

WEB SCRAPING E CIÊNCIA DE DADOS NA PESQUISA APLICADA EM COMUNICAÇÃO: um estudo sobre avaliações online

WEB SCRAPING AND DATA SCIENCE IN APPLIED RESEARCH IN COMMUNICATION: a study on online reviews

WEB SCRAPING Y CIENCIA DE DATOS EM LA INVESTIGACIÓN APLICADA EN COMUNICACIÓN: un estudio sobre evaluaciones online

Marcello Tenorio de Farias

Mestre em Comunicação pela Universidade Municipal de São Caetano do Sul (USCS). Especialista em Gestão de Marketing e Comunicação pela Escola de Comunicações e Artes da Universidade de São Paulo. Professor na Universidade São Judas Tadeu. cello.farias@hotmail.com.

 0000-0001-7453-9461

Alan César Belo Angeluci

Doutor em Ciências com ênfase em Sistemas Eletrônicos pela Escola Politécnica da Universidade de São Paulo (USP), com período sanduíche na University of Brighton (Inglaterra). Possui estudos de Pós-Doutorado na University of Texas at Austin e na Escola de Comunicações e Artes da USP. Docente permanente do PPGCOM e PPGE da Universidade Municipal de São Caetano do Sul (USCS). angeluci@gmail.com.

 0000-0002-4093-0590

Brasilina Passarelli

Professora Titular pela Escola de Comunicação e Artes (ECA) da Universidade de São Paulo (USP) e Diretora (2021-2024). Coordenadora Científica do NACE Escola do Futuro - USP. Vice-Presidente da Agência USP de Gestão da Informação Acadêmica - AGUIA. Docente permanente do PPGCOM ECA/USP. Doutora em Comunicação pela ECA/USP com período sanduíche na Michigan State University. Realizou estudos de pós-doutorado na Universidad Carlos III de Madrid. Foi Diretora de Marketing para a Educação da Apple Computer Brasil. linapassarelli2@gmail.com.

 0000-0003-2949-7914

Correspondência: Universidade Municipal de São Caetano do Sul, Pró-Reitoria de Pós-graduação e Pesquisa, Rua Conceição, 321 - Santo Antônio - São Caetano do Sul, São Paulo, Brasil - CEP: 09530-060.

Recebido em: 16.02.2021.

Aceito em: 16.04.2021.

Publicado em: 01.07.2021.

RESUMO:

Com a difusão do acesso e uso da informação pela *web* e redes sociais, a recuperação de informação em grandes volumes de dados tem se tornado desafiadora para métodos manuais. Neste estudo aplicado, relatou-se a contribuição do desenvolvimento e uso de um protótipo de ferramenta de raspagem automática de dados de avaliações online feitas no Google Maps – a *Discovery Stars*. Os dados recuperados permitiram investigar como essas avaliações podem ter potencial para influenciar comportamentos dos usuários da plataforma. Dentre os resultados, observou-se que a leitura e postagem de avaliações impactam na formação de opinião e nas motivações dos usuários do Google Maps.

PALAVRAS-CHAVES: Avaliações Online; Raspagem de Dados; Web scraping; Ciência de Dados; Comunicação.

Introdução

Os grandes avanços que ocorreram nos últimos anos relacionados às Tecnologias de Informação e Comunicação (TIC), amparados pela Internet e pela sociedade hiperconectada propiciou o surgimento de plataformas digitais que possuem como objetivo principal coletar opiniões de usuários, em formato de avaliações, sobre locais públicos e privados, estabelecimentos comerciais, produtos e serviços. Essas plataformas funcionam como um grande catálogo que reúnem opiniões deixadas pelos usuários que as utilizam – *reviews* -, sobre locais, produtos ou serviços, e são consultadas por outros usuários que buscam obter mais informações. Nesse contexto, emerge um cenário

altamente colaborativo, em que opiniões diversas são expostas e consumidas em um ritmo acelerado, típico da Internet.

Como destacou Passarelli (2007, p. 11), “na sociedade globalizada em rede, permeada pelo imediatismo e pela transitoriedade dos eventos, algum tipo de permanência precisa ser criado no bojo das comunidades virtuais”. Segundo a autora, essa permanência se manifesta por meio de folksonomias: rastros digitais de usuários em sites de relacionamento, *wikis* e outros dispositivos de comunicação digital contemporâneos que permitem aos usuários dessas redes identificar informações que possam ser relevantes para suas experiências online – influenciando fortemente suas tomadas de decisão. O *boom* das plataformas móveis na última década tem tornado esse fenômeno cada vez mais marcante, suscitando adequações metodológicas para dar conta dos desafios da coleta e análise de massivos dados digitais.

A partir da experiência pregressa de estudos sobre comportamentos e hábitos de usuários nas redes (Angeluci; Rosa; Passarelli, 2020; Passarelli; Angeluci, 2018; Passarelli; Junqueira; Angeluci, 2014), o estudo aqui apresentado visou colaborar com o entendimento de possíveis novos hábitos possibilitados pelo comportamento de realizar e consumir avaliações nessas plataformas. Dentre as plataformas que permitem a realização de avaliações, a escolhida foi o Google Maps, por ser um serviço de geolocalização e mapas em crescente popularidade e de alta adesão do uso, lançada em 2005, pela Google.

Além de ser uma plataforma mundialmente utilizada, passou por diversas modificações e atualizações ao longo dos anos, tornando-se um ambiente de alto fluxo de colaboração e troca de informações na Internet. Ao dar ênfase às avaliações sobre locais que estão inseridos na plataforma, o sistema permite ao usuário escrever uma avaliação, compartilhar fotos, vídeos e responder perguntas sobre o local em questão.

Para realizar uma análise sobre o comportamento dos usuários da plataforma de avaliações do Google Maps com o objetivo de entender as suas motivações em utilizá-la, o primeiro desafio a ser enfrentado é extrair os dados sobre hábitos e usos online. A netnografia é a base metodológica desse estudo, mas o trabalho de investigação também demandou a aplicação de estratégias pouco convencionais nos estudos da área de comunicação e que permitam a extração de um volume de dados massivos desses ambientes, demandando um esforço de inovação no método. Entende-se que essa *big data*, muito além de questões quantitativas, revela uma fotografia genuína de um fenômeno e precisa ser observada a partir de sua natureza massiva.

Portanto, técnicas de *web scraping* parecem se tornar relevantes nesse contexto: por meio de inteligência computacional, é possível coletar um número massivo de dados

de ambientes digitais de maneira automática usando um robô ou algoritmo programado para determinada tarefa. Observando-se sempre os aspectos éticos e legais, a raspagem de dados pode ser uma importante estratégia nas pesquisas contemporâneas que se debruçam sob fenômenos digitais e precisa ser abraçada pelo campo da comunicação em benefício da diversidade e do aprimoramento de suas práticas metodológicas. Como bem posiciona Santos (2018, p. 20):

Não se defendem aqui mudanças bruscas ou nenhum tipo de reformulação geral e muito menos a desvalorização de todo o esforço acadêmico de gerações de pesquisadores que, com seu trabalho, construíram um campo de conhecimento brasileiro e legítimo, mas sim a possibilidade de uma expansão, de um arejamento e de liberdade para propor e experimentar novas conexões, com áreas pouco usuais até então como a Ciência da Computação, o Design, as Engenharias e as Ciências Cognitivas, por exemplo.

