

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

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

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

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.

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.

INTRODUÇÃO

O Estado do PiauÍ é o terceiro maior do Nordeste em extensão e um dos mais biodiversos do Brasil (ICMBIO, 2023). O seu território é composto por uma flora diversa e rica, o sudeste do estado está inserido no domínio da Caatinga, ao norte há predomínio da vegetação litorânea e ao Sul, do Cerrado. O valor e abundância da flora do PiauÍ podem ser comprovados pela presença de zonas de transição entre Caatinga/Amazônia e Caatinga/Cerrado (ICMBIO, 2023).



A porção sul do estado do Piauí guarda, além de importantes patrimônios ambientais como nascentes, planaltos, serras e vegetação, um dos maiores acervos arqueológicos do mundo cujas descobertas a partir dos vestígios materiais encontrados mudaram toda a concepção sobre as primeiras ocupações humanas nas Américas (Guidon, 1995). Os vestígios arqueológicos podem ser encontrados de norte a sul do estado, com importantes sítios na região do PARNA de Sete Cidades (Cavalcante, 2013). Portanto, embora o estado apresente relevante diversidade paisagística e cultural, o Piauí permanece desconhecido para grande parte dos turistas brasileiros e internacionais (Braga, 2021), ainda que seja necessário fazer destaque à matéria publicada no início de 2022 pelo Jornal The New York Times com recomendação de lugares para turismo onde os viajantes “podem fazer parte da solução”. A matéria é intitulada “52 lugares para um mundo mudado” e cita o Parque Nacional da Serra da Capivara na posição 45, o único representante brasileiro da lista (The New York Times, 2022). Faz-se necessário evidenciar que a atividade turística, assim como as demais atividades econômicas, precisa ser planejada, publicizada e praticada nessa região de forma a valorizar e preservar os patrimônios naturais e culturais para não ser mais um fator de pressão sobre eles.

As áreas naturais, seus potenciais recursos de atração e os atuais atrativos turísticos no Piauí têm sido pesquisados, mas até o presente momento não há estudos com o uso de *big data* (grande conjunto de dados) proveniente de mídias sociais (conteúdo gerado pelos usuários) com vistas a compreender aspectos da paisagem, dos territórios, por meio da análise do discurso dos viajantes sobre os destinos turísticos. Esse conjunto de informações associadas a técnicas inovadoras ao turismo podem auxiliar na compreensão da demanda, apoiando o planejamento turístico, marketing turístico, gestão territorial focada no turismo, entre outras atividades associadas, seja em unidades de conservação ou qualquer outra tipologia de atrativo turístico. Os critérios defensáveis e reproduzíveis aplicados à metodologia são detalhadamente descritos neste trabalho.



O trabalho apresenta números a respeito da visitação dos Parques Nacionais do Piauí e traz conceitos importantes para compreensão do trabalho, discutindo mídias sociais e *crowdsourcing*. Em seguida apresenta o *TripAdvisor* justificando o uso das informações provenientes desta fonte, e expõe caracterização dos Parques Nacionais do Piauí, a metodologia aplicada, os resultados e as considerações finais. Diante desse contexto, o trabalho tem como objetivo aplicar 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 para apoio às atividades de planejamento, marketing e gestão territorial do turismo. A seguir serão apresentados os fundamentos teóricos que subsidiam este trabalho, iniciando pela apresentação dos aspectos relativos à caracterização da mídia social e o *crowdsourcing*.

MÍDIA SOCIAL E O CROWDSOURCING

Por mídia social entende-se um grupo de aplicações baseadas na internet que se constrói sobre a tecnologia web 2.0 e que permite a criação de conteúdo gerado pelos usuários (Kaplan e Haenlein, 2010; Izaias, Soares e Mondo, 2022). Dados gerados pelos usuários das mídias sociais podem ser extraídos e oferecer informações para variados propósitos, bem como acrescentar uma nova perspectiva de se estudar a imagem do destino turístico (Perinotto, 2013; Morais, 2022). As publicações online são compostas por metadados associados (Campagna, 2014 e Perinotto & Soares, 2019), como *tags*, textos, geolocalização, entre outros. Neste contexto, as pessoas passam a ser vistas como sensores, contribuintes e criadores de informações e não são mais vistas apenas como consumidores (Silva & Davis JR., 2008; Goodchild, 2007).



De acordo com Mou *et. al.* (2020) com a crescente popularidade das plataformas de compartilhamento de mídia social e sites de viagens (onde os turistas trocam fotos e comentários durante a viagem), dados espaciais e temporais exclusivos sobre a mobilidade turística tornaram-se recentemente disponíveis. Este tipo de dados é comumente referido como “pegada digital” (*digital footprint*), ou rastros digitais dos turistas, segundo Mou *et. al.* (2020) a “pegada digital” dos turistas, como um traço eletrônico, não só oferece uma nova maneira de coletar dados sobre os fluxos turísticos, mas também fornece novas perspectivas de pesquisa para a pesquisa de mobilidade turística (de onde saíram e para onde estão indo). Com o rápido desenvolvimento da Tecnologia da Informação e Comunicação (TIC) e a crescente popularidade da rede onipresente global, a utilização dos dados da pegada digital dos turistas tornou-se uma orientação dominante no desenvolvimento e pesquisa do turismo (Perinotto, 2013).

O termo *crowdsourcing* tem origem na palavra *outsourcing*, que em inglês significa terceirização. Mas o vocábulo *out* (terceiros, externo) é substituído por *crowd* (multidão). O primeiro uso do termo se deu aplicado a ampla terceirização de fotografias nas áreas de publicidade e propaganda, e jornalismo em consequência da ampliação do acesso aos equipamentos fotográficos (Howe e Robinson, 2006). De acordo com Borges *et. al.* (2015) “*Crowdsourcing* é o conjunto de técnicas que permitem a criação de conjuntos de dados por meio da coleta e junção de contribuições de cidadãos sem treinamento prévio ou especialização”. Neste contexto, os autores apontam que as informações coletadas em mídia social compõem um banco de dados construído de forma colaborativa e abrem oportunidades para novas investigações. Assim, o papel do pesquisador amplia-se para se tornar um decodificador de valores coletivos no processo de tomada de decisões do planejamento (Cullen, 1983; Moura, 2014; Borges, 2017).



Os depoimentos dos usuários do TripAdvisor como subsídio ao planejamento e gestão dos destinos turísticos é relevante porque o monitoramento do comportamento pós-compra, ou pós-consumo, é de grande importância para os profissionais de gestão, organização e marketing (Mondo *et. al.*, 2022). O monitoramento desse comportamento tornou-se mais acessível com o uso das Tecnologias de Informação e Comunicação. Ele permite maior agilidade para monitorar o feedback do cliente ou o conteúdo gerado pelo usuário (CGU). Além disso, os gestores podem resolver problemas resultantes de falhas de serviço e melhorar seus produtos ou serviços (Ayeh *et. al.*, 2013; Buhalis & Law, 2008; Lu & Stepchenkova, 2015). O feedback expresso pelos clientes-visitantes através do conteúdo gerado pelo usuário envolve outros potenciais consumidores e influencia as decisões finais de outros usuários (Mondo *et. al.*, 2022 e Soares *et. al.*, 2022). A sessão seguinte discutirá a defesa do uso do *TripAdvisor* como fonte de informações relevantes ao planejamento turístico.

O TRIPADVISOR E SUA HABILITAÇÃO PARA APLICAÇÕES NO PLANEJAMENTO TURÍSTICO

De acordo com o LinkedIn do TripAdvisor, o site “ajuda centenas de milhões de pessoas todos os meses a se tornarem melhores viajantes, apoiando-os no planejamento, reserva e durante a viagem”. A audiência usa o site para descobrir onde ficar, o que fazer e onde comer com base nas orientações de quem já esteve lá antes. As orientações de viagens estão disponíveis em 43 mercados e 22 idiomas. (Fonte: <https://www.linkedin.com/company/TripAdvisor/about/> acesso em 08/03/2023).

Em relação ao planejamento estratégico, a missão do TripAdvisor é “inspirar e permitir que as pessoas explorem, experimentem e compartilhem nosso mundo com confiança” e a visão é “em nosso compromisso de fornecer um recurso confiável para viajantes com conselhos valiosos e insights da



comunidade TripAdvisor, com base na sabedoria das multidões. Sabemos que nossos usuários usam o TripAdvisor com bom senso e tomam uma decisão fundamentada com base nas opiniões de muitos". Os valores são: "Adoramos viajar. Nós nunca paramos de aprender. Aja como dono. Estamos juntos. Acreditamos na Transparência. Velocidade ganha." Fonte: <https://www.comparably.com/companies/TripAdvisor> acesso em 08/03/2023, tradução nossa. Em resumo, de acordo com Silva *et. al.* (2018) o TripAdvisor é uma empresa que utiliza de novos processos de mediação das práticas turísticas, aliadas à complexidade da comunicação empresarial e turística.

A "TripAdvisor, Inc." é a maior plataforma de orientações de viagem do mundo. A relevância atribuída pela popularidade de acessos que possui justifica o uso das informações. Segundo Perinotto *et. al.* (2023) para que uma comunidade online seja estudada, ela precisa ter um segmento importante, muitos membros interagindo no grupo, comentários significativos e orientados pela comunidade. A rede escolhida é importante no contexto das avaliações de consumo turístico, conta com mais de 60 milhões de membros em todo o mundo, 200 milhões de avaliações e avaliações de consumo, abrangendo praticamente todos os países do mundo, além de ter mais de 315 milhões de visitas por mês em seu endereço na internet (Perinotto *et. al.*, 2023).

