The history, evolution, and future of big data & analytics

Batistič, Sasa; van der Laken, Paul

Published in:
British Journal of Management

DOI:
10.1111/1467-8551.12340

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

BJM BRITISH JOURNAL of MANAGEMENT BAM British Academy of Management

British Journal of Management, Vol. 30, 229–251 (2019)
DOI: 10.1111/1467-8551.12340

History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations

Saša Batistič and Paul van der Laken
Tilburg University, School of Social and Behavioral Sciences, Department of Human Resource Studies, The Netherlands (E-mail: paulvanderlaken@gmail.com)
Corresponding author email: s.batistic@uvt.nl

Big data and analytics (BDA) are gaining momentum, particularly in the practitioner world. Research linking BDA to improved organizational performance seems scarce and widely dispersed though, with the majority focused on specific domains and/or macro-level relationships. In order to synthesize past research and advance knowledge of the potential organizational value of BDA, the authors obtained a data set of 327 primary studies and 1252 secondary cited papers. This paper reviews this body of research, using three bibliometric methods. First, it elucidates its intellectual foundations via co-citation analysis. Second, it visualizes the historical evolution of BDA and performance research and its substreams through algorithmic historiography. Third, it provides insights into the field's potential evolution via bibliographic coupling. The results reveal that the academic attention for the BDA–performance link has been increasing rapidly. The study uncovered ten research clusters that form the field's foundation. While research seems to have evolved following two main, isolated streams, the past decade has witnessed more cross-disciplinary collaborations. Moreover, the study identified several research topics undergoing focused development, including financial and customer risk management, text mining and evolutionary algorithms. The review concludes with a discussion of the implications for different functional management domains and the gaps for both research and practice.

Introduction

Big data and analytics (BDA) continue to spark interest among scholars and practitioners. Organizations are increasingly aware that they may process and analyse their large data volumes to capture value for their businesses and employees (George, Haas and Pentland, 2014). With the advent of more computational power, machine learning – particularly deep learning through neural networks – has become more broadly deployable in organizations. Academic research on the topic also skyrocketed. Searching for the term 'big data', the Web of Science Core Collection yields 3347 hits in 2015, and over 4000 in both 2016 and 2017.

Several studies have discussed how BDA influences organizational performance, arguing that firms with data-driven strategies tend to be more productive and profitable than their competitors (Brynjiolfsson, Hill and Kim, 2011; LaValle *et al.*, 2011). Scholars have argued that novel machine learning capabilities may realize the predictive value of big data, unleashing its strategic potential to transform business processes and

Both authors contributed equally to the paper.
The copyright line for this article was changed on May 11, 2019 after original online publication.

providing the organizational capabilities to tackle key business challenges (Fosso Wamba *et al.*, 2015). Yet, very few attempts have been made to consolidate the plethora of BDA research and explore the underlying theoretical foundations. Although some attempts have been made to review and theorize how organizational value can be derived from BDA, these attempts have mostly taken on a narrow information systems and technology perspective (for some exceptions, see Grover and Kar, 2017; Günther *et al.*, 2017; Fosso Wamba *et al.*, 2015). Calls to explore the organizational impact of BDA from other functional management perspectives (e.g. marketing, human resource; Angrave *et al.*, 2016) remain largely unanswered to date.

A more comprehensive review of the implications of BDA for the management of performance in and of organizations seems warranted. Synthesizing past research findings is one of the most important tasks for advancing a field of research, particularly one characterized by an extensive growth of publications, such as BDA research (Garfield, 2004; Zupic and Čater, 2015). An overview of the BDA–performance debate may (a) delineate the subfields that constitute the intellectual foundation of the debate and how these subfields relate to one another, (b) unveil and explore the evolution and roots of the debate, and (c) provide insight into the future development of the debate. Moreover, a review could stimulate cross-fertilization of best practices, research designs and theoretical frameworks by unveiling discrepancies in the maturity of BDA of different functional management domains and their research streams.

A bibliometric review using science mapping could be particularly valuable, providing several advantages over classical qualitative and meta-analytical methods. First, a bibliometric approach is more macro-oriented, because it allows the analysis of a comprehensive field of research. Researchers do not need to specify the exact relationship they wish to explore, which offers increased objectivity in reviewing literature (Garfield, 1979). Second, science mapping consists of a classification and visualization of previous research (Small, 1999). This produces a spatial representation analogous to a geographic map that can demonstrate how knowledge domains and individual studies relate to one another. This seems particularly useful for BDA research, which may span different research domains (Günther *et al.*, 2017). Here, science mapping could provide the bigger picture of the state of the art of these domains combined. Third, multiple, complementary bibliometric methods can be easily combined in a single study. Via document co-citation analysis and algorithmic historiography, we explore respectively the past intellectual structure/foundations and the evolution of the BDA–performance debate whereas bibliographic coupling facilitates an objective exploration of the possible future state of research.