No que se refere ao lócus desta pesquisa, o local escolhido para analisar as avaliações feitas no Google Maps foi uma praça pública localizada no bairro da Bela Vista, região central da cidade de São Paulo: Franklin Roosevelt. Por ser um local público de alto fluxo, a praça tem como principal característica a diversidade cultural da cidade, visível pelo intenso fluxo de pessoas de diversas idades, etnias, culturas e interesses. Isso já se revela em uma pré-observação das avaliações deixadas sobre o local, o que torna um objeto de estudo complexo e repleto de controvérsias, com possibilidade de análise de dados de diferentes perfis de usuários da plataforma.

Esse texto estrutura-se da seguinte forma: no primeiro momento, apresentam-se os métodos para a coleta de dados, partindo da abordagem netnográfica de Kozinets e detalhando o desenvolvimento e uso do *script* em Python para *web scraping* – o protótipo *Discovery Stars*. A análise de dados mais quantitativos é apresentada na sequência e foram amparadas e discutidas a partir de dados obtidos por entrevistas semiestruturadas. Nos resultados, são evidenciadas as motivações, a responsabilidade ética, a privacidade e a formação de opinião dos usuários.

Materiais e métodos: desenvolvimento e aplicação da ferramenta de *web scraping* na netnografia

Já como parte do percurso netnográfico, a coleta dos dados envolveu o desenvolvimento de um robô *crawler* para a raspagem dos dados da página da praça Franklin Roosevelt no Google Maps. Sobre a técnica de *web scraping*, Santos (2015, p. 7) observa:

A coleta automatizada de dados, também conhecida como raspagem (*scraping*) ou mineração é um recurso cada vez mais comum (...) podendo, no caso do trabalho acadêmico, ser utilizada tanto para a execução de rotinas repetitivas, permitindo ao pesquisador mais tempo para as tarefas de maior complexidade, como para identificar padrões e tendências em grandes volumes de informação.

Capaz de realizar uma extração automatizada da dados em sites da Internet que não fornecem, naturalmente, esses dados para *download* automático, o *web scraping* possibilita a extração de dados, por meio de *scripts* capazes de identificar, coletar e armazenar metadados para análises posteriores.

É necessário observar a capacidade inovadora que o *web scraping* proporciona aos estudos comunicacionais, contribuindo com perspectivas que se conectam com a ciência de dados, a inteligência artificial e a computação, em um exercício de colaboração de áreas de conhecimento:

É importante destacar que a mineração de dados busca extrair conhecimento inovador, ou seja, algo anteriormente não conhecido e que tenha valor para o domínio em que é aplicada. (FACELI, K., LORENA, A. C., GAMA, J., e CARVALHO, A., 2011, p. 331)

Para a criação de *scripts* capazes de realizar o *web scraping*, de modo geral, utiliza-se da linguagem de programação Python, que escreve os algoritmos para a recuperação da informação. Essa extração, nesta pesquisa, proporcionou a coleta das avaliações feita sobre a Praça Franklin Roosevelt, no Google Maps, e, portanto, foi necessário o desenvolvimento de um *script* específico nessa linguagem para que fosse possível extrair os dados diretamente para uma planilha de dados estruturados.

Esse *script*, que se pode entender como um conjunto de códigos de comando desenvolvido para que fosse possível realizar a raspagem de dados, foi incorporado a uma ferramenta de *web scraping* para desempenhar essa função – batizada de *Discovery Stars*¹ e criada por autores deste estudo. O código desenvolvido utilizou a combinação da linguagem Python com o Selenium, que funciona como um ambiente integrado de desenvolvimento para *scripts* de testes automatizados. Quando acionado, permite que o *script* desenvolvido em linguagem Python possa controlar um navegador, simulando

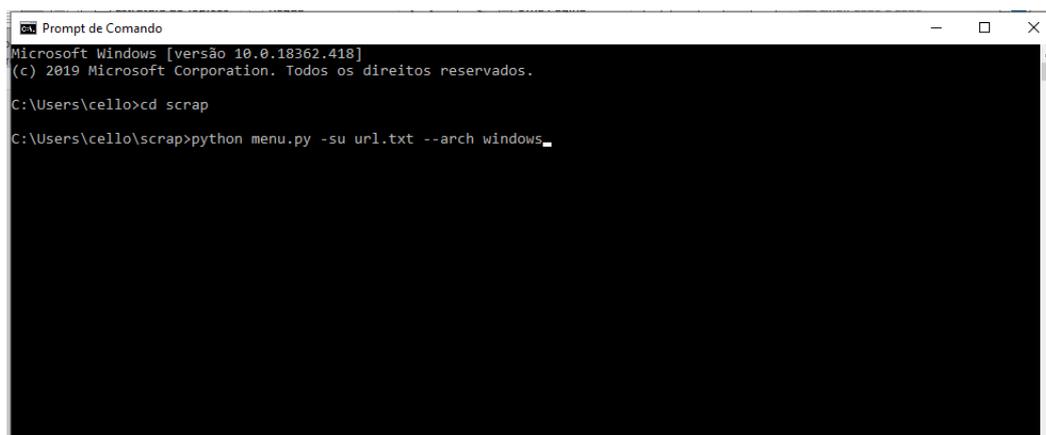
¹ A ferramenta desenvolvida, Discovery Stars, foi disponibilizada para download, para fins exclusivamente acadêmicos, no endereço <https://sites.google.com/view/discoverystars/página-inicial>. Os códigos de programação também podem ser solicitados via formulário na página indicada.

uma ação humana, porém com um número de dados inviável de ser analisado manualmente em tempo hábil. Essa técnica é comumente utilizada em cenários que necessitam de agilidade para realizar tarefas repetidas - como no caso do comando "copiar e colar" - em grandes bases de dados. Nesse sentido, o *script* desenvolvido visa copiar avaliações feitas sobre determinado local no Google Maps e colar para uma planilha, de forma ágil e com o mínimo de ocorrência de falhas.

Após o desenvolvimento do *script* e a sua instalação no computador, o primeiro passo foi acessar o Google Maps e encontrar, na plataforma, o local desejado que seria feita a extração de dados - a praça. Após identificar a URL, o endereço foi hospedado em um arquivo específico, criado na pasta do código instalado, com o objetivo de direcionar o *script* para extrair os dados daquela determinada URL.

Sendo assim, para iniciar o funcionamento da ferramenta, executou-se o "Prompt de Comando"² do sistema de operação do Windows, sendo possível iniciar o processo de extração de dados, conforme a Figura 1.

Figura 1 Prompt de Comando.



```

Microsoft Windows [versão 10.0.18362.418]
(c) 2019 Microsoft Corporation. Todos os direitos reservados.

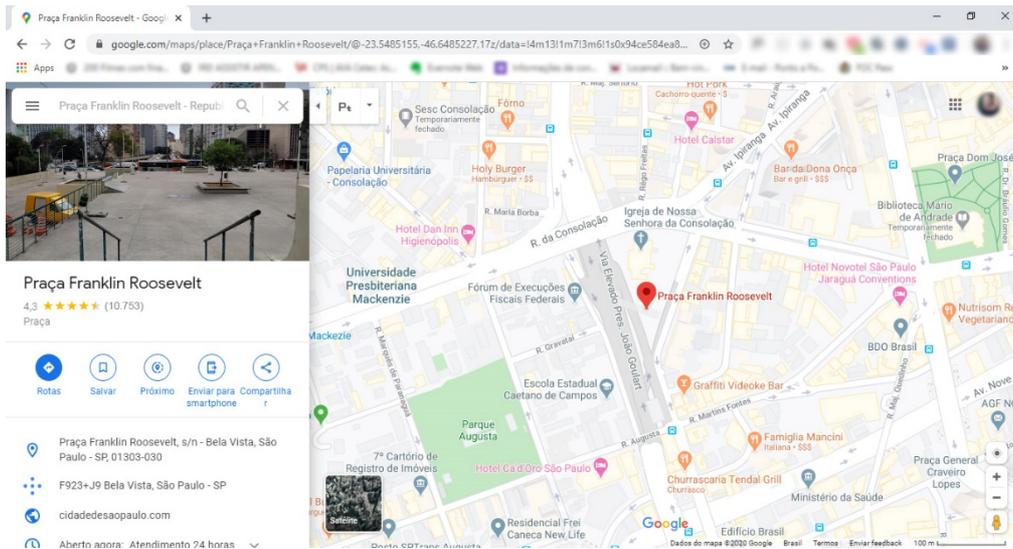
C:\Users\cello>cd scrap
C:\Users\cello\scrap>python menu.py -su url.txt --arch windows_
    
```

Fonte: Elaboração dos Autores.

