

## **INTELIGÊNCIA ANALÍTICA: UM ESTUDO BIBLIOMÉTRICO SOBRE A PRODUÇÃO CIENTÍFICA**

**PEDRO HENRIQUE GOMES RUGGIERO**

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

**GUSTAVO HERMÍNIO SALATI MARCONDES DE MORAES**

UNIVERSIDADE ESTADUAL DE CAMPINAS (UNICAMP)

**FERNANDO DE SOUZA MEIRELLES**

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

# INTELIGÊNCIA ANALÍTICA: UM ESTUDO BIBLIOMÉTRICO SOBRE A PRODUÇÃO CIENTÍFICA

## 1 INTRODUÇÃO

As ferramentas de *Business Intelligence & Analytics* e o campo relacionado à *Big Data Analytics* têm se tornado importantes para comunidades acadêmicas e também para o mercado (Chen, Chiang, & Storey, 2012). A área de ciência dos dados envolve termos como *Analytics*, *Big Data*, *Big Data Analytics*, *Business Analytics*, *Business Intelligence*, *Business Intelligence & Analytics* e Inteligência Analítica, e ainda não há um consenso entre a separação dos termos, com interpretações distintas entre os autores (Meirelles, 2018; Wamba *et al.*, 2015).

A partir do início dos anos 2000, a publicação de textos na área de ciências dos dados, com a utilização de métodos analíticos e tecnologia da informação (TIC), vem se consolidando e centenas de textos acadêmicos e profissionais foram publicados (Goes, 2014).

O uso extensivo de dados utilizando modelos estatísticos e análises quantitativas, modelos de análise preditiva e exploratória, decisões e ações gerenciais baseadas em fatos, pode ser definido como *Analytics*, que é um subconjunto do que é intitulado *Business Intelligence* (Davenport & Harris, 2007).

*Big Data* pode ser definido como o fenômeno do aumento da oferta e demanda de dados e de ferramentas para manejá-los para apoio a tomada de decisão (Luvizan, Meirelles, & Diniz, 2014), o que parece bastante relacionado com o conceito de *Business Intelligence*, que é um conjunto de tecnologias e processos utilizados para entender e analisar a performance de um negócio (Davenport & Harris, 2007). Enquanto o conceito *Business Analytics*, pode ser considerado como o componente analítico de *Business Intelligence* (Chen *et al.*, 2012).

Os termos *Analytics*, *Big Data* e *Business Intelligence* têm, frequentemente, se misturado, sendo muitas vezes utilizados em conjunto em pesquisas sobre o fenômeno (Chen, 2012). *Big Data* é um conceito relativo, pois está relacionado a grandes volumes de dados, mas para que um volume de dados seja considerado grande, o contexto (como empresa, projeto e até momento) em que está sendo utilizado tem que ser levado em consideração (Rousseau, 2012).

Além de *Business Analytics*, outros termos são utilizados no mercado e também em pesquisas científicas. Na pesquisa anual de Administração e Uso da Tecnologia da Informação (TI) realizada pela Fundação Getúlio Vargas (FGV) com mais de 2.500 empresas participantes, o termo *Business Intelligence* é destacado como um termo utilizado comercialmente (Meirelles, 2018). Desde os anos 2000, com a crescente popularização do termo *Big Data*, uma das questões centrais desse tema é o esclarecimento sobre a sua definição, pois é possível perceber que existem várias interpretações sendo apresentadas (Wamba *et al.*, 2015).

A grande quantidade de termos dificulta uma separação entre os rótulos comerciais da taxonomia já consolidada, e em uma tentativa de consolidar todos esses termos, Meirelles (2018) define Inteligência Analítica como um componente que inclui: *Analytics*, *Business Analytics*, *Business Intelligence*, *Business Simulation*, *Customer Relationship Management*, *Executive Information System*, *Extreme Analytics*, *Predictive Analytics*, *Scenario Modeling*, entre outros rótulos comerciais. Na presente pesquisa, o termo Inteligência Analítica (IA) será utilizado como integrando os conceitos envolvidos nessa nova área de pesquisa.

Já existem estudos que sugerem relação entre a *performance* das grandes empresas e o uso de Inteligência Analítica (Trkman, McCormack, Oliveira, & Ladeira, 2010; LaValle, Lesse, Shockley, Hopkins, & Kruschwitz, 2011). As oportunidades associadas à utilização de dados em diferentes organizações geram um crescente interesse em Inteligência Analítica, e essa utilização pode estar associada a técnicas, tecnologias, sistemas, práticas, metodologias e

aplicativos, que auxiliam na análise de dados críticos, contribuindo para a tomada de decisões com um maior entendimento do negócio (Chen *et al.*, 2012).

Em relação ao mercado de inteligência analítica, as vendas globais de software de *Business Intelligence & Analytics* alcançaram US\$ 18,3 bilhões em 2017, o que representa um aumento de 7,3% em relação a 2016, e espera-se que até o final de 2020 o mercado total seja de US\$ 22,8 bilhões (Sallam *et al.*, 2017). O mercado de plataformas analíticas para Inteligência Analítica é altamente competitivo, com um número de fornecedores cada vez maior, com produtos e tecnologias com funcionalidades cada vez mais ricas, e com os preços das soluções tradicionais de BI em queda (Klisarova-Belcheva, Ilieva, & Yankova, 2017).

## 2 PROBLEMA DE PESQUISA E OBJETIVO

Dada a importância do tema e a abrangência de assuntos correlatos, o objetivo deste artigo é mapear, por meio de um estudo bibliográfico, artigos, autores e periódicos que estão colaborando cientificamente no tema da Inteligência Analítica. A pergunta de pesquisa é “Quais são as principais contribuições teóricas do campo de Inteligência Analítica?” Para isso, a pesquisa tem como objetivos específicos, identificar:

- a) O *Main Path* da temática;
- b) Os artigos e os autores mais citados no tema;
- c) Os principais periódicos que publicam artigos sobre o tema.

Em uma pesquisa bibliométrica, no campo da análise de citações, identificar o *Main Path* significa explicitar os textos essenciais de uma área de conhecimento (Georgiou, 2014). O *Main Path* (caminho principal cronologicamente ordenado) pode ser visto como um método que rastreia o conhecimento que se mantém em uma literatura ao longo do tempo (Hummon & Doreian, 1989). Em uma literatura há textos de diferentes tipos e com diferentes objetivos. Calcular o *Main Path* é uma maneira de definir os textos mais relevantes em uma literatura a partir das citações que esses textos receberam.

