

Big data na academia: um dos desafios contemporâneos nas pesquisas em marketing

Academic big data: one of the contemporary challenges of the marketing research

MARIA AMÁLIA DUTRA MACHADO

Mestranda em Administração (PUCRS) e bolsista CAPES
e-mail: maria.amalia@acad.pucrs.br

1 Introdução

“What significant insights are gained through big data?
Why are we so poor at gaining insights from big data?”
(Marketing Science Institute, 2016, p. 13)

Resumo:

O fenômeno Big Data tem atraído a atenção dos pesquisadores em marketing nos últimos anos, os quais buscam entender como extrair informações relevantes desse aumento exponencial de disponibilidade de dados. As prioridades de pesquisa elencadas pelo *Marketing Science Institute* demonstram essa preocupação ao solicitar o desenvolvimento de novas ferramentas de coleta e análise de informações. Entretanto, até o momento, não houve um debate sobre as implicações acadêmicas desse fenômeno. Neste sentido, o presente artigo tem como objetivo compreender como o Big Data pode afetar as pesquisas acadêmicas na área de marketing. Para isto, com base nos estudos em várias áreas do conhecimento, apresenta uma discussão dos possíveis conceitos e das características de volume, velocidade, variedade, veracidade, valor e outros V's. Depois, elencam-se diversas possibilidades de estudos como as ferramentas que facilitam a visualização e a interpretação das informações. Por fim, os desafios que o fenômeno impõe às pesquisas são apresentados, sobretudo com relação à pressão por publicações e à falta de tempo, às questões éticas no uso dos dados e à acessibilidade das informações.

Palavras-chave: Big Data; Pesquisa Acadêmica; Marketing

Abstract:

The Big Data phenomenon has attracted the attention of marketing researchers in recent years, who seek to understand how to extract relevant information from this exponential increase in data availability. The research priorities for the Institute of Marketing Sciences demonstrate this concern to the request of development of new tools of collection and analysis of information. However, to date, there has been no debate about the academic implications of this phenomenon. In this way, the present article aims at how Big Data can affect academic research in the field of marketing. For this, based on studies in several areas of knowledge, it presents a discussion of the possible concepts and characteristics of volume, velocity, variety, veracity, value and other V's. Then, several possibilities of studies are listed as tools that facilitate analysis and interpretation of the information. Finally, the challenges that the research imposes are published, especially with the publications pressure and lack of time, ethical issues about de use of the data and the accessibility of information.

Keywords: Big Data; Academic Research; Marketing

Esse questionamento demonstra a preocupação dos empresários de marketing sobre o fenômeno Big Data, que também tem atraído a atenção dos pesquisadores da área nos últimos anos (Sheng, Amankwah-Amoah, & Wang, 2017). Para o *Marketing Science Institute*, é necessário investigar as rápidas mudanças que ocorrem no mundo contemporâneo e desenvolver novas abordagens analíticas para extrair informações relevantes dessa grande quantidade de dados disponíveis (Marketing Science Institute, 2016). Esse aumento exponencial de informações numéricas (Singh Jain, Mehta, Mitra, & Agrawal, 2017) exige modelos matemáticos e novos indicadores e métricas para analisar tendências (Wedel & Kannan, 2016). Ao mesmo tempo, vídeos, textos e imagens, disponibilizados pelos consumidores na internet e acessíveis a todos os pesquisadores, oferecem uma riqueza de dados qualitativos, que também precisam de ferramentas específicas para serem analisados.

Vários estudos têm buscado investigar a diferença entre a análise de marketing tradicional (*Traditional Marketing Analysis*) e a análise do Big Data (*Big Data Analysis*) (Xu, Frankwick, & Ramirez, 2016) big data analytics (BDA, a aplicação das suas dimensões no varejo (Bradlow, Gangwar, Kopalle, & Voleti, 2017) e na análise do comportamento do consumidor (Erevelles, Fukawa, & Swayne, 2016) unprecedented volume, velocity, and variety of primary data, Big Data, are available from individual consumers. To better understand the impact of Big Data on various marketing activities, enabling firms to better exploit its benefits, a conceptual framework that builds on resource-based theory is proposed. Three resources-physical, human, and organizational capital-moderate the following: (1. Entretanto, até o momento não foi encontrada nenhuma publicação na área de marketing que discuta as implicações da utilização dos dados provenientes do Big Data nas pesquisas acadêmicas. Para Wilkie e Moore (2003), a academia em marketing precisa pensar sobre as implicações, problemas e oportunidades apresentadas pela globalização e pela internet. Nesse sentido, este trabalho buscou discutir como o fenômeno do Big Data pode afetar as pesquisas acadêmicas na área de marketing.

Para isso, o presente artigo contribui para o entendimento dessa problemática ao apresentar as características do fenômeno, as possibilidades e desafios que poderão ser enfrentados pelos pesquisadores da área. Essa discussão está organizada da seguinte forma: primeiramente, os conceitos de Big Data apresentados na literatura são relacionados com os aspectos do

estudo científico. Depois, são descritas as características do fenômeno, e discutidas as possibilidades e os desafios que se apresentam com relação à utilização do Big Data nas pesquisas acadêmicas em marketing.

2 O Fenômeno Big Data

Nos últimos anos, os avanços tecnológicos possibilitaram a diminuição dos custos de coleta e de armazenamento de informações, gerando uma quantidade enorme de dados com altos níveis de aprofundamento e detalhamento (Wedel & Kannan, 2016). Esse fenômeno de aumento exponencial de dados disponíveis é denominado Big Data (Singh Jain *et al.*, 2017) e exige uma capacidade de extrair e explorar as informações por meio de interpretação avançada (Erevelles *et al.*, 2016) *unprecedented volume, velocity, and variety of primary data, Big Data, are available from individual consumers. To better understand the impact of Big Data on various marketing activities, enabling firms to better exploit its benefits, a conceptual framework that builds on resource-based theory is proposed. Three resources—physical, human, and organizational capital—moderate the following: (1. Por ser um fenômeno recente, ainda há muitas divergências com relação à sua conceituação. Por isso, apresentam-se a seguir os principais conceitos definidos na literatura.*

2.1 Os Conceitos de Big Data na Literatura

A quantidade de informações disponíveis é sempre a característica mais citada do Big Data, mas isso não expressa as possibilidades de uso desses dados. Para Jacobs (2009), é exatamente o tamanho que permite ultrapassar métodos já consolidados na atualidade para vislumbrar tendências futuras. O registro constante de observações repetidas, ao longo do tempo ou do espaço, fornecem dados que exigem novas tecnologias para extrair valor das informações e permitir a coleta e a análise dessa ampla variedade de dados em alta velocidade (Gantz & Reinsel, 2012; Jacobs, 2009).