Assim, o código inserido executou automaticamente a tela do Google Maps com o local que já foi previamente escolhido, iniciando a cópia automática de cada um dos comentários, conforme as Figuras 2 e 3.

Figura 2 Local do Google Maps no navegador.

² Ferramenta disponível no sistema Windows que permite executar comandos inseridos a partir de códigos executáveis.



Fonte: Elaboração dos Autores.

Figura 3 *Script* extraíndo comentários.

```

C:\ Prompt de Comando - python menu.py -su url.txt --arch windows
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements HTTP/1.1" 200 643
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element {"using": "css selector", "value": "html body.keynav-mode-off.screen-mode-jsl div#app-container.vasquette.pane-open-mode div#content-container div#pane div.widget-pane.widget-pane-visible div.widget-pane-content.scrollable-y div.widget-pane-content-holder div.section-layout.section-layout-root div.section-layout.section-scrollbox.scrollable-y.scrollable-show"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element HTTP/1.1" 200 82
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value {"text": "\ue00f", "value": ["\ue00f"], "id": "377be584-fb15-42f6-a01c-f3f8eb5d6601"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value HTTP/1.1" 200 14
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements {"using": "css selector", "value": "div.section-review.ripple-container div div.section-review-content"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements HTTP/1.1" 200 643
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element {"using": "css selector", "value": "html body.keynav-mode-off.screen-mode-jsl div#app-container.vasquette.pane-open-mode div#content-container div#pane div.widget-pane.widget-pane-visible div.widget-pane-content.scrollable-y div.widget-pane-content-holder div.section-layout.section-layout-root div.section-layout.section-scrollbox.scrollable-y.scrollable-show"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element HTTP/1.1" 200 82
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value {"text": "\ue00f", "value": ["\ue00f"], "id": "377be584-fb15-42f6-a01c-f3f8eb5d6601"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value HTTP/1.1" 200 14
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
    
```

Fonte: Elaboração dos Autores.

Por fim, após a raspagem de dados ter sido finalizada pelo *script*, todos os comentários foram extraídos e salvos em formato CSV, conforme Figura 4. Os dados extraídos foram separados por colunas, que contêm "Name", que é o nome utilizado pelo perfil que realizou o comentário; "Star", sendo a nota que o usuário deu para aquele local, classificando de 1 a 5; "Rating", que significa a quantidade de comentários que aquele usuário já fez com o seu perfil no Google Maps; e "Text", que é o conteúdo da avaliação escrita pelo usuário. Essa organização da informação em metadados e atributos vêm da própria arquitetura informacional do Google Maps.

Figura 4 Dados em formato CSV.

Name	Star	Rating	Text
			Ambiente tranquilo e calmo para dar uma caminhada ao ar livre ou praticar esportes como skate e patinação! Também cai bem levar o dog para dar um passeio por lá!
		5	Quem nunca foi , tem que ir, e quem foi nunca mais vai deixar de ir ; é uma vibe muito louca varios tipos de gente, vários níveis sociais, mais em único sentido que curti a noite Paulista e a Praça.
		5	Só precisa de mais segurança ????
		5	Uma das principais praças e espaços de lazer do centro de São Paulo. Wi-fi livre e fácil acesso pra Consolação e Augusta. Ótima opção de lazer pra quem deseja conhecer a região do centro, mais pro lado da Consolação.
		4	Ideal para publico jovem ,moderno e desenganado, pessoas co mundo todo passa por aqui. Ideal para varias acoes, esportiva , ponto de encontro, tem os barzinhos ao redor, lugar massa!
		4	Praça simples, porém enorme, com pessoas interessantes, alternativas... skatistas e tal. Ponto de passagem importante. Há presença da policia e algumas lojinhhas.
		3	Espaço multiuso, mas as vezes é impossível passar pelo local, pois são tantos cachorro correndo e pulando em você que mais parece um canil.

Fonte: Elaboração dos Autores.

Pela netnografia, segundo Kozinets (2014, p. 95), pode-se classificar esses dados extraídos como dados arquivais:

Dados que o pesquisador copia diretamente de comunicações mediadas por computador preexistentes dos membros de uma comunidade online, dados em cuja criação ou estimulação ele não está diretamente envolvido.

Foram extraídas, no dia 18 de agosto de 2019, 4.008 avaliações sobre a Praça Franklin Roosevelt deixadas no Google Maps até aquele momento, que trouxeram uma grande amostra de dados com as variáveis das avaliações feitas. Vale ressaltar que o Discovery Stars realizou essa tarefa de raspagem de mais de 4 mil dados em alguns minutos com grande precisão. Essa coleta certamente levaria semanas se feita de maneira ininterrupta por humanos, e provavelmente estaria sujeita a uma alta taxa de erros.

Dados extraídos: entrevistando informações

Após extrair os dados dos comentários feitos sobre a Praça Roosevelt no Google Maps usando a Discovery Stars, seguiu-se o itinerário clássico da ciência de dados. Foi feita a limpeza da base, arrumando caracteres especiais e reorganizando os dados pelos próprios atributos de categoria da plataforma, já mencionados anteriormente (*Name / Star / Rating / Text*). Depois, por meio de estatística descritiva básica, fez-se a definição dos possíveis recortes analíticos com os dados extraídos e quais as relações poderiam

ser estabelecidas diante desse cenário para compreender o comportamento dos usuários que avaliam na plataforma.

Dentro das variáveis disponíveis, foi possível analisar a relação entre número de avaliações já feitas pelo usuário (*rating*) e nota que ele deu (*star*). Inferiu-se se os usuários com um *rating* maior, ou seja, com mais avaliações em seu histórico, normalmente têm um comportamento de avaliar os locais com notas maiores ou menores, e analisar os usuários com menor *rating*. Essa abordagem parecia ser promissora para entender a influência das avaliações online no comportamento de outros usuários.

Sendo assim, utilizou-se a separação na base de dados extraídos de todos os usuários que deram nota 1, nota 2, nota 3, nota 4 e nota 5. Em seguida, foi feita uma análise de *ratings* dentro desses perfis, realizando uma investigação sobre qual a quantidade de avaliações que os usuários, divididos por notas, possui em seus perfis. Ou seja, foi possível visualizar dentro de cada nota dada para a Praça Franklin Roosevelt quantas avaliações os usuários tinham em seus históricos.

