Daniel Costa
Vianna Mucciolo
Universidade Federal de
Santa Catarina

Resenha

**Dados digitais e as
novas abordagens para
estudar o comporta-
mento humano**

**Digital data and new
approaches to studying
human behavior**

**Datos digitales y
nuevos enfoques para
estudiar el comporta-
miento humano**

Submissão: 12-8-2019
Decisão editorial: 21-7-2020

Stephens-Davidowitz, Seth. **Todo mundo mente**: O que a internet e os dados dizem sobre quem realmente somos. Rio de Janeiro: Alta Books Editora, p. 352. 2018.

Este intrigante livro traz um compilado de informações obtidas através da análise de dados digitais resultantes do uso da internet. O conhecimento trata da realidade da vida humana em várias áreas: política, preconceito, relacionamentos, hábitos, sexualidade e outras, e por se tratar de uma metodologia nova, essas informações em muitos casos contradizem aspectos que eram considerados verdades dadas da condição humana contemporânea.

A importância de se produzir uma resenha sobre esse livro se deve à originalidade do trabalho, pois o autor mostra o resultado de pesquisas com diversos bancos de dados recentes de usuários da internet, os quais poucas pessoas têm acesso, sendo possível ter *insights* sobre o comportamento dos usuários, também pelo peculiar currículo do autor que possui graduação em Filosofia pela Universidade de Stanford, doutorado em Economia pela Universidade de Harvard, trabalhou para o Google e estudou programação e análise de dados, demonstrando, assim, a importância do caráter transdisciplinar na produção do conhecimento.

O autor introduz a temática do livro falando sobre seu contato com a ferramenta Google Trends, que apresenta graficamente a relação entre a quantidade de pesquisas de um termo no buscador do Google por um determinado espaço de tempo. E, ao ver as potencialidades para compreender os hábitos dos usuários da internet, passou a usá-la para tentar responder perguntas que o inquietavam, inclusive utilizando-a em sua tese de doutorado. As pesquisas que uma pessoa faz no Google podem dizer muito sobre ela, pois a pessoa pode pesquisar assuntos que não têm coragem de falar nem para aqueles que são mais íntimos. Então, a combinação de dados de uma população nesses buscadores, podem revelar informações que nunca foram acessadas sobre esse determinado grupo, e são essas descobertas que o autor traz ao longo do livro.

Entre algumas vantagens que superam as análises de dados digitais, com relação a metodologias convencionais de pesquisa, está a de poder trabalhar com uma amostra da população que seria impraticável acessar entrevistando cada uma pessoalmente, e também porque, retira-se a presença de um entrevistador que em algum grau constrange a liberdade de resposta do entrevistado. Nesse sentido, o autor enfatiza a potencialidade de trabalhar com uma amostra grande de dados, o que vem sendo chamado hoje de *Big Data*. Trata-se da existência de grandes volumes de dados gerados e que as empresas de todos os ramos vêm investindo muitos recursos para transformar esses dados brutos em conhecimentos úteis. É evidente que existem questões a serem consideradas sobre a veracidade desses resultados, pois essa produção de dados pode estar sendo influenciada pela existência

de *bots*, robôs que podem impulsionar determinado tema. Mas, de fato, vivemos numa época em que a produção de dados digitais produzidos cresce espantosamente, é só pensarmos a quantidade de vídeos, imagens e textos publicadas na internet diariamente para ter uma ideia da dimensão do que estamos produzindo, e podemos nos beneficiar de análises de todos esses dados para gerar novos conhecimentos sobre diversos aspectos da sociedade.

Para mostrar a importância de tentar outras abordagens com dados para responder problemas, ele apresenta o caso de um comerciante de cavalos de corrida. Até então, o preço do cavalo esteve sempre muito relacionado historicamente ao *pedigree*, à linhagem genética do animal. Entretanto, depois de coletar e analisar dados de vários cavalos de corrida, esse comerciante chegou à conclusão de que é o tamanho do ventrículo esquerdo um dos maiores preditores de sucesso nas do animal nas corridas, e com isso, fez uma previsão excelente com um dos cavalos que analisou. Sendo assim, os dados da internet também são uma novidade de nossa época, e nos possibilitam novas abordagens para entender o ser humano.