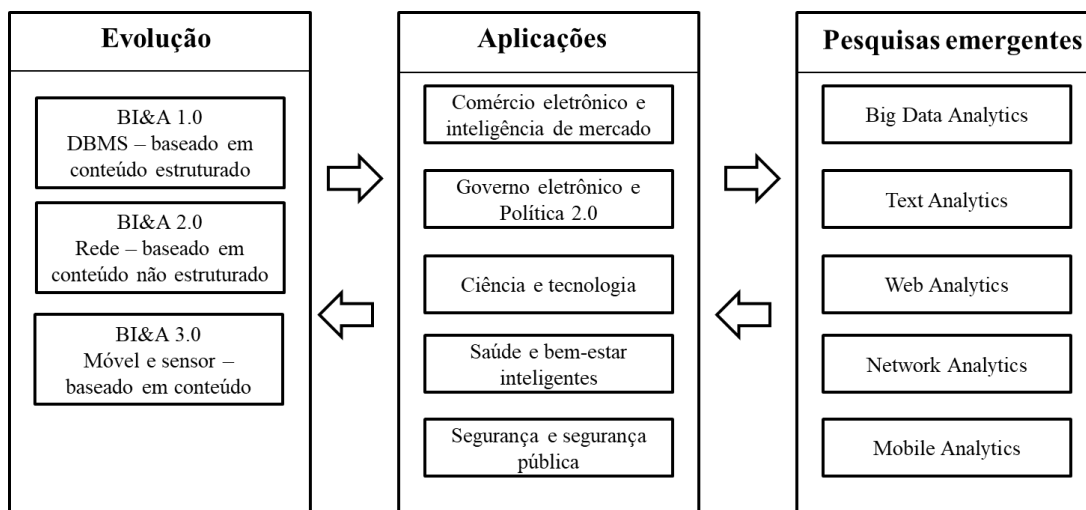
## 3 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica do artigo está concentrada em dois pontos, sendo o primeiro uma apresentação da evolução dos conceitos e aplicações em Inteligência Analítica e o segundo a apresentação de tecnologias fundamentais e pesquisas emergentes sobre o tema.

### 3.1 A evolução da Inteligência Analítica

Para apresentar a evolução da Inteligência Analítica, é necessário começar com o termo mais conhecido, que é o *Big Data*. Apesar da nomenclatura remeter ao tamanho dos dados, o que é mais importante nesse conceito é a diversidade dos dados existentes. Os dados existentes em uma análise de *Big Data* podem ser brutos, dispersos em diversos formatos, estruturados, semiestruturados e desestruturados (Kimball & Ross, 2013).

A evolução dos sistemas e aplicações de *Big Data*, bem como das pesquisas emergentes sobre o tema, foi apresentada por Chen *et al.* (2012), e a Figura 1 apresenta um resumo dessas informações.



**Figura 1.** A Evolução do *Business Intelligence & Analytics*  
**Fonte:** Chen et al. (2012)

O *Business Intelligence & Analytics* 1.0 (BI&A 1.0) teve início com as primeiras aplicações que coletavam e concentravam informação que eram originadas de sistemas legados distintos. No BI&A 1.0, a tecnologia utilizada foi desenvolvida nas décadas anteriores, com técnicas analíticas, de *Data Mining* e modelos estatísticos. Os sistemas existentes utilizavam tecnologia de extração, carga e transformação para integrar os dados, com ferramentas para acompanhamento de métricas e de *Data Mining*, permitindo análises por associação, separação por *clusters*, classificação e análise de regressão, detecção de anomalias em padrões e criação de modelos de análise preditiva.

Em um segundo momento, surge o *Business Intelligence & Analytics* 2.0 (BI&A 2.0), desenvolvido a partir da utilização da internet, com o início do comércio eletrônico e a coleta de informações a partir dos sites, no início dos anos 2000. As fontes de informações do BI&A 2.0 eram os *cookies*, históricos de navegação e de compras. Com a evolução da utilização da internet, surgem também os sistemas para análise das informações das mídias sócias, *blogs* e fóruns (Chen *et al.*, 2012).

Por fim, surge o *Business Intelligence & Analytics* 3.0 (BI&A 3.0), com o crescimento do número de *smartphones* e *tablets*. A quantidade de dispositivos móveis no Brasil já chega a 284 milhões, e no mundo esse número é de 8,6 bilhões (Meirelles, 2018). Esses novos dispositivos aumentaram a quantidade de informações disponíveis e a necessidade de tecnologias mais sofisticadas para análise. Atualmente, os dados utilizados também são provenientes de dispositivos de *Internet of Things* (IoT) e *Radio-Frequency IDentification* (RFID), e as pesquisas existentes sobre BI&A 3.0 ainda estão em estágios iniciais (Chen *et al.*, 2012).

### 3.2 Tecnologias fundamentais e pesquisas emergentes em Inteligência Analítica

Para que os avanços nas pesquisas sobre Inteligência Analítica continuem, é necessário estimular pesquisas sobre as principais tecnologias que fazem parte dessa área. Existe uma grande quantidade de tópicos a serem abordados pelos pesquisadores, mas como forma de auxiliar os estudiosos do tema a se concentrarem nos principais, Chen *et al.* (2012) listou as cinco áreas técnicas críticas para o desenvolvimento da Inteligência Analítica, que também são apresentadas na Figura 1, na coluna de pesquisas emergentes. São elas: *Big data analytics*; *Text analytics*; *Web analytics*; *Network analytics*; e *Mobile Analytics*. Essas pesquisas podem contribuir para o BI&A 1.0, BI&A 2.0 e BI&A 3.0. A seguir, é apresentada a classificação

destas áreas técnicas, destacando as principais características e possibilidades de pesquisas para cada área.

- a) *Big Data Analytics*: refere-se às tecnologias de Inteligência Analítica baseadas em mineração de dados e análise estatística (Chen *et al.*, 2012). A maioria dessas técnicas é baseada em tecnologias comerciais maduras de SGBD relacional, armazenamento de dados, ETL, OLAP e BPM (Chaudhuri *et al.* 2011).