Para realizar essa análise, foram testados três diferentes intervalos de avaliações: de 10 em 10, de 50 em 50 e de 100 em 100. A primeira tentativa, de 10 em 10, trouxe uma análise com muitos dados e com intervalos muito grandes. Já a tentativa de 100 em 100 trouxe uma análise pouco precisa, onde não era possível compreender o impacto dos intervalos. Por fim, o espaçamento escolhido para esse estudo foi a divisão com um intervalo de 50 em 50 avaliações, que dá melhores indícios sobre quais são as maiores concentrações de avaliações no histórico dos usuários. Dessa forma, foi possível entender se a maioria dos usuários que deram nota 1 tinham menos de 50 avaliações em seu histórico, por exemplo. Isso foi feito para os perfis que também deram nota dois, três, quatro e cinco. A visualização gráfica dessas informações é explorada nas seções seguintes.

Entrevistas semiestruturadas: compreendendo melhor o comportamento dos usuários

Foram selecionados dos dados tratados diferentes perfis que fizeram avaliações demonstrando opiniões contundentes sobre a Praça Franklin Roosevelt, sendo feita uma pesquisa dos respectivos nomes no Facebook. O objetivo era tentar encontrar os usuários e contatá-los. Após encontrá-los por similaridade de nomes, foi feito contato por meio do Messenger, confirmados se se tratava da mesma pessoa e, então, convidados para a realização de entrevistas semiestruturada por chamada de vídeo na própria plataforma. Sobre a técnica de entrevista semiestruturada:

Podemos entender por entrevista semiestruturada, em geral, aquela que parte de certos questionamentos básicos, apoiados em teorias e hipóteses, que interessam à pesquisa, e que, em seguida, oferecem amplo campo de interrogativas, fruto de novas hipóteses que vão surgindo à medida que se recebem as respostas do informante. Desta maneira, o informante, seguindo espontaneamente a linha de seu pensamento e de suas experiências dentro do foco principal colocado pelo investigador, começa a participar na elaboração do conteúdo da pesquisa (TRIVINOS, 1987, p. 146).

Para a realização das entrevistas, fez-se previamente um roteiro de perguntas com a finalidade de relacionar os objetivos desta pesquisa com os dados coletados pelos participantes durante a raspagem e análise dos dados. A realização da entrevista em vídeo via Facebook facilitou o contato com os usuários e estava alinhada aos pressupostos da netnografia. Sobre essa técnica, Kozinets aponta:

Usar Skype ou um meio semelhante para uma entrevista online pode significar que você tem um contato face a face com alguém. Em muitas circunstâncias, isso pode ser quase tão bom quanto uma entrevista pessoal em relação à leitura e ao registro de indicadores sociais, como linguagem corporal (KOZINETS, 2014, p. 107).

Dentre os usuários selecionados, foram escolhidos os com um maior número de realização de avaliações em seu histórico na plataforma (dado denominado como *rating*) e outros com poucas avaliações em seus históricos, visando selecionar personagens com diferentes características de utilização do Google Maps. Após essa seleção prévia, foram buscados no Facebook os usuários que tinham os nomes mais incomuns, que consequentemente seriam mais fáceis de serem encontrados.

Após os personagens serem encontrados, foram adicionados no Facebook, e então, foram enviadas mensagens de apresentação, explicando que eles foram encontrados por terem deixado avaliações no Google Maps e, por fim, perguntando se eles aceitariam participar de uma pesquisa, que seria realizada por meio de entrevista em vídeo, com questões sobre as suas interações com o Google Maps.

Dessa forma, a técnica utilizada para definir a amostra da coleta de entrevistas foi a de saturação (Glaser e Strauss, 1967). As entrevistas foram realizadas até ocorrer a saturação, que aconteceu com oito entrevistados. Das oito, foram selecionadas seis entrevistas para realizar a análise e transcrição. Duas foram descartadas por não trazerem contribuições significativas.

Os seis perfis escolhidos para a análise (Quadro 1) trazem uma pluralidade de relatos, pois possuem características diferentes de idade, região que moram, resultados

em números variados na plataforma e diferentes formas de utilização do Google Maps. Vale ressaltar que os entrevistados deram anuência para uso de seus dados anonimizados para fins da pesquisa.

Quadro 1 Perfis dos participantes das entrevistas.

Pseudônimo	Sexo	Local	Idade	Profissão	Nível no Google Maps ³
Entrevistado 1	Homem	Morador de São Paulo (SP)	33 anos	Profissional da área de transportes	Nível 8
Entrevistado 2	Mulher	Morador de Mogi Mirim (SP)	60 anos	Diretora de escola e Youtuber	Nível 8
Entrevistado 3	Mulher	Moradora de São Paulo (SP)	32 anos	Profissional da área de e-commerce	Nível 4
Entrevistado 4	Homem	Morador de São Paulo (SP)	29 anos	Produtor cultural	Nível 5
Entrevistado 5	Homem	Morador de São Paulo (SP)	45 anos	Historiador e técnico em museologia	Nível 6
Entrevistado 6	Homem	Morador do Rio de Janeiro (RJ)	29 anos	Engenheiro químico	Nível 7

Fonte: Elaboração dos Autores.

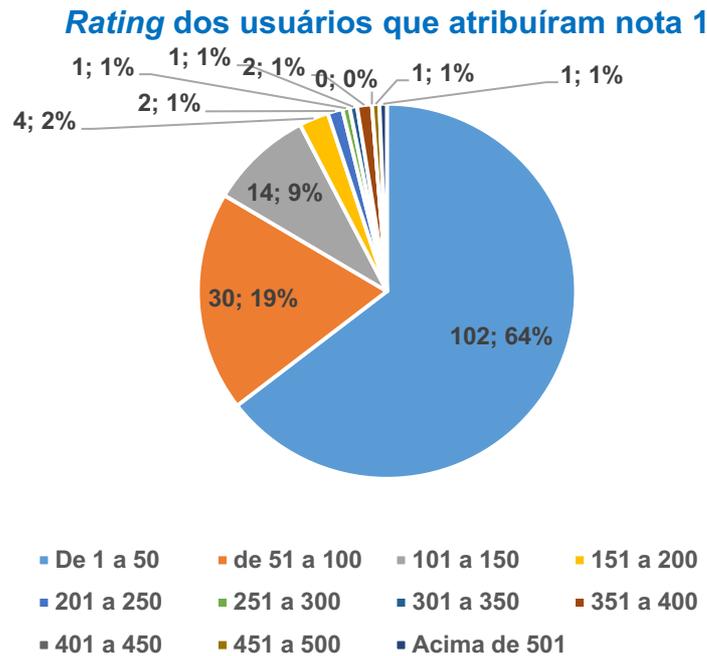
Formação de opinião

Um dos objetivos desse estudo foi investigar de que forma o Google Maps influenciou na formação de opinião dos seus usuários. Dessa forma, os dados analisados buscaram evidências exploratórias sobre esse comportamento.

Partindo de dados quantitativos, fez-se uma análise sobre o *rating* dos usuários que avaliaram a Praça Franklin Roosevelt e a nota que eles atribuíram (*star*), a fim de identificar possíveis relações. Sendo assim, as notas foram divididas, de 1 a 5, para que fosse possível analisar a faixa de *rating* dos usuários, considerando um intervalo de 50 em 50. Pode-se observar essa análise nos Gráficos 1, 2, 3, 4 e 5.

Gráfico 1 Faixa de *rating* de usuários que atribuíram nota 1.