O autor discorre sobre diversos casos onde essa metodologia vem sendo aplicada, na economia para prever os indicadores econômicos, linguística para processar textos e postagens da internet para prever ideologia política, na saúde pública tem a possibilidade de identificar epidemias através da consulta em buscadores da internet pelos sintomas, e aponta que realmente as aplicações são infindáveis e que estamos diante de uma grande ferramenta para ajudar a entender nossos comportamentos.

Ele traz alguns exemplos peculiares, por exemplo, analisa a validade da teoria freudiana, para isso se vale de dados de erros de digitação para verificar se a ocorrência de erros envolvendo palavras com temática sexual apareciam com maior frequência, aspecto importante nessa teoria. O autor do livro não achou dados que reforçassem a teoria, mas ao investigar as 100 pesquisas mais ocorrentes de um site pornográfico ele constatou que 16 delas tratavam de temáticas incestuais, o que pode indicar a relevância da teoria de Freud ao abarcar tais temas. Ele também mostra que os dados envolvendo sites pornográficos trazem informações peculiares sobre uma população, como no caso da Índia, onde existe a prevalência de interesse pela população masculina em material pornográfico envolvendo lactantes.

O autor segue narrando diversas pesquisas, em uma delas analisa o ano de nascimento de perfis de usuários masculinos do Facebook que curtiram a página de um time de baseball. Os dados mostram que se o time dessa pessoa tivesse sido campeão da *World Series*, que é o campeonato mais importante desse esporte, quando ela tivesse oito anos de idade a probabilidade dela ser um fã do esporte quando adulta seria a mais elevada. Aponta-se uma correlação e a importância de uma experiência marcante nessa fase da vida para determinar o comportamento na fase adulta, algo que os teóricos do desenvolvimento humano estudam e discutem sempre, que são os limites da influência da formação nesses estágios iniciais da vida. Ele também narra outro estudo americano sobre preferências política, que diz que as preferências políticas partidárias de uma pessoa tendem a ser influenciadas mais fortemente pela popularidade

do presidente em exercício quando ela tem a idade de 14 e 24 anos.

Uma outra pesquisa narrada no livro fala sobre a possibilidade de uma pessoa que nasce em uma família pobre se tornar rica nos Estados Unidos, para verificar a fama daquele país como uma nação das oportunidades. O estudo que analisou a trajetória de várias pessoas demonstrou que essa possibilidade é maior em alguns estados e menor em outros, e em países como Dinamarca e Canadá essa ascensão é mais fácil do que nos Estados Unidos.

Em outro estudo, o autor faz um levantamento sobre local de nascimento das pessoas da geração "baby boomers" que obtiveram sucesso na vida. Para isso, analisou os dados da Wikipedia, uma vez que é esperado que pessoas famosas e bem-sucedidas tenham uma página nessa enciclopédia. Uma característica comum de todos os estados com elevado número de pessoas que tiveram páginas criadas foi o fato de o estado em que elas nasceram possuir uma grande universidade.

Uma outra ferramenta que o autor utiliza é a de autocompletar. Quando se digita algo no campo de pesquisa do buscador do Google, ele sugere uma série de opções para completar a pesquisa, sendo estas as mais populares pela região em que o usuário se encontra. Então, por meio das sugestões é possível saber sobre o que uma população anda pesquisando. O autor mostra diversos resultados sobre dúvidas, por exemplo, do que uma grávida pode fazer ou comer, em que a dúvida varia muito pela região. Na Austrália, a maior dúvida é se uma grávida pode ingerir *cream cheese* e nos Estados Unidos se elas podem comer camarão.

No final do livro, ele tenta contar sua tentativa de descobrir a porcentagem das pessoas que terminam de ler um livro, para isso, ele se utiliza da quantidade de citações que as pessoas fazem nos aparelhos leitores de ebook, e constata que existe um maior número de citações nas páginas iniciais do que nas finais, indicando que poucas pessoas vão até o final dos livros. Neste livro, específico devido à grande gama de exemplos que o autor usa, pela originalidade e pela forma leve que apresenta todas essas informações, é válido contrariar essa estatística e realizar a leitura até o final.

DADOS DO AUTOR

Daniel Costa Vianna Mucciolo

ORCID iD <http://orcid.org/0000-0002-3980-2161>

Universidade Federal de Santa Catarina

É psicólogo, mestre em Psicologia pela Universidade Federal Rural do Rio de Janeiro, doutorando em Psicologia na Universidade Federal de Santa Catarina. E-mail: daniel.mucciolo@posgrad.ufsc.br.

Currículo Lattes: <http://lattes.cnpq.br/9827685950079560>