Crowdsourcing, *Big Data* and Network analysis techniques applied to Tourist Demand

O uso de técnicas de Crowdsourcing, Big Data e análise de Redes aplicadas à Demanda Turística

Júnia Lúcio de Castro Borges
Universidade Federal do Maranhão, Brasil
borgesjunia@gmail.com

André Riani Costa Perinotto
Universidade Federal do Delta do Parnaíba, Brasil
perinotto@ufpi.edu.br

Solano de Souza Braga
Universidade Federal de Ouro Preto, Brasil
solanobraga@yahoo.com.br

ABSTRACT

This study applied text data mining techniques, statistical text analysis and network analysis in a qualitative database with information resulting from comments from social media specialized in tourism in order to identify the perception of visitors to the National Parks of Piauí. These conservation units are important tourist assets because they house a great natural heritage with extreme historical relevance with thousands of archaeological sites. The methodology developed in this article presents defensible and reproducible criteria to be replicated in any tourist attraction present on TripAdvisor, and greatly expands the understanding about the perception of visitors and the essence of places, facilitating the decision making of planners and managers of tourist destinations. The presented results were achieved from the use of automation techniques and computational programming to extract a large volume of data - Big Data - from digital footprints left by travelers on TripAdvisor, regarding the tourist attractions of interest. The qualitative database was mined using free software aimed at textual analysis, statistical treatment and network analysis. Based on the results, it was possible to identify key aspects regarding

the tourist destinations such as the centrality of themes regarding archaeological aspects and the monitoring of local guides.

Keywords: TripAdvisor, Tourist Demand, Big data, Tourism Planning, digital footprint.

RESUMO

O presente estudo aplicou técnicas de mineração de dados textuais, análise textual estatística e análise de redes em um banco de dados qualitativo com informações resultantes de comentários provenientes de mídia social especializada em turismo com vistas a analisar o discurso dos visitantes dos Parques Nacionais do Piauí. Estas unidades de conservação são importantes ativos turísticos porque além de abrigarem grande patrimônio natural, são de extrema relevância histórica por conterem milhares de sítios arqueológicos. A metodologia desenvolvida neste artigo apresenta critérios defensáveis e reproduzíveis para ser replicada em qualquer atrativo turístico presente no TripAdvisor, e amplia sobremaneira a compreensão acerca da percepção dos visitantes e da essência dos lugares, facilitando a tomada de decisões de planejadores e gestores de destinos turísticos. Os resultados apresentados foram alcançados a partir do uso de técnicas de automação e programação computacional para extrair um grande volume de dados - Big Data - de rastros (pegadas digitais) deixados pelos viajantes no TripAdvisor, a respeito dos atrativos turísticos de interesse. O banco de dados qualitativos foi minerado com o uso de softwares livres voltados à análise textual, tratamento dos dados e análise das redes. A partir dos resultados foi possível identificar aspectos fundamentais a respeito dos destinos turísticos ora destacados como, por exemplo, a centralidade das temáticas a respeito dos aspectos arqueológicos e o acompanhamento de guias.

Palavras-chave: TripAdvisor, Demanda Turística, Big Data, Planejamento Turístico, pegadas digitais.

INTRODUCTION

Piauí State is the third largest in the Northeast in terms of extension and one of the most biodiverse in Brazil (ICMBIO, 2023). Its territory is made up of a diverse and rich flora, the southeast of the state is part of the Caatinga domain, to the north there is a predominance of coastal vegetation and to the south, the Cerrado. The value and abundance of Piauí's flora can be proven by the presence of transition zones between Caatinga/Amazonia and Caatinga/Cerrado (ICMBIO, 2023).

The southern portion of the state of Piauí holds, in addition to important environmental heritage such as springs, plateaus, mountains and vegetation, one of the largest archaeological collections in the world whose discoveries based on the material remains found changed the entire conception of the first human occupations in the Americas (Guidon, 1995). Archaeological remains can be found from north to south of the state, with important sites in the PARNA region of Sete Cidades (Cavalcante, 2013). Therefore, although the state presents relevant landscape and cultural diversity, Piauí State remains unknown to large part of Brazilian and international tourists (Braga, 2021), even though it is necessary to highlight the article published in early 2022 by The New York Times with recommendation of places for tourism where travelers "can be part of the solution". The article is titled "52 places for a changed world" and mentions Serra da Capivara National Park in position 45, the only Brazilian representative on the list (The New York Times, 2022). It is necessary to highlight that tourist activity, as well as other economic activities, needs to be planned, publicized and practiced in this region in order to value and preserve natural and cultural heritage so as to no longer be a factor of pressure on them.

Natural areas, their potential attraction resources and current tourist attractions in Piauí have been researched, but to date there are no studies using *big data* (large set of data) from social media (user generated content) with a view to understanding aspects of the landscape and territories, through the analysis of travelers' speech about tourist destinations. This set of information associated with innovative tourism techniques can help understand demand, supporting tourism planning, tourism marketing, territorial management focused on tourism, among other associated activities, whether in conservation units or any other type of tourist attraction. The defensible and reproducible criteria applied to the methodology are described in detail in this work.

The work presents numbers regarding visitation at Piauí's National Parks and brings important concepts to understand the work, discussing social media and crowdsourcing. It then presents TripAdvisor justifying the use of information

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n. 1, 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

from this source, and presents the characterization of the National Parks of Piauí, the methodology applied, the results and final considerations. Given this context, the work aims to apply textual data mining techniques, statistical textual analysis and network analysis in a qualitative database with information resulting from comments from social media specialized in tourism with a view to analyzing visitors' speech of Piauí's National Parks to support tourism planning, marketing and territorial management activities. Next, the theoretical foundations that support this work will be presented, starting with the presentation of aspects related to the characterization of social media and crowdsourcing.

SOCIAL MEDIA AND CROWDSOURCING

Social media is understood as a group of internet-based applications that are built on web 2.0 technology and that allow the creation of user-generated content (Kaplan and Haenlein, 2010; Izaias, Soares and Mondo, 2022). Data generated by social media users can be extracted and offer information for various purposes, as well as adding a new perspective to studying the image of the tourist destination (Perinotto, 2013; Morais, 2022). Online publications are composed of associated metadata (Campagna, 2014; Perinotto & Soares, 2019), such as tags, texts, geolocation, among others. In this context, people start to be seen as sensors, contributors and creators of information and are no longer seen just as consumers (Silva & Davis JR., 2008; Goodchild, 2007).

According to Mou *et. al.* (2020) with the increasing popularity of social media sharing platforms and travel websites (where tourists exchange photos and comments while traveling), unique spatial and temporal data on tourist mobility have recently become available. This type of data is commonly referred to as a "digital footprint", or tourists' digital footprints, according to Mou *et. al.* (2020) the "digital footprint" of tourists, as an electronic trace, not

only offers a new way of collecting data on tourist flows, but also provides new research perspectives for tourist mobility research (where they left from and where to they are going). With the rapid development of Information and Communication Technology (ICT) and the increasing popularity of the global ubiquitous network, the use of tourists' digital footprint data has become a dominant orientation in tourism development and research (Perinotto, 2013).

The term *crowdsourcing* originates from the word outsourcing where the word out (third parties, external) is replaced by crowd. The first use of the term was applied to the widespread outsourcing of photographs in the areas of advertising and journalism as a result of increased access to photographic equipment (Howe and Robinson, 2006). According to Borges *et. al.* (2015) "Crowdsourcing is the set of techniques that allow the creation of data sets through the collection and joining of contributions from citizens without prior training or specialization". In this context, the authors point out that information collected on social media makes up a collaboratively constructed database and opens up opportunities for new investigations. Thus, the role of the researcher expands to become a decoder of collective values in the planning decision-making process (Cullen, 1983; Moura, 2014; Borges, 2017).

TripAdvisor user testimonials as a subsidy for the planning and management of tourist destinations is relevant because monitoring post-purchase, or post-consumption, behavior is of great importance for management, organization and marketing professionals (Mondo et. al., 2022). Monitoring this behavior has become more accessible with the use of Information and Communication Technologies. It allows greater agility to monitor customer feedback or user-generated content (UGC). Furthermore, managers can solve problems resulting from service failures and improve their products or services (Ayeh *et. al.*, 2013; Buhalis & Law, 2008; Lu & Stepchenkova, 2015). Feedback expressed by customer-visitors through user-generated content engages other potential consumers and influences the final decisions of other users (Mondo *et. al.*, 2022 and Soares *et.al.*, 2022). The following session

will discuss the defense of using TripAdvisor as a source of information relevant to tourism planning.

TRIPADVISOR *AND* ITS QUALIFICATION FOR APPLICATIONS IN TOURISM PLANNING

According to TripAdvisor's LinkedIn, the site "helps hundreds of millions of people every month become better travelers by supporting them in planning, booking and during their trip." Audiences use the site to discover where to stay, what to do and where to eat based on advice from those who have been there before. Travel guidance is available in 43 markets and 22 languages. (Source: https://www.linkedin.com/company/TripAdvisor/about/ accessed on 03/08/2023).

Regarding strategic planning, TripAdvisor's mission is "to inspire and enable people to explore, experience and share our world with confidence" and the vision is "our commitment to providing a trusted resource for travelers with valuable advice and community insights TripAdvisor, based on the wisdom of crowds. We know that our users use TripAdvisor with common sense and make an informed decision based on the opinions of many." The values are: "We love to travel. We never stop learning. Act like an owner. We are together. We believe in Transparency. Speed wins." Source: https://www.comparably.com/companies/TripAdvisor accessed on 03/08/2023, our translation. In summary, according to Silva *et. al.* (2018) TripAdvisor is a company that uses new processes of mediatization of tourist practices, combined with the complexity of business and tourist communication.