Esse grande número de dados dificulta, e até mesmo impossibilita, seu gerenciamento e análise. Não é possível examinar a veracidade das informações, os cálculos levam muito tempo para serem processados e novos dados podem ser constantemente agregados à base de análise. Além disso, os diferentes formatos de dados, como imagens, textos e áudios, exigem outras ferramentas de análise e síntese (Fisher, DeLine, Czerwinski, & Drucker, 2012). Portanto, o Big Data também precisa ser pensado com relação ao desenvolvimento de habilidades de alto nível e ao uso de ferramentas para coletar, armazenar, organizar, extrair, analisar e gerar informações de diversas fontes (Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015).

Como a frequência e os tipos de dados coletados podem variar, o conceito de Big Data também precisa ser fluído para acompanhar os avanços tecnológicos (Gandomi & Haider, 2015). O fenômeno tem despertado a atenção da investigação acadêmica (Miah, Vu, Gammack, & McGrath, 2017) por representar, além de tecnológica, uma revolução científica (Ann Keller, Koonin, & Shipp, 2012). Atualmente, é possível medir múltiplas variáveis ao mesmo tempo, registrar imagens em alta definição e gerar milhares de *gigabytes* de dados para serem examinados em um único estudo (Jacobs, 2009).

Nesse sentido, Boyd e Crawford (2012) contribuem para o entendimento da relação entre o conceito de Big Data e as pesquisas acadêmicas ao definir o fenômeno como cultural, tecnológico e acadêmico baseado na interação (1) da tecnologia, ferramentas computacionais que reúnem, analisam, vinculam e comparam grandes conjuntos de dados; (2) da análise, que utiliza esses conjuntos para identificar padrões e interpretar fatores econômicos, sociais, técnicos e legais; e (3) da mitologia, crença de que esses dados são uma forma mais elevada de inteligência e de conhecimento que pode gerar resultados inéditos (Boyd & Crawford, 2012). Todavia, o Big Data exige a gerência, o processamento e a análise de dados com várias características (Wamba *et al.*, 2015), descritas a seguir.

2.2 Três, Cinco ou Oito V's? As Características do Big Data

Além das discrepâncias na conceituação do fenômeno, também há divergência com relação às características dos “grandes dados”. A maior parte dos estudos elenca como principais características Três V's: o volume, a velocidade e a variedade (Bradlow *et al.*, 2017; McAfee & Brynjolfsson, 2012; Wamba *et al.*, 2015). A junção dessas características é vista como uma forma de definir Big Data e, de fato, facilita a compreensão da evolução das informações disponíveis nos últimos anos (Dessureault, 2016). A **Figura 1** ilustra essa evolução.

A seguir, descreve-se cada uma dessas três características.

2.2.1 Volume

O volume de informações geradas pelo fenômeno Big Data é uma das principais características citadas nos estudos e se refere à magnitude do tamanho das medidas de armazenamento dos dados (Erevelles *et al.*, 2016; Gandomi & Haider, 2015). São dezenas de *petabytes*, *exabytes* ou *zettabytes* gerados diariamente (Wedel & Kannan, 2016) que, se fossem impressos em formato de texto, seria necessário bilhões de armários de arquivos para guardá-los (Erevelles *et al.*, 2016;

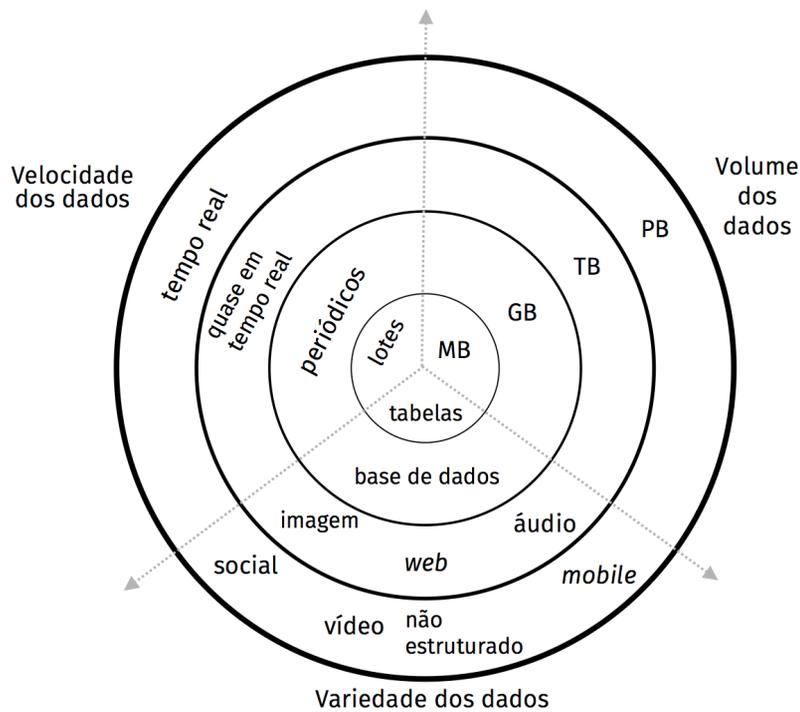


Figura 1. Evolução dos 3 V's do Big Data

Fonte: Adaptado de Rouse, M. (2016). Big Data. Recuperado de <http://search-cloudcomputing.techtarget.com/definition/big-data-Big-Data>

McAfee & Brynjolfsson, 2012). A quantidade é tão grande que, em 2011, o volume de dados gerados em dois dias (cerca de cinco *exabytes*) era equivalente ao montante de informações registradas em toda a história da humanidade até o ano de 2003 (Elragal, 2014).