Ainda há espaço para utilização e pesquisa em novas técnicas de análise de dados que exploram e aproveitam características exclusivas dos dados, com mineração de dados para fluxos de dados de alta velocidade e dados de sensores, por exemplo. Outra oportunidade é em relação aos problemas de privacidade, que preservam a mineração de dados como uma área emergente (Gelfand, 2012).

- b) *Text Analytics*: a grande maioria dos conteúdos não estruturados coletados por uma organização está em formato textual, que podem ser, por exemplo, de comunicação por e-mail, documentos corporativos para páginas da Web e conteúdo de mídia (Chen *et al.*, 2012). Os modernos mecanismos de pesquisa, sistemas de busca de bibliotecas e sistemas de pesquisa corporativa, se utilizam de tecnologia de *text analytics* (Salton, 1989).

As tecnologias de análise de texto também oferecem oportunidades de pesquisa e desafios significativos em várias áreas mais focadas, incluindo *web stylometric*, análise para atribuição de autoria, análise multilíngue para documentos da web e visualização de texto em larga escala.

- c) *Web analytics*: baseada nos fundamentos de mineração de dados e análise estatística de análise de dados, além de modelos de recuperação de informações e PNL na análise de texto (Chen *et al.*, 2012). A análise da web se tornou ainda mais interessante com a maturidade e a popularidade dos serviços da web e dos sistemas da Web 2.0 em meados dos anos 2000 (O'Reilly, 2005).

As principais oportunidades de pesquisa de análise da web estão no desenvolvimento de plataformas e serviços de computação em nuvem, que incluem aplicativos, software de sistema e hardware entregues como serviços pela internet. Outras possibilidades estão na pesquisa social e mineração, sistemas de reputação, análise de mídia e visualização na web.

- d) *Network analytics*: é uma área de pesquisa nova que evoluiu da análise bibliométrica baseada em citações para incluir novos modelos computacionais para comunidade *online* e análise de rede (Chen *et al.*, 2012). Fundamentado na análise bibliométrica, a citação de redes e de redes de coautoria foram adotadas há muito tempo para examinar o impacto científico e a difusão do conhecimento. As pesquisas recentes têm se concentrado em mineração de link e detecção de comunidade. Ferramentas como o UCINet (Borgatti, Everett, & Freeman, 2002) e Pajek (Batagelj & Mrvar, 1998) foram desenvolvidas e são amplamente utilizadas para análise de rede em grande escala e para visualização de redes.

- e) *Mobile analytics*: a computação móvel oferece formas para o crescimento profissional de TI, à medida em que cada vez mais organizações desenvolvem suas aplicações. Com sua grande e crescente base de instalação global, o Android tem sido classificado como a melhor plataforma móvel desde 2010 (Chen *et al.*, 2012).

As pesquisas nessas tecnologias móveis abordam diferentes áreas, como: aplicativos de sensoriamento sensível à localização e sensível a atividades; inovação social móvel para saúde e aprendizagem; redes sociais móveis; *crowdsourcing*; visualização móvel; e personalização; modelagem comportamental para aplicativos móveis; modelos sociais, comportamentais e econômicos para a gamificação; publicidade móvel; e marketing social.

## 4 ASPECTOS METODOLÓGICOS

A pesquisa foi desenvolvida por meio de uma revisão bibliométrica, que é uma técnica que explora e analisa estatisticamente, com métodos quantitativos, a literatura e publicações acadêmicas sobre um termo em um determinado espaço de tempo (Alvarado, 2007; Silva, Hayashi, & Hayashi, 2011). A base teórica para a construção do conhecimento e a indicação dos trabalhos que estão sendo produzidos em uma área é encontrada na literatura por meio da pesquisa bibliométrica (Alvarado, 2007; Silva, Hayashi, & Hayashi, 2011), que compreende em um processo replicável, transparente e detalhado, que auxilia os pesquisadores da área nos trabalhos futuros (Crossan & Apaydin, 2010).

Assim, a pesquisa pode ser caracterizada como um estudo bibliométrico, de natureza exploratória, descritiva e transversal, permitindo avaliar as seguintes características das publicações selecionadas: *Main Path* da temática; artigos e autores mais citados no tema; principais periódicos que publicam artigos sobre o tema.

O primeiro passo foi definir os termos chave que representavam a literatura a ser estudada. “*Business Intelligence*”, “*Business Analytics*” e “*Big Data*” são termos considerados como principais descritores da área de Inteligência Analítica (Chen *et al.*, 2012). O primeiro registro de uso do termo *Business Intelligence* (‘BI’) é de 1958 (Luhn, 1958) no *IBM Journal of Research and Development*. Neste texto define-se que BI é um conjunto de atividades relacionadas à habilidade de perceber inter-relacionamentos entre fatos apresentados de forma a guiar ações em direção a determinado objetivo.

Para entender os relacionamentos dos textos na base estudada realizou-se uma análise de citações. Uma análise de citações consiste em verificar todas as seções de referências de um conjunto de textos que caracteriza uma área de conhecimento, para encontrar que texto cita o outro com o objetivo de construir uma rede que permita entender esta área. Segundo Garfield (1970), o objetivo de uma análise de citações é realizar medidas em determinada área científica, um processo geralmente lento. Com um número reduzido de textos (até 50 textos) é possível realizar este processo manualmente. Entretanto, o processo manual torna-se inadequado quando há mais textos a analisar, e uma área de conhecimento pode reunir centenas ou milhares de trabalhos científicos. Para Garfield (1979), uma análise de citações é uma medida genérica do nível de contribuição individual de um pesquisador, texto ou *journal* em determinado assunto, não respondendo perguntas como quem é mais importante, ou quem é melhor. Citações contêm importantes informações de como o conhecimento se dissemina (Liu, Lu, Lu, & Lin, 2013).

Ao estudar quais textos foram mais influentes ao longo do tempo e o porquê, tem-se uma boa estimativa de quais devem ser lidos para uma imersão mais informada na área de Inteligência Analítica.

Os procedimentos metodológicos da análise de citações seguiram os passos propostos por Cobo *et al.*, (2011): a) definir uma base de dados para extrair as informações sobre os textos; b) definir critérios de busca nessa base; c) examinar os resultados da busca nessa base; d) verificar e corrigir inconsistências de forma a definir a base de textos; e) escolher um *software* para análise de citações; f) construir o *input* para este *software* e realizar o processamento.