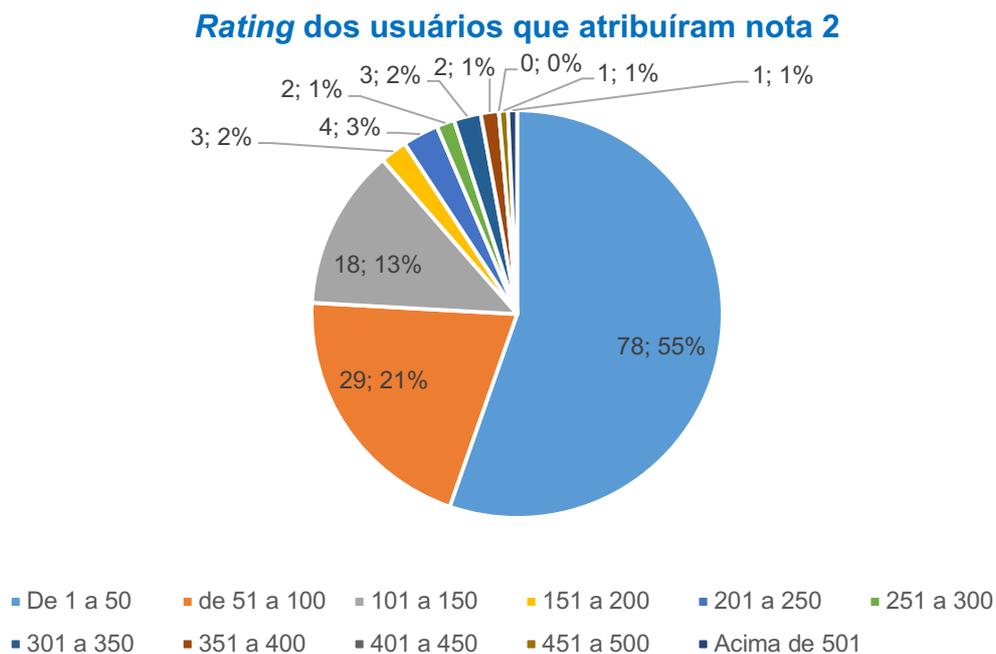
³ Refere-se aos pontos de *Local Guides*, ou seja, usuários que recebem pontos ao contribuir com a avaliação de locais no Google Maps. Sua escala vai de 1 a 10. Mais informações em: <https://support.google.com/local-guides/answer/6225851?hl=pt-BR>.



Fonte: Elaboração dos Autores.

Segundo essa análise, foi possível observar que, dos usuários que deram nota 1, 64% tem um *rating* de 1 a 50, ou seja, já fizeram de 1 a 50 avaliações em seu perfil, seguido de 19% com *rating* de 51 a 100, 9% com *rating* de 101 a 150, e a partir disso, com *ratings* que variam de 0% a 2% até *ratings* acima de 500.

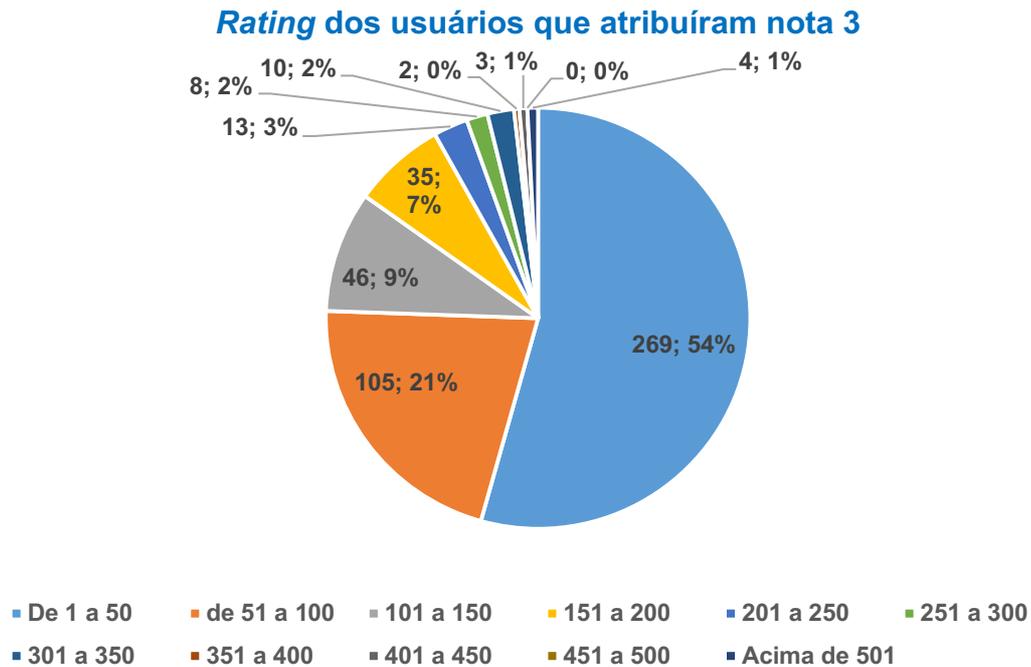
Gráfico 2 Faixa de *rating* de usuários que atribuíram nota 2.



Fonte: Elaboração dos Autores.

Já os usuários que atribuíram para a Praça Franklin Roosevelt nota 2, a maioria, 55%, tiveram *rating* de 1 a 50, 21% obtinha *rating* de 51 a 100, 13% de 101 a 150, e a partir disso, também apresentaram variações de *ratings* que variaram de 0% a 3% até *ratings* acima de 500.

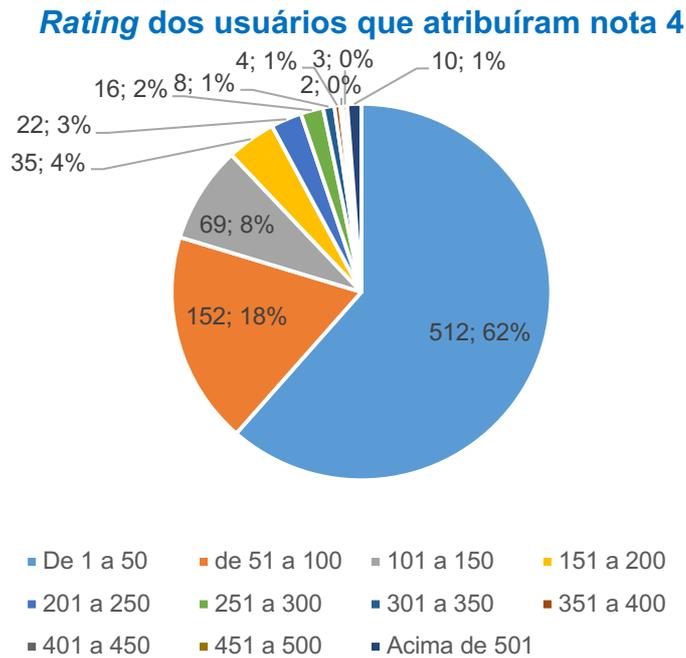
Gráfico 3 Faixa de *rating* de usuários que atribuíram nota 3.



Fonte: Elaboração dos Autores.

Os usuários que atribuíram nota 3 também apresentam comportamento similar: 54% também tem *rating* de 1 a 50, 21% tem *rating* de 51 a 100, 9% de 101 a 150, 7% de 151 a 200, e a partir disso, também apresenta variações de *ratings* que variam de 0% a 3% até *ratings* acima de 500.