Outro aspecto relacionado ao volume de dados é o acesso proporcionado, principalmente, pela internet. As mídias sociais disponibilizam diversos tipos de informações geradas pelos próprios usuários e, no caso do marketing, pelos consumidores e pelas empresas. A Wikipédia, por exemplo, tinha 2,5 bilhões de páginas armazenadas em 19,9 GB, em janeiro de 2008. Em dezembro de 2013, esse número havia aumentado para 8,3 bilhões de páginas e 66,2 GB de dados armazenados. Na rede social *Twitter*, em setembro de 2008, os usuários haviam feito 9,45 milhões de *tweets*, totalizando 1,49 GB de dados. Nesse mesmo mês, no ano de 2013, o número de *tweets* chegava a 1,28 bilhões, armazenados em 601,49 GB (Liu, Singh, & Srinivasan, 2016).

Por necessitarem de tecnologias específicas de armazenamento, geralmente essas informações ficam localizadas na nuvem ou em armazéns de dados (Grover & Kar, 2017). Além disso, tipos de informações diferentes dificultam a definição de um limite específico para o volume de dados a serem gravados (Gandomi & Haider, 2015). É possível observar, no âmbito acadêmico, essa

mesma tendência. Atualmente, existem vários bancos de periódicos, livros e dados disponíveis aos pesquisadores, que armazenam as informações na nuvem. O Scopus, por exemplo, pode ser considerada a maior base de dados científicos do mundo, considerando-se o período de 2000–2011 (Guerrero-Bote & Moya-Anegón, 2012). A **Figura 2** resume a quantidade de informações disponíveis nesta base de dados.

Essa grande quantidade de itens e referências disponíveis aos pesquisadores pode exemplificar o fenômeno Big Data na academia. O aumento de publicações nos últimos anos também reforça a ideia do tamanho de informações às quais os acadêmicos do marketing têm acesso. O *SCImago Journal & Country Rank* é uma lista do Scopus que agrupa os periódicos em 313 categorias de assuntos específicos (SCImago, 2007). Uma busca feita nos *journals* de Negócios, Gestão e Contabilidade (*Business, Management and Accounting*) selecionou apenas os *rankings* da subárea de marketing, entre 1999 e 2016, para identificar o número de publicações por ano. O resultado foi resumido no **Gráfico 1**.

É possível observar que houve um aumento significativo no número de publicações científicas internacionais de marketing nos últimos anos. Para os pesquisadores, ao mesmo tempo em que essa disponibilidade de informações pode ajudar a melhorar a qualidade dos trabalhos também pode atrapalhar a tarefa de

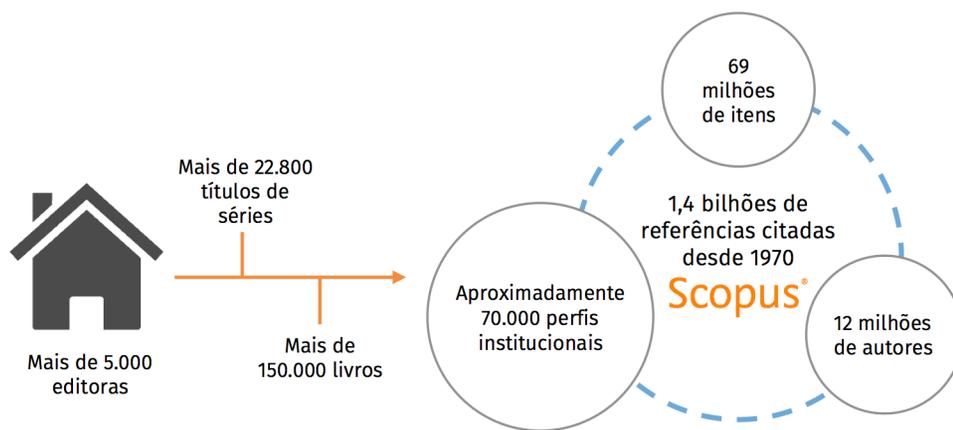
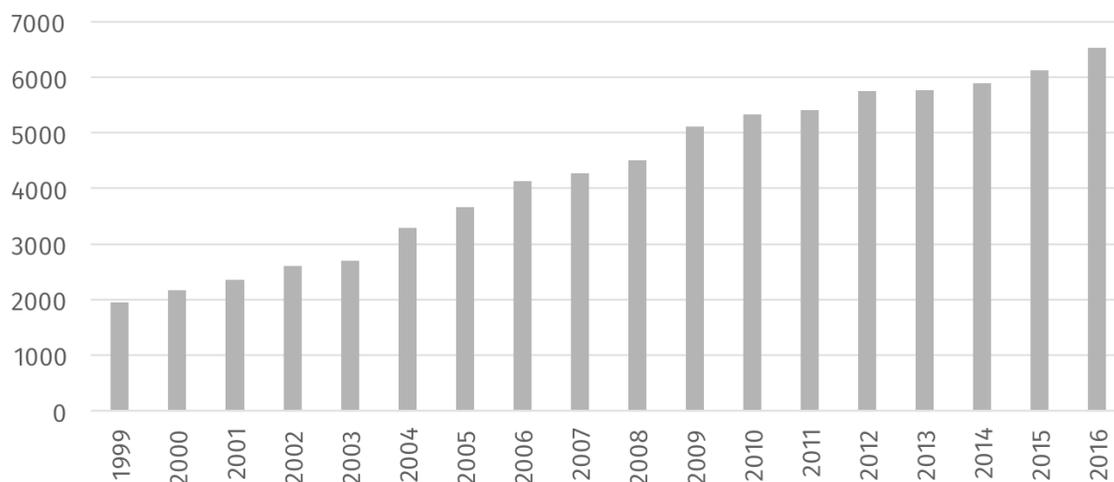


Figura 2. Esquema visual do conteúdo do Scopus

Fonte: Adaptado de Scopus (<https://www.elsevier.com/solutions/scopus/content>, recuperado em 25 de agosto, 2017).