"TripAdvisor, Inc." is the largest travel guidance platform in the world. The relevance attributed by the popularity of accesses justifies the use of the information. According to Perinotto *et. al.* (2023) for an online community to be studied, it needs to have an important segment, many members interacting in

the group, significant and community-oriented comments. The chosen network is important in the context of tourist consumption evaluations, it has more than 60 million members worldwide, 200 million evaluations and consumption evaluations, covering practically all countries in the world, in addition to having more than 315 million of visits per month to your internet address (Perinotto *et. al.*, 2023).

The behavior of TripAdvisor.com.br users can also be observed based on the origin, that is, what is the initial source that takes people to the address. It is observed that organic search is the most relevant, reaffirming its notoriety among users. In other words, this information demonstrates that people access TripAdvisor.com.br intentionally and do not need stimuli such as advertisements (Figure 1).
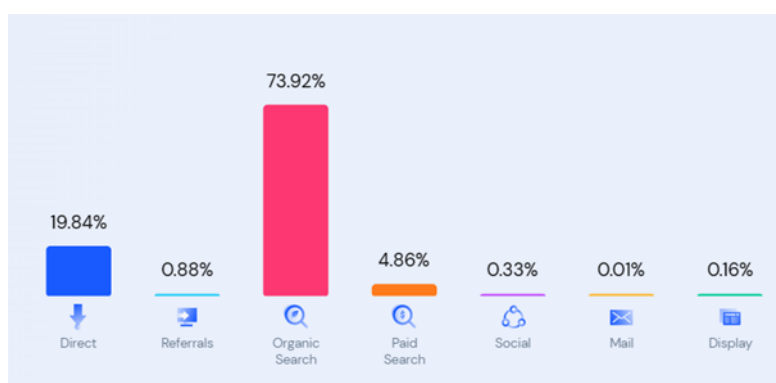

Figure 1: TripAdvisor.com.br traffic sources from left to right: direct, link access, organic, paid search, social media, email, desktop ads. Source: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb accessed on 03/08/2023.

Figure 2 shows other social media that direct people to TripAdvisor.com.br. The online video platform YouTube is observed as the main source followed by the news site and forum with socially curated content and promoted by members through voting, Reddit.
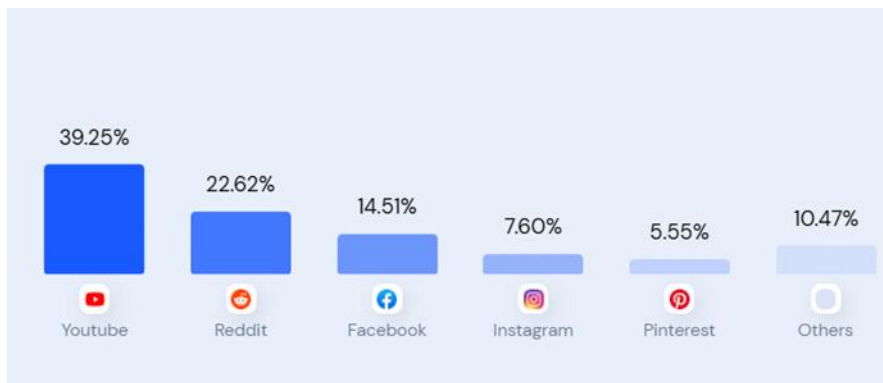
Figure 2: Social Media that sends the most traffic to TripAdvisor.com.br. Source: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb accessed on 03/08/2023.

It is interesting to note that TripAdvisor.com.br is mostly accessed by people in Brazil, but there is emphasis on other countries according to the table presented by Similarweb.com (Figure 3).

Traffic share by country ⓘ
Dec 2022 - Feb 2023 · All traffic

	Country (86)	↓ Traffic Share	Change	Country rank	Visit Duration	Pages / Visit	Bounce Rate
1	🇧🇷 Brazil	93.38%	↓ 23.82%	#155	00:03:25	3.34	55.58%
2	🇵🇹 Portugal	1.93%	↓ 11.31%	#554	00:02:11	7.31	62.53%
3	🇺🇸 United States	1.28%	↓ 19.52%	#19,346	00:03:51	6.43	49.29%
4	🇫🇷 France	0.40%	↑ 6.45%	#8,560	00:03:29	23.44	57.18%
5	🇦🇷 Argentina	0.32%	↓ 23.02%	#6,828	00:01:30	5.49	58.17%
6	🇮🇹 Italy	0.25%	↑ 75.42%	#29,604	00:01:37	4.97	55.27%
7	🇲🇿 Mozambique	0.18%	↓ 57.65%	#1,185	00:04:09	5.95	44.58%
8	🇬🇧 United Kingdom	0.17%	↑ 3.13%	#38,469	00:02:06	7.70	58.63%
9	🇦🇴 Angola	0.15%	↑ 46.24%	#355	00:01:58	5.93	47.40%
10	🇮🇳 India	0.14%	↓ 70.97%	#191,597	00:01:37	1.54	77.65%
11	🇨🇦 Canada	0.14%	↓ 41.38%	#63,886	00:02:22	3.14	62.56%
12	🇨🇭 Switzerland	0.14%	↑ 11.31%	#6,702	00:02:51	10.89	63.23%
13	🇨🇴 Colombia	0.13%	↑ 8.08%	#18,200	00:00:57	1.83	53.77%
14	🇪🇸 Spain	0.13%	↓ 20.62%	#24,813	00:03:10	11.00	49.09%
15	🇲🇽 Mexico	0.12%	↓ 64.88%	#84,926	00:01:45	1.29	74.54%
16	🇩🇪 Germany	0.11%	↓ 2.38%	#71,402	00:01:57	8.04	55.16%
17	🇺🇾 Uruguay	0.10%	↓ 95.55%	#27,421	00:00:58	9.60	77.84%
18	🇯🇵 Japan	0.09%	↑ 104.78%	#126,730	00:01:06	2.71	60.49%

Figure 3: Accesses made to TripAdvisor.com.br. Source: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb accessed on 03/08/2023.

The first column presents the origin of the countries of access, followed by the volume of traffic, the variation in the volume of traffic in the last month, the ranking that TripAdvisor.com.br presents in the country, the average duration of the visit, the number of pages visited and the percentage of people who visit only one page and leave the site (Source: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb accessed on 03/08/2023). Brazil, in this case, has a low percentage of people who visit the website and spend only a short time accessing it.

Despite the broad scope and relevance of TripAdvisor highlighted by the argument presented, and the growing tendency to expand tools like this, as digital literacy advances in the 21st century, there are still researchers, as Soares et.al. (2022) that pointed out, who criticize the use of this data for "representing only a very specific section of tourists, who have the profile of posting comments and reviews on websites about their travel experiences". The following section will present the characterization of the area of the present study.

CHARACTERIZATION OF THE STUDY AREA

The Piauí State has four (Figure 4) strictly protected area units such as National Park category: Sete Cidades, Serra da Capivara, Serra das Confusões and Nascentes do Rio Parnaíba. Below we will present the characteristics of each park.

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n. 1, 2024
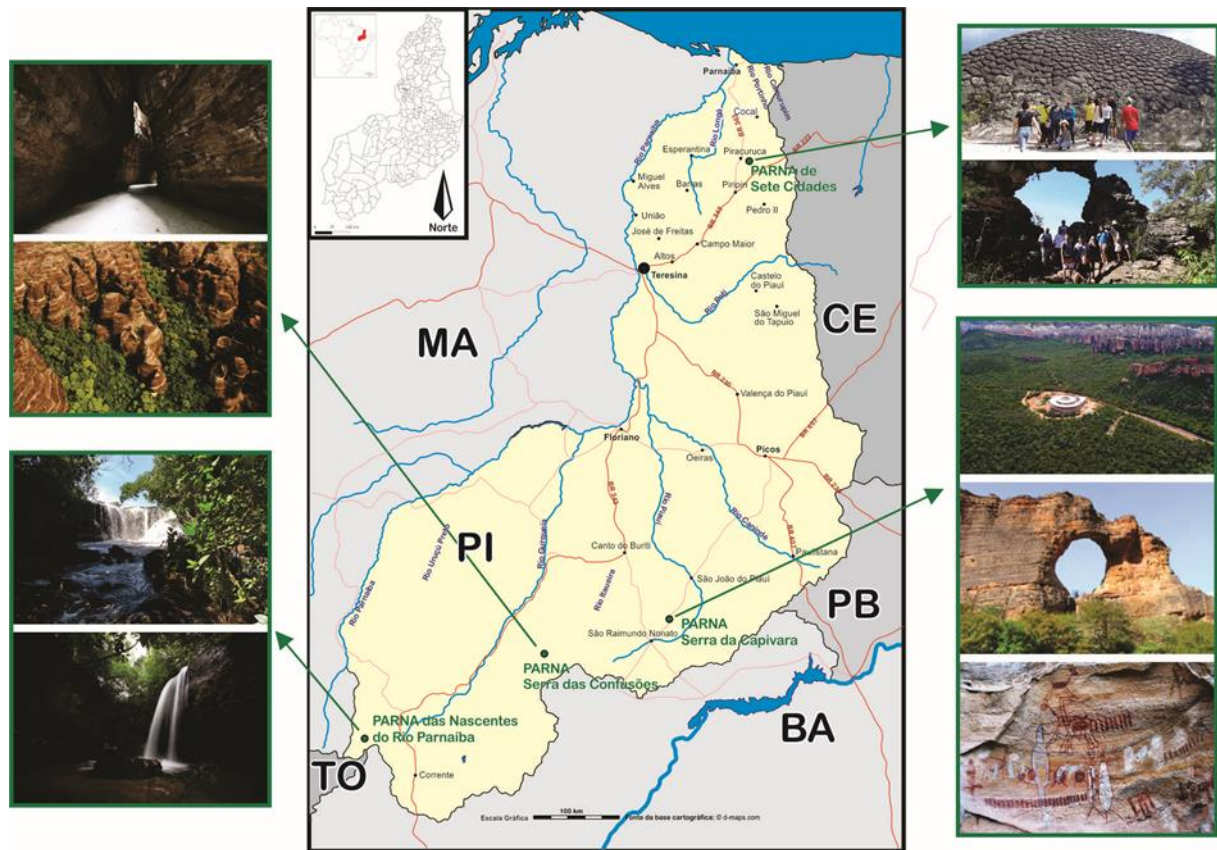**NEECIM TUR** • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

Figure 4: Location of the National Parks of Piauí State and the indication of the surrounding states: MA – Maranhão, CE – Ceará, PB – Paraíba, BA – Bahia, TO – Tocantins.