O comportamento dos usuários do TripAdvisor.com.br também pode ser observado a partir da origem, ou seja, qual a fonte inicial que leva as pessoas até o endereço. Observa-se que a pesquisa orgânica é a mais relevante, reafirmando a notoriedade entre os usuários. Em outras palavras, essa informação demonstra que as pessoas acessam o TripAdvisor.com.br intencionalmente e não precisam de estímulos como por exemplo propagandas (Figura 1).



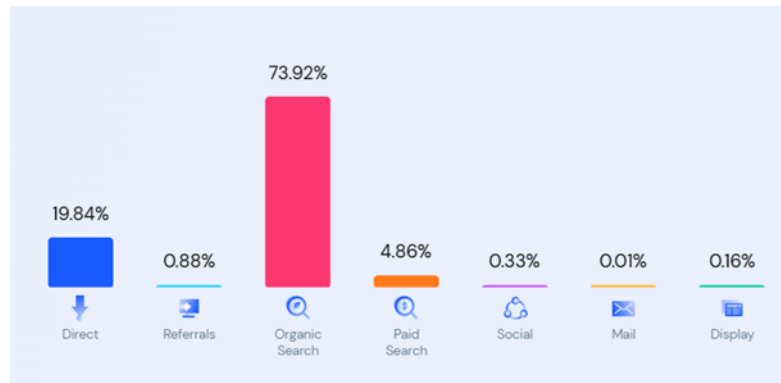


Figura 1: As fontes de tráfego do TripAdvisor.com.br da esquerda para direita: direto, acesso com link, orgânico, pesquisa paga, mídias sociais, email, anúncios no desktop. Fonte: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb acesso em 08/03/2023.

A figura 2 mostra outras mídias sociais que encaminham as pessoas ao TripAdvisor.com.br. Observa-se a plataforma de vídeos online YouTube como a principal origem seguida pelo site de notícias e fórum com conteúdo curado socialmente e promovido pelos membros por meio de votação, Reddit.

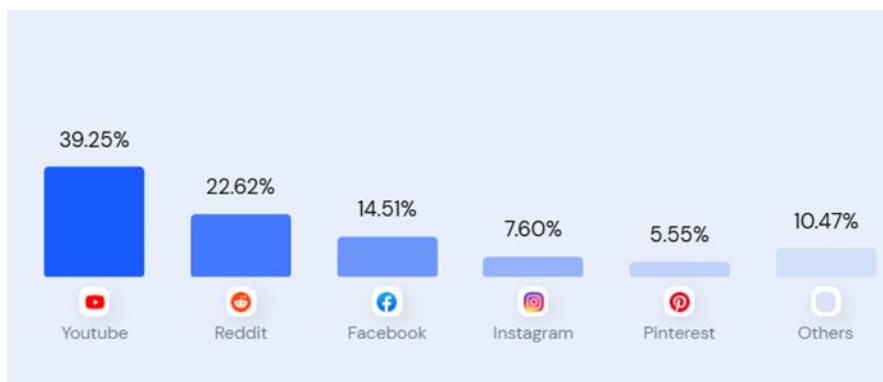


Figura 2: Mídias Sociais que mais encaminham tráfego ao TripAdvisor.com.br. Fonte: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb acesso em 08/03/2023.

Interessante observar que o TripAdvisor.com.br é acessado maioritariamente por pessoas no Brasil, mas há destaque para outros países conforme tabela apresentada pelo Similarweb.com (Figura 3).

Traffic share by country 📅 Dec 2022 - Feb 2023 🌐 All traffic 🔍 +

🔍 Search...

	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%

Figura 3: Acessos realizados ao TripAdvisor.com.br. Fonte: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb acesso em 08/03/2023.

A primeira coluna apresenta a origem dos países do acesso, seguida pelo volume de tráfego, a variação do volume de tráfego no último mês, o ranking que o TripAdvisor.com.br apresenta no país, a duração média da visita, o número de páginas visitadas e a porcentagem de pessoas que visitam somente uma página e saem do site (Fonte: TripAdvisor.com.br Traffic Analytics & Market Share | Similarweb acesso em 08/03/2023). O Brasil neste caso, tem uma porcentagem baixa de pessoas que visitam o site e que ficam pouco tempo em acesso.

Apesar da ampla abrangência e relevância do TripAdvisor apontada pela argumentação apresentada, e crescente tendência de ampliação de ferramentas como esta, conforme avança o letramento digital no século XXI, ainda existem pesquisadores, como apontou Soares *et.al.* (2022), que criticam o uso destes dados por “representarem apenas um recorte muito específico

dos turistas, que possuem o perfil de postarem comentários e avaliações em websites, sobre suas experiências de viagens". A seção seguinte apresentará a caracterização da área do presente estudo.

CARACTERIZAÇÃO DA ÁREA DE ESTUDO

O Estado do Piauí possui um total de quatro (Figura 4) Unidades de Proteção Integral na categoria de Parque Nacional: Sete Cidades, Serra da Capivara, Serra das Confusões e o das Nascentes do Rio Parnaíba. Apresentaremos a seguir as características de cada parque.

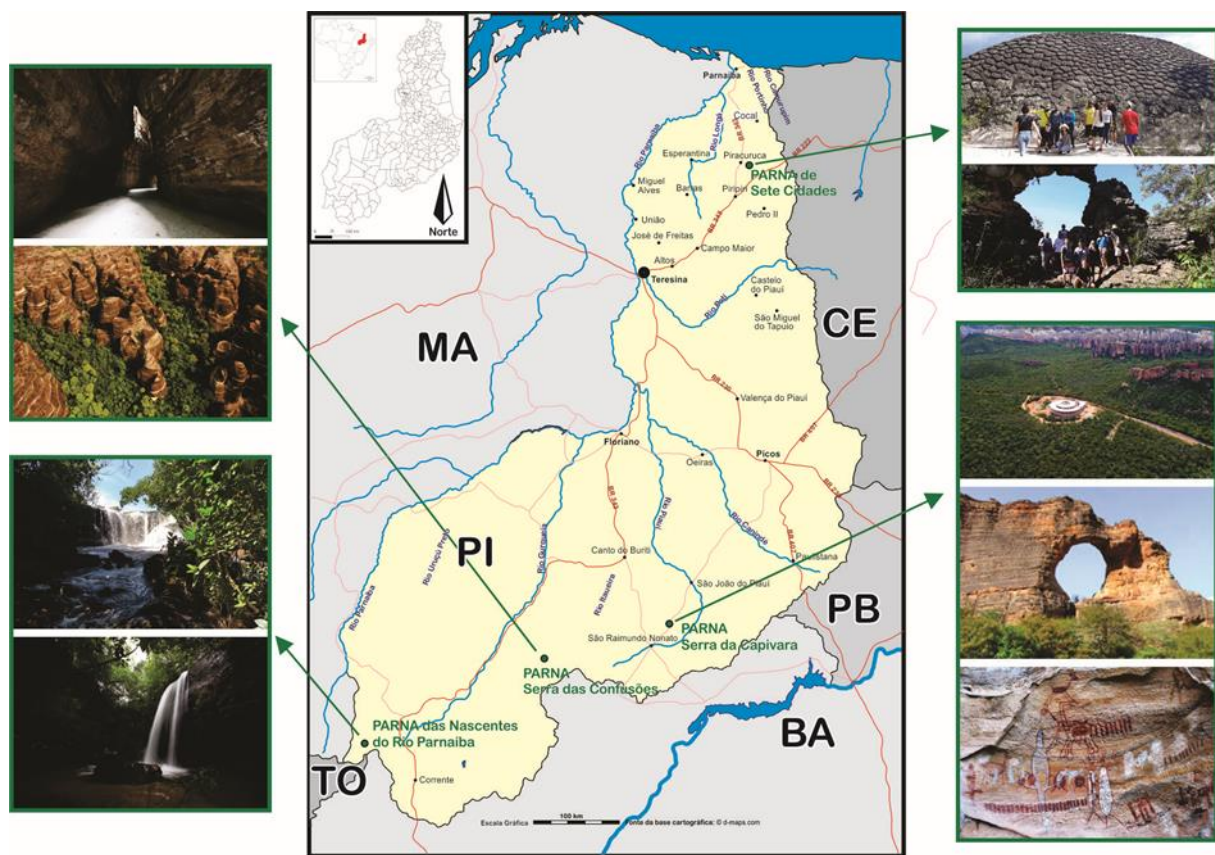


Figura 4: Os Parques Nacionais do Estado do Piauí.

O Parque Nacional Serra da Capivara ocupa 130 mil hectares e está localizado na porção sudeste do Estado do Piauí (FUMDHAM, 2023) e ocupa parte dos municípios de São Raimundo Nonato, João Costa, Brejo do Piauí e Coronel José Dias. Pela importância dos seus sítios arqueológicos, o PARNA foi

inscrito pela Unesco na Lista do Patrimônio Mundial em 13 de dezembro de 1991, e também na Lista Indicativa brasileira como patrimônio misto (IPHAN, 2023). Segundo dados da FUMDHAM (2023) até o ano de 2018 foram registrados mais de mil sítios com pinturas e gravuras rupestres pré-históricas na área do Parque, dados que comprovam que a área possui uma das maiores concentrações de sítios pré-históricos do mundo por quilômetro quadrado (FUMDHAM, 2023). Atualmente o PARNA da Serra da Capivara é o mais visitado do Piauí. Desde a abertura do Museu da Natureza, o Parque ganhou mais visibilidade e ultrapassou o Parque Nacional de Sete Cidades em número de visitantes.