Gráfico 1

Total de documentos publicados por ano



Fonte: Elaborado pelo Autor (2017).

fazer uma revisão de literatura abrangente. Quanto maior o número de documentos, maior devem ser os filtros aplicados para retornar uma quantidade de artigos que seja humanamente possível de analisar. Um exemplo disso pode ser encontrado no **Gráfico 2**, o qual mostra que o número de referências citadas nos documentos publicados no período de 1999 a 2016 também aumentou.

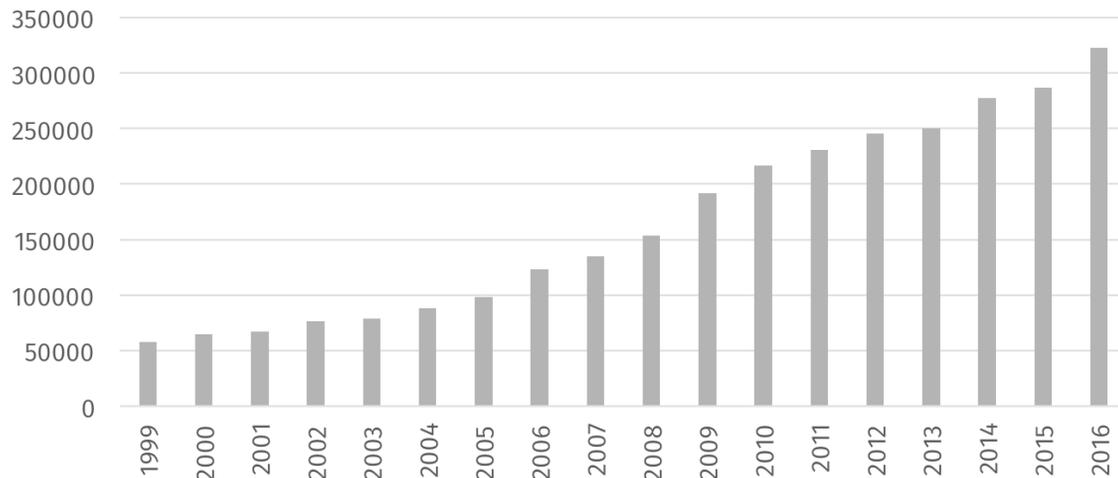
Apesar do volume ser considerado uma característica distintiva primária do Big Data, os dados só poderão ser considerados “grandes” se possuírem as demais características (Erevelles *et al.*, 2016).

2.2.2 Velocidade

A implacável rapidez com que os dados são constantemente gerados é considerada outra característica do fenômeno Big Data (Erevelles *et al.*, 2016; Lam, Sleep, Hennig-Thurau, Sridhar, & Saboo, 2017; McAfee & Brynjolfsson, 2012). Essa velocidade em que os dados são coletados e processados exige uma rapidez, também, na sua análise (Gandomi & Haider, 2015; Grover & Kar, 2017), e, com frequência, essas informações estão disponíveis em tempo real (Chintagunta, Hanssens, & Hauser, 2016). Além disso, a rapidez na transmissão dos dados facilitou o acesso. No ano de 2016, por exemplo, estimava-se 18,9 bilhões de conexões de rede e quase 2,5 conexões por pessoa no mundo todo (Herschel & Miori, 2017).

Gráfico 2

Total de documentos referenciados em artigos publicados, por ano



Fonte: Elaborado pelo Autor (2017).

Essa velocidade também pode ser identificada na academia. Uma busca feita em agosto de 2017, na base *SCImago*, em *journals* da área de Negócios, Gestão e Contabilidade, no idioma inglês, retornou 37 artigos com data de publicação de 2018. Dessa lista de artigos, o periódico *Cities* (<https://www.journals.elsevier.com/cities>, recuperado em 25 de agosto, 2017) chamou a atenção por já permitir o acesso aos volumes; 69, com data de publicação em setembro de 2017; 70, em outubro de 2017; 71, em novembro de 2017; e 72—Parte A, em fevereiro de 2018. Todos os artigos desses volumes tinham suas versões finais disponíveis para citação. Além disso, tinham outros 25 artigos *in Press*, que já haviam sido aceitos e revisados, mas não estavam organizados em algum volume.

A velocidade nas publicações acadêmicas também pode ser observada na multiplicidade de redes sociais disponíveis aos pesquisadores para divulgação de suas investigações. Vários sites, como o *ResearchGate* (<https://www.researchgate.net/home>, recuperado em 25 de agosto, 2017) e o *Mendeley* (<https://www.mendeley.com/research-network/community>, recuperado em 25 de agosto, 2017), permitem criar perfis e comunicar aos colegas quais temas estão sendo estudados, além de compartilhar artigos e livros em formato digital.

Outro exemplo que demonstra a velocidade da geração de informações são as diversas ferramentas online e gratuitas que apresentam dados em tempo real (ou quase). É possível acessar medidas (<http://www.worldometers.info/pt/>, recuperado em 25 de agosto, 2017) e estatísticas demográficas do mundo todo (<http://www.world-statistics.com/>, recuperado em 25 de agosto, 2017). O comportamento do consumidor na internet (<https://www.webpagefx.com/internet-real-time/>,

recuperado em 25 de agosto, 2017) e nas redes sociais (<https://www.tweetping.net/#/>, recuperado em 25 de agosto, 2017) também pode ser visualizado ao vivo.

2.2.3 Variedade

A diferença entre uma grande quantidade de informações, já disponíveis aos pesquisadores de marketing (e.g. métricas de vendas, relatórios de inventários e fluxo de clientes na loja), e o Big Data é a variedade (Bradlow *et al.*, 2017; Wedel & Kannan, 2016). Esses dados armazenados possuem uma rica diversidade e podem estar organizados de maneira estruturada, semiestruturada ou não estruturada (Erevelles *et al.*, 2016; Gandomi & Haider, 2015). As tabelas e planilhas, que representam apenas 5% do total, são os dados estruturados, enquanto os textos, as imagens, os áudios e os vídeos são dados não estruturados (Gandomi & Haider, 2015).

As informações não estruturadas podem ser textos postados no *Twitter* ou no Facebook, *reviews* escritos em sites especializados e *blogs*, além de fotos e vídeos publicados no Instagram e no Youtube (Lam *et al.*, 2017). Em um estudo recente, Liu *et al.* (2016) demonstraram como utilizam as informações das redes sociais para compreender o comportamento do consumidor. Além disso, os autores utilizaram a plataforma Google Trends (<https://trends.google.com.br/trends/>, recuperado em 28 de agosto, 2017) para comparar os assuntos postados nas redes sociais e as buscas feitas na internet.