The Serra da Capivara National Park occupies 130 thousand hectares and is located in the southeastern portion of Piauí (FUMDHAM, 2023) and it occupies part of the municipalities of São Raimundo Nonato, João Costa, Brejo do Piauí and Coronel José Dias. Due to the importance of its archaeological sites, The National Park (PARNA) was inscribed by UNESCO on the World Heritage List on December 13, 1991, and also on the Brazilian Indicative List as mixed heritage (IPHAN, 2023). According to data from FUMDHAM (2023), until 2018, more than a thousand sites with prehistoric cave paintings and engravings were recorded in the park area, data that proves that the area has one of the largest concentrations of prehistoric sites in the world per square kilometer (FUMDHAM, 2023). Currently, the Serra da Capivara PARNA is the

most visited in Piauí. Since the opening of the Nature Museum, the Park has gained more visibility and surpassed Sete Cidades National Park in number of visitors.

The Serra das Confusões National Park is also located in the southeast of Piauí and is relatively close to Serra da Capivara. The municipality of Brejo do Piauí has areas of these two Parks within its territory. The Serra das Confusões PARNA covers 12 municipalities: Brejo do Piauí, Alvorada do Gurguéia, Bom Jesus, Canto do Buriti, Caracol, Cristino Castro, Curimatá, Guaribas, Jurema, Redenção do Gurguéia, Santa Luz and Tamboril do Piauí (ICMBio, 2023). The Serra das Confusões National Park draws attention for its numbers: its area of 823,854.54 hectares is home to the largest Caatinga reserve in Brazil, it is the largest National Park in the Northeast and the sixth largest in area in Brazil. Serra das Confusões is the third most visited park in Piauí.

The Sete Cidades National Park was created on June 8th, 1961, by Federal Decree 50,744 (Braga, *et. al.*, 2022). Its main attractions are "rock formations sculpted by water and wind over 190 million years, with a rich and diverse flora and fauna, recounting prehistory for its visitors" (ICMBio, 2023). When visiting the seven "stone cities" it is possible to identify the marks of civilizations from 6 thousand years ago, recorded in cave paintings (Braga *et. al.*, 2022). In addition to the tour through the sandstone rock formations, there are also two natural pools and a waterfall in the park that can be visited during the rainy season (*Op. cit.*). The park's trails total 12 km and much of it can be covered by car, bicycle or by foot. The Sete Cidades National Park is located in the municipalities of Brasileira and Piracuruca, in Piauí State. It has a demarcated area of 6,221.48 ha and a perimeter of 36.2 km, according to its Management Plan, dated 1979 (ICMBio, 2023). Interestingly, Piauí has one of the largest PARNAs in Brazil, Serra das Confusões, and one of the smallest Brazilian National Park: Sete Cidades. The Sete Cidades National Park is the second most visited in the state.

The Nascentes do Parnaíba National Park is located on the divider of three of the largest Brazilian river basins: the Parnaíba River Basin, the São Francisco River Basin and the Tocantins River Basin (ICMBio, 2023). Located on the border between the states of Piauí, Bahia, Tocantins and Maranhão, the Conservation Unit covers areas of nine municipalities: Alto Parnaíba (MA), Barreiras do Piauí, Corrente, Gilbués and São Gonçalo do Gurguéia (PI), Formosa do Rio Preto (BA), São Félix do Tocantins, Mateiros and Lizarda (TO). Nascentes do Parbaíba National Park was created by Federal Decree on July 16th, 2002 and expanded, also by Federal Decree, on January 12th, 2015, with an area of 749,848 ha (ICMBio, 2023). Currently the Park does not have a management plan and is not officially open for visitors. In addition to the access difficulties, unlike the other National Parks located in Piauí, Nascentes do Parnaíba still does not offer services to assist with visitation, a fact that is reflected in the low visitation and the number of evaluations, both lower in relation to the other analyzed parks.

In Brazil, Conservation Units (CUs) are categorized according to their natural characteristics and the type of management implemented, arising from law 9,985/2000 (BRASIL, 2012), which implemented the National System of Conservation Units (SNUC), and which organizes these areas into two significant groups: full protection CUs and sustainable use CUs, which are subdivided into several categories according to their specific characteristics. SNUC defines criteria and standards for the creation, implementation and management of conservation units. Adventure tourism and ecotourism are the most practiced in Brazilian Conservation Units, especially Full Protection areas such as national parks and ecological stations.

In a diagnosis carried out on the PARNA of Sete Cidades, Santos *et. al.* (2021, p. 105) stated that "the park has a high tourist potential, with natural attractions such as unique geological formations and archaeological sites that include cave paintings" and that these are reasons for the park to have "an increase in activities linked to different segments of tourism, such as geotourism,

sports tourism, study tourism, ecotourism and cultural tourism". It is possible to affirm, based on the attractions of each Conservation Unit, that this scenario extends to the Serra da Capivara and Serra das Confusões Parks.

ICMBio recorded a record number of visits to Brazilian conservation units in 2021. Visits to National Parks reached the mark of 6.9 million people in 2021, compared to 4.4 million in the previous year, 2020 which may have still been influenced of visitation restrictions due to the COVID-19 pandemic. The ten most visited national parks in 2021 represented 65% of total visits among the 74 units currently existing in the category (ICMBio, 2023).

Piauí's CU's followed the national trend of relative increase in visitation between 2020 and 2021. The variations in the number of visitors between the studied national parks, according to data made available by ICMBio (2023), were: in Sete Cidades 15,736 visitors in 2019 ; 3,329 in 2020 and 5,461 in 2021; Serra da Capivara: 29,733 in 2019, 8,486 visitors in 2020 and 12,238 in 2021; Serra das Confusões, the only data available is 2,700 visitors in 2021 and in the PARNA das Nascentes do Rio Parnaíba there is also only visitation data for the year 2021, which had 10 visitors recorded (ICMBio, 2022).

The methodological process applied in this work is presented below.

METHODOLOGICAL PROCESS

The use of information from social media to identify collective values, also called "mass value assessment" has already been widely studied (Roche 2013:26 *apud* Graham and Zook 2011 in: Sui *et.al.* 2013) with a view to provide *insights* and reshape the presentation and understanding of places. The process has already been demonstrated by several authors (Borges, 2017; Borges, Jankowski and Davis Junior, 2015; Borges and Zyngier, 2014; Borges et. al. 2017, Campagna, 2013, Goodchild, 2007, *Haklay*, 2015, Jankowski *et. al.* 2010, among others) using different sources and methodologies, and proving the effectiveness of this type of information (crowdsourced, big data) for

identifying collective values. The aforementioned work addresses the use of information collected from contributions made by citizens to support decision-making associated with both urban planning as planning and management of territory. The work carried out by Jankowski *et. al.* in 2010 about discovering preferences for landmarks in the urban landscape through displacement patterns analyzed through photo posts stands out. By confirming the hypothesis of citizen inclusion in urban planning through the identification of collective values using various tools and data associated with social media, Borges (2017) proved that it is possible to identify the "soul/essence of the place, its unique characteristics, which enables values identification representing affectivity", also called *genius loci*, from the analysis of social media posts. Therefore, to support planning, management and marketing of tourist destinations, the choice of TripAdvisor.com.br is justified, as a platform based on sharing information with and between users, which allows search and review of accommodation, services and tourist attractions.

For this work, all comments available until February 20th, 2023 were collected in the links relating attractions: National Parks in Piauí State territory. The comments' structure available and required by TripAdvisor.com.br is presented by a title that must have a minimum of characters, as well as the review itself, which also requires a minimum of characters to be accepted by the website. The links used are presented in table 1 as follows:

Atrativo	Link	Quantidade de comentários
Parque Nacional Serra da Capivara	https://www.TripAdvisor.com.br/Attraction_Review-g3844605-d318724-Reviews-Serra_da_Capivara_National_Park-Serra_da_Capivara_National_Park_State_of_Piaui.html	357
Parque Nacional das Nascentes do Rio Parnaíba	https://www.TripAdvisor.com.br/Attraction_Review-g3844594-d4465156-Reviews-Parque_Nacional_das_Nascentes_do_Rio_Parnaiba-Alto_Parnaiba_State_of_Maranhao.html	11
Parque Nacional Serra das Confusões	https://www.TripAdvisor.com.br/Attraction_Review-g4481684-d4474629-Reviews-National_Park_of_Serra_das_Confusoes-Caracol_State_of_Piaui.html	48
Parque Nacional Sete Cidades	https://www.TripAdvisor.com.br/Attraction_Review-g303462-d313045-Reviews-Parque_Nacional_de_Sete_Cidades-State_of_Piaui.html	48

Table 1: Attractions, links and number of comments analyzed

It is important to highlight that although there is quantitative supremacy of comments about the Serra da Capivara National Park in relation to the other conservation units, the methodology ensures that this difference is relativized, as the input data was divided into text segments per unit.