A bibliometric review of the relationship between BDA and organizational performance contributes to the literature in two ways. First, our bibliographic methods complement earlier qualitative reviews. Compared with previous reviews (see Fosso Wamba *et al.*, 2015; Grover and Kar, 2017; Günther *et al.*, 2017), we take a broader scope and include a larger sample of documents. Hence, we provide a more comprehensive and objective exploration of the history and past evolution of the BDA–performance debate, while also unveiling more specialized topics within BDA research. Second, our bibliometric approach provides a more objective perspective on the potential future of BDA research. Via bibliographic coupling, we hope to shift attention from traditions to future trends, highlighting the current and future development areas for continued evolution of the BDA debate. This review aims to demonstrate: what BDA applications have been, are being, and will be studied in relation to organizational performance; how distant, disconnected perspectives could be linked via theory or empirical application; how emerging research fields may learn from more established domains; what the current rate and topics of development of BDA are; and how these can be stimulated further into the 21st century.

Big data and performance

In management literature, at least, a cross-disciplinary overview of the BDA discussion is lacking. Hence, it remains unclear whether and how BDA applications in the different domains overlap, how these domains perceive BDA, and what theories have been used to ground potential BDA–performance linkages (Sheng, Amankwah-Amoah and Wang, 2017; Sivarajah *et al.*, 2017).

Past reviews of BDA–performance research have mostly adopted information technology (IT) perspectives (Günther *et al.*, 2017). These

IT studies on BDA frequently used macro-level strategic management theories to ground their hypotheses. Particularly, the resource-based view is often cited in relation to the BDA–performance linkage, postulating that resources (such as capital or information) can provide organizations with the competitive advantage and greater performance (Barney, 1991). From an IT perspective, three main organizational resources are considered: (1) the tangible resources related to the physical IT infrastructure; (2) the human IT resources (e.g. technical and managerial IT skills); and (3) the intangible IT resources (e.g. knowledge or culture) (Bharadwaj, 2000). The extent to which an organization is able to develop, mobilize and exploit resources are called their organizational capabilities (Russo and Fouts, 1997). Thus, BDA can contribute to organizational performance functioning as both an organizational resource and an organizational capability.

Additionally, research suggests that the combination of resources and capabilities matters. For instance, scholars have theorized that there are dependencies with other internal resources: BDA can only add value if the right IT infrastructure is in place when the organizational culture is there, or when the workforce is skilled enough (Fosso Wamba *et al.*, 2015; Gupta and George, 2016). Moreover, grounded in strategic management literature, the dynamic capabilities perspectives suggests that competitive advantage is achieved and sustained by the right use of capabilities (Bowman and Ambrosini, 2003; Brandon-Jones *et al.*, 2014). This shifts the attention from the organization itself to the external organizational environment and the actions required to reshape and align business operations in light of constantly changing global demands (Easterby-Smith, Lyles and Peteraf, 2009; Gunasekaran *et al.*, 2017). Again, the IT perspective is dominant, focusing on how IT-infused organizational capabilities such as BDA help organizations to renew and reconfigure their existing operational mode (Mikalef and Pateli, 2017). Based on this theory, BDA does add value and relate to performance, but only if continuous adaptation and change is considered.

Overall, BDA can be considered both a resource and a capability that can enable efficient and effective business operations, if leveraged appropriately considering the internal and external organizational context. Authors have argued that BDA can now be considered 'a major differentiator between high performing and low-performing organizations' (Liu, 2014), allowing organizations to become more proactive and future-oriented, while decreasing customer acquisition costs and increasing revenue. In general, BDA will add business value, as it stimulates data-driven decision-making capabilities, in which case judgements are often more precise than when they are based solely on intuition or experience (McAfee *et al.*, 2012).

General methods

Sample

To identify the primary research papers on BDA and performance, we contacted 47 prominent scholars and practitioners who either published on BDA in general or on BDA in management research (e.g. business studies, human resource management). These experts were asked to elicit ten keywords describing the relationship between BDA and performance at various levels (i.e. organizational, business unit, team, individual). Ten experts (21.3%) responded and, based on the most frequently proposed keywords (e.g. big data, machine learning, deep learning, data science, analytics, artificial intelligence), we obtained 54 keyword combinations (e.g. 'big data' AND 'organizational performance'). On 7 September 2017, we searched the ISI Web of Knowledge bibliographic database – acknowledged as the most reliable database (Bar-Ilan, 2008; Jacso, 2008) – for these keyword combinations and extracted the results of the relevant work-related domains (i.e. operation research, management science, business, business finance, psychology, psychology applied, management, sport sciences, economics). This retrieved data set included 324 primary documents, which, in turn, provided 14,767 unique secondary (cited) documents. To reduce the complexity of this large data set of secondary documents, we determined a citation threshold – the minimum number of citations a secondary document had to have in order to be included. Via an iterative approach (Zupic and Čater, 2015), a minimum threshold of two citations reduced our sample of secondary documents to 1252 papers[1]. Table 1 demonstrates