O Parque Nacional da Serra das Confusões também está localizado no Sudeste do Piauí e é relativamente próximo ao PARNA da Serra da Capivara. O município de Brejo do Piauí abriga no seu território áreas desses dois Parques. O PARNA da Serra das Confusões abrange 12 municípios: Brejo do Piauí, Alvorada do Gurguéia, Bom Jesus, Canto do Buriti, Caracol, Cristino Castro, Curimatá, Guaribas, Jurema, Redenção do Gurguéia, Santa Luz e Tamboril do Piauí (ICMBio, 2023). O PARNA da Serra das Confusões chama a atenção pelos seus números: na sua área de 823.854,54 hectares abriga a maior reserva de Caatinga do Brasil, é o maior Parque Nacional do Nordeste e o sexto maior em extensão do Brasil. A Serra das Confusões é o terceiro parque mais visitado no Piauí.

O Parque Nacional de Sete Cidades foi criado em 8 de junho de 1961 pelo Decreto Federal Nº 50.744 (Braga, *et. al.*, 2022). Tem como principais atrativos “formações rochosas esculpidas pela água e pelo vento ao longo de 190 milhões de anos, com uma rica e diversificada flora e fauna e reporta a pré-história para seus visitantes” (ICMBio, 2023). Ao percorrer as sete “cidades de pedra” é possível identificar as marcas de civilizações de 6 mil anos AP, registradas em pinturas rupestres (Braga *et. al.*, 2022). Além do passeio pelas formações rochosas de arenito, há ainda duas piscinas naturais e uma cachoeira do Parque que podem ser visitadas durante o período chuvoso



(*Op. cit.*). As trilhas do parque totalizam 12 km e grande parte pode ser percorrida de carro, bicicleta ou a pé. O PARNA localiza-se nos municípios de Brasileira e Piracuruca, no estado do Piauí. Possui uma área demarcada de 6.221,48 ha e um perímetro de 36,2 km, segundo seu Plano de Manejo, datado de 1979 (ICMBio, 2023). Curiosamente o Piauí tem um dos maiores PARNAs do Brasil, o da Serra das Confusões, e um dos menores parques nacionais brasileiros, o de Sete Cidades. O parque é o segundo mais visitado do estado.

O Parque Nacional Nascentes do Parnaíba está localizado no divisor de três das maiores bacias hidrográficas brasileiras: Bacia do Rio Parnaíba, Bacia do Rio São Francisco e da Bacia do Rio Tocantins (ICMBio, 2023). Localizado na divisa entre os estados do Piauí, Bahia, Tocantins e Maranhão a UC abrange áreas de nove municípios: Alto Parnaíba (MA), Barreiras do Piauí, Corrente, Gilbués e São Gonçalo do Gurguéia (PI), Formosa do Rio Preto (BA), São Félix do Tocantins, Mateiros e Lizarda (TO). O PARNA foi criado por meio de Decreto no dia 16 de julho de 2002 e ampliado, também por Decreto, em 12 de janeiro de 2015, com 749.848 ha de área (ICMBio, 2023). Atualmente o Parque não possui plano de manejo e não está oficialmente aberto para visitação. Além das dificuldades de acesso, ao contrário dos demais Parques Nacionais localizados no Piauí, o Nascentes do Parnaíba ainda não oferece serviços para auxiliar na visitação, fato que se reflete na baixa visitação e nos números de avaliações, ambas menores em relação aos parques analisados.

No Brasil, as Unidades de Conservação (UCs) se categorizam conforme suas características naturais e o tipo de gestão implementado, advindas da lei 9.985/2000 (BRASIL, 2012), que implementou o Sistema Nacional de Unidades de Conservação (SNUC), e que organiza essas áreas em dois grupos significativos: as UCs de proteção integral e as UCs de uso sustentável, as quais se subdividem em várias categorias de acordo com as suas características específicas. O SNUC define critérios e normas para a criação, implantação e gestão das unidades de conservação. O turismo de aventura e ecoturismo são os mais praticados nas Unidades de Conservação Brasileiras,



especialmente as Proteção integral como os parques nacionais e as estações ecológicas.

Em um diagnóstico realizado sobre o PARNA de Sete Cidades, Santos *et al.* (2021, p. 105) afirmaram que “o parque apresenta um alto potencial turístico, com atrativos naturais como formações geológicas singulares e sítios arqueológicos que contemplam pinturas rupestres” e que estes são motivos para o parque ter “um incremento de atividades vinculadas a diferentes segmentos do turismo, como o geoturismo, o turismo esportivo, o turismo de estudos, o ecoturismo e o turismo cultural”. É possível afirmar, com base nos atrativos de cada Unidade de Conservação, que este cenário se estende aos Parques da Serra da Capivara e da Serra das Confusões.

O ICMBio contabilizou um recorde nos números de visitação nas unidades de conservação brasileiras em 2021. As visitas aos PARNAs atingiram a marca de 6,9 milhões de pessoas em 2021, contra 4,4 milhões no ano anterior, 2020 que pode ter ainda sofrido influência de restrições de visitação por causa da pandemia de COVID-19. Os dez parques nacionais mais visitados em 2021, representaram 65% do total de visitas entre as 74 unidades atualmente existentes na categoria (ICMBio, 2023).

As UCs do Piauí acompanharam a tendência nacional de aumento relativo da visitação entre os anos de 2020 e 2021. As variações do número de visitantes entre os PARNAs estudados, conforme dados disponibilizados pelo ICMBio (2023), foi: no Sete Cidades 15.736 visitantes em 2019; 3.329 em 2020 e 5.461 em 2021; Serra da Capivara: 29.733 em 2019, 8.486 visitantes em 2020 e 12.238 em 2021; Serra das Confusões o único dado disponível é de 2.700 visitantes em 2021 e no PARNA das Nascentes do Rio Parnaíba também só existem dados de visitação para o ano de 2021 que teve 10 visitantes contabilizados (ICMBio, 2022). A seguir apresenta-se o processo metodológico aplicado neste trabalho.

PROCESSO METODOLÓGICO



O uso de informações provenientes de mídias sociais para identificação de valores coletivos, também chamado de “avaliação de valor em massa” já vem sendo amplamente estudado (Roche 2013:26 *apud* Graham and Zook 2011 in: Sui *et.al.* 2013) com vistas a fornecer *insights* e remodelar a apresentação e compreensão dos lugares. O processo já foi demonstrado por diversos autores (Borges, 2017; Borges, Jankowski e Davis Junior, 2015; Borges e Zyngier, 2014; Borges *et. al.* 2017, Campagna, 2013, Goodchild, 2007, Haklay, 2015, Jankowski *et. al.* 2010, entre outros) utilizando-se fontes e metodologias diversas, e comprovando a eficácia dessa tipologia de informações (*crowdsourced, big data*) para identificação de valores coletivos. Os trabalhos supracitados abordam o uso de informações coletadas a partir de contribuições feitas por cidadãos para o embasamento da tomada de decisão associada tanto ao planejamento urbano quanto ao planejamento e gestão do território. Destaca-se o trabalho realizado por Jankowski *et. al.* em 2010 sobre a descoberta de preferências por marcos na paisagem urbana através de padrões de deslocamento analisados através de postagens de fotos. Ao confirmar a hipótese de inclusão do cidadão no planejamento urbano através da identificação de valores coletivos usando diversas ferramentas e dados associados às mídias sociais, Borges (2017) descobre que a partir da análise de postagens em mídia social é possível identificar a “alma/essência do lugar, ou seja, características que dão singularidade, possibilitando a identificação dos valores que representam essa afetividade”, também denominado *genius loci*. Assim, para o suporte ao planejamento, gestão e marketing de destinos turísticos se justifica a escolha do TripAdvisor.com.br, uma plataforma baseada em compartilhamento de informações com e entre usuários, que permite buscas e avaliações sobre hospedagem, serviços e atrativos turísticos.

Para este trabalho foram coletados todos os comentários disponíveis até o dia 20 de fevereiro de 2023 nos *links* relativos aos atrativos: Parques Nacionais no território do Estado do Piauí. A estrutura dos comentários disponíveis e exigidos pelo TripAdvisor.com.br apresenta-se em formato de



título que deve ter um número mínimo de caracteres, assim como a avaliação propriamente dita, que também necessita um número mínimo de caracteres para ser aceita pelo *website*. Os *links* utilizados são apresentados na tabela 1 a seguir:

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

Tabela 1: Atrativos e links analisados

É importante destacar que embora haja supremacia quantitativa de comentários sobre o Parque Nacional da Serra da Capivara em relação às demais unidades de conservação, a metodologia de tratamento garante que essa diferença seja relativizada, pois os dados de entrada foram divididos em segmentos de textos por unidade.

Após a seleção dos *links*, foi feita extração das informações deixadas pelos usuários do TripAdvisor.com.br utilizando-se a técnica da área da ciência da computação denominada “*web scraping*” ou “*web crawling*” que se refere ao procedimento de extração automática de dados de sites usando *software* (Khder, 2021). De acordo com Khder (2021) o processo permite extração de dados estruturados em formato de texto, escritos em linguagem computacional. De acordo com Glez-Peña et. al. (2013) o *web scraping* pode ser definido como processo de extração e combinação de conteúdos de



interesse na rede de forma sistemática. O *web scraping* ocorre em três passos: acesso ao site, análise da linguagem de programação para extração de conteúdo e construção do banco de dados de saída (Glez-Peña et.al., 2013 e Khder, 2021).