After selecting the links, the information left by TripAdvisor.com.br users was extracted using the computer science technique called "web scraping" or "web crawling", which refers to data automatic extraction procedure from using software (Khder, 2021). According to Khder (2021), the process allows structured data extraction in text format, written in computational language. According to Glez-Peña et. al. (2013) web scraping can be defined as the systematically extracting and combining content of interest from the web. Web *scraping* occurs in three steps: accessing the website, analyzing the programming language to extract content and building the output database (Glez-Peña et.al., 2013 and Khder, 2021).

Currently, structures and tools are available on the internet, called API (Application Programming Interface) in which little knowledge of programming

language is required to master them. The API used to extract the data is available at www.octoparse.com (accessed in February 2023).

Data manipulation began with database preparation and organization in to the format requested by IRAMUTEQ software, a text format qualitative data analysis tool, which helps understanding statistical structure phenomena in linguistic data sets according to Zipf's law, defined by prof. George Kingsley Zipf (1902-1950), Harvard University. Iramuteq statistical analysis is carried out with the support of R software. Both must be installed to perform analysis. Both software are also free and available for download on internet.

Iramuteq was developed by the *Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales* (LERASS) at the University of Toulouse, France. Through textural data mining, Iramuteq allows the analysis of the "linguistic body" using tools such as word counting, hierarchical classification, similarity and word cloud. In this work, two techniques were used to support the analysis of the discourse of tourists from Piauí's PARNAs: word cloud and similarity.

The Word Cloud presents words with different sizes, that is, the largest words are those that have greater importance in the textual *corpus*, based on the frequency indicator or other chosen statistical score (Salviati, 2017). Similarity is based on graph theory (Ratinaud and Marchand, 2012), and identifies co-occurrences between words. The result of similarity points to the connection between each of the words, helping visualize general structure. The output data from Iramuteq was taken to Gephi software *(Bastian et. al.*, 2009), which is another free software created for exploring and manipulating networks through the creation of graphs that allow network analysis. Gephi makes it easy to understand networks from dynamic visualizations.

It is important to highlight the difficulty presented in this work to define graph images, since Gephi presents dynamic data, moving, and overlaps information, such as node labels, nodes itself and data sets (clusters). Therefore, dynamic analysis, carried out directly using in the software, is most suitable and

allows one to identify features in the graph, by selection. In this process, the graph "erases" features and only highlights the relationships that are suited for the screen analysis, as presented below. Note that the central word "park" can only be read because other words have been "deleted" (Figure 5).
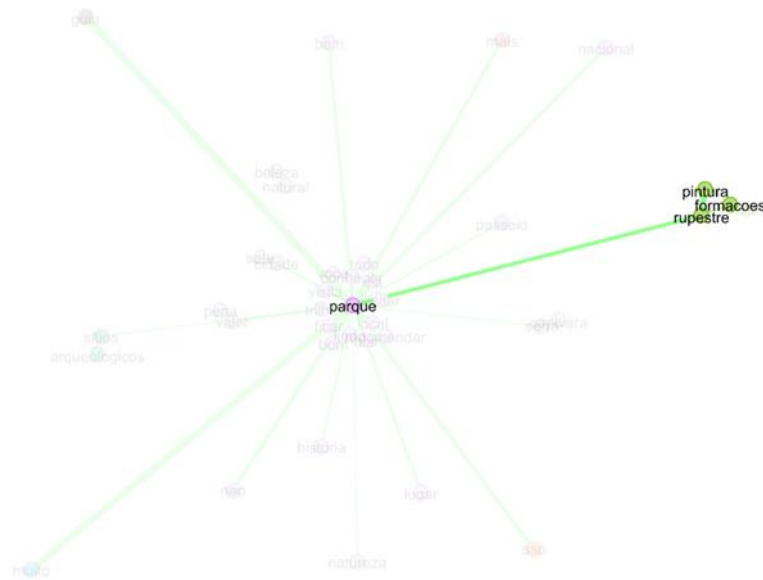

Figure 5: Example of dynamic graph analysis using Gephi.

The comment titles' database contained 2,317 words, which 519 words repeats and 280 words appeared only once. The data referring to comments has 23,629 words, where 2,350 words repeats and 1,105 words appeared once. In order to facilitate the graphs' reading, simplified versions were created, excluding least mentioned words, both in the comments and in the comments' titles. Words that had less than 10 repetitions in the title database were eliminated. In the comments database, words with less than 50 repetitions were eliminated. In order to clean the graph presentation to a static format and improve its visualization words were removed based on visual inspection. Analyzes were carried out in dynamic format and without word cuts.

Thus, the methodological proposal presented is based on network analysis, an emerging scientific field, which examines physical, information, biological, cognitive and social interconnections, and whose purpose is to search for new postulates, algorithms and tools that govern behavior of networks. Therefore, graph visualization supports analysis, extraction data and structure classification of patterns (https://gephi.org/users/publications/ accessed March 2023).

Results of applied methodology described its inspirations and analyzes are presented below.

COMMENTED RESULTS

Textual data mining influx information allowed to observe peculiar aspects about visitors' speeches regarding Piauí's National Parks. Figures 6, 7, 8 presents graphs networks pointing out interconnections, patterns and structures. It is possible to observe words clusters that shows associations, and its connectors, demonstrated by lines between nodes, expressed through its thickness. In other words, the wider it is, the greater is the connection between lexicons.
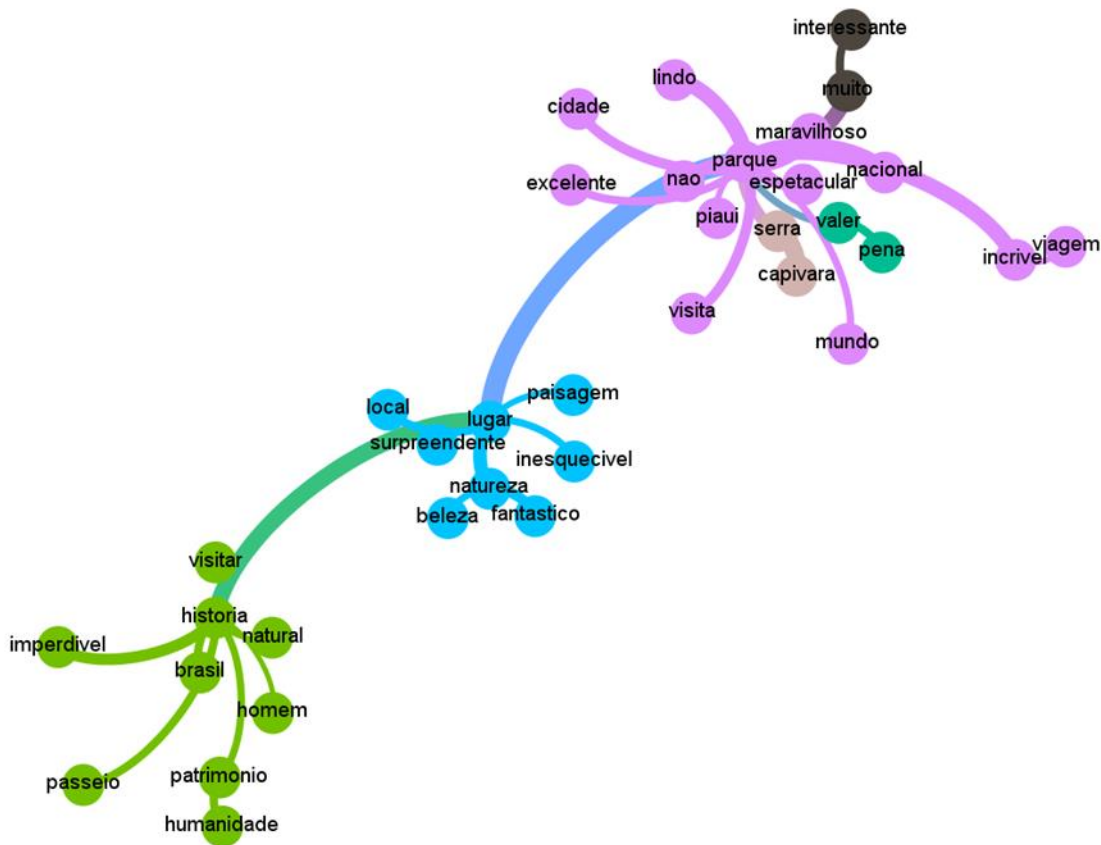
Figure 6: Graph made from comments' titles about Piauí's National Parks available on TripAdvisor. Graph shows words words with more than 10 repetitions.

The comments' title graph (Figure 6) reveals evidence through three clusters: the word "lugar" which means "place"(is a peculiarity among words commented by tourists), 'história' (history) and "patrimônio da humanidade" (human heritage) (reference beyond fauna, flora and waterfalls that one would expect from a national park environment) and positive adjectives linked to the word "parque"(park), denoting not only aspects of remarkable landscape, but also its structure, services and competitive edges from Piauí's National Parks).