As postagens nas redes sociais também possibilitam a investigação do comportamento do consumidor por meio de imagens e dados de geolocalização. Miah *et al.* (2017) utilizaram essas informações postadas no

site Flickr (<https://www.flickr.com/>, recuperado em 28 de agosto, 2017) para identificar o comportamento de turistas na Austrália. A **Figura 3** retrata a forma como os resultados da pesquisa foram analisados.

Outros autores, entretanto, acrescentam outras duas características para diferenciar uma grande quantidade de dados do Big Data: a veracidade e o valor (Elragal, 2014; Erevelles *et al.*, 2016; Wedel & Kannan, 2016) big in variety (structured; semi-structured; unstructured). As características são descritas a seguir.

2.2.4 Veracidade

Com a disponibilidade de um grande volume de dados, em grande velocidade e variedade, passou-se a questionar a veracidade das informações coletadas (Gandomi & Haider, 2015; Herschel & Miori, 2017). Para apresentar resultados relevantes, é preciso examinar a consistência, a precisão e a utilidade dos dados analisados (Lam *et al.*, 2017). A credibilidade e a capacidade de interpretação fornecem validade e confiabilidade aos estudos (Lukoianova & Rubin, 2013; Wedel & Kannan, 2016). Além disso, a velocidade na qual os dados são processados pode gerar ruídos, tendências, *outliers* e anormalidades, exigindo uma limpeza antes da análise (Grover & Kar, 2017).

A falta de credibilidade inerente às fontes de informação também desafia a veracidade (Gandomi & Haider, 2015) e exige tecnologias específicas para analisar (Grover & Kar, 2017; McAfee & Brynjolfsson, 2012). Dessa forma, é fundamental verificar a procedência dos dados (Elragal, 2014) para não serem coletados de maneira incompleta ou imprecisa (Herschel & Miori, 2017). Outro aspecto relacionado às fontes de informação é a sua natureza incerta. Os sentimentos dos consumidores expressos nas publicações das redes sociais, por exemplo, são baseados no julgamento humano e exigem ferramentas de análise específicas para gerenciar e minerar informações (Erevelles *et al.*, 2016; Gandomi & Haider, 2015).

No contexto acadêmico, o volume e a velocidade das publicações nem sempre são sinônimos de qualidade dos estudos. Os *journals Quantitative Marketing and Economics (QM&E), Marketing Science (MS), Journal of the Academy of Marketing Science (JAMS), Journal of Supply Chain Management (JSCM), Journal of Public Administration Research and Theory (JPART), Journal of Marketing Research (JMR), Journal of Marketing (JM), Journal of Consumer Research (JCR), Journal of Consumer Psychology (JCP) e Academy of Management Perspectives (AMP)* figuraram entre as primeiras cinco posições do *ranking SCImago*, entre 1999 e 2016. O **Gráfico 3** mostra que, até o ano de 2010, houve um aumento no número de documentos publicados por esses periódicos. Entretanto, de 2011 até 2016, houve um decréscimo.

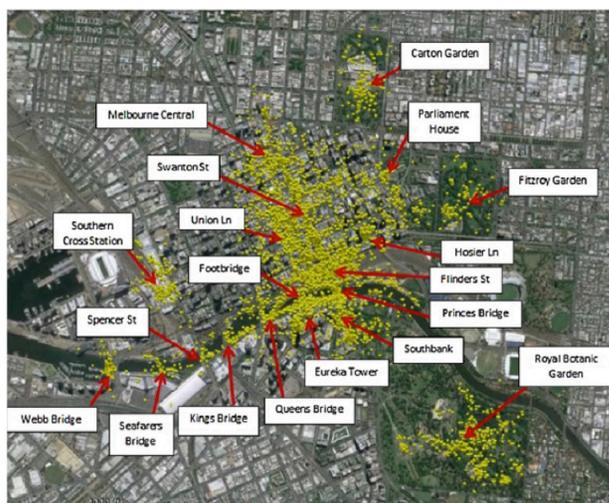
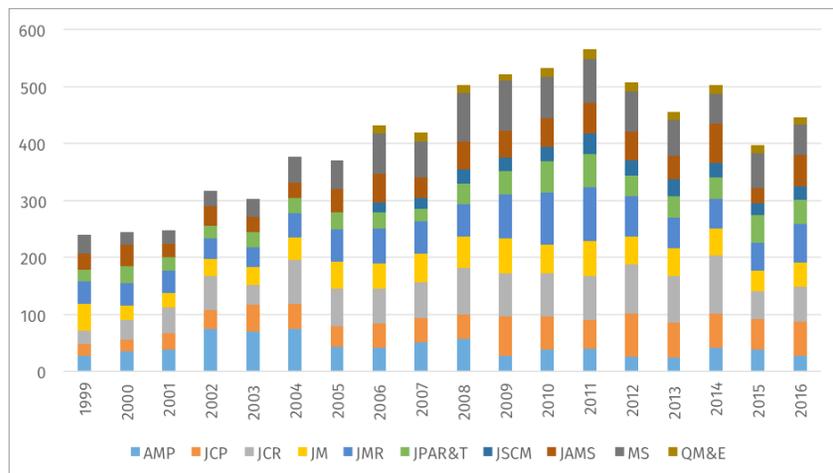


Figura 3. Exemplo de representação visual de dados de geolocalização e imagens

Fonte: Adaptado Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771-785. <https://doi.org/10.1016/j.im.2016.11.011>