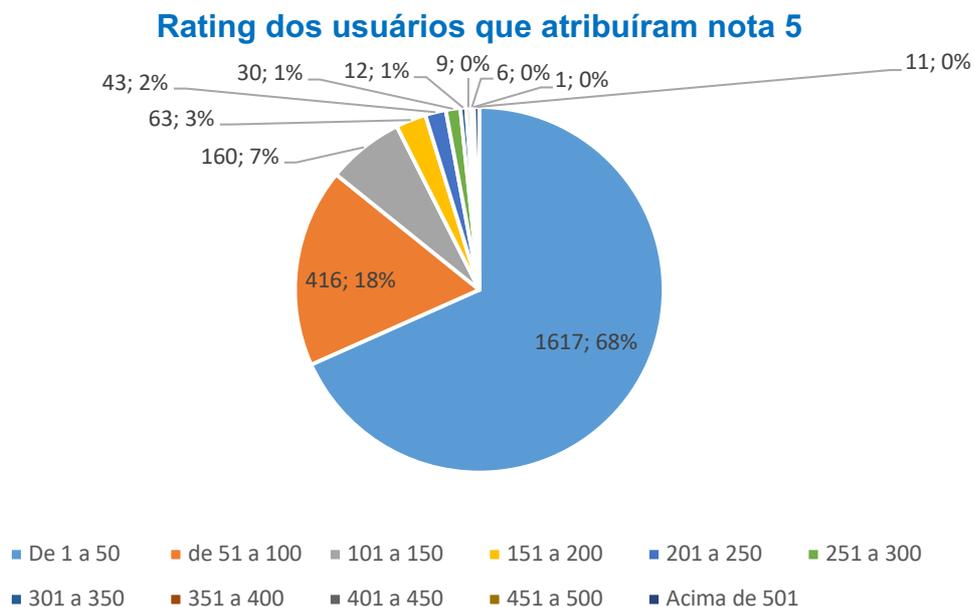
Gráfico 4 Faixa de *rating* de usuários que atribuíram nota 4.



Fonte: Elaboração dos Autores.

Analisando os usuários que atribuíram nota 4, pode-se observar que aumenta a porcentagem de usuários que tem *rating* de 1 a 50, se aproximando dos usuários que atribuíram nota 1: 62% têm *rating* de 1 a 50, 18% têm *rating* de 51 a 100, 8% de 101 a 150, e a partir disso, também apresenta variações de *ratings* que variam de 0% a 4% até *ratings* acima de 500.

Gráfico 5 Faixa de *rating* de usuários que atribuíram nota 5.



Fonte: Elaboração dos Autores.

Por fim, analisando os usuários que atribuíram nota 5, observa-se um aumento significativo dos usuários que possuem *rating* de 1 a 50, representando 68% deles. Na sequência, 18% com *rating* de 51 a 100, 7% com *rating* de 101 a 150, e a partir disso, com *ratings* que variam de 0% a 3% até *ratings* acima de 500.

Portanto, é possível notar uma variação de 4% entre os usuários de *rating* 1 a 50 que deram nota 1 e nota 5, que pode apontar uma leve inclinação a investigação de que usuários com menos quantidade de avaliações em seu histórico têm uma tendência a avaliar com nota maior o local, fomentando a suspeita de que usuários que avaliam negativamente são mais críticos e por isso avaliam mais lugares.

Essa análise sobre a visão crítica dos usuários que utilizam a plataforma foi aprofundada nas entrevistas semiestruturadas, as quais comprovaram que alguns consideram que o Google Maps ampliou a percepção e a opinião deles sobre os locais que visitam depois que começaram a utilizar a plataforma:

Eu acho que sim. Porque o Google Maps faz uma série de perguntas, ele costuma perguntar "neste lugar tem estacionamento?", "pode fazer happy hour?", "este lugar, você pode ir com criança?", "ele é pet friendly?", "é isso, é aquilo?", então depois que você responde várias, quando você entra em algum lugar você já mapeia tudo na sua mente. Ele faz esse monte de pergunta, então você já fica condicionado a... quando você olha o lugar, se ele já cai naqueles critérios ou não, em quais critérios ele se encaixa. Então, sim. Acho que eu me tornei um pouco mais criteriosa. (Entrevistada 3, informação verbal)

Demais, muito. Já me fiz essa autoavaliação [...] é uma coisa tão assim... que eu já vou [nos lugares] com esse olhar, entendeu? [...] fiquei muito mais crítica. Muito mais. (Entrevistada 2, informação verbal)

Desse ponto de vista, é importante ressaltar que a utilização da plataforma e o seu mecanismo de perguntas pode se relacionar diretamente com a construção de um novo olhar sobre os locais que os usuários frequentam. Isso parece ampliar a visão de mundo deles, uma vez que se tornam mais questionadores em relação a diversos pontos que a plataforma questiona e que, intrinsecamente, se torna perguntas que eles refletem em suas experiências vividas. O Google Maps, portanto, pode se tornar um fator de interferência na construção da opinião dos seus usuários, os fazendo perceber espaços com um novo olhar: mais crítico, mais questionador e mais curioso.

A plataforma de avaliações do Google Maps tende a estimular os usuários para a sua utilização por meio de técnicas gamificação, que gratificam os usuários com pontos a cada atividade realizada por eles, fazendo-os evoluir de nível e conquistar “selos” – técnicas essas que visam estimular o engajamento dos usuários a se manterem ativos na plataforma. Porém, além desse fator, os usuários apresentaram motivações próprias para escreverem avaliações na plataforma, trazendo aspectos pessoais. Nesse sentido, podemos observar outras motivações que ocorrem nesse comportamento, influenciados pela mudança social que a Internet resulta. Hjarvard nos ajuda a compreender esse contexto ao pontuar que “a sociedade contemporânea está permeada pela mídia de tal maneira que ela não pode mais ser considerada como algo separado das instituições culturais e sociais” (2012, p. 53).

Quando questionados sobre suas motivações para a utilização da plataforma, primeiramente foi comum entre os usuários entrevistados apontarem a vontade de contribuir com outros usuários e ajudarem, de forma colaborativa, outras pessoas. Podemos observar na fala do Entrevistado 1:

[Faço avaliações] na intenção de ajudar, não para querer difamar um local, mas para alguém que tá procurando alguma informação do local poder ir sem estar no escuro, saber como é. Porque eu tiro foto de vários lugares, então eu quero fazer com que a pessoa tenha uma experiência de como é antes de ir, para que a pessoa tenha um parâmetro. (Entrevistado 1, informação verbal)

O Entrevistado 5 também demonstrou, na sua forma de analisar as motivações para realizar avaliações, argumentos em concordância com o mesmo raciocínio apresentado:

Eu acho que a possibilidade de você poder falar sobre o lugar dá para a gente um senso de participação, de uma importância para que outras pessoas possam ter parâmetros. Eu achava que as minhas avaliações não eram lidas, até que começaram a pipocar no meu celular várias perguntas em posts e o Google começou a me mandar mensal falando “bom, você teve 15 mil visualizações”, e eu falava “mas como assim, eu tive 15 mil visualizações, né? (Entrevistado 5, informação verbal)

Castells (2015), ao utilizar o conceito de comunicação consciente reflete que ela é de extrema importância para a nossa vivência e é o que faz com que sejamos humanos. Sendo assim, pensar a responsabilidade que se tem ao comunicar é fundamental para refletir a importância desse ato na sociedade cada vez mais conectada e midiaticada, afinal, a opinião exposta em rede ultrapassa esferas e possui um alcance inimaginável.

Diante disso, Castells coloca que “a comunicação consciente e significativa é o que nos torna humanos. Por isso, qualquer transformação importante na tecnologia e na organização da comunicação é de grande relevância para a mudança social” (CASTELLS, 2015, p. 20, tradução nossa).