It is outstanding that the word "place" emerges as graph's central feature in the comments' titles analysis. Yazigi (2001) discusses aspects that brings uniqueness and denote the term "place" as a reference to a peculiar landscape, spatial identity and personality as opposed to the term "non-

place". For Carlos (2007) "Place is life's reproduction's basis and can be analyzed by the inhabitant - identity - place triad." These references shows that people who talk about Piauí's National Parks on TripAdvisor quality experience.

Another node in the same graph refers to positive adjectives associated with the word park, thus presenting a challenge to marketing to maintain strategies that report visitor perceptions. UGC (User Generated Content) impacts internet users in their consumption choices (Lu & Stepchenkova, 2015; Mondo et al, 2022), appropriating praised characteristics to aid managers and suppliers to maintain and improve services, products, segments and landscapes. When analyzing the comments' title the low, or almost zero, level of negative criticism regarding the attractions analyzed is evident, generating an idea of Piauí's National Parks qualities (Figure 6).

The following graph (Figure 7) presents the comments' complete database. Although this visualization is confusing for reading and understanding, four clusters stand out respectively in the positions: central, northwest, southwest and northeast.
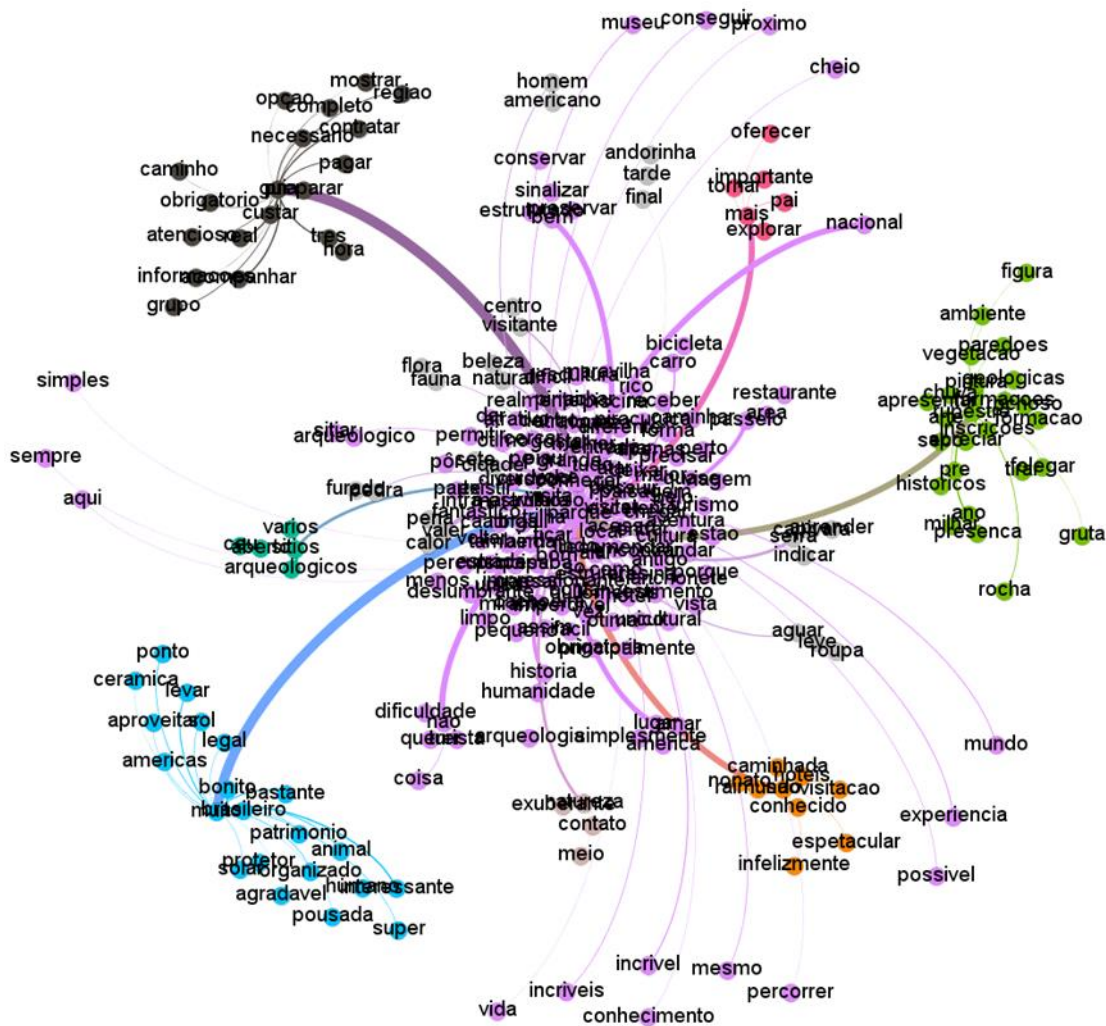
Figure 7: General graph made based on comments about Piauí's National Parks on TripAdvisor, without eliminating words.

Figure 8 shows the same database as Figure 7, but cleaned, which eliminated words with less than 50 repetitions, allowing to visualize the main words of the four clusters highlighted previously in the positions: central, northwest, southwest and northeast. In the central position, although it is not possible to see it, in this visualization, the word "parque" (park) appears as the most prominent. This was also the most cited word in the database, 337 times.
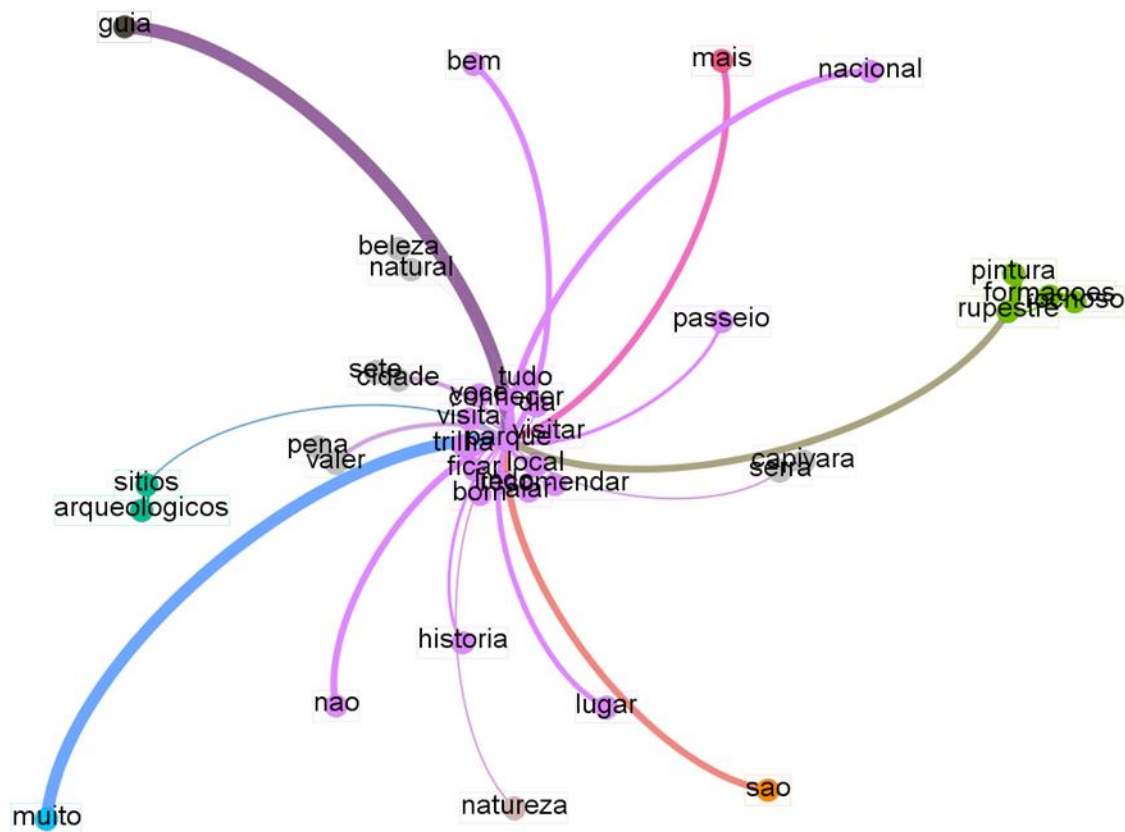
Figure 8: Graph made from comments about Piauí's National Parks on TripAdvisor, eliminating words with less than 50 repetitions.

In the northwest portion (Figures 7 and 8) there is the word "guia" (guide), third most frequent in the database with 210 occurrences. "Guide" is presented as a centrality or node, and the associated words (three, hours, group, show, path, region, hire, information, pay, mandatory, option) show the related discourse. This result is understood as a consistent response to Piauí's National Parks reality, as in the three of them open to visitors: Sete Cidades, Serra da Confusão and Serra da Capivara, visits are carried out accompanied by local operators and guides. The operators, in addition to acting as environmental interpretation agents, also protect the archaeological heritage from possible visitors' vandalism. In a survey carried out with Sete Cidades National Park visitors', Santos *et. al.* (2021) highlighted that 90% rated the experience with

local operators as "great", recommending the importance of guides in presenting the park and its attractions. Thus, within the graph analysis, when highlighting the cluster represented by the word "guia" (guide), it is also important to emphasize the mandatory service of these professionals at Piauí's National Parks, especially for the protection and conservation of rock painting sites, archaeological and geological heritage, competitive advantage amongst those parks (Santos *et. al.*, 2021).