Atualmente estão disponíveis estruturas e ferramentas na internet, denominadas API (*Application Programming Interface*) nas quais são necessários poucos conhecimentos em linguagem de programação para seu domínio. Utilizou-se a API disponível em www.octoparse.com (acesso em fevereiro de 2023) para a extração dos dados.

A manipulação dos dados iniciou a partir da preparação e organização do banco de dados no formato solicitado para a entrada no *software* de texto IRAMUTEQ, uma ferramenta de análise de dados qualitativos em formato de texto, e que auxilia na compreensão de fenômenos da estrutura estatística em conjuntos de dados linguísticos, de acordo com a lei de Zipf, definida pelo prof. George Kingsley Zipf (1902-1950), da Universidade de Harvard. A análise estatística no Iramuteq é feita com apoio do *software* R. Ambos os *softwares* são livres e disponíveis na internet para *download* gratuito e devem estar instalados na máquina para performance das análises.

O Iramuteq foi desenvolvido pelo *Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales* (LERASS) da Universidade de Toulouse, na França. Através da mineração de dados textuais o Iramuteq permite a análise do “corpo linguístico” utilizando-se de ferramentas como a contagem de palavras, classificação hierárquica, similitude e nuvem de palavras. Neste trabalho, foram utilizadas duas técnicas de apoio à análise do discurso dos turistas dos PARNAs do Piauí: nuvem de palavras e similitude.

A Nuvem de Palavras apresenta palavras com tamanhos diferentes, ou seja, as palavras maiores são aquelas que detêm maior importância no *corpus* textual, a partir do indicador de frequência ou outro escore estatístico escolhido (Salviati, 2017). A similitude se baseia na teoria dos grafos (Ratinaud e Marchand, 2012), e identifica as co-ocorrências entre as palavras. O



Figura 5: Exemplo de análise dinâmica do grafo no Gephi.

O banco de dados a respeito dos títulos dos comentários continha 2.317 palavras, sendo 519 palavras que se repetem e 280 palavras que aparecem somente uma única vez. Já os dados referentes aos comentários apresentaram 23.629 palavras, sendo 2.350 palavras que se repetiram e 1.105 palavras que só apareceram uma vez. No sentido de facilitar a leitura dos grafos, foram elaboradas versões simplificadas, excluindo-se as palavras menos citadas, tanto nos comentários como nos títulos dos comentários. Foram eliminadas as palavras que tiveram menos de 10 repetições no banco de dados referente aos títulos. Já no banco de dados relativo aos comentários foram eliminadas as palavras com menos de 50 repetições. A quantidade de palavras retiradas baseou-se em inspeção visual com vistas a melhoria da visualização dos grafos, de modo a oferecer limpeza do grafo para apresentação em formato estático, embora as análises tenham sido realizadas em formato dinâmico e sem cortes de palavras.

Assim, a proposta metodológica que se apresenta é baseada na análise das redes, um campo científico emergente, que examina as interconexões físicas, de informação, biológicas, cognitivas e sociais, e cujo propósito é buscar novos postulados, algoritmos e ferramentas que governam o comportamento das redes. Portanto, a visualização dos grafos apoia as análises, extração e classificação de padrões em estruturas e dados (<https://gephi.org/users/publications/> acesso em março de 2023). A seguir são apresentados os resultados da aplicação da metodologia descrita e as inspirações e análises.

RESULTADOS COMENTADOS

O influxo de informações oriundas da mineração de dados textuais permitiu observar aspectos peculiares acerca do discurso dos visitantes dos

PARNAs do Piauí. As Figuras 6, 7, 8 apresentam os grafos das redes que apontam as interconexões e indicam padrões e estruturas, nelas pode-se observar aglomerados de palavras que mostram a associação dos vocábulos, e os conectores, ou linhas entre os nós, que demonstram a conexão, que é expressada através da espessura. Ou seja, quanto mais largo, maior a conexão entre os léxicos.

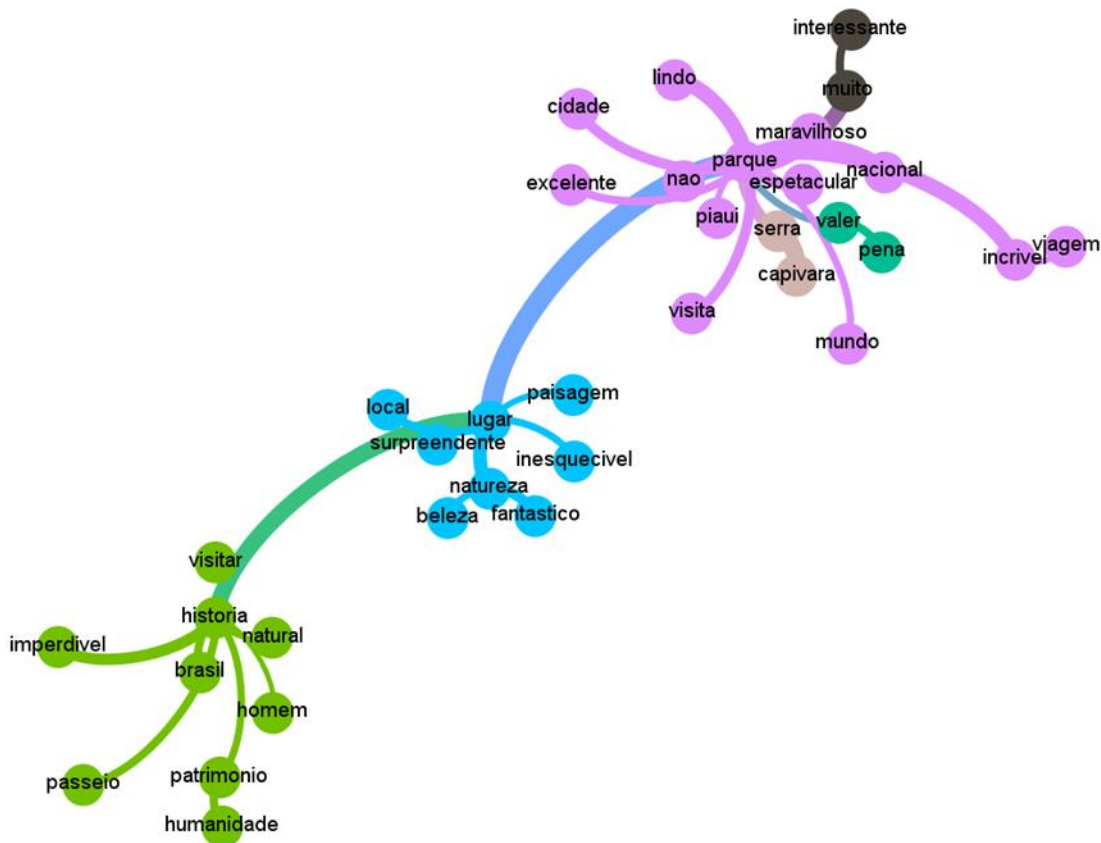


Figura 6: Grafo feito a partir dos títulos dos comentários sobre os parques nacionais do Piauí no TripAdvisor, eliminando-se as palavras com menos de 10 repetições.

O grafo de títulos dos comentários (Figura 6) revela evidências através de três aglomerados: a palavra “lugar” (peculiaridade entre as palavras comentadas por turistas), “história” e “patrimônio da humanidade” (referência para além da fauna, flora e cachoeiras que se espera em um parque nacional) e adjetivos positivos ligados à palavra “parque” (denotando

não somente os aspectos da paisagem notável, mas de estrutura, serviços e diferenciais dos PARNAs do Piauí).

É notável o surgimento da palavra “lugar” como centralidade no grafo que representa os títulos dos comentários. Yazigi (2001) discorre sobre os aspectos que trazem singularidade e denotam o termo “lugar” como referência àquele que tem paisagem peculiar, identidade espacial e personalidade em oposição ao termo “não-lugar”. Para Carlos (2007) “O lugar é a base da reprodução da vida e pode ser analisado pela tríade habitante - identidade - lugar.” Essas referências nos remetem à qualidade da experiência das pessoas que discursam sobre os PARNAs do Piauí no TripAdvisor.

Outro nó do mesmo grafo se reporta a adjetivos positivos associados a palavra parque, deste modo apresenta um desafio ao marketing para manutenção de estratégias que reportam a percepções de visitantes. O CGU (Comentário Gerado pelo Usuário) impacta nos usuários da internet para escolhas de consumo (Lu & Stepchenkova, 2015; Mondo et al, 2022), assim como se apropriar do que é elogiado por parte dos gestores e fornecedores é valioso para manter e melhorar serviços, produtos, segmentos e paisagens. Fica evidente o baixo, ou quase nulo, ponto de críticas negativas sobre os atrativos analisados, gerando na análise dos títulos dos comentários uma noção das qualidades dos PARNAS do Piauí (Figura 6).

O grafo a seguir (Figura 7) apresenta o banco de dados completo acerca dos comentários. Embora essa visualização seja confusa para leitura e entendimento, destacam-se respectivamente quatro aglomerados nas posições: central, noroeste, sudoeste e nordeste.