Gráfico 3

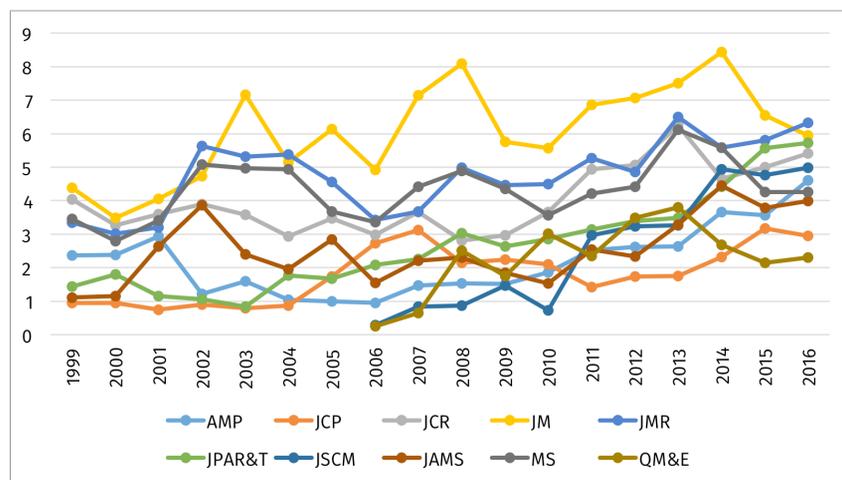
Quantidade de documentos publicados pelos journals que figuram nas cinco primeiras posições do SCImago, por ano



Fonte: Elaborado pelo Autor (2017).

Gráfico 4

Evolução do SJR dos journals que figuram entre as cinco primeiras posições



Fonte: Elaborado pelo Autor (2017).

Com relação ao impacto dessas publicações, calculado pelo SJR, o aumento não foi igual para esses mesmos *journals*. Dessa forma, o **Gráfico 4** demonstra que o período até 2010 (de aumento no número de documentos publicados) não significou crescimento do impacto das publicações.

2.2.5 Valor

O valor é uma característica do Big Data que surge quando os custos de obtenção das informações são menores que os benefícios alcançados (Elragal, 2014). Assim, o valor pode ser entendido como a capacidade de se analisar e interpretar grande volume e variedade de dados, em grande velocidade (Erevelles *et al.*, 2016). Entretanto, quanto maior é o volume de informações, menor é a capacidade de análise e menor é seu valor. Da mesma forma, quanto menor a veracidade dos dados, menor será a sua relevância (Gandomi & Haider, 2015; Sheng *et al.*, 2017).

Para extrair valor dos dados, é preciso identificar informações relevantes (Erevelles *et al.*, 2016; Grover & Kar, 2017). Contudo, a variedade dos dados e a heterogeneidade das fontes dificultam o processamento, pois cada base armazena campos, tipos de dados e metadados diferentes (Grover & Kar, 2017). Dentro do contexto acadêmico, o valor gerado tem relação com as necessidades dos atores envolvidos (outros pesquisadores, empresas, consumidores e sociedade) (Markus & Topi, 2015). Assim, as informações precisam ser transparentes, acessíveis com frequência, mais detalhadas e precisas para poder embasar o desenvolvimento de novas ferramentas, tecnologias e teorias (Elragal, 2014).

2.2.6 Outros V's

Além dos cinco V's (volume, velocidade, variedade, veracidade e valor) elencados na maior parte dos estudos, alguns autores ainda acrescentam como características do Big Data a volatilidade, a visualização e o vocabulário. Apesar de ser, algumas vezes, confundido com a variedade das informações (Grover & Kar, 2017), a volatilidade se refere à variação nas taxas de fluxos de dados (Gandomi & Haider, 2015; Miah *et al.*, 2017). Para Borne (2014), a inconstância desses fluxos também está relacionada aos aspectos dinâmicos, evolutivos, espaço temporais e sazonais dos dados.

A visualização é uma característica que pode ser associada à aplicação de inteligência artificial e algoritmos de aprendizagem para analisar e interpretar as informações (Grover & Kar, 2017). Além disso, alguns autores utilizam essa nova forma de representação gráfica e sintetização dos dados para definir o Big Data. Essas novas maneiras de visualização dos resultados facilitam a análise e a interpretação e podem ter diversas formas. Os mapas digitais, por exemplo, permitem

analisar a renda *per capita* (http://143.107.85.55/demograficas/mapa_renda.asp, recuperado em 27 de agosto, 2017), as etnias (http://143.107.85.55/demograficas/mapa_etnia.asp, recuperado em 27 de agosto, 2017) e as classes socioeconômicas (http://143.107.85.55/demograficas/mapa_estratos.asp, recuperado em 27 de agosto, 2017) da população brasileira. No âmbito acadêmico, é possível localizar publicações e autores por meio do mapeamento (<https://www.authormapper.com/>, recuperado em 27 de agosto, 2017).

O vocabulário é uma característica que tem relação com os modelos e esquemas dos dados, a semântica, a ontologia, as taxonomias e os outros metadados baseados em conteúdo e contexto que descrevem a estrutura, a sintaxe, o conteúdo e a proveniência das informações (Borne, 2014). As informações textuais, *tags* e descrições de imagens e vídeos exigem dicionários específicos para contextualizar e classificar as informações. Entretanto, nenhuma dessas técnicas é capaz de captar completamente o significado contido nos dados não estruturados (Wedel & Kannan, 2016).

Ao identificar as oito principais características identificadas na literatura sobre Big Data (sim, existem mais V's), vislumbram-se as possibilidades que surgem a partir da coleta e da análise dessas informações. Contudo, é possível elencar também os desafios que poderão ser enfrentados pelos pesquisadores ao utilizar esses dados. Na próxima seção, serão apresentadas e discutidas essas possibilidades e desafios.

2.3 Big Data e as Possibilidades de Pesquisa

A primeira possibilidade que surge, de maneira pertinente, da discussão das características do Big Data são as ferramentas de coleta e análise que poderão ser utilizadas. O acesso a informações estatísticas da população tem sido ampliado com a ajuda de governos federais e instituições internacionais. Os Estados Unidos (<https://www.data.gov/>, recuperado em 30 de agosto, 2017) e as Nações Unidas (<https://unstats.un.org/home/>, recuperado em 30 de agosto, 2017) possuem bases de dados que permitem identificar comportamentos dos consumidores e das empresas no mundo todo.