As mentioned, the edges colors also facilitate the word cluster identification, just as the thickness reinforces the strength of the connection between them. In the upper right corner (northeast position) the word "rupestre" (rupestrian) appears associated with painting, formation, geological, vegetation, wall, rocky, cave, environment, pre and historical. This set of words denotes one of the main objectives of Piauí's National Park visits, focused on issues of rock painting sites, geological formations in natural environments that are protected, conserved and have tourism potential for (geotourism, cultural tourism, historical tourism, tourism scientific, educational tourism and perhaps the main one is ecotourism, amongst others). These words aggregate mirrors the main products, services and tourist segments associated with Piauí's National Parks advantage, where archaeological, historical, anthropological, cultural and geological issues qualify products, services and tourist segments (Braga *et. al.*, 2022*)*.

The archaeological sites, which appear prominently in the evaluation discourse, are fundamental for the interpretation of the history of humanity and have records between 12,000 BP, in the Serra da Capivara and Confusões south of Piauí Nationa Park region and remote 6,000 BP, at Sete Cidades National Park, this one located in the north of the State (Braga, 2021). Piauí has more than one thousand and eight hundred archaeological sites registered by the National Historical and Artistic Heritage Institute (IPHAN) and, despite being well evaluated and arousing the interest of visitors, most of the sites still do not have visiting infrastructure, and are unknown from the public (IPHAN, 2021).

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n. 1, 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

In the southwest position of Figures 7 and 8, words associated with positive or superlative adjectives can be seen. However, these terms stand out: "pousada, patrimônio, cerâmica" (inn, heritage, ceramics) as seen in Figure 7. These associated occurrences confirm the positive discourse in relation to visitors' experience to Piauí National Parks and, in addition, expands positive interpretation for accommodation and *souvenirs*. It is necessary to highlight, however, that it would be necessary to carry out an evaluation of the comments and evaluations present on social media specialized in accommodation to confirm this information, as the focus and initial object of analysis is the discourse regarding Piauí's National Parks on TripAdvisor, as demonstrated previously in Table 1, where is presented the comments' origin.

Both in the comments graph (Figures 7 and 8) and in the word cloud (Figure 9), many compliments are highlighted and no negative criticism was observed. This highlights uniqueness amongst Piauí's National Parks and projects future experience (expectation) to those who access the comments generated by users on TripAdvisor to find qualities while practicing tourist activity. It also challenges managers, public and private sector planners, to meet and exceed the expectations of those who access the comments (Mondo *et. al.*, 2021).
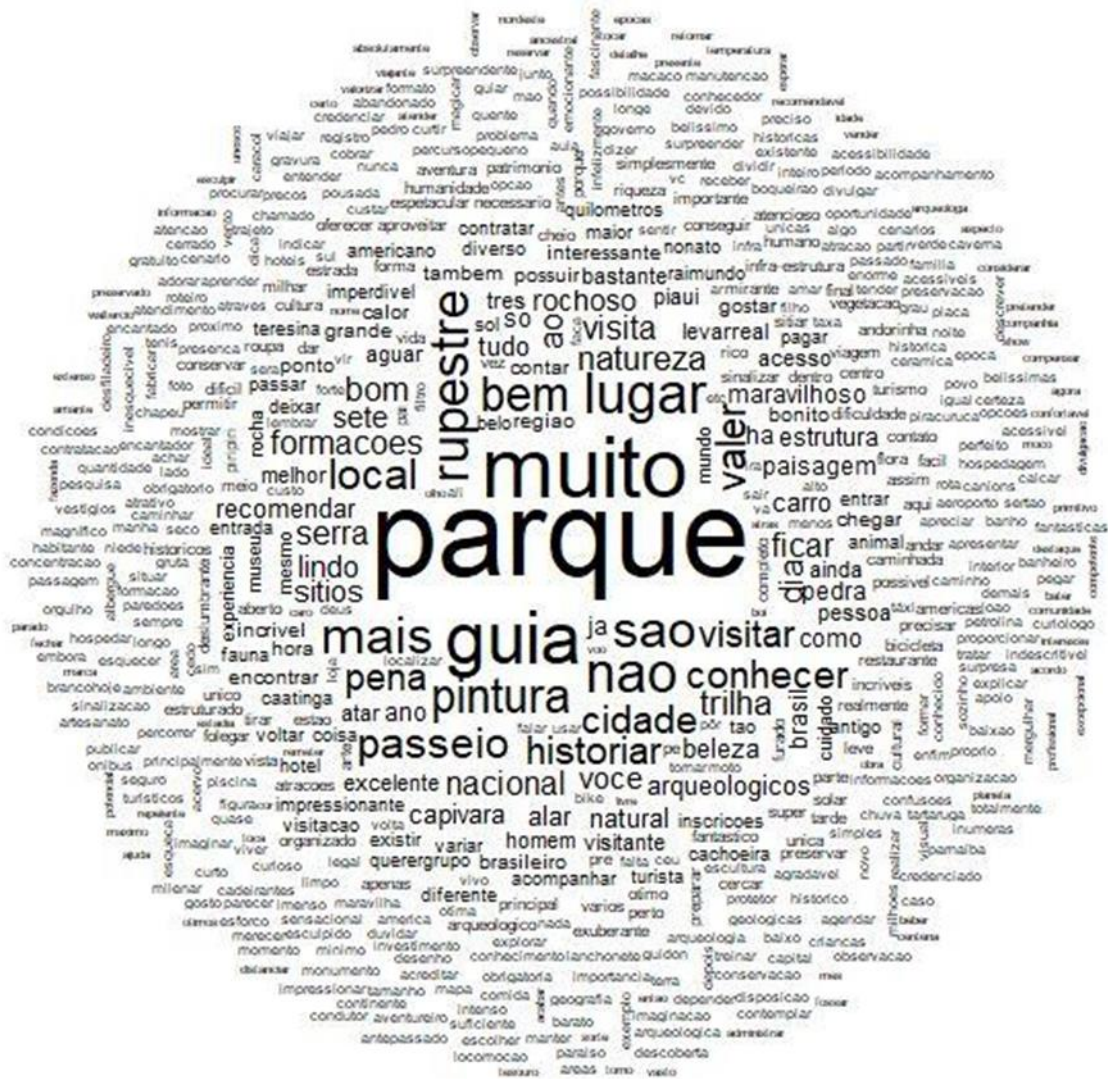
Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n.1 , 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

Figure 9: Word cloud made from comments about Piauí's National Parks on TripAdvisor.

The difficulty to find information about conservation units located in Piauí State was highlighted by an analysis entitled "Tourist potential of conservation units in Piauí" carried out by Ivanov & Lemos (2020), , in addition to the fact that they are not yet fully structured to receive tourists, and lack references on internet prevents potential tourists (and even researchers) from getting information and organizing visits to these places. Despite the difficulties reported by the authors, it did not negatively impact the experience of those who visited the parks, as predominance of positive evaluations was observed.

It is necessary to consider that a large number of reviews were not available on TripAdvisor about the National Parks of Serra das Confusões, Nascentes do Parnaíba and Sete Cidades.

Finally, Ivanov & Lemos (2020) concluded that tourism potential of Conservation Units can be better explored and lack of structure limits tourist access. Despite access difficulties and limited infrastructure around the National Parks Serra das Confusões, Serra da Capivara and Sete Cidades, services and infrastructure are available to serve visitors. When reporting the behavior and evaluation of tourists who visit Serra da Capivara, Mesquita *et. al.* (2015) described that evaluations regarding the National Park are extremely positive and contrast with the negative evaluation of the infrastructure surrounding it.

Another survey about the visiting experience of Sete Cidades National Park, published in 2021, reports that 78% of those interviewed considered the visit to be "great", followed by 19% who considered the visit "good", counting almost 100% of positive impressions and 81% of interviewed stated desire to return to the National Park (Santos *et. al.*, 2021). Observing previous research about Piauí's the National Parks tourists' perception and comparing it with the results demonstrated is observed that the methodological strategy adopted and demonstrated helps planning actions, marketing strategies and management of those conservation units, as well as tourism management at surrounding municipalities.

In relation to Piauí's state tourism, the analyzed evaluations reinforce Braga *et. al.* (2022) observation on great potential and competitiveness of Piauí's natural attractions for the Adventure and Ecotourism segments within national scene. When compared with other states of Brazilian northeast, the authors emphasize that Piauí is not very competitive in the Sun and Beach Tourism market, but can stand out in the adventure, ecotourism and cultural tourism segments and diversify the offer of attractions in northeast region.

The intermediation, disintermediation, reintermediation processes have caused stress in the tourism market and had been causing structural changes as well as directly impacting businesses and the way in which businesses gain market share. Therefore, technological advances, social media and tourism's big techs cause amazement and repulsion among a significant part of academics and market fellows, a natural reaction to changes. However, there is no return to old patterns, it is necessary to understand the new mechanisms and work with them. More and more internet users, who are tourists, use online tools to make decisions about tourism, and in addition, over time, evaluation's observation grows. A product or service potential user wants to know how others felt and perceived the consumption.

Therefore, enabling identification of tourism trends and patterns through the collection of social media data networks can provide valuable information about evolution of travel behavior, popular destinations, emerging attractions and customer preferences. Thus, network and social media data scraping and data analytics aid to collect and analyze tourism related social media "conversations". By observing what internet users are saying about their travel experiences, researchers can understand feelings, interests and travel motivations. Furthermore, it is possible to extract and analyze geolocation data from social media posts to identify popular tourist spots and destinations, as demonstrated by Jankowski *et. al.* (2010). Mapping social media activity can provide insights into tourist flow and crowd patterns. Or, analyze user generated content to gain insights into what attracts and delights tourists to specific destinations or experiences.