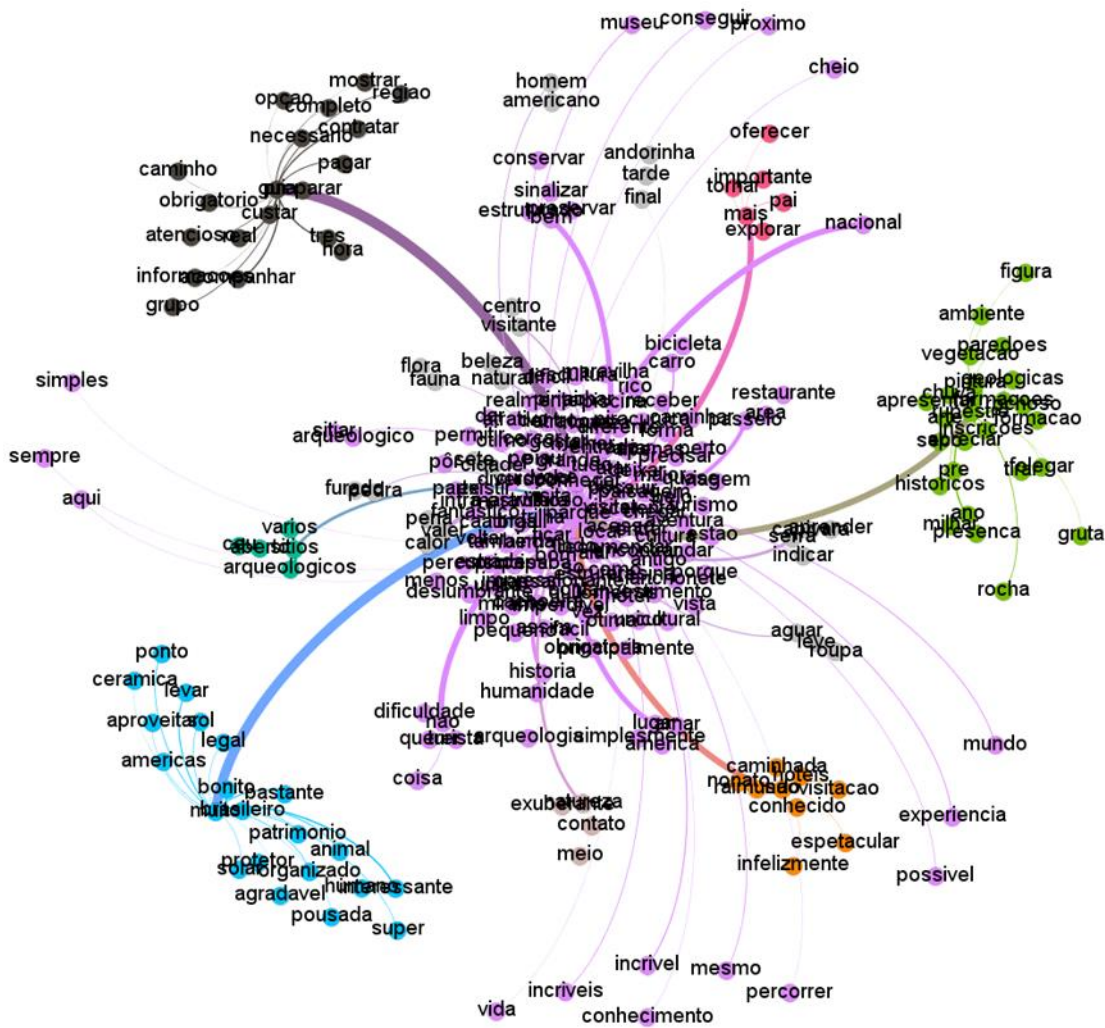


Figura 7: Grafo geral feito a partir dos comentários sobre os parques nacionais do Piauí no TripAdvisor, sem eliminar palavras.

A Figura 8 expõe o mesmo banco de dados da Figura 7, porém com limpeza, que eliminou as palavras com menos de 50 repetições, permitindo visualizar as palavras principais dos quatro aglomerados destacados anteriormente nas posições: central, noroeste, sudoeste e nordeste. Na posição central, embora não seja possível ver, nesta visualização, a palavra



“parque” aparece como a de maior destaque. Essa também foi a palavra mais citada no banco de dados, 337 vezes.

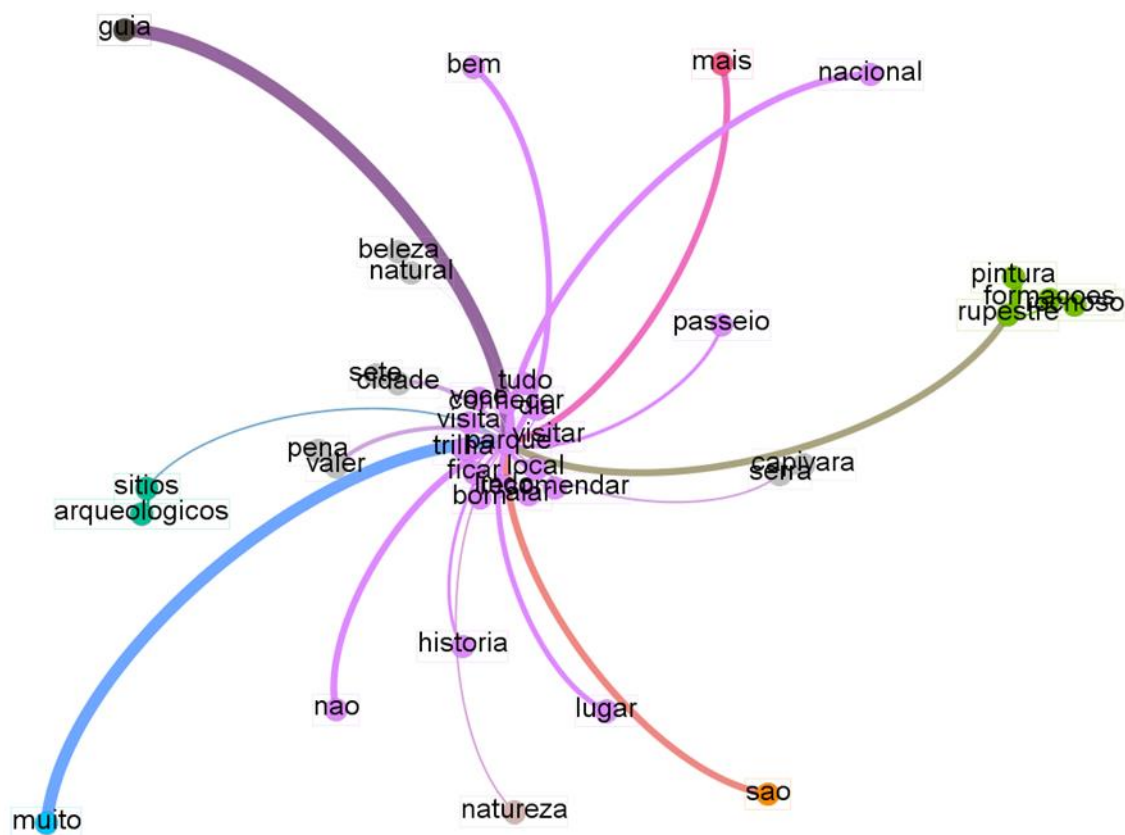


Figura 8: Grafo feito a partir dos comentários sobre os parques nacionais do Piauí no TripAdvisor, eliminando-se as palavras com menos de 50 repetições.

Na porção noroeste (Figuras 7 e 8) há a palavra “guia”, terceira mais frequente no banco de dados com 210 ocorrências. “Guia” se apresenta como uma centralidade ou nó, e as palavras (três, horas, grupo, mostrar, caminho, região, contratar, informações, pagar, obrigatório, opção) associadas mostram o discurso relacionado à palavra. Entende-se este resultado como uma resposta coerente com a realidade dos PARNAs piauienses, pois nos três abertos para a visitaç o, Sete Cidades, Serra da Confus o e Serra da Capivara as visitas s o realizadas com os acompanhamentos de condutores e guias locais. Os condutores, al m de

atuarem como agentes de interpretação ambiental também protegem o patrimônio arqueológico de possíveis ações de vandalismo dos visitantes. Em pesquisa realizada com visitantes no Parque Nacional de Sete Cidades, Santos *et. al.* (2021) destacaram que 90% avaliaram como “ótima” a experiência com os condutores locais, ressaltando a importância dos guias na apresentação do parque e de seus atrativos. Assim, dentro da análise do grafo, ao se destacar o aglomerado representado pela palavra “guia” também é importante enfatizar a obrigatoriedade do serviço destes profissionais nos PARNAs do Piauí, em especial para proteção e conservação da arte rupestre e do patrimônio arqueológico e geológico, diferencial destes parques (Santos *et. al.*, 2021).

Como mencionado, as cores das arestas também facilitam a identificação desses grupos ou clusters de palavras, assim como as espessuras reforçam a força da conexão entre elas. No canto superior direito (posição nordeste) a palavra “rupestre” aparece associada a pintura, formação, geológica, vegetação, paredão, rochoso, gruta, ambiente, pré e histórico. Este conjunto de palavras denota um dos principais objetivos das visitas aos PARNAs do Piauí, que são voltados para questões da arte-rupestre e das formações geológicas em ambientes naturais protegidos, conservados e com potencial turístico para (geoturismo, turismo cultural, turismo histórico, turismo científico, turismo pedagógico e talvez o principal seja o ecoturismo, entre outros). Este agregado de palavras espelha o principal diferencial dos produtos, serviços e segmentos turísticos associados aos PARNAs do Piauí, onde questões arqueológicas, históricas, antropológicas, culturais e geológicas qualificam produtos, serviços e segmentos turísticos (Braga *et. al.*, 2022).