[1]The full list of proposed and selected keywords can be provided upon request. Additional Supporting

Table 1. The most important primary and secondary journals in the big data and performance debate

	Primary papers		Secondary (cited) papers	
	Journal	Frequency	Journal	Frequency
1	Expert Systems with Applications	61	MIS Quarterly	196
2	Decision Support Systems	27	Harvard Business Review	172
3	International Journal of Sports Science & Coaching	18	MIT Sloan Management review	80
4	European Journal of Operational Research	14	Journal of Management Information Systems	49
5	International Journal of Production Research	8	Academy of Management Journal	44
6	Journal of Knowledge Management	8	California Management Review	39
7	Journal of Business Research	6	Journal of Marketing	38
8	International Journal of Production Economics	6	Academy of Management Review	34
9	Frontiers in Human Neuroscience	6	Journal of the Association for Information Systems	31
10	Journal of Management Information Systems	6	Journal of Machine Learning Research	29

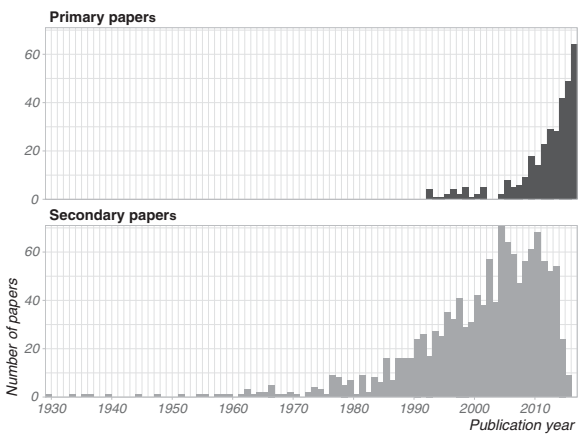

Figure 1. Histograms of the years in which the retrieved papers were published

which journals published our primary and secondary papers; Figure 1 demonstrates when they were published.

Analyses

Three bibliometric analyses were conducted. Document co-citation analysis and algorithmic historiography were applied to the sample of secondary papers whereas bibliographic coupling was applied to the sample of primary papers. These three methods are explained in detail later.

Modularity optimization algorithms are often used to cluster nodes in a (citation) network. Detecting clusters in a network requires the partitioning of a network into communities of densely connected nodes. Here, one expects the nodes belonging to different communities to be only sparsely connected. The quality of the partitioning can thus be quantified via the modularity of the network – a value that represents the density of links within communities as compared with links between communities. In the best clustering solution, the modularity is optimized, and this solution can thus be identified algorithmically (Blondel *et al.*, 2008). Because iterative clustering algorithms use a random starting point, we confirmed the robustness of solution by running the algorithm 50 times (using Gephi's default resolution settings; i.e. 1.0) for Study 1 and 3 and taking the number of clusters closest to the average optimal number (respectively, 10.38 and 8.02). For Study 2, we had to cluster publications in CitNetExplorer, which includes only an older modularity algorithm (see Newman, 2004 for a detailed explanation). Here, strongly connected nodes are grouped and assumed to represent an evolutionary stream over time (Waltman and van Eck, 2012). We again ran the algorithm 50 times (using CitNetExplorer default resolution settings; i.e. 1.0) and retrieved the number of clusters closest to the average optimal number (6.12).

The cluster interpretation followed the suggestion of Zupic and Čater (2015). After running the cluster analyses (Study 1 and 3), the two authors independently explored the content of each cluster by reading through the abstracts and full text of the 25 publications with the highest weighted degree and recording any relevant keywords or topics. In a subsequent session, the authors compared and discussed their keywords, topics and interpretations, after which the current cluster names were determined.

Information may be found in the online version of this article at the publisher's website.

Measures

Several network statistics were calculated during the analyses. The weighted degree centrality represents the number of edges (i.e. citation relationships) a node (i.e. document) has to other nodes, weighted for the edges' importance. Both incoming and outgoing edges are included in this measure. In general, the higher the weighted degree, the more important a document is to the network. Closeness centrality represents a node's distance to all other network nodes, inversed. The higher the closeness, the more central a document's location in the network. Finally, betweenness centrality represents a node's uniqueness in connecting other unconnected nodes. The higher the betweenness, the more a document functions as an important pathway connecting other documents (for more information see Nooy, Mrvar and Batagelj, 2011).