FINAL CONSIDERATIONS

Research limitations are recognized, such as the low number of evaluations in some analyzed parks and collection date (until February 2023) imposed restriction. Update and even data collection and its analysis from

other sources, such as Instagram, are necessary to make inferences about the state of tourism in the studied region. The exclusive use of online data from a single source stands out: TripAdvisor, as a deficiency highlighted even by the academic community, as pointed out by Soares *et.al.* (2022).

It was evident from the results of this study that Piauí's National Parks have strong and unique anthropological, archaeological, historical and cultural biases. The presented results confirm the potential already demonstrated by Borges (2017) to identify the essence/soul/ *genius loci* of places as well as collective values from the analysis of data coming from the crowd, "crowdsourced". Public managers, destination planners and service administrators from different tourism segments linked to Piauí's National Parks have different challenges and opportunities from other National Parks in Brazil. The spread between Piauí's National Parks represented by the rock painting sites, where human history is intertwined with the biological and geomorphological aspects demonstrate relevant characteristics for marketing and publicity campaigns rationale. It is necessary to give priority to the essential aspects of Piauí's National Parks (Braga, Guzzi, Perinotto and Malta, 2022) corroborated by travelers' core discourse.

Regional scale and/or segmented similar research can support federal public policies that understand the uniqueness of each set of data and carry out individualized action plans that respect specific characteristics and peculiarities. The strong historical appeal of Piauí's National Parks demands creation and maintenance of museums, implementation of interpretive plans focused on archeology, and visitation monitoring and follow-up to support the specificities of those parks.

TripAdvisor.com.br has a significant amount of access in Brazil, but numbers related to use in the North American territory, TripAdvisor.com (www.similarweb.com/website/TripAdvisor.com/#overview accessed in March 2023), appear to be significantly superior. It is expected that the use of this tool will be expanded throughout Brazil, as internet access and digital

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n.1 , 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

literacy amongst Brazilian population increases. Furthermore, when looking both at access' statistics, possibility of instant translation, and use of the apparatus in 22 different languages, TripAdvisor presents itself as a fundamental tool for tourist activities dissemination in Brazil, when it comes to of foreign tourists. The appliance is capable of increasing security, length of stay and visiting possibilities for ordinary people who would hardly venture into foreign lands if it weren't for this precious information already available online.

This paper methodology has defensible and reproductible criteria and, therefore, has the potential to be applied to other regions, destinations and tourist attractions. It is believed that increasing understanding of visitors' online discourse, perception of tourist attractions, as well as identifying the essence of places will facilitate the work of tourist destinations planners and managers.

The study presents contributions regarding the application of these methods and techniques in the area of tourism as a contribution and subsidy. It advances the knowledge frontier by providing through results and analyzes carried out, an opportunity to understand the demand, contributing to different sectors in the tourism sphere.

The research expanded the understanding regarding user generated content and tourist activity in the attractions that it was proposed to analyze. However, to broaden the the demand understanding for the attractions under analysis, it is recommended that future research focus on verification using other techniques, such as textual, qualitative evaluation of the set of evaluations carried out, and expand the discussion started here about attractions, brands (image) and marketing of Piauí's National Parks. Given this, Mazurek (2018) defended the importance of psychographic factors for creating demand. Because, although the analysis of user generated content promotes real demand perception understanding, a systematic analysis of the psychographic profile should help business owners and its managers to reach content generating consumers.

The qualitative big data database manipulation regarding Piauí's National Parks speech of visitors, collected on social media, using textual data mining, statistical textual analysis and network analysis techniques have potential to support tourism planning, marketing and territorial management decisions. The remotely collected and analysis regarding travelers' speech about tourist destinations, clarified specific characteristics of context, making possible to understand landscape and territories aspects.

Additionally, there are countless opportunities for future research. It is possible to carry out the analysis of each of the parks separately, expanding textual analysis using other functions of Iramuteq software. Understand tourist demand analyzing users' location to understand sending regions to a specific destination. Use of cartographic modeling of attractions' supply location potential and origin of tourist demand to verify territorial interaction potential. Destination facilities and attractions territorial analysis of its hierarchy available in a given territory. Comparative cartographic and textual studies amongst attractions, destinations and tourist regions. Attractions' questions textual analysis.

Tourism trends analysis numerous possibilities, even in real time, are provided by exploring how network and especially social media data can be combined to provide insights. It becomes necessary to analyze social media posts, *hashtags* and discussions (reviews), to identify emerging travel interests.

Future studies may consider analog assessment, carried out through face-to-face interviews in comparison to a digitally made, to confirming or not digital over analog methods and techniques. However, as observed by the research, the methodological strategy adopted corroborates with other researches that adopted analog strategies about Piauí's National Park visitors' perception. Therefore, the presented results can be used for marketing strategies for those conservation units, as most visitors indicated through their comments visiting highlights. These characteristics can be used in advertising campaigns to attract visitors interested in the highlighted aspects: *souvenirs*,

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n.1 , 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

archaeological aspects, amongst others. Likewise, the planning of both conservation units, tourism support equipment and municipalities involved in visiting those parks must be attentive to offering favorable conditions that enhance these elements found.

Work about other attractions can also be carried out with to compare results, confirm and calibrate the technique potential presented in this study as a method of analyzing tourist demand in tourist destinations. Furthermore, comparative research between destinations should be carried out to demonstrate recurring aspects in segment typologies, verifying patterns, data structures, postulates that govern network behavior, mental models and tourists' expectations by segment.

The world wide web big data collected about user interaction on various social media offers possibilities for combining variables, opening up a thriving field of investigation.

The autonomy that information and communication technologies make possible increases the speed of business, favors economic activity, and boosts the tourism sector in an innovative and creative way. To suppliers of products and services, entrepreneurs in the tourism sector, for being able to reach consumers, and for consumers for accessing direct information, offering information from "consumers to consumers".

REFERENCES:

Ayeh, J. K., Au, N., & Law, C. H. R. (2013). Do we believe in TripAdvisor? Examining credibility perceptions and online travelers' attitude toward using user-generated content. Journal of Travel Research, 52(4), 437–452. https://doi.org/10.1177/0047287512475217

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n. 1, 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

Borges, JLC (2017). The Potential of Voluntary Geographic Information to Support Democracy in Territorial Planning and Management. 2017. 227 fl. Thesis (Doctorate) - Postgraduate Program in Architecture and Urbanism, UFMG School of Architecture, Belo Horizonte, 2017.

Borges, J., Jankowski, P.; Davis Junior, C. A. (2015). Crowdsourcing for Geodesign: Opportunities and Challenges for Stakeholder Input in Urban Planning. In: ROBBISLUTER, C.; CRUZ, CBM; MENEZES, PML (Org.). Cartography - Maps Connecting the World. 1.ed. Switzerland: Springer International Publishing, 2015, vol. 1, p. 361-373.

Borges, J., Jankowski, P.; Davis Jr. (2016). A Study On The Use Of Crowdsourced Information For Urban Decision-Making. Brazilian Cartography Magazine. n.68/4, Special Edition Geoinformation and Spatial Analysis, p. 695-703, 2016. https://doi.org/10.14393/rbcv68n4-44276 Available at: <http://www.lsie.unb.br/rbc/index.php/rbc/article/download/1313/953 >. Accessed on: 22 Sep. 2017.

Borges, J.; Zyngier, C. (2014). Crowdsourcing, a Citizen Participation Challenge. Journal of Land Use Mobility and Environment. INPUT 2014.

Braga, SS, de Oliveira, SA, Melo, RS, Nascimento Neto, WJ, & Santos, MR (2022). Reflections on the implementation of a school hostel in the Sete Cidades National Park (PI). Brazilian Ecotourism Magazine (RBEcotur), 15(5). https://doi.org/10.34024/rbecotur.2022.v15.12891

Braga, SS, (2021). Characterization and evaluation of territories and tourist attractions on the coast of Piauí. Thesis (Doctorate) - Federal University of Piauí, Center for Natural Sciences, Regional Postgraduate Program in Development and Environment, Teresina, 2021.

Braga, SS, Guzzi, A., Perinotto, ARC, & Malta, GAP (2022). Analysis of the tourist attractiveness of the coast of Piauí: update of the evaluation of tourist attractions, between 2010 and 2020. Revista Turismo em Análise, v. 33, no. 1, p. 29-49, 2022. https://doi.org/10.11606/issn.1984-4867.v33i1p29-49

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the internet: The state of e-tourism research. Tourism Management, 29(4), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Cavalcante, Luis Carlos Duarte (2013). Sete Cidades National Park, Piauí, Brazil: biodiversity, archeology and conservation of rock art. Mneme-humanities magazine, v. 14, no. 32, 2013.

Campagna, M. (2013). Place, I care! Crowdsourced planning information. AESOP – Association of European Schools of Planning, conference. Dublin, 2013.

Campagna, M. (2014). The Geographic Turn in Social media: Opportunities for Spatial Planning and Geodesign. In: Mugarte, B.; et al. 2014. Computational Science and Its Applications – ICCSA 2014.14th International Conference. Switzerland:Springer International Publishing Switzerland, 2014, p. 598-610.

Carlos, Ana Fani Alessandri.(2007). The place in/of the world. São Paulo: FFLCH, 2007, 85p.

Cullen, G. (1993). Urban landscape. São Paulo: Martins Fontes, 1983.