Os sítios arqueológicos, que aparecem com destaque no discurso das avaliações, são fundamentais para a interpretação da história da humanidade e possuem registros entre 12.000 AP, na região dos PARNAs Serra da Capivara e das Confusões ao sul do Piauí e 6.000 AP, PARNA de Sete



Cidades, esse localizado no norte do Estado (Braga, 2021). O Piauí possui mais de mil e oitocentos sítios arqueológicos registrados pelo Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) e, apesar de bem avaliados e de despertar o interesse dos visitantes, a maioria dos sítios ainda não possui infraestrutura de visitação, e é desconhecida do público (IPHAN, 2021).

Na posição sudoeste das Figuras 7 e 8 observam-se palavras associadas a adjetivos positivos ou superlativos. Porém, destacam-se os termos: pousada, patrimônio, cerâmica como pode ser observado na Figura 7. Essas ocorrências associadas confirmam o discurso positivo em relação à experiência dos visitantes dos PARNAs do Piauí e, além disso, ampliam essa interpretação positiva para a hospedagem e os *souvenirs*. É preciso destacar, entretanto, que seria necessário realizar avaliação sobre os comentários e avaliações presentes em uma mídia social especializada em hospedagem para confirmar essa informação, pois o foco e objeto inicial de análise é o discurso sobre os PARNAs do Piauí no TripAdvisor, conforme demonstrado anteriormente na Tabela 1, que apresenta a origem dos comentários.

Tanto no grafo dos comentários (Figuras 7 e 8) quanto na nuvem de palavras (Figura 9), muitos elogios são apontados e não foi observada crítica negativa. O que evidencia um diferencial dos PARNAs do Piauí e cria uma projeção de experiência futura (expectativa) para quem acessa os comentários gerados pelos usuários no TripAdvisor de encontrar qualidades na atividade turística. Isso também cria um desafio aos gestores, planejadores do setor público e privado, para atender e superar as expectativas de quem acessa os comentários (Mondo *et. al.*, 2021).

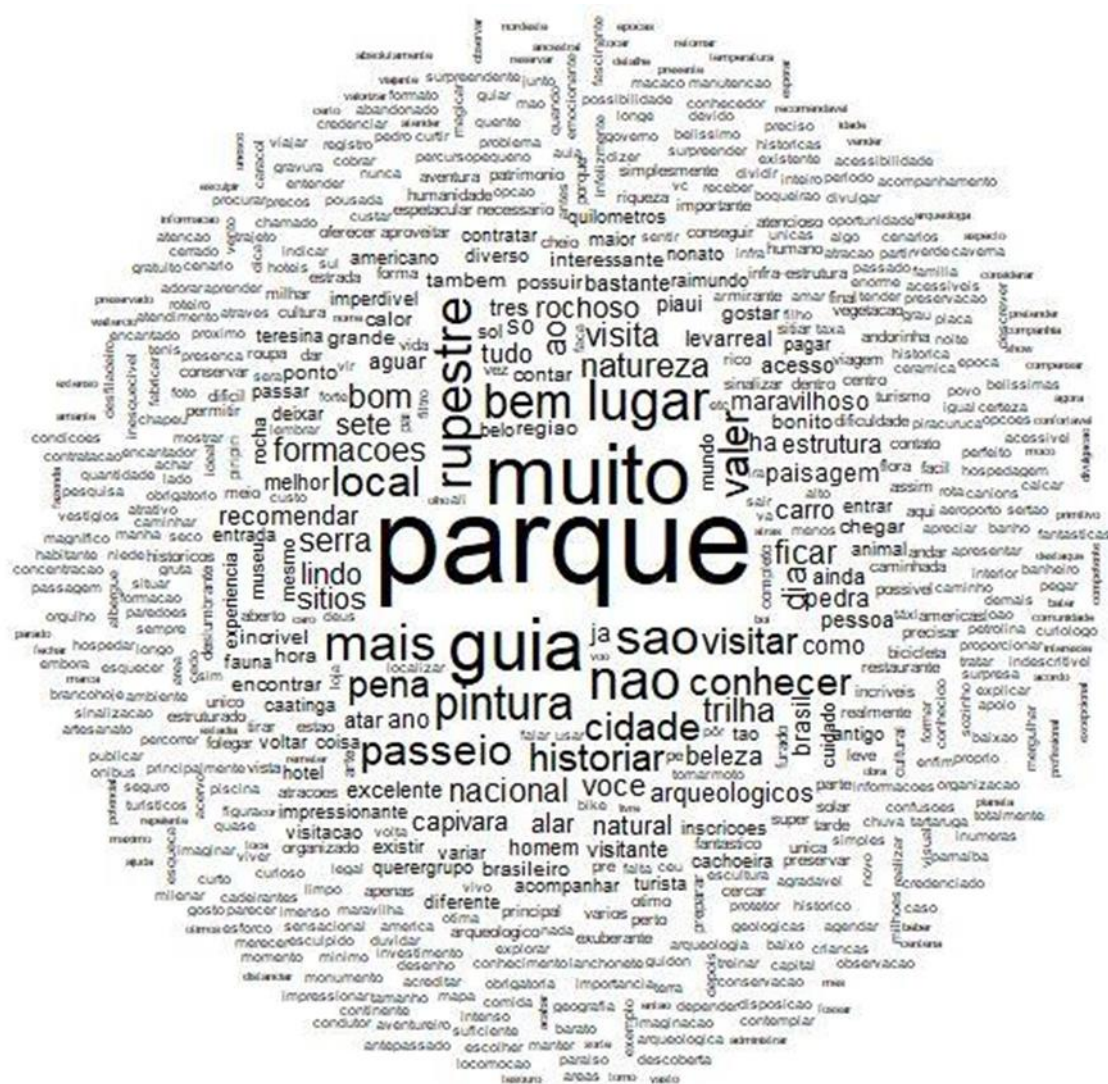


Figura 9: Nuvem de palavras geral feita a partir dos comentários sobre os parques nacionais do Piauí no TripAdvisor.

Em uma análise generalista intitulada “Potencial turístico das unidades de conservação piauienses” realizada por Ivanov & Lemos (2020) destacou-se a dificuldade em encontrar informações sobre as unidades de conservação localizadas no Piauí, além do fato delas ainda não estarem plenamente estruturadas para receber os turistas, e a falta de referências disponíveis na internet impedindo os turistas em potencial (e até pesquisadores) de se informarem e organizarem visitas a esses locais. Apesar das dificuldades relatadas pelos autores, isso não impactou negativamente a experiência das pessoas que visitaram os parques, pois foi observado no estudo aqui analisado

o predomínio das avaliações positivas. De qualquer forma, é preciso considerar que não existia grande número de avaliações disponíveis no TripAdvisor sobre os PARNAs da Serra das Confusões, Nascentes do Parnaíba e do Sete Cidades.

Por fim, Ivanov & Lemos (2020) concluíram que o potencial turístico das UCs pode ser mais bem explorado e a falta de estrutura limita o acesso de turistas. Apesar das dificuldades de acesso e da limitação de infraestrutura no entorno dos PARNAs, na Serra das Confusões, Serra da Capivara e no Sete Cidades existe disponibilidade de serviços e infraestrutura para o atendimento dos visitantes. Ao relatar o comportamento e a avaliação dos turistas que visitam a Serra da Capivara, Mesquita *et. al.* (2015) descreveram que as avaliações em relação ao PARNA são extremamente positivas e contrastam com a avaliação negativa da infraestrutura do entorno do parque.

Outra pesquisa sobre a experiência sobre a visitaç o no Parque Nacional de Sete Cidades, publicada em 2021, relata que 78% dos entrevistados consideraram como " tima a visitaç o", seguidos de 19% que consideraram "boa a visitaç o", totalizando quase 100% de impress es positivas e 81% dos entrevistados afirmaram o desejo de retornar ao PARNA (Santos *et. al.*, 2021). Observando as pesquisas anteriores sobre a percepç o dos turistas sobre os Parques Nacionais no Piau , e comparando-as com os resultados demonstrados nas an lises dos textos contidos nas avaliaç es dispon veis no TripAdvisor, observa-se que a estrat gia metodol gica adotada e demonstrada neste trabalho auxilia nas a ç es de planejamento, estrat gias de marketing e gest o dessas UCs, bem como na gest o do turismo nos munic pios do entorno.

Em rela o ao turismo no Estado do Piau , as avaliaç es analisadas reforçam tamb m os apontamentos de Braga *et. al.* (2022) sobre o grande potencial e competitividade dos atrativos naturais do Piau  para os segmentos de Turismo de Aventura e o Ecoturismo no cen rio nacional. Quando comparado com os demais estados no nordeste do Brasil, os autores ressaltam



que o Piauí é pouco competitivo no mercado de Turismo de Sol e Praia, mas pode se destacar nos segmentos de aventura, ecoturismo e turismo cultural e diversificar a oferta desse tipo de atrativos na região nordeste.