Em linha, o discurso destacado do Entrevistado 1 e do Entrevistado 5, exemplificam a importância da comunicação consciente ao demonstrar suas motivações em ajudar diferentes usuários, demonstrando que eles levam em consideração, enquanto estão avaliando, a boa vontade em ajudar outras pessoas que acessam a plataforma também. Juntamente com esse apontamento, podemos caracterizar que as principais motivações dos usuários são oriundas da gamificação e da vontade de contribuir com outras pessoas que utilizam o Google Maps para consultar avaliações, o que retrata a preocupação com uma comunicação consciente.

Do ponto de vista da consciência e da responsabilidade ética na criação de avaliações, a Entrevistada 3 expôs a forma como se sente ao avaliar um local, considerando a responsabilidade que tem com essa ação:

Eu me sinto de duas formas muito distintas: eu me sinto julgando algo, então eu fico pensando “como que o dono desse lugar vai se sentir quando ele ler isso”? Importante, a escolha de palavras é importante, porque tudo tem a ver com pessoas. Apesar de você estar em um ambiente digital, apesar de não ter tanto isso, o cara a cara com o outro, você tá lidando com pessoas de fato, é muito difícil a gente esquecer isso. Eu penso muito em como que o dono do lugar, ou a pessoa que administra aquele lugar, vai se sentir, se ela vai achar que foi legal ou não, mas eu penso muito mais em quem está pesquisando para ir naquele lugar. Eu penso muito mais, uns 80% nesse cara [que está pesquisando], e uns 20% no administrador. (Entrevistada 3, informação verbal)

Contribuindo com essa visão, alguns aspectos interessantes aparecem quando questionados sobre os tipos de local que os usuários avaliam. A Entrevistada 2 apontou a responsabilidade de avaliar lugares, trazendo informações sobre acessibilidade para pessoas com deficiência:

[Eu avalio] todos, todos. [...] porque assim, se você pensar, hoje uma pessoa com uma deficiência, né? Ele tem... ele é um cidadão e ele tem direito a uma vida social, então é tudo, né? É todos os lugares, desde uma loja que eu vou, uma sorveteria que eu vá, supermercado. (Entrevistada 2, informação verbal)

Em sequência, após mencionar que entende o ato de avaliar como uma grande responsabilidade, o Entrevistado 4 levantou o questionamento sobre a forma como outros usuários utilizam a plataforma:

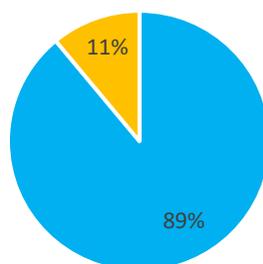
Eu acho que a maioria das avaliações não são feitas com esse tipo de parâmetro, tá ligado? Eu não posso levar em consideração pessoas que eu sei que... mano, tem uma galera que às vezes, você vê que a pessoa tá fazendo uma avaliação por fazer, tipo "estive nesse lugar, blá blá blá, muito legal recomendo", você vê que são coisas superficiais, e aí eu penso "mano, eu tô tentando fazer algo que agregue um pouco mais, então não posso ser mordido por isso". Esse é meu pensamento, sendo bem honesto. (Entrevistado 4, informação verbal)

De acordo com a visão exposta, o Entrevistado 4 questionou a forma como outros usuários utilizam a plataforma, subjetivamente indagando sobre valores e crenças individuais de cada usuário na Internet. Isso acontece porque, na sociedade midiaticizada, indivíduos tendem a encontrar na Internet uma forma de expressar suas opiniões e percepções do mundo, de maneira simplificada e muitas vezes com alta exposição. O questionamento do Entrevistado 4, então, trouxe a reflexão sobre a Internet ser um espaço de encontro de diferentes opiniões e percepções, e instiga o pensamento sobre como as pessoas se expõem.

Sobre a exposição, fez-se uma análise dos dados extraídos das avaliações da Praça Franklin Roosevelt no Google Maps, com o objetivo de verificar a quantidade de usuários que utilizam nomes falsos ou não identificáveis no momento de realizar uma avaliação. O resultado dessa análise pode ser observado no Gráfico 6.

Gráfico 6 Análise da identificação dos usuários.

Identificação dos usuários



- Usuários com nomes
- Usuários com apelidos ou nomes não identificáveis

É possível observar que dentre a amostra analisada, 11% dos usuários utilizam de apelidos ou nomes não identificáveis. Sendo assim, os 11% de usuários que escreveram avaliações, utilizando nomes que dificultam serem encontrados, podem estar resguardando a sua privacidade. Por outro lado, essa prática tende a gerar escrita de avaliações sem responsabilidade ética: falsas, sem informações relevantes, desconexas ou não contributivas. De qualquer forma, os usuários, ao escreverem uma avaliação, estão expostos ao mundo digital: é o momento em que eles expõem a percepção que possuem para que outras incontáveis pessoas leiam, de maneira intencionalmente enviesada ou não. Vale lembrar que nomes não identificáveis só estão blindados a uma primeira camada de privacidade, já que indivíduos podem ser rastreados caso a plataforma assim deseje, como nos casos de determinação judicial.

Por um lado, ter o nome completo explícito na expressão de uma opinião sobre determinado local, para qualquer pessoa que acessar aquele conteúdo, mostra uma certa legitimidade ao que está escrito, pois denota que foi um usuário real que a escrever e que pode ser identificado e confrontado facilmente. Por outro lado, torna a privacidade do usuário vulnerável, uma vez que é possível encontrá-lo em outras redes sociais, como um rastro deixado. Também, o uso de um nome reconhecível pode ser feito de maneira desautorizada por outra pessoa. Portanto, a cautela na identificação do sujeito é necessária.

Sendo assim, reflete Kerckhove sobre a privacidade no mundo digital:

Mas é claro que já somos totalmente prisioneiros dos dados que deixamos nas nossas pistas. Ainda não sabemos até que ponto somente a lei ou a tecnologia são capazes de proteger a privacidade. Mas se trata realmente de privacidade? (KERCKHOVE, 2016, p. 4).

Dessa forma, o ato de escrever avaliações online não se resume a apenas escrever opiniões na Internet. Fica claro que esse comportamento extrapola para questões de responsabilidade ética, de privacidade e de pertencimento ao meio social em que o usuário se propõe ao interagir e avaliar: quando um usuário escreve uma avaliação online, ele expressa sua visão de mundo, a sua vivência, a forma como analisou e processou a experiência de conhecer determinado local, e narra, de maneira pessoal, o que aquilo representou para ele.

É importante ressaltar o fato de que esses usuários entendem a responsabilidade que lhes compete, e que percebem que aquilo não é apenas um comentário, mas sim uma forma de contribuir com outras pessoas que buscam por essa informação, o que de

certa forma se relaciona com uma responsabilidade ética e cidadã, por meio de uma comunicação consciente. Por outro lado, também entendem que a plataforma esbarra em limites éticos, que precisariam ser melhor analisados em seus mecanismos.

Considerações finais

O presente estudo, em caráter de pesquisa aplicada, revelou diversas percepções que permeiam diferentes âmbitos relacionados aos ambientes digitais contemporâneos e as suas utilizações do ponto de vista dos usuários. A Internet é responsável por transformar a forma da sociedade de se relacionar, trazendo mudanças em nossos comportamentos, reflexões que ficaram evidentes diante dos resultados apresentados na presente pesquisa e do exposto pelas análises quantitativas e qualitativas.