Daniel Glez-Peña, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato and Florentino Fdez-Riverola. (2013). BRIEFINGS IN BIOINFORMATICS. Web scraping technologies in an API world. VOL 15. NO 5. 788 ^797 doi:10.1093/bib/bbt026 Advance Access published on 30 April 2013

Museum of American Man Foundation – Fumdham. Available at: http://fumdham.org.br/parque/ Accessed on 03/15/2023

Goodchild, M. F., (2007). Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. International Journal of Spatial Data Infrastructures Research, v.2, p.24-32, 2007.

Guidon, N. (1995). Archaeological research in São Raimundo Nonato (Piauí): new data. Revista de Arqueologia, São Paulo, 8(2):37-46, 1994-95.

Marketing & Tourism Review • Belo Horizonte - MG - Brazil • v. 9, n. 1, 2024
NEECIM TUR • Center for Studies and Strategies in Integrated Marketing and Tourism Communication •
UFMG

Haklay, Y. M. (2015). Citizen Science and Policy: An European Perspective. 9Washington, DC: Woodrow Wilson International Center for Scholars, 2015.

Howe, J.; Robinson, M. (2006). The Rise of Crowdsourcing. Available at: <https://www.wired.com/2006/06/crowds/>. Accessed on: 04 Apr. 2017.

ICMBio. Chico Mendes Institute for Biodiversity Conservation. (2022). Federal conservation units reach new visitation record in 2021. Available at: https://www.gov.br/pt-br/noticias/viagens-e-turismo/2022/04/unidades-de-conservacao-federais-atingem -new-record-for-visitation-in-2021 Accessed on 03/22/2023

ICMBio. Chico Mendes Institute for Biodiversity Conservation. (2023). Sete Cidade National Park. Available at: https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-de-sete-cidades /information-about-visitation-2013-parna-de-sete-cidades Accessed on 03/20/2023

ICMBio. Chico Mendes Institute for Biodiversity Conservation. (2023). Sete Cidades National Park. https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-de-sete-cidades Available at: Accessed on 03/20/2023

ICMBio. Chico Mendes Institute for Biodiversity Conservation. (2023). Serra das Confusões National Park. https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-da-serra-das-confusoes Available in: Accessed on 03/20/2023

Institute of National Historical and Artistic Heritage - Iphan. Available at: http://portal.iphan.gov.br/pagina/detalhes/42 Accessed on 03/15/2023

Ivanov, MMM, & Lemos, JR (2020). Conservation units in the state of Piauí. Teresina: EDUFPI.

Izaias, AS, Soares, PF, & Mondo, TS (2022). Assessment of the Quality of Services Perceived by Visitors to the Museum of Tomorrow – Rio de Janeiro. Marketing & Tourism Review, 7(1). https://doi.org/10.29149/mtr.v7i1.7084

Jankowski, P., Adrienko, N. Andrienko, G. Kisilevich, S. (2010). Discovering Landmark Preferences and Movement Patterns from Photo Postings. Transactions in GIS, v.14, n.6, p.833–852, 2010. https://doi.org/10.1111/j.1467-9671.2010.01235.x

Kaplan, AM; Haenlein, M. (2010). Users of the word, unite! The challenges and opportunities of Social Media. Business Horizons, v.53, Issue 1, p.59-68, 2010. Available at: <http://www.sciencedirect.com/science/article/pii/S0007681309001232>. Accessed on: March 21, 2023.

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research model in tourism and hospitality applications: Topics, methods, and software. Journal of Hospitality Marketing & Management, 24(2), 119-154. https://doi.org/10.1080/19368623.2014.907758

Mazurek, M. (2018). Smart management systems (triple helix model) in Waterloo, Canada. Journal of Global Business Insights, 3(2), 12-26. https://doi.org/10.5038/2640-6489.3.2.1033

Moaiad Ahmad Khder. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. Int. J. Advance Soft Compu. Appl, Vol. 13, No. 3, November 2021 Print ISSN: 2710-1274, Online ISSN: 2074-8523 Copyright © Al-Zaytoonah University of Jordan (ZUJ). DOI: 10.15849/IJASCA.211128.11.

Mondo, TS, Perinotto, AR, & Souza-Neto, V. (2022). A user-generated content analysis on the quality of restaurants using the TOURQUAL model. Journal of Global Business Insights, 7(1), 1-15. https://www.doi.org/10.5038/2640-6489.7.1.1172

Morais, PAP (2022). Opinion extraction and sentiment analysis: the image of the tourist destination in the case of the Algarve (Doctoral dissertation). Master's in Marketing Management. University of Algarve.

Mou, Naixia, Zheng, Yunhao; Makkonen, Teemu; Yang, Tengfei; Tang, Jinwen & Song, Yan. (2020). Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China. Tourism Management, Volume 81, 2020, ISSN 0261-5177. https://doi.org/10.1016/j.tourman.2020.104151.

Moura, ACM; Santana, SA (2014). From Authorial Drawings to The Parametric Modeling of Territorial Occupation: Representation and Modeling Influences in The Process of Designing the Urban Space. Brazilian Cartography Magazine, n.66/7, p.451-1463, 2014. https://doi.org/10.14393/rbcv66n0-44739

Mesquita, RF, Sousa, LRM, Matos, FRN, & do Monte, ALA (2015). Tourist behavior of visitors to the Serra da Capivara National Park. Tourism & Management Studies, 11(2), 78-85.

Perinotto, ARC (2013). Investigating tourist communication in Parnaíba/PI-Brazil: Internet and social networks, description and analysis. Turydes: magazine for research into tourism and local development, v. 6, no. 15.

Perinotto, AR C; Soares, JRR (2019). Photographic Image, Credibility, and Consumption of Tourism in the Digital Era. In: / Santos, Jose Duarte; Silva, Oscar Lima (editors). Digital marketing strategies for tourism, hospitality, and airline industries. IGI Global, 2019. DOI: 10.4018/978-1-5225-9783-4.ch005

Perinotto, ARC, Belfort Simões, S., Mesquita de Sousa, S. ., & de Souza Braga, S. (2021). Impact of digital platforms on the supply of accommodation facilities: sales comparison between booking.com and Airbnb. Marketing & Tourism Review, 5(2). https://doi.org/10.29149/mtr.v5i2.5940

Perinotto, ARC; Camarço, JCF; Braga, SDS; Gonçalves, MF (2023). Perceptions on Services in Ceará-Brazil Luxury Hotels Registered on TripAdvisor. J. Glob. Sch. Mark. Sci., 1–20. https://doi.org/10.1080/21639159.2020.1808841

Ratinaud, P.; Marchand, P. (2012). Application of the ALCESTE method to the «gross» corpus and stabilization of «worlds lexicaux»: analysis of «CableGate» with IraMuTeQ. In Actes des 11 eme Journées internationales d'Analyse statistique des Données Textuelles (pp. 835–844). Presented at the 11th Journées internationales d'Analyse statistics des Données Textuelles. JADT

Salviati, M. E. (2017). Iramuteq Application Manual. Version 0.7 Alpha 2 and R Version 3.2.3. Available at (http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati accessed in March 2023)

Santos, MRD, Braga, SDS, Vieira, VB, Araujo, ROP, Nascimento Neto, WJD, & Oliveira, SAD (2020). Tourist at Sete Cidades National Park, Piauí. In. Tourism in protected areas [electronic resource] / Thaise Sutil, Nilzo Ivo Ladwig, José Gustavo Santos da Silva (organizers). - Criciúma, SC: UNESC, 2021.

Silva, Ermelinda Lopes da; Paulo, Eveline Alexandre; Coelho-Costa, Ewerton Reubens; Perinotto, André Riani Costa. (2018). The TripAdvisor Application and Online Complaints Made by Tourists: A View of Gastronomic Enterprises in Fortaleza-CE. PODIUM Sport, Leisure and Tourism Review v. 7, no. 3 (2018) https://doi.org/10.5585/podium.v7i3.214

Silva, JCT; Davis Junior, C. A. (2008). A framework for collecting and filtering voluntarily provided geographic data. In: X Brazilian Symposium on GeoInformatics. Proceedings of the X Brazilian Symposium on GeoInformatics. Porto Alegre (RS): SBC - Brazilian Computing Society, 2008.

Soares, RAM da C., Albuquerque, TV de, Mendes-Filho, L., & Alexandre, ML (2022). Systematic review of Brazilian scientific production on tourism and information and communication technology (ICT). Brazilian Journal of Tourism Research, 16, 2629. https://doi.org/10.7784/rbtur.v16.2629

Sui, Daniel. Elwood, Sarah. Goodchild, Michael. (2013). Crowdsourcing Geographic Knowledge. Dordrecht, Heidelberg, New York and London: Ed: Springer. 2013. 396 p.

Tuan, Y. F. (1980). Topophilia: a study of perception, attitudes and values of the environment. São Paulo: DIFEL, 1980.

Yázigi, Eduardo Abdo. (2001). The soul of the place: tourism, planning and everyday life on coasts and mountains. São Paulo: Context. Accessed on: 21 Mar. 2023.

Websites consulted:

The New York Times. (2022). accessed via www.nytimes.com/interactive/2022/travel/52-places-travel-2022.html - accessed in March 2023

https://gephi.org/users/publications/ - accessed March 2023

www.similarweb.com/pt/website/TripAdvisor.com.br/#overview - accessed in March 2023

www.similarweb.com/website/TripAdvisor.com/#overview - accessed March 2023

www.octoparse.com - accessed February 2023