Algumas bases de dados são consideradas clássicas no contexto de análise de citações, pois contêm grande parte dos textos de interesse acadêmico (seja no padrão acadêmico ou em padrões menos rigorosos) e permitem exportar informações. As principais são *ISI Web of Science*, *Scopus*, *Google Scholar* e *NLM’s MEDLINE* (Cobo *et al.*, 2011).

Diversos aspectos devem ser considerados no momento da escolha de qual base ou quais bases serão utilizadas, mas há dois que merecem destaque. O primeiro deles é que tipo de textos a base abrange. Nem todas as bases possuem todos os textos, por isso é apropriado buscar na literatura quais bases são mais adequadas a cada tema, de forma a obter um conjunto representativo de textos. Trabalhar com um conjunto incompleto de textos pode permitir

conclusões que não correspondem à realidade da área. O segundo deles é que tipo de exportação a base permite: no momento de definir quais textos representam uma literatura, pode-se chegar a um número alto de textos que inviabilize o processo manual de verificação de citações.

Assim como a seleção incoerente da base, a seleção incoerente de critérios de busca pode excluir textos importantes para determinada área, invalidando análises e conclusões. Novamente, indica-se a busca em literatura para encontrar os melhores termos que definam determinada área.

O próximo passo é a realização de análises de consistência e correções, conhecido como pré-processamento de dados (Cobo *et al.*, 2011). Essa etapa consiste em detectar informações duplicadas ou incompletas e pode demandar um grande esforço dependendo da base de dados de origem. Cada base de dados apresenta particularidades. O próximo passo é a seleção de *software* para análise de citações.

Há diversidade de *software* para realizar análises de citações. Alguns deles são (Cobo *et al.*, 2011): *Bibexcel*, *CiteSpace II*, *Sci2Tool*, *VantagePoint*, etc. Além dos programas específicos para análises de citações, existem outros mais matemáticos, utilizados em diversos tipos de redes (não apenas de citações), por exemplo o *software Pajek*. A escolha do *software* deve ser orientada não somente pela facilidade de uso e pelas operações disponíveis, mas também pela capacidade de representação gráfica das redes (que permite que o ser humano entenda as relações de interesse). Utilizar um *software* matemático como o *Pajek* pode ser interessante, pois permite operações matemáticas avançadas e tratamentos mais específicos em uma rede (Batagelj & Mrvar, 1998). Isso garante algumas vantagens ao pesquisador, como ter maior imparcialidade ao realizar operações sobre a rede.

Na sequência, realizou-se a leitura na íntegra dos textos do *Main Path* e o resumo dos tópicos principais tem a finalidade de indicar ao interessado em Inteligência Analítica que textos devem ser lidos e por que motivos.

O *Main Path* pode ser utilizado para a representação de uma área de conhecimento (Georgiou, 2014), que consiste em explicitar textos essenciais. O *Main Path* é um ponto de partida para pesquisadores, pois é um método que rastreia o conhecimento que se mantém em uma literatura ao longo do tempo (Hummon & Doreian, 1989). Uma alternativa à construção de uma rede de citações e a realização da sua análise a partir dessa rede é utilizar o *Science Citation Index*, já existente e bem conhecido (Garfield, 1970). Porém, os resultados são mais genéricos e menos conclusivos dos que os obtidos a partir de uma rede específica do assunto.

#### 4.1 Definição da base de dados e dos critérios de busca

Para este trabalho optou-se pelo uso da base *ISI Web of Science*. O uso dessa base é justificado por Liu *et al.*, (2013) e atende ao propósito de testar o método de pesquisa numa base definida e de fácil extração dos dados. A *ISI Web of Science* contém mais de 10.000 *journals* e materiais de mais de 120.000 congressos. Adicionalmente, exportaram-se informações importantes para a verificação automatizada de citações (através de programação), como a lista de referências de um texto, por exemplo.

O artigo “*Business Intelligence and Analytics: from Big Data to Big Impact*” (Chen *et al.*, 2012), publicado em 2012 no periódico *MIS Quarterly*, foi escolhido como referência para a definição das palavras-chave que orientaram a busca dos textos. Havia a percepção prévia de que era um artigo importante, muito citado e durante a realização deste trabalho confirmou-se ser também o mais citado da rede construída. Nesse trabalho, os autores fizeram uma busca em diversas bases, utilizando como critério os termos “*Business Intelligence*”, “*Business Analytics*” ou “*Big Data*” (através dos campos título, *abstract* e *keywords*) e obtiveram cerca de 3600 registros. Assim, para refinar a pesquisa, foi adicionado o termo “*Analytics*” aos termos anteriores, o que resultou no seguinte critério:  $TS=(“BUSINESS ANALYTICS”) OR TS=(“BIG$

*DATA" AND ANALYTICS) OR TS=("BUSINESS INTELLIGENCE" AND ANALYTICS)*, em textos publicados até 31 de Março de 2016, sem restrição inicial de tempo. Foram considerados todos os registros de textos encontrados. Com essa segunda estratégia, foram obtidos 1560 registros, dos quais 1234 estão na Principal Coleção *Web of Science* (contém todas as informações necessárias para automatizar parte da construção da rede), que foi o conjunto considerado para as próximas etapas.

#### **4.2 Exame dos resultados da busca e verificação e correção de inconsistências**

Para realizar a análise, foram consideradas as informações exportadas do *ISI Web of Science*: título, local de publicação, lista de autores, lista de referências, *DOI* (*digital object identifier*) e ano de publicação. Entre os textos, 1234 continham título, local de publicação, lista de autores e ano de publicação. Por outro lado, 799 continham *DOI* (identificador único na base) e 1155 continham lista de referências. A ausência da lista de referências indica que o texto não é acadêmico.

Sempre que um texto da Principal Coleção *Web of Science* possui *DOI*, ele aparece nas listas de citação de outros textos com esse identificador, sendo esse o campo utilizado para cruzamentos. Porém, para os casos em que o *DOI* não existe, o padrão das listas de citação é PRIMEIRO\_AUTOR, ANO\_DE\_PUBLICAÇÃO, sendo esta informação a segunda maneira de realizar cruzamentos. Através do *MS Excel* cruzou-se os 1234 textos, totalizando 1.522.756 comparações.