O processo de intermediação, desintermediação, reintermediação causou estresse no mercado de turismo e vem provocando mudanças estruturais e impactando diretamente os negócios e a forma como os negócios ganham o mercado. Por isso, o avanço tecnológico, as mídias sociais e as *big techs* no turismo causam espanto e repulsa por parte significativa de acadêmicos e do mercado. Uma reação natural às mudanças. Porém, não há retorno a velhos padrões, é preciso entender os novos mecanismos e trabalhar com eles. Cada vez mais os usuários de internet, que são os turistas, usam ferramentas on-line para tomada de decisões sobre turismo, e além disso, com o passar do tempo cresce a observação sobre a avaliação. O usuário potencial de um produto ou serviço quer saber como outros que já passaram por lá se sentiram e perceberam o consumo.

Deste modo, oportunizar a identificação de tendências e padrões de turismo por meio de coleta na rede de dados de mídia social pode fornecer informações valiosas sobre a evolução dos comportamentos de viagem, destinos populares, atrações emergentes e preferências do cliente. Assim, a raspagem de dados na rede e a análise de dados de mídia social podem ser usadas para coletar e analisar “conversas” de mídia social relacionadas ao turismo. Ao observar o que os usuários de internet estão dizendo sobre suas experiências de viagem, os pesquisadores podem entender sentimentos, interesses e motivações de viagem. Além disso, é possível extrair e analisar dados de geolocalização de postagens de mídia social para identificar pontos turísticos e destinos populares, como demonstrado por Jankowski *et. al.* 2010. O mapeamento da atividade de mídia social pode fornecer informações sobre o fluxo de turistas e os padrões de multidão. Ou ainda,



analisar o conteúdo gerado pelo usuário para obter *insights* sobre o que atrai e encanta turistas para destinos ou experiências específicas.

CONSIDERAÇÕES FINAIS

As limitações da pesquisa são reconhecidas, como a baixa quantidade de avaliações em alguns parques analisados e a restrição imposta pela data de coleta até fevereiro de 2023. Atualização e até mesmo coleta de dados e análises de outras fontes, como Instagram por exemplo, são necessários para fazer inferências sobre o estado do turismo na região estudada. Destaca-se o uso exclusivo de dados online de uma única fonte: o TripAdvisor, como deficiência apontada inclusive pela comunidade acadêmica, como apontou Soares *et.al.* (2022).

Ficou evidente com os resultados deste estudo que os PARNAS do Piauí têm vieses antropológicos, arqueológicos, históricos e culturais fortes e singulares. Os resultados apresentados confirmam o potencial já demonstrado por Borges (2017) de identificar a essência/*alma/genius loci* dos lugares assim como valores coletivos a partir da análise de dados provenientes da multidão, "crowdsourced". Desse modo, gestores públicos, planejadores de roteiros e administradores de serviços de diversos segmentos do turismo ligados aos PARNAs do Piauí possuem desafios e oportunidades distintas de outros PARNAs do Brasil. O diferencial dos PARNAs do Piauí representado pelas artes rupestres, onde a história humana se confunde com os aspectos biológicos e geomorfológicos demonstraram-se relevantes características para a fundamentação de campanhas de marketing e divulgação. É necessário dar prioridade aos aspectos essenciais dos PARNAs do Piauí (Braga, Guzzi, Perinotto e Malta, 2022) corroborado pelo âmago do discurso dos viajantes.

Pesquisas similares em escala regional e/ou segmentadas podem subsidiar políticas públicas federais que entendam a singularidade de cada conjunto de dados e realizar planos de ações individualizados que respeitem características específicas e peculiaridades. O forte apelo histórico dos



Parques Nacionais piauienses demanda a criação e manutenção de museus, a implantação de planos interpretativos voltados para a arqueologia, o monitoramento e o acompanhamento da visitaç o para apoiar as especificidades desses parques.

O TripAdvisor.com.br tem relevante quantidade de acesso no Brasil, por m os n meros relativos ao uso no territ rio norte americano, TripAdvisor.com (www.similarweb.com/website/TripAdvisor.com/#overview acesso em mar o de 2023), mostram-se expressivamente superiores. Espera-se que o uso desta ferramenta seja ampliado no territ rio brasileiro,   medida que aumenta o acesso a internet e a alfabetiza o digital da popula o brasileira. Al m disso, ao observar em conjunto as estat sticas de acesso, a possibilidade de tradu o instant nea, e o uso da ferramenta j  em 22 idiomas diversos, o TripAdvisor se apresenta como uma ferramenta fundamental para a difus o das atividades tur sticas no Brasil, quando se tratando dos turistas estrangeiros. A ferramenta   capaz de aumentar a seguran a, o tempo de perman ncia e as possibilidades de visita o de pessoas comuns que dificilmente iriam se aventurar em terras estrangeiras se n o fossem essas preciosas informa o j  dispon veis online.

A metodologia deste artigo tem cr terios defens veis e reproduz veis e, por isso, ostenta potencial para ser aplicada em outras regi es, destinos e atrativos tur sticos. Acredita-se que o aumento da compreens o acerca do discurso e percep o on-line dos visitantes sobre os atrativos tur sticos, e a identifica o da ess ncia dos lugares facilitar  o trabalho dos planejadores e gestores de destinos tur sticos.

O estudo apresenta contribui es a respeito da aplica o desses m todos e t cnicas na  rea de turismo como forma de contribui o e subs dio    rea. Avan a na fronteira do conhecimento, ao proporcionar atrav s dos resultados e an lises realizadas oportunidade para compreens o da demanda, contribuindo com diferentes setores na esfera do turismo.



A pesquisa ampliou a compreensão a respeito do CGU e da atividade turística nos atrativos que se propôs analisar. Porém, para alargar a compreensão da demanda dos atrativos em análise é recomendável que pesquisas futuras se concentrem na verificação utilizando outras técnicas, como a avaliação textual, qualitativa do conjunto de avaliações realizadas, e amplie a discussão aqui iniciada sobre os atrativos, marcas (imagem) e marketing dos Parques Nacionais do Piauí. Diante disso, Mazurek (2018) defendeu a importância dos fatores psicográficos para a criação de demanda. Pois, embora a análise do conteúdo gerado pelo usuário promova uma compreensão da percepção sobre a demanda real, a análise sistemática sobre o perfil psicográfico poderá ajudar proprietários e gestores de empresas a atingir consumidores geradores de conteúdo.

A manipulação de grande conjunto de dados (*big data*), em um banco de dados qualitativo a respeito do discurso dos visitantes, PARNAs do Piauí, coletados em mídia social, utilizando-se técnicas de mineração de dados textuais, análise textual estatística e análise de redes tem potencial para apoiar decisões do planejamento, marketing e gestão territorial do turismo. Pois, como demonstrado neste estudo, esclareceu características específicas do contexto, possibilitando compreender aspectos da paisagem e dos territórios, por meio da análise de coleta remota do discurso dos viajantes sobre os destinos turísticos.

Além disso, há incontáveis oportunidades para pesquisas futuras. É possível realizar a análise de cada um dos parques em separado, ampliar a análise textual utilizando outras funções do *software* Iramuteq. Usar a localização dos usuários que realizaram comentários para compreender as regiões emissoras de turistas a um destino específico, demanda turística. Uso de modelagem cartográfica para verificar o potencial de interação territorial baseando-se na localização da oferta de atrativos e origem da demanda turística. Análise territorial sobre a hierarquia da oferta turística disponível em determinado território. Estudos cartográficos e textuais comparativos entre



atrativos, destinos e regiões turísticas. Análise textual sobre as perguntas a respeito dos atrativos.

Inúmeras possibilidades em análises de tendências do turismo, até em tempo real, são proporcionadas ao explorar como os dados da rede e, especialmente da mídia social, podem ser combinados para fornecer informações. Torna-se necessário analisar postagens de mídia social, *hashtags* e discussões (avaliações), para identificar interesses de viagem emergentes.

Estudos futuros podem contemplar a avaliação analógica, feita por entrevistas presenciais em comparação à avaliação feita de forma digital com vistas a confirmar ou não métodos e técnicas digitais sobre analógicas. Entretanto, conforme observado pela pesquisa, a estratégia metodológica adotada corrobora com outras pesquisas que adotaram estratégias analógicas sobre a percepção dos visitantes dos PARNAs do Piauí. Por isso, os resultados apresentados podem ser utilizados para estratégias de marketing das unidades de conservação, pois a maior parte dos visitantes indicou através de seus comentários quais as características de destaque da visita. Essas características podem ser utilizadas em campanhas de publicidade e propaganda para atrair visitantes interessados nos aspectos destacados: *souvenirs*, aspectos arqueológicos, entre outros. Da mesma maneira, o planejamento tanto das unidades de conservação, dos equipamentos de apoio ao turismo como das municipalidades envolvidas com a visita desses parques devem estar atentos para oferecer condições favoráveis e que exaltem esses elementos encontrados.

Trabalhos sobre outros atrativos também podem ser realizados com vistas a comparar resultados, confirmar e calibrar a potencialidade de aplicação da técnica apresentada neste estudo como método de análise de demanda turística de destinos turísticos. Além disso, pesquisas comparativas entre destinos podem ser realizadas com vistas a demonstrar aspectos recorrentes em tipologias de segmentos, verificando padrões, estruturas de



dados, postulados que governam o comportamento das redes, modelos mentais e expectativas dos turistas por segmento.

A grande quantidade de dados - *Big Data* - coletada na rede mundial de computadores sobre a interação dos usuários nas diversas mídias sociais oferece possibilidades de combinações de variáveis, abrindo-se um campo próspero de investigação.

A autonomia que as tecnologias de informação e comunicação possibilitam incrementar a velocidade dos negócios, favorecer a atividade econômica, e dinamizar o setor de turismo de modo inovador e criativo. Aos empreendedores da área de turismo, fornecedores de produtos e serviços, por poderem alcançar os consumidores, e por parte dos consumidores por acessarem informações diretas, com oferta de informações de “consumidores a consumidores”.