O Google tem várias ferramentas que auxiliam a coletar e analisar os mais diferentes dados. O Analytics, por exemplo, é uma ferramenta que ajuda a encontrar padrões ocultos em dados (Erevelles *et al.*, 2016). Já o Google Trends permite examinar o que os usuários da internet estão buscando (<https://trends.google.com.br/>, recuperado em 30 de agosto, 2017). O Google Correlates (<https://www.google.com/trends/correlate/search?e=beer&t=monthly&p=us>, recuperado em 30 de agosto, 2017), por sua vez, retorna correlações entre os assuntos de maior interesse da população.

O próprio site da *American Marketing Association* (AMA) tem uma lista de softwares que podem ser usados pelos pesquisadores (<https://www.ama.org/academics/Pages/research-tools.aspx>, recuperado em 30 de agosto, 2017). Paralelamente, a internet também tem possibilitado que os próprios autores compartilhem os dados de suas pesquisas e obtenham ajuda de outros pesquisadores para identificar resultados relevantes de suas análises. Um exemplo é a plataforma de compartilhamento de dados do Mendeley (<https://www.elsevier.com/authors/author-services/research-data>, recuperado em 30 de agosto, 2017).

Para extrair informações do Big Data são necessários dois processos: o gerenciamento e a análise. O gerenciamento dos dados requer tecnologias para adquirir e armazenar as informações, que serão preparadas e organizadas para o processo seguinte. Na fase de análise, são usadas técnicas para examinar as informações, adquirir conhecimento e gerar valor (Gandomi & Haider, 2015). Esses dois processos de extração das informações do Big Data são muito semelhantes a duas fases das etapas de qualquer pesquisa acadêmica: a coleta e a análise dos dados (Gil, 2008). A **Figura 4** associa as etapas da investigação científica com os processos do Big Data.

Por fim, mas não menos importante, a seção seguinte deste artigo buscou elencar os principais desafios que o Big Data impõe sobre as pesquisas acadêmicas em marketing.

2.4 Os Desafios do Big Data

As restrições no acesso às tecnologias, a urgência na publicação dos estudos, a falta de treinamento específico dos pesquisadores e as barreiras organizacionais para analisar e interpretar as informações podem apresentar desafios para a utilização do Big Data (Wedel & Kannan, 2016). O grande paradoxo desse fenômeno é que, ao mesmo tempo em que aumenta a disponibilidade de dados, diminui o percentual de dados que podem ser processados e analisados (Dessureault, 2016). Nesse sentido, é preciso salientar que está ocorrendo uma revolução científica e uma alteração na definição de conhecimento (Boyd & Crawford, 2012). Por isso, os pesquisadores de marketing precisam compreender as mudanças nas práticas e nos valores acadêmicos, no papel da teoria e da predição, nas formas de amostragem estatística e teste de hipóteses, na eficácia da revisão por pares e nas políticas de utilização dos dados (Markus & Topi, 2015).

Atualmente, o acesso limitado ao Big Data cria novas divisões digitais (Boyd & Crawford, 2012) e gera disputa entre as empresas e os pesquisadores de marketing. Por outro lado, o grande volume de dados disponíveis, gerados com rapidez, impossibilita uma análise profunda e o conhecimento do estado da arte de um assunto. Por isso, é preciso compreender de maneira holística quais são as possibilidades de uso e os desafios a serem enfrentados com relação ao Big Data. Esses dados podem ser usados por empresas, governos e pesquisadores acadêmicos, mas também podem ser utilizados por qualquer pessoa. Ao mesmo tempo que isso pode ser importante para que os usuários contribuam com informações pessoais e ideias, também existe a possibilidade de usos ilegais, antiéticos e socialmente inaceitáveis (Markus & Topi, 2015).

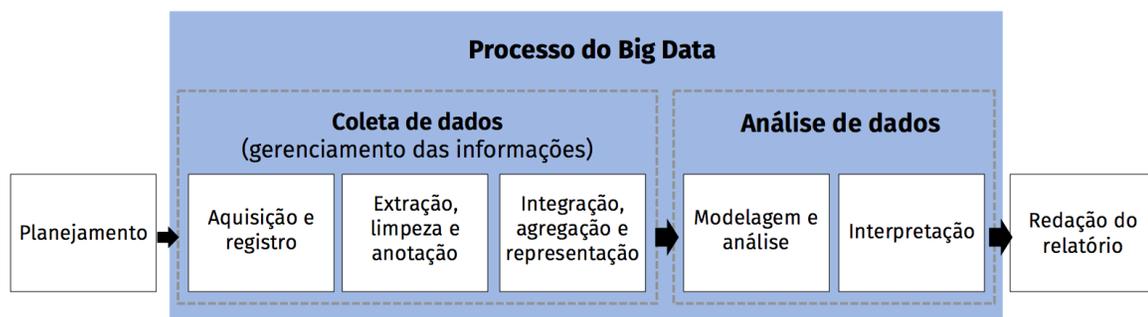


Figura 4. Etapas da pesquisa acadêmica e Big Data
 Fonte: Elaborado pelo Autor (2017).

Os usuários que disponibilizam informações, de maneira pública (ou semipública) nas redes sociais, não permitem, antecipadamente, a utilização irrestrita dos dados. A maioria deles não tem consciência da multiplicidade de agentes e algoritmos que estão reunindo e armazenando suas informações para serem usadas no futuro. É preciso atentar-se para os direitos e para o bem-estar dos informantes. Contudo, a maioria dos conselhos de ética não conhecem os processos de mineração dos dados e o anonimato, não levando em consideração os possíveis erros que podem permitir a identificação dos participantes (Boyd & Crawford, 2012).

Mais especificamente com relação ao âmbito acadêmico, a pressão por publicações pode diminuir a qualidade da produção científica e limitar o tempo dos pesquisadores para aprender novas ferramentas de análise ou desenvolver novos métodos de coleta. Todavia, essa pressão por publicações não melhora a performance acadêmica em marketing (Richard, Plimmer, Fam, & Campbell, 2015). Semelhantemente, a limitação de tempo dos pesquisadores exige a limitação do escopo a ser estudado, levando à especialização em um determinado assunto. Apesar do aprofundamento em um assunto facilitar a análise de Big Data e contribuir para avanços significativos para a ciência, impede que os pesquisadores conheçam novos campos (Wilkie & Moore, 2003).