Houve 700 ocorrências de citação, ou seja, quase 700 relacionamentos “um citou o outro”. Algumas ocorrências eram inválidas por duplicações nas citações de textos identificados pela mesma dupla PRIMEIRO\_AUTOR, ANO\_DE\_PUBLICAÇÃO. Após esta análise foram mantidas 634 ocorrências válidas.

A seguir, verificou-se que 444 dos 1234 textos foram responsáveis pelas 634 ocorrências de citação, sendo que apenas esses foram considerados para construção da rede de citações. Os demais não citam outros textos e nem são citados.

#### **4.3 Escolha do *software*, construção do *input* e processamento**

Optou-se pelo uso do *software Pajek*, por permitir operações matemáticas avançadas e tratamentos mais específicos em uma rede (Batagelj & Mrvar, 1998), como o cálculo do *Main Path*, por exemplo, além da representação gráfica.

O *input* foi construído considerando-se os 444 textos e os 634 relacionamentos de citação que foram encontrados. O processamento no *software* consistiu em dois grupos de operações: análise dos textos mais citados e cálculo do *Main Path*. Os mais citados foram encontrados através da criação de uma partição com o comando *Pajek Net/Partitions/Degree/Input*. O *Main Path* foi encontrado com o comando *Pajek Net/Citation Weights/Search Path Count (SPC)*.

Explicar a maneira que tais métodos matemáticos funcionam é complexo e não é escopo deste trabalho, que foca no conceito e na utilização.

#### **4.4 Leitura dos textos *Main Path***

A partir do *Main Path* obtido na etapa anterior realizou-se uma leitura detalhada de cada texto que a ele pertence na tentativa de resumir para os leitores o conteúdo e alguns tópicos relevantes.

Os textos mais citados também foram analisados. A maioria desses textos aparece no *Main Path*. Os 4 textos entre os 10 mais citados e que não aparecem no *Main Path*, também foram lidos na íntegra e resumidos.



Em seguida, foram analisados os *journals* mais citados, de forma a evidenciar a quais fontes de textos os pesquisadores devem dar prioridade quando pesquisando no tema de Inteligência Analítica.

## 5 APRESENTAÇÃO DOS RESULTADOS

A matriz de citações é muito esparsa: de  $444 \times 444 = 197.136$  possíveis relacionamentos de citação, apenas 634 foram verificados, ou seja, aproximadamente 0,32% das combinações possíveis. Esse fato indica que os pesquisadores da área de Inteligência Analítica ainda não estabeleceram um diálogo por meio da literatura no tema. Ao invés disso, parecem estar se utilizando de teorias e conhecimentos de outras áreas.

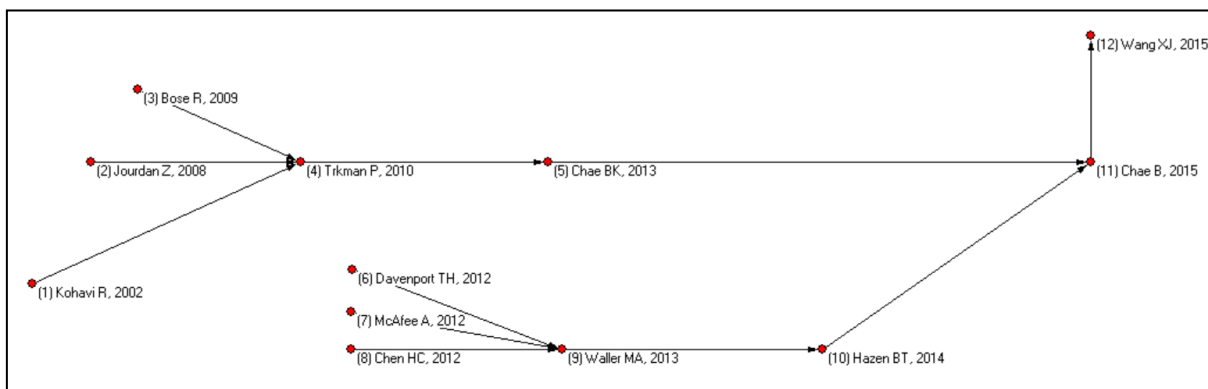
O *Main Path* calculado consiste em 24 textos. Foram lidos os 24 textos do *Main Path* em profundidade, de forma a entender o conteúdo e as convergências e divergências dos textos. Adicionalmente, foram lidos 4 textos que estão entre os 10 mais citados e que não apareceram no *Main Path* (os outros 6 fazem parte do *Main Path* e também do grupo dos 10 mais citados).

### 5.1 *Main Path*

O cálculo *Main Path* resultou em 24 textos. Dos 24 textos, 12 foram desconsiderados, apesar de terem relevância matemática na rede de citações. Os 12 desconsiderados se dividem em dois grupos: 8 por terem aparecido no *Main Path*, por serem dos mesmos *journals* e por se co-citarem, de forma que tornaram importantes matematicamente, mas sem relevância no contexto substantivo de Inteligência Analítica; 4 que citaram os textos possuem relevância matemática, mas que não possuem relevância para o objetivo do trabalho, ou seja, matematicamente importantes por citarem os textos específicos que geram mais peso no cálculo.

É importante destacar que a desconsideração dos 12 textos ocorreu após o cálculo do *Main Path*, uma vez que a remoção destes textos da base de dados traria alterações que distorceriam o resultado, pois com uma base de dados diferente, há uma grande chance do cálculo do *Main Path* resultar em outro conjunto, principalmente se forem eliminados textos com grande peso em termos de citações, que é o caso.

O *Main Path* de Inteligência Analítica, então, é composto de 12 textos (Figura 2). Seus títulos, locais de publicação e anos de publicação estão listados na Tabela 1.



**Figura 2.** *Main Path* das pesquisas de Inteligência Analítica

**Fonte:** elaborado pelos autores

Conforme ilustrado, o *Main Path* de Inteligência Analítica possui dois caminhos principais (caminho superior e caminho inferior). Ambos se iniciam com 3 textos cada e

somente se encontram em 2015, o que pode ser interpretado como uma das poucas evidências de tendência futura de diálogo na literatura.