A ferramenta desenvolvida possibilitou extrair dados que foram capazes de construir um percurso metodológico netnográfico capaz de analisar dados quantitativos e qualitativos, por meio de análises em gráficos e entrevistas, sobre as diversas percepções dos usuários ao utilizarem a plataforma do Google Maps. A exploração de técnicas oriundas da ciência de dados, como o *web scraping*, contribuiu de forma significativa para a inovação metodológica presente no estudo. Essa prática possibilitou a extração de dados massivos de forma automatizada. Com base nesses dados, foi possível realizar análises quantitativas imprescindíveis para a geração de gráficos e análises que permitiram relacionar hipóteses e suscitar novos questionamentos. Também possibilitou encontrar personagens para a realização das entrevistas semiestruturadas, que foram determinantes para a investigação.

A partir da ferramenta Discovery Stars, foram extraídos e analisados comentários que foram deixados no Google Maps sobre a praça Franklin Roosevelt, local público em São Paulo. A partir desses dados, identificou-se uma variação de 4% de diferença quando se compara os usuários de *rating* 1 a 50 que deram nota 1, representados por 54%, e dos usuários de *rating* 1 a 50 que deram nota 5, representados por 68%. Esse dado, apesar de não ser conclusivo, reforça a inferência de que possa existir uma tendência aos usuários com menor quantidade de avaliações em seu histórico terem uma predisposição a avaliarem com nota maior o local; também, que usuários que avaliam com notas mais baixas tendem a ser mais críticos, e que, portanto, avaliam mais lugares.

Ao realizar o comparativo desses dados com a fala dos entrevistados, percebeu-se que a utilização frequente do Google Maps contribuiu na percepção que os usuários constroem, tomando como base o que leem e o que escrevem na plataforma. Os usuários apontaram que se tornaram mais críticos após a utilização da plataforma, o que demonstrou que ela foi capaz de ampliar a opinião de seus usuários sobre espaços que

eles frequentam. É possível refletir que, a partir da utilização da ferramenta de avaliações online do Google Maps, os usuários começaram a perceber diversos aspectos sobre os locais que visitaram, como fatores relacionados a acessibilidade, localização, preço, atendimento e ambiente, desenvolvendo uma visão mais crítica que os acompanha nos locais que visitam, além da plataforma.

Esta pesquisa também revelou que o fator motivador dos usuários ao escreverem avaliações é o sentimento de contribuição. Os entrevistados relataram que se sentiram úteis quando deram as suas opiniões na plataforma, pois sentiram que estavam ajudando outros usuários que iriam acessar e consultar o que escreveram ali. A importância dessa ajuda para eles se tornou ainda mais evidente quando eles relatam que, quando receberam notificações do Google Maps avisando a quantidade de pessoas que já visualizaram os seus comentários, ficaram ainda mais motivados e tendo a sensação de que estavam contribuindo realmente com outras pessoas.

Quando se analisa os dados contidos no Gráfico 6, onde foi possível identificar a relação dos usuários com a forma que eles expõem seus nomes e identificações, percebeu-se que apenas 11% deles escondem, de alguma forma, a sua identidade real. É possível inferir, com algum cuidado, que a maioria, representada por 89%, não se preocupou com a exposição de seu nome completo. Esse comportamento pode ser justificado pelos usuários talvez não se preocuparem com a rastreabilidade dos seus comentários, ou até mesmo que não tenham se atentado de que podem ser encontrados a partir da opinião deixada.

Por fim, o presente estudo contribuiu, a partir dos dados coletados por meio da ferramenta de *web scraping* Discovery Stars, com novas percepções sobre o comportamento dos usuários da Internet que realizam avaliações online, apontando características desse grupo, incitando outras questões sobre os impactos que esse hábito causa em suas vivências, e abrindo espaço para futuros estudos desse mesmo tema, que tem vazão para ser mais aprofundado e discutido.

Referências

- ANGELUCI, A. C. B.; ROSA, B. B.; PASSARELLI, B. Podcasts sobre Covid-19: o projeto# MDDFcontraocorona. *Comunicação & Educação*, v. 25, n. 1, p. 186-199, 2020.
- CASTELLS, M. *The impact of the internet on society: a global perspective*, 2015. Disponível em: <https://http://www.bbvaopenmind.com/en/article/the-impact-of-the-internet-on-society-aglobal-perspective/?fullscreen=true>. Acesso em: fev. 2021
- SANTOS, M. C. Métodos digitais e a memória acessada por APIs: Desenvolvimento de ferramenta para extração de dados de portais jornalísticos a partir da WayBack Machine. *Revista Observatório*, v. 1, n. 2, p. 23-41, 8 dez. 2015.

- SANTOS, M. C. Pesquisa Aplicada em Comunicação - O estranhamento da interdisciplinaridade que nos assombra. *Comunicação & Inovação*, v. 19, n. 41, 2018.
- FACELI, K., LORENA, A. C., GAMA, J., e CARVALHO, A. (2011). Inteligência artificial: Uma abordagem de aprendizado de máquina. Livros Técnicos e Científicos.
- GLASER, B.G.; STRAUSS, A.L. The discovery of grounded theory: strategies for qualitative research. New York: Aldine de Gruyter; 1967.
- HJARVARD, S. Mídiação: teorizando a mídia como agente de mudança social e cultural. *Matrizes*: São Paulo, ano 5, n. 2, p. 53-91, jan./jun. 2012.
- KERCKHOVE, D. Ética de transparência na era do Big Data. *Comunicação, Cultura e Mídias Sociais*. São Paulo, 2016.
- KOZINETS, R. V. Netnografia: Realizando pesquisa etnográfica online. Porto Alegre: Penso, 2014.
- PASSARELLI, Brasilina–Interfaces Digitais na Educação:@ lucin [ações] consentidas. São Paulo: Escola do Futuro da USP, 2007.
- PASSARELLI, B.; ANGELUCI, A. C. B. Conectividade contínua e acesso móvel à informação digital: jovens brasileiros em perspectiva. *Informação & Sociedade*, v. 28, n. 2, 2018.
- PASSARELLI, B.; JUNQUEIRA, A. H.; ANGELUCI, A. C. B. Os nativos digitais no Brasil e seus comportamentos diante das telas. *Matrizes*, v. 8, n. 1, p. 159-178, 2014.
- TRIVINOS, A. N. S. Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.

ABSTRACT:

With the spread of access and use of information through the web and social networks, information retrieval in large volumes of data has become unfeasible by manual methods. In this applied study, the contribution of the development and use of a prototype tool for automatic data scraping from online evaluations made on Google Maps – Discovery Stars – was reported. The retrieved data allowed us to investigate how these assessments can have the potential to influence the behavior of the platform's users. Among the results, it was observed that the reading and posting of reviews impact the formation of opinion and motivations of Google Maps users.

KEYWORDS: Online Evaluation; Data Scraping; Web scraping; Data Science; Communication.

RESUMEN:

Con la difusión del acceso y el uso de la información a través de la web y las redes sociales, la recuperación de información en grandes volúmenes de datos se ha convertido en un desafío para los métodos manuales. En este estudio aplicado, se informó la contribución del desarrollo y uso de una herramienta prototipo para el raspado automático de datos de evaluaciones en línea realizadas en Google Maps - Discovery Stars. Los datos recuperados nos permitieron investigar cómo estas evaluaciones pueden tener el potencial de influir en el comportamiento de los usuarios de la plataforma. Entre los resultados, se observó que la lectura y publicación de reseñas impacta en la formación de opinión y motivaciones de los usuarios de Google Maps.

PALABRAS-CLAVES: Evaluaciones en línea; Raspado de datos; Web scraping; Ciencia de los datos; Comunicación.