Dessa forma, é preciso que a academia em marketing discuta as implicações do Big Data para as pesquisas científicas com relação ao desenvolvimento de habilidades para analisar os dados; à melhoria dos métodos de investigação; às questões éticas, jurídicas e sociais no uso das informações; às aplicações comerciais, privacidade, segurança e não discriminação; às melhores práticas para o investimento, implementação e adoção de Big Data, e aos padrões, métodos e responsabilidades de funções de auditoria de algoritmos (Markus & Topi, 2015).

3 Considerações Finais

Este artigo teve como objetivo discutir como o fenômeno do Big Data pode afetar as pesquisas acadêmicas na área de marketing. Ao apresentar as características, as possibilidades e os desafios, foi possível ter uma visão ampla do fenômeno e suas implicações para as investigações no campo. São várias as limitações que podem ser elencadas neste estudo, como por exemplo a falta de uma revisão sistemática de literatura sobre Big Data no marketing. Contudo, este trabalho não utilizou essa abordagem porque precisava de uma liberdade maior para identificar estudos em outras áreas de conhecimento. Futuras pesquisas poderão empreender outras técnicas de revisão da literatura para analisar o estado da arte da utilização de Big Data no marketing.

Outra limitação tem relação com a utilização, principalmente, de artigos internacionais de idioma inglês. Dessa forma, futuras pesquisas poderão aprofundar a discussão sobre o tema na academia de marketing brasileira e as implicações desse fenômeno para os programas de pós-graduação em administração e negócios. A discussão sobre as possibilidades e os desafios do Big Data na academia foi iniciada, e é preciso, a partir de agora, que os pesquisadores debatam sobre as implicações éticas, as limitações técnicas e as aplicações práticas da utilização desses dados.

4 Referências

- Ann Keller, S., Koonin, S. E., & Shipp, S. (2012). Big data and city living — what can it do for us? *Significance*, 9(4), 4–7. <https://doi.org/10.1111/j.1740-9713.2012.00583.x>
- Borne, K. (2014). Top 10 Big Data Challenges—A Serious Look at 10 Big Data V's. Retrieved August 30, 2017, from <https://mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs/>
- Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, 93(1), 79–95. <https://doi.org/10.1016/j.jretai.2016.12.004>
- Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Editorial—Marketing Science and Big Data. *Marketing Science*, 35(3), 341–342. <https://doi.org/10.1287/mksc.2016.0996>
- Dessureault, S. (2016). Understanding big data. *CIM Magazine*, 11(1).
- Elragal, A. (2014). ERP and Big Data: The Inept Couple. *Procedia Technology*, 16(December 2014), 242–249. <https://doi.org/10.1016/j.protcy.2014.10.089>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with Big Data Analytics. *Interactions*, 50–59. Retrieved from <https://pdfs.semanticscholar.org/de29/c51d169cdb5066f6832e9a8878900c43f100.pdf>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

- Gantz, J., & Reinsel, D. (2012). *The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east*. Retrieved from <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social* (6th ed.). São Paulo: Atlas.
- Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, 18(3), 203–229. <https://doi.org/10.1007/s40171-017-0159-3>
- Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, 6(4), 674–688. <https://doi.org/10.1016/j.joi.2012.07.001>
- Herschel, R., & Miori, V. M. (2017). Ethics & Big Data. *Technology in Society*, 49, 31–36. <https://doi.org/10.1016/j.techsoc.2017.03.003>
- Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM*, 52(8), 36. <https://doi.org/10.1145/1536616.1536632>
- Lam, S. K., Sleep, S., Hennig-Thurau, T., Sridhar, S., & Saboo, A. R. (2017). Leveraging Frontline Employees' Small Data and Firm-Level Big Data in Frontline Management. *Journal of Service Research*, 20(1), 12–28. <https://doi.org/10.1177/1094670516679271>
- Liu, X., Singh, P. V., & Srinivasan, K. (2016). A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing. *Marketing Science*, 35(3), 363–388. <https://doi.org/10.1287/mksc.2015.0972>
- Lukoianova, T., & Rubin, V. L. (2013). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances in Classification Research*. Retrieved from https://s3.amazonaws.com/academia.edu.documents/32744189/Lukoianova_Rubin_ADVANCEES_In_CLASS_Workshop_Draft_120113_VR_FINAL.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1504086360&Signature=0dlkLpsAOWP4%2BnXv0tC-72JPhji8%3D&response-content-disposition=inline%3Bfilename%3DVeracity_Roadmap_Is_Big_Data_Objective_T.pdf
- Marketing Science Institute. (2016). Research Priorities 2016–2018. Cambridge: Marketing Science Institute. Retrieved from www.msi.org
- Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business*. New York. Retrieved from <https://www.bentley.edu/files/2015/10/08/Big-DataWorkshopFinalReport.pdf>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, 90(10), 60–68.
- Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771–785. <https://doi.org/10.1016/j.im.2016.11.011>
- Richard, J., Plimmer, G., Fam, K.-S., & Campbell, C. (2015). Publishing success of marketing academics: antecedents and outcomes. *European Journal of Marketing*, 49(1/2), 123–145. <https://doi.org/10.1108/EJM-06-2013-0311>
- Rouse, M. (2016). Big Data. Retrieved August 28, 2017, from <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
- SCImago. (2007). SJR — SCImago Journal & Country Rank. Retrieved August 24, 2017, from <http://www.scimagojr.com>
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191, 97–112. <https://doi.org/10.1016/j.ijpe.2017.06.006>
- Singh Jain, A. D., Mehta, I., Mitra, J., & Agrawal, S. (2017). Application of Big Data in Supply Chain Management. *Materials Today: Proceedings*, 4(2), 1106–1115. <https://doi.org/10.1016/j.matpr.2017.01.126>
- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How “big data” can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234–246. <https://doi.org/10.1016/j.ijpe.2014.12.031>
- Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6), 97–121. <https://doi.org/10.1509/jm.15.0413>
- Wilkie, W. L., & Moore, E. S. (2003). Scholarly research in marketing: Exploring the “4 Eras” of thought development. *Journal of Public Policy & Marketing*, 22(2), 116–146. <https://doi.org/10.1509/jppm.22.2.116.17639>
- Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566. <https://doi.org/10.1016/j.jbusres.2015.10.017>