REFERÊNCIAS:

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*.

Borges, J. L. C. (2017). O Potencial da Informação Geográfica Voluntária como Suporte à Democracia no Planejamento e Gestão Territorial. 2017. 227 fl. Tese (Doutorado) - Programa de Pós-Graduação em Arquitetura e Urbanismo, Escola de Arquitetura da UFMG, 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, C.B.M.; MENEZES, P.M.L. (Org.).



Cartography - Maps Connecting the World. 1.ed. Switzerland: Springer International Publishing, 2015, v. 1, p. 361-373.

Borges, J., Jankowski, P.; Davis Junior. (2016). A Study On The Use Of Crowdsourced Information For Urban Decision-Making. Revista Brasileira de Cartografia. n.68/4, Edição Especial Geoinformação e Análise Espacial, p. 695-703, 2016. <https://doi.org/10.14393/rbcv68n4-44276> Disponível em: <<http://www.lsie.unb.br/rbc/index.php/rbc/article/download/1313/953>>.

Acesso em: 22 Set. 2017.

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

Braga, S. S., de Oliveira, S. A., Melo, R. S., Nascimento Neto, W. J., & Santos, M. R. (2022). Reflexões sobre a implantação de uma pousada-escola no Parque Nacional de Sete Cidades (PI). Revista Brasileira de Ecoturismo (RBEcotur), 15(5). <https://doi.org/10.34024/rbecotur.2022.v15.12891>

Braga, S. S., (2021). Caracterização e avaliação dos territórios e atrativos turísticos do litoral piauiense. Tese (Doutorado) - Universidade Federal do Piauí, Centro de Ciências da Natureza, Programa Regional de Pós-Graduação em Desenvolvimento e Meio Ambiente, Teresina, 2021.

Braga, S. S., Guzzi, A., Perinotto, A. R. C., & Malta, G. A. P. (2022). Análise da atratividade turística do litoral piauiense: atualização da avaliação dos atrativos turísticos, entre 2010 e 2020. Revista Turismo em Análise, v. 33, n. 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). Parque Nacional de Sete Cidades, Piauí, Brasil: biodiversidade, arqueologia e conservação de arte rupestre. Mneme-revista de humanidades, v. 14, n. 32, 2013.



Campagna, M. (2013). Place, I care! Crowdsourcing 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. Suíça:Springer International Publishing Switzerland, 2014, p. 598-610.

Carlos, Ana Fani Alessandri.(2007). O lugar no/do mundo. São Paulo: FFLCH, 2007, 85p.

Cullen, G. (1993). Paisagem urbana. 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

Fundação Museu do Homem Americano – Fumdam. Disponível em: <http://fumdam.org.br/parque/> Acessado em 15/03/2023

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). Pesquisas arqueológicas em São Raimundo Nonato (Piauí): novos dados. Revista de Arqueologia, São Paulo, 8(2):37-46, 1994-95.

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. Disponível em: <<https://www.wired.com/2006/06/crowds/>>. Acesso em: 04 Abr.2017.

ICMBio. Instituto Chico Mendes de Conservação da Biodiversidade. (2022). Unidades de conservação federais atingem novo recorde de visitação em 2021. Disponível em: <https://www.gov.br/pt-br/noticias/viagens-e->



turismo/2022/04/unidades-de-conservacao-federais-atingem-novo-recorde-de-visitacao-em-2021 Acessado em 22/03/2023

ICMBio. Instituto Chico Mendes de Conservação da Biodiversidade. (2023). Parque Nacional de Sete Cidade. Disponível em: <https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-de-sete-cidades/informacoes-sobre-visitacao-2013-parna-de-sete-cidades> Acessado em 20/03/2023

ICMBio. Instituto Chico Mendes de Conservação da Biodiversidade. (2023). Parque Nacional de Sete Cidades. <https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-de-sete-cidades> Disponível em: Acessado em 20/03/2023

ICMBio. Instituto Chico Mendes de Conservação da Biodiversidade. (2023). Parque Nacional da Serra das Confusões. <https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/caatinga/lista-de-ucs/parna-da-serra-das-confusoes> Disponível em: Acessado em 20/03/2023

Instituto do Patrimônio Histórico e Artístico Nacional - Iphan. Disponível em: <http://portal.iphan.gov.br/pagina/detalhes/42> Acessado em 15/03/2023

Ivanov, M. M. M., & Lemos, J. R. (2020). Unidades de conservação do estado do Piauí. Teresina: EDUFPI.

Izaias, A. S., Soares, P. F., & Mondo, T. S. (2022). Avaliação da Qualidade de Serviços Percebida por Visitantes do Museu do Amanhã – 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, A. M.; Haenlein, M. (2010). Users of the word, unite! The challengers and opportunities of Social Media. *Business Horizons*, v.53, Issue 1, p.59-68, 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0007681309001232>>.

Acesso em: 21 Mar. 2023.

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode 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, T. S., Perinotto, A. R., & 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, P. A. P. (2022). Extração de opiniões e análise de sentimento: a imagem do destino turístico no caso do Algarve (Doctoral dissertation). Mestrado em Gestão de Marketing. Universidade do 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, A. C. M.; Santana, S. A. (2014). From Authorial Drawings to The Parametric Modeling of Territorial Occupation: Representation and Modeling Influences in The Process of Designing the Urban Space. *Revista Brasileira de Cartografia*, n.66/7, p.451-1463, 2014. <https://doi.org/10.14393/rbcv66n0-44739>

Mesquita, R. F., Sousa, L. R. M., Matos, F. R. N., & do Monte, A. L. A. (2015). Comportamento turístico dos visitantes do Parque Nacional da Serra da Capivara. *Tourism & Management Studies*, 11(2), 78-85.

Perinotto, A. R. C. (2013). Investigando a comunicação turística de Parnaíba/PI-Brasil: Internet e redes sociais, descrição e análise. *Turydes: revista de investigación en turismo y desarrollo local*, v. 6, n. 15.

Perinotto, A. R. C; Soares, J. R. R. (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, A. R. C., Belfort Simões, S., Mesquita de Sousa, S. ., & de Souza Braga, S. (2021). Impacto das plataformas digitais na oferta de meios de hospedagem: comparativo de vendas entre booking.com e Airbnb. *Marketing & Tourism Review*, 5(2). <https://doi.org/10.29149/mtr.v5i2.5940>

Perinotto, A. R. C.; Camarço, J. C. F.; Braga, S. D. S.; Gonçalves, M. F. (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 de la méthode ALCESTE à de «gros» corpus et stabilité des «mondes lexicaux»: analyse du «CableGate» avec IramuTeQ. Em *Actes des 11 eme Journées internationales d'Analyse statistique des Données Textuelles* (pp. 835–844). Presented at the 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT

Salviati, M. E. (2017). Manual do Aplicativo Iramuteq. Versão 0.7 Alpha 2 e R Versão 3.2.3. Disponível em



(<http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati> acesso em março de 2023)

Santos, M. R. D., Braga, S. D. S., Vieira, V. B., Araujo, R. O. P., Nascimento Neto, W. J. D., & Oliveira, S. A. D. (2020). Turista do Parque Nacional de Sete Cidades, Piauí. In: Turismo em áreas protegidas [recurso eletrônico] / Thaise Sutil, Nilzo Ivo Ladwig, José Gustavo Santos da Silva (organizadores). - Criciúma, SC: UNESC, 2021.

Silva, Ermelinda Lopes da; Paulo, Eveline Alexandre; Coelho-Costa, Ewerton Reubens; Perinotto, André Riani Costa. (2018). O Aplicativo TripAdvisor e as Reclamações Online Realizadas Pelos Turistas: Uma Visão dos Empreendimentos Gastronômicos em Fortaleza-CE. *PODIUM Sport, Leisure and Tourism Review* v. 7, n. 3 (2018) <https://doi.org/10.5585/podium.v7i3.214>

Silva, J. C. T.; Davis Junior, C. A. (2008). Um framework para coleta e filtragem de dados geográficos fornecidos voluntariamente. In: X Brazilian Symposium on Geoinformatics. Proceedings of the X Brazilian Symposium on Geoinformatics. Porto Alegre (RS): SBC - Sociedade Brasileira de Computação, 2008.

Soares, R. A. M. da C., Albuquerque, T. V. de, Mendes-Filho, L., & Alexandre, M. L. (2022). Revisão sistemática da produção científica brasileira sobre turismo e tecnologia da informação e comunicação (TIC). *Revista Brasileira De Pesquisa Em Turismo*, 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). *Topofilia: um estudo de percepção, atitudes e valores do meio-ambiente*. São Paulo: DIFEL, 1980.

Yázigi, Eduardo Abdo. (2001). *A alma do lugar: turismo, planejamento e cotidiano em litorais e montanhas*. São Paulo: Contexto. Acesso em: 21 mar. 2023.



Sites consultados:

The New York Times. (2022). acesso via www.nytimes.com/interactive/2022/travel/52-places-travel-2022.html - acesso em março de 2023

<https://gephi.org/users/publications/> - acesso em março de 2023

www.similarweb.com/pt/website/TripAdvisor.com.br/#overview - acesso em março de 2023

www.similarweb.com/website/TripAdvisor.com/#overview - acesso em março de 2023

www.octoparse.com - acesso em fevereiro de 2023