O caminho superior é composto de 5 textos, dos quais 3 podem ser vistos como iniciais e estão em padrão acadêmico. Note-se que o texto #1 (Kohavi, Rothleder, & Simoudis, 2002), que é o mais antigo da base de dados, aparece nesse caminho. Como o *Main Path* é um caminho cronológico na literatura, o primeiro que foi citado deve aparecer no resultado. Este texto é o mais citado entre o início de 2000 e o final de 2009, o que destaca sua importância para a literatura de Inteligência Analítica.

Sobre os 3 textos iniciais do caminho superior, o texto #1 relata o passado recente (em relação a 2002) sobre o tema Inteligência Analítica, o que se considera importante para conhecimento da história do assunto; e relata possibilidades que os autores do texto enxergavam para o futuro da área: é interessante perceber que houve evolução, mas que muitas tendências apontadas à época ainda não se concretizaram. O texto #2 (Jourdan *et al.*, 2008) é uma análise de literatura com o tema BI que contempla 167 artigos científicos dos principais *journals* da área de *Information Systems* entre os anos de 1997 e 2006, sendo esta análise também valiosa para o entendimento da história do tema. O texto #3 (Bose, 2009) é uma extensa revisão de literatura, e exibe um esquema para ajudar gestores de negócios a entenderem as tendências e avanços de ferramentas ligadas ao Inteligência Analítica.

Sobre os outros 2 textos do caminho superior, o texto #4 (Trkman *et al.*, 2010) traz uma operacionalização de Inteligência Analítica na área de *Supply Chain* através de uma *survey* com 310 empresas de diversos países. Neste texto tenta-se relacionar a performance do negócio com a utilização de Inteligência Analítica. O texto #5 (Chae & Olson, 2013) é similar, pois trata de uma proposta para gerentes de *Supply Chain* maximizarem chances de entendimento de Inteligência Analítica e de obterem bons resultados.

Os textos #1, #2, #3 #4 e #5 ilustram diferentes momentos da literatura, que se inicia na definição de conceitos e continua em aplicações e necessidades para maximização de resultados.

Similarmente, o caminho inferior possui 3 textos iniciais. Porém, como os textos são mais recentes (todos de 2012), há neles uma mescla de conceitos e de formas de operacionalização desses conceitos. O texto #6 (Davenport & Patil, 2012) contém informações sobre *Data Scientists* e alerta sobre a ausência de profissionais qualificados para tal posição. O texto #7 (Mcafee & Brynjolfsson, 2012) traz explicações de conceitos relacionados ao Inteligência Analítica, porém com linguagem mais gerencial; são esclarecidos os conceitos básicos de *Big Data* e discutidos os principais mitos ligados ao seu uso. Com maior destaque por ser o mais citado de toda a rede de citações, o texto #8 (Chen *et al.*, 2012) pode ser considerado como a principal recomendação de leitura para aqueles que estão investigando o tema.

**Tabela 1.**

Textos que compõem o Main Path

#	Informações sobre o texto	Nº de vezes em que citou	Nº de vezes em que foi citado
1	<b>Kohavi R, 2002.</b> Emerging Trends in Business Analytics. <i>COMMUNICATIONS OF THE ACM.</i>	0	15
2	<b>Jourdan Z, 2008.</b> Business Intelligence: An Analysis of the Literature 1. <i>INFORMATION SYSTEMS MANAGEMENT.</i>	0	11
3	<b>Bose R, 2009.</b> Advanced analytics: opportunities and challenges. <i>INDUSTRIAL MANAGEMENT &amp; DATA SYSTEMS.</i>	0	10
4	<b>Trkman P, 2010.</b> The impact of business analytics on supply chain performance. <i>DECISION SUPPORT SYSTEMS.</i>	4	14

5	<b>Chae BK, 2013.</b> Business Analytics for Supply Chain: A Dynamic - Capabilities Framework. <i>INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY &amp; DECISION MAKING.</i>	3	2
6	<b>Davenport TH, 2012.</b> Data Scientist: The Sexiest Job of the 21st Century. <i>HARVARD BUSINESS REVIEW.</i>	0	30
7	<b>McAfee A, 2012.</b> Big data. The Management Revolution. <i>HARVARD BUSINESS REVIEW.</i>	0	31
8	<b>Chen HC, 2012.</b> Business Intelligence and Analytics: From Big Data to Big Impact. <i>MIS QUARTERLY.</i>	0	77
9	<b>Waller MA, 2013.</b> Data Science, Predictive Analytics, and Big Data: a Revolution That Will Transform Supply Chain Design and Management. <i>JOURNAL OF BUSINESS LOGISTICS.</i>	3	15
10	<b>Hazen BT, 2014.</b> Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. <i>INTERNATIONAL JOURNAL OF PRODUCTION ECONOMICS.</i>	3	3
11	<b>Chae B, 2015.</b> Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. <i>INTERNATIONAL JOURNAL OF PRODUCTION ECONOMICS.</i>	5	2
12	<b>Wang XJ, 2015.</b> Big data research for the knowledge economy: past, present, and future. <i>INDUSTRIAL MANAGEMENT &amp; DATA SYSTEMS.</i>	12	0

**Fonte:** Elaborado pelos autores

O texto #8 contém a história cronológica de Inteligência Analítica, aplicações e macro áreas do mercado, as oportunidades para quem utiliza as tecnologias vinculadas a esta área, possíveis temas para estudo, um extenso estudo bibliométrico e ainda analisa o aspecto educacional sobre a formação de profissionais para atuarem nessa área em termos de habilidades e conhecimento. Além de conter esses assuntos, permite que o leitor conheça, através de apenas um texto, a história, oportunidades e desafios sobre Inteligência Analítica.

Os textos #9 e #10 exploram Inteligência Analítica no contexto de *Supply Chain*. Como nos dois caminhos (superior e inferior) do *Main Path* encontra-se esse tema, pode-se concluir que essas áreas são pioneiras na operacionalização de Inteligência Analítica. O texto #9 (Waller & Fawcett, 2013) também é de *Supply Chain*, porém com maior ênfase nos profissionais que operam as tecnologias (no texto definidos como *Data Scientists*) similar ao texto #6. O texto #10 (Hazen, Boone, Ezell, & Jones-Farmer, 2014) aborda a qualidade de dados na operacionalização de Inteligência Analítica em *Supply Chain*. Para isso, conscientiza sobre os problemas a serem resolvidos e sugere pesquisas futuras e aplicações.

