the language types of online reviews in restaurants. Figure 3 shows that French-speaking customers are the most frequent, followed, respectively, by English, Italians, and Indians speaking customers. From all the restaurants, 47 restaurants received online comments in French, and 21 restaurants attracted English-speaking customers. Among them, Le loft is the most popular restaurant, accounting for 9.98%, followed by Da Guido (9.52%) and Badaloge (7.82%).

For foreigners, restaurants of their country’s flavor seem to be their first dining choice. For example, Annapurna, the most popular Indian restaurant that

Table 4. Loading scores and indices for each component.

| No. | Index                        | Component 1 | Component 2 | Component 3 |
|-----|------------------------------|-------------|-------------|-------------|
| 1   | Hotels’ Stars                | 0.413       | 0.747       | -0.056      |
| 2   | Hotels’ Styles               | 0.375       | 0.007       | 0.689       |
| 3   | Hotels’ price                | 0.245       | 0.491       | 0.412       |
| 4   | No. of Reviews               | 0.850       | 0.233       | 0.190       |
| 5   | Positive opinions            | 0.954       | 0.005       | 0.098       |
| 6   | Owners Nationalities         | 0.049       | 0.079       | -0.884      |
| 7   | Security Level               | -0.125      | 0.855       | -0.091      |
attracted 25.00% of India people in Bamako. Their second and third choices for India people were, respectively, Appaloosa (18.75%) and Soukhothai (12.00%), whose owner comes from Thailand and offers Thai food. At the same time, Italians firstly chose an Italian restaurant (Da Guido) which got 23.33% as cumulative percentage. For English-speaking people, the Le loft restaurant attracts 12.22% English-speaking customers and ranked first; followed by Appaloosa, an American restaurant that attracted 11.31% of customers; followed by Soukhotai, a Thai restaurant hosted by Thai citizens.

6. Discussion and conclusions

The social media data have shown its value as a kind of novel but trustworthy data source for tourism industry research. In this paper, we offer an in-depth insight of Bamako tourism industry by analyzing the travel-related social media data from TripAdvisor and Facebook. From the user-generated travel information, we can understand the tourists’ behavior and emotions. These findings

| Table 5. Loading scores for indices in each component. |
|-----------------------------------------------------|
| No | Index                      | Component 1 | Component 2 |
|----|----------------------------|-------------|-------------|
| x_1| Cuisine Type               | 0.724       | 0.189       |
| x_2| Price                      | 0.769       | -0.111      |
| x_3| No. Reviews                | 0.722       | 0.363       |
| x_4| Restaurant Rating          | 0.007       | 0.700       |
| x_5| Restaurant Type            | -0.100      | -0.827      |
| x_6| Positive opinions          | 0.459       | 0.243       |
| x_7| Owners nationalities       | -0.698      | 0.160       |

Note: The bold fonts mean the significant indices.
contribute to a thoughtful understanding of Bamako’s hospitality industry and help Mali’s tourism participators to make corresponding policies.

After collecting the travel and hospitality data from about 2000 TripAdvisor and Facebook users between 2006 and 2017, we studied the hospitality industry’s structures and characters of Bamako. We used PCA method to study the key factors that direct tourists’ choices on hotels and restaurants, and we found out that the guests would pay much attention on the reviews, positive opinions, security level, hotel’s star rating, and hotel’s style (Table 3). The NLP method and Natural Language toolkit are applied to analyze the customers’ positive and negative opinions from their comments on the hotels and restaurants, which leads a better understanding of tourists’ preferences and flavors. And we surprisingly find out that most tourists were satisfied with the accommodation condition and service in Bamako and wrote down positive words in their reviews. Despite of the critical security situation and the economic decline since the 2012 Northern Mali conflict, the hospitality industry kept good services for tourists all over the world and earned a good reputation.

Business tourism generally includes activities such as congress, fairs, and exhibitions, seminars, conventions, incentives, team building, etc. The leisure tourism can be defined as time available to an individual when work, sleep, and other basic needs have been met. It is how someone chooses to spend their free time. In Mali, the purpose of travel is recorded from tourist when they land at airport. They have to fill up the arrival form including name, age, jobs, and the purpose of travel, type of accommodation. Yearly, tourism administration provides statistical reports on tourism including propose of travel. In this study, All users (families, friends, couples, solo) whose purpose of travel such as pure leisure and business were gathered, while business remain the same. The statistics from DNTH on the travels in Bamako indicated that business is the main purpose of travel in Mali with 67% of cumulative percentage of travel goal. Leisure tourism areas are located in the middle, e.g. Dogon Country, the ancient cities of Djenne and Timbuktu, and in the north where the immense Sahara Desert is located. Unfortunately, the security crisis have deteriorated the tourism and especially leisure tourism. Leisure tourism has thus declined even more. However, the business tourism continues to increase. Indeed, some politicians and entrepreneurs have treated the crisis as a boon. Since the crisis began, the number of seminars and training sessions on security questions has increased. The number of security equipment providers visiting the country has increased as well. The foreign tourists care about the services and room conditions, and the online reviews would affect their decisions in a great extent. Particularly, for the business tourists, they concerned more about the security situation.

We also found out that there exists a connection between customers’ languages and tourism features owners’ languages in Mali. As shown in Figure 3, the customers who speak different languages have different preferences.

This study illustrates the importance and significance of its theory and practice of tourism research from social media data. We discussed the significance of customer review contents from social media in studying tourism industry, especially, in hotel and catering industry. Also, we introduced the official statistics data, i.e. the DNTH tourist statistic data in evaluating the social media online review data. Still, the potential of social media data in tourism research have not been fully exploited. For the underdevelopment country which lack of information infrastructure, e.g. Mali, the social media could be a great tool and resource to develop knowledge about tourism industry. In the future research, we would expand our research scopes, for instance, tourist attractions. Meanwhile, we would try to analyze tourists’ preferences and traces in Mali from social media data.

Acknowledgments
This work was supported by the National Key R&D Program of China (grant number 2017YFB0503700) and the National Nature Science Foundation of China (grant number 41501439).

Notes on contributors
Sanogo Bruno, Geographer Administrator of Agence pour la Promotion Touristique du Mali (APTM), is a PhD candidate of Geographic Information System (GIS) at China University of Geosciences, Wuhan. His research interests cover the areas of Big data analysis, text mining, social media, spatial analysis.

Chao Yang, PhD, is Associate Professor in the School of Geography and Information Engineering at China University of Geosciences, Wuhan. He received PhD degree in Geography Information System(GIS) from Wuhan University at 2012. His research interests include data mining, text analytics, social media analytics for human geography studies.

Wenwen Tian is faculty in the School of Geography and Information Engineering at China University of Geosciences, Wuhan. She received Ph.D. degree in Geography Information System(GIS) from Wuhan University at 2013. Her research interests include data mining, geospatial analytics, and Geocomputation.

Zhong Xie, PhD, is Professor in the School of Information Engineering at China University of Geosciences, Wuhan. His research interests include geospatial analytics and Geocomputation.

Yuanzheng Shao is Associate Professor in the Collaborative Innovation Center of Geospatial Technology, Wuhan University. He received PhD degree in Geography Information System(GIS) from Wuhan University at 2011. His research interests include geospatial analytics and Geocomputation.