Study 1: Document co-citation

Co-citation analysis (McCain, 1990) uses the frequency with which two documents are cited together to determine their semantic similarity. The underlying assumption is that secondary papers that are co-cited (i.e. both referred to in the same primary document) share content-wise similarities and are thus semantically related. Co-citation count would thus indicate to what extent papers represent related key concepts, theories or methods that a certain field or fields have or have drawn from (Small, 1973). Co-citation is a dynamic measure, because it changes over time as documents accumulate citations (Batistič, Černe and Vogel, 2017). Therefore, it can reflect both the state of a certain intellectual field as well as the shifts in schools of thought (Pasadeos, Phelps and Kim, 1998). Additionally, co-citations can reveal the intellectual roots of a scientific domain through the identification of its core, most cited works.

Via document co-citation analysis, we aimed to explore the intellectual structure/foundations of the BDA–performance debate. The previously described database of secondary papers was normalized for association strength in VOSviewer (van Eck and Waltman, 2014b), a software tool for constructing and visualizing bibliometric networks and the relationship between documents, thereby acknowledging that certain nodes (secondary papers) are more important to the network because they have more connections. Subsequently, the normalized data were loaded into Gephi (Bastian, Heymann and Jacomy, 2009), a leading open-source visualization and exploration software for graphs and networks, which allows for more flexibility in refinement and visualization. Using a force-directed network layout (Hu, 2005), the program displays nodes (i.e. papers) in a two-dimensional space in such a way that more related nodes are co-located, whereas weakly related nodes are distant from each other.

Results

The 1252 documents in the co-citation network stabilized into ten clusters. The content of these clusters was assessed by examining the full texts of the most important papers by weighted degree. Consequently, the clusters could be named (1) BDA Foundation, (2) Statistical Algorithms, (3) Marketing Analytics, (4) Customer Analytics, (5) Knowledge and Innovation, (6) Information Technology (IT) and Supply Chain (SC), (7) Adoption and Integration, (8) Corporate Social Responsibility, (9) Sports Analytics and (10) Brain-Computer Interfaces (BCI). Table 2 provides an overview of these clusters and their papers.

The structure of the co-citation network (Figure 2) provided several insights. First, a large cluster of papers ($N = 324$), very central to the network, covers various topics that are seemingly the foundation for research linking BDA to performance in organizations. Popular publications explain how BDA and data-driven strategies provide organizations a competitive advantage (Barton and Court, 2012; Davenport, 2006; Davenport and Harris, 2007; Davenport, Barth and Bean, 2012; Fosso Wamba *et al.*, 2015; LaValle *et al.*, 2011), whereas other publications focus on the impact of IT for organizational performance (Devaraj and Kohli, 2003; Melville, Kraemer and Gurbaxani, 2004; Mithas, Ramasubbu and Sambamurthy, 2011; Santhanam and Hartono, 2003; Tippins and Sohi, 2003). In either case, the resource-based view is a theory that explains the impact (Barney, 1991; Bharadwaj, 2000). Other publications cover more methodological topics, such as structural equation modelling and partial least squares regression (Fornell and Larcker, 1981; Hair, Ringle and Sarstedt, 2011; Wetzels, Odekerken-Schröder and van Oppen, 2009), mediation (Baron and Kenny,

Table 2. Statistics of the clusters and papers in the document co-citation network

ID	Cluster (*N*)	First author, year	Weighted degree	Closeness	Betweenness
67	1. Big Data and Analytics Research Foundation (324)	Barney, 1991	582	0.542	0.058
393		Fornell, 1981	548	0.496	0.012
895		Podsakoff, 2003	493	0.490	0.015
97		Bharadwaj, 2000	467	0.484	0.007
985		Santhanam, 2003	455	0.489	0.007
130	2. Algorithms (264)	Breiman, 1996	371	0.443	0.020
23		Altman, 1968	354	0.464	0.055
132		Breiman, 1984	300	0.450	0.030
1180		West, 2000	236	0.398	0.003
131		Breiman, 2001	208	0.428	0.033
1170	3. Marketing Analytics (131)	Webster, 2005	87	0.421	0.001
422		Germann, 2013	80	0.414	0.002
1133		Vargo, 2004	72	0.420	0.001
789		Michaelidou, 2011	70	0.408	0.001
869		Pauwels, 2009	70	0.406	0.002
492	4. Customer Analytics (124)	Hanley, 1982	145	0.422	0.002
305		Delonger, 1988	127	0.422	0.004
664		Lariviere, 2005	121	0.412	0.001
913		Prinzie, 2008	111	0.416	0.001
1122		Van den Poel, 2005	111	0.417	0.001
743	5. Knowledge & Innovation (116