O texto #11 (Chae, 2015), trata de *Analytics* e é a união dos caminhos superior e inferior do *Main Path*, ou seja, pode-se interpretar como uma pequena amostra de diálogo na literatura.

Finalmente, o texto #12 (Wang, White, & Chen, 2015) é o final do *Main Path*. Este texto consiste em uma revisão de literatura sobre *Big Data* em diversas esferas (econômica, social, científica, etc). Trata-se de um grande resumo do que houve até o ano de 2015 e o que se espera para o futuro.

Como a matriz de citações é muito esparsa, apesar da análise do *Main Path* ser conclusiva, considera-se necessária a realização de uma análise sobre os textos mais citados.

## 5.2 Textos mais citados

Os 10 textos mais citados, que não estão na *Main Path*, foram listados na Tabela 2. Desses 10 textos, 6 estão no *Main Path* e já foram analisados na seção anterior. Esta seção discorre sobre esses 4 textos restantes.

**Tabela 2.**Textos entre os dez mais citados não pertencentes ao *Main Path*

#	Informações sobre o texto
13	<b>LaValle S, 2011.</b> Big data, Analytics and the Path from Insights to Value. <i>MIT SLOAN MANAGEMENT REVIEW.</i>
14	<b>Boyd D, 2012.</b> Critical Questions for Big Data. <i>INFORMATION, COMMUNICATION AND SOCIETY.</i>
15	<b>Demirkan H, 2013.</b> Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. <i>DECISION SUPPORT SYSTEMS.</i>
16	<b>Barton D, 2012.</b> Making Advanced Analytics Work For You. <i>HARVARD BUSINESS REVIEW.</i>

**Fonte:** Elaborado pelos autores

O texto #13 (LaValle et al., 2011) foi escrito em linguagem gerencial com o objetivo de compartilhar experiências de empresas que utilizaram técnicas de Inteligência Analítica para coletar informações e obter *insights* para tomada de decisão. O material foi obtido em um estudo anual focado em BI. São definidos três estágios de adoção que uma empresa pode ter e suas respectivas oportunidades. O texto ainda contém uma seção com recomendações para obter bons resultados em BI. O sucesso do texto em termos de citação parece estar ligado ao fato de ser um texto esclarecedor e aplicado.

Uma diferenciada visão do fenômeno *Big Data* é apresentada no texto #14 (Boyd & Crawford, 2012) sob as perspectivas cultural, tecnológica e educacional. Este texto é particularmente interessante por ser o único com estas perspectivas. Há conteúdo sobre como o *Big Data* altera conceitos antes bem definidos, como por exemplo o conhecimento.

O texto #15 (Demirkan & Delen, 2013) contém informações sobre Inteligência Analítica em *cloud* (computação em nuvem), de uma forma que o diferencia de sistemas de suporte à decisão orientado a serviços. Este texto demonstra o que é exclusivo de Inteligência Analítica em *cloud* e fornece vários conceitos sobre o tema. O último texto, o #16 (Barton & Court, 2012), é um manual para gerentes que desejam adotar Inteligência Analítica e precisam de um ponto de partida. O manual aborda desde a escolha dos dados até as transformações necessárias para o ganho de benefícios a partir da Inteligência Analítica. O texto fornece tópicos de atenção para os gestores.

### 5.3 Periódicos mais citados

A análise dos periódicos mais citados permite uma estimativa da importância de determinado *journal* na literatura de determinada área. Para a elaboração da Tabela 3 foram considerados todos os 444 textos e somaram-se as citações de todos os textos de cada *journal*.

Apesar dos *journals* listados serem bem conhecidos pelos pesquisadores de SI, destaca-se que a quantidade de textos não reflete a ordem de importância em termos de citação. O *journal MIS QUARTERLY*, por exemplo, apesar de ter apenas um texto (de acordo com o critério adotado), é o mais citado. Isso significa que, apesar de não ser um *journal* específico para assuntos acerca de tomada de decisão, ele é amplamente lido e utilizado como referência para esse assunto.

**Tabela 3.**

Journals mais citados

Ranking	Nome do jornal	Quant. Textos	Quant. Citações (soma)
1	<i>MIS QUARTERLY</i>	1	77
2	<i>MIT SLOAN MANAGEMENT REVIEW</i>	3	61
3	<i>HARVARD BUSINESS REVIEW</i>	5	60
4	<i>DECISION SUPPORT SYSTEMS</i>	16	50

5	<i>INFORMATION COMMUNICATION &amp; SOCIETY</i>	1	24
6	<i>JOURNAL OF BUSINESS LOGISTICS</i>	6	21
7	<i>COMMUNICATIONS OF THE ACM</i>	4	21

**Fonte:** Elaborado pelos autores

Em termos de quantidade de textos por *journal*, destacam-se o *DECISION SUPPORT SYSTEMS* e o *JOURNAL OF BUSINESS LOGISTICS*. Esses números mostram que esses *journals* são os que possuem maior foco no tema de Inteligência Analítica.

Inteligência Analítica tem se destacado na área de *Supply Chain*, sendo que a maioria de textos no assunto foram publicados no *JOURNAL OF BUSINESS LOGISTICS*. Já o *DECISION SUPPORT SYSTEMS* tem como objetivo principal discutir sistemas de apoio a tomada de decisão, o que pode indicar por que ele é o mais procurado para autores que publicam textos de Inteligência Analítica.

Por fim, é importante destacar os *journals MIT SLOAN MANAGEMENT REVIEW* e *HARVARD BUSINESS REVIEW*, mais focados na utilização prática (gerencial) dos conceitos. Esses *journals* são particularmente interessantes para pesquisadores que desejam entender o uso de ferramentas de Inteligência Analítica, suas dificuldades e requisitos nas empresas.

## 6 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi apresentar as principais contribuições teóricas do campo de Inteligência Analítica, facilitando a imersão de pesquisadores na área, por meio da indicação de textos para leitura. A partir do conceito de *Main Path* e com uma análise de citações, o trabalho evidenciou um conjunto de doze textos pelo qual se sugere que o interessado no tema, pesquisador ou profissional, comece a leitura. Outros quatro textos estão entre os mais citados.

Os resultados deste estudo evidenciam que até o final de março de 2016 houve pouco diálogo na literatura.

A indicação dos textos principais permite, inclusive, um início de diálogo com outros pesquisadores, uma vez obtido o panorama da área. Na leitura, sugere-se seguir a ordem definida neste trabalho: *Main Path* da literatura e os mais citados. É relevante o conhecimento do conjunto de *journals* que se dedicam ao tema.

Há textos não acadêmicos que constituem o *Main Path* e o conjunto de textos mais citados. Logo, há interesse pelo assunto no mundo corporativo, o que demonstra a atualidade do tema e a relevância para a prática. Sublinha-se que o mundo corporativo está produzindo referências no tema inclusive para a academia, evidenciando a lacuna no campo e a oportunidade de geração de conhecimento por pesquisadores.

Percebe-se que, para o assunto Inteligência Analítica, a realidade acadêmica busca acompanhar o que acontece na realidade das empresas, de forma a abordar fenômenos e tentar explicá-los depois que acontecem.

Em relação às pesquisas futuras, um tópico foi dedicado às cinco áreas técnicas críticas para o desenvolvimento da Inteligência Analítica, de acordo com as sugestões de Chen *et al.* (2012), que podem ser exploradas em pesquisas futuras: *Big data analytics*; *Text analytics*; *Web analytics*; *Network analytics*; e *Mobile Analytics*.

Este trabalho apresenta algumas limitações. A primeira é em relação ao tempo e a maturidade da área de Inteligência Analítica: como a área é relativamente nova e ainda há pouco diálogo na literatura, as análises de citações permitem menos conclusões do que em uma área mais consolidada. Espera-se que dentro de alguns anos tenha-se uma evolução e a realização de novas análises de citação poderá mostrar novas tendências e movimentos.

Apesar dessas limitações, o trabalho apresenta aspectos relevantes para essa área promissora de pesquisa e pode ajudar o início de futuros trabalhos de pesquisa acadêmica e aplicada sobre Inteligência Analítica.

## REFERÊNCIAS BIBLIOGRÁFICAS

Alvarado, R., 2007. A bibliometria: história, legitimação e estrutura. Para entender a ciência da informação. Salvador: EDUFBA, p. 185-217.

Barton, D., & Court, D. (2012). Making Advanced Analytics Work For You. *Harvard Business Review*, 90(10), 78-83.

Batagelj, V., & Mrvar, A. (1998). Pajek - Program For Large Network Analysis. *Connections*, 21(2), 47-57.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCInet for Windows: Software for Social Network Analysis*, Harvard, MA: Analytic Technologies.

Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155-172.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication and Society*, 15(5), 662-679.

Chae, B. K. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247-259.

Chae, B., & Olson, D. L. (2013). Business Analytics for Supply Chain: A Dynamic - Capabilities Framework. *International Journal of Information Technology & Decision Making*, 12(1), 9-26.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology, *Communications of the ACM* (54:8), pp. 88-98.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382-1402.

Crossan, M. M., & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: a systematic review of the literature. *Journal of Management Studies*, v. 47, n. 6.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press, edition 1.

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, 90, 70-76.

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.

Garfield, E. (1970). Citation Indexing for Studying Science. *Nature*, 227, 669-671.

- Garfield, E. (1979). Is Citation Analysis a Legitimate Evaluation Tool?. *Scientometrics*, 1(4), 359-375.
- Gelfand, A. (2012). Privacy and Biomedical Research: Building a Trust Infrastructure—An Exploration of Data-Driven and Process-Driven Approaches to Data Privacy. *Biomedical Computation Review*, Winter, pp. 23-28.
- Georgiou, I. (2014). Seeing the Forest for the Trees: An Atlas of the Politics–Administration Dichotomy. *Public Administration Review*, 74(2), 156-175.
- Goes, P. B. (2014). Editor's Comments: Big Data and IS Research. *MIS Quarterly*, 38(3), iii-viii.
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72-80.
- Hummon, N. P., & Dereian, P. (1989). Connectivity in a Citation Network: The Development of DNA Theory. *Social Networks*, 11(1), 39-63.
- Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business Intelligence: An Analysis of the Literature 1. *Information Systems Management*, 25(2), 121-131.
- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. (PP. 527 – 537). John Wiley & Sons.
- Klisarova-Belcheva, S., Ilieva, G., & Yankova, T. (2017). Business intelligence and analytics – contemporary system model. *Trakia Journal of Sciences*, 15(1), 298-304.
- Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging Trends in Business Analytics. *Communications of the ACM*, 45(8), 45-48.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review*, 52(2), 21.
- Liu, J. S., Lu, L. Y., Lu, W. M., & Lin, B. J. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, 41(1), 3-15.
- Luhn, H. P. (1958). A Business Intelligence System. *IBM journal of Research and Development*, 2(4), 314-319.
- Luvizan, S., Meirelles, F. S., & Diniz, E. H. (2014). Big Data: Publication Evolution and Research Opportunities. 11th CONTECSI-International Conference on Information Systems and Technology Management. Sao Paulo, 30.
- McAfee, A., & Brynjolfsson, E. (2012). Big data. The Management Revolution. *Harvard Business Review*, 90(10), 61-67.
- Meirelles, F. S. (2018). Pesquisa Anual – Administração de Recursos de Informática. GVcia – Centro de Tecnologia de Informação Aplicada da FGV-EAESP, 29ª edição, São Paulo.
- O'Reilly, T. (2005). *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*, September 30, (<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>).
- Rousseau, R. (2012). A view on big data and its relation to Informetrics. *Chinese Journal of Library and Information Science*, 5(3), 12-26.

- Sallam, R. L., Howson, C., Idoine, C. J., Oestreich, T. W., Richardson, J. R., & Tapadinhas, J. (2017). *Magic Quadrant for Business Intelligence and Analytics Platforms*. pp 126, Gartner.
- Salton, G. 1989. *Automatic Text Processing*, Reading, MA: Addison Wesley.
- Silva, M. R., Hayashi, C. R. M., & Hayashi, M. C. P. I. (2011). Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. *Revista de Ciência da Informação e Documentação*, Ribeirão Preto, v. 2, n. 1, p. 110-129, jan/jun.
- Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318-327.
- Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: a Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84.
- 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), pp. 234-246.
- Wang, X., White, L., & Chen, X. (2015). Big data research for the knowledge economy: past, present, and future. *Industrial Management & Data Systems*, 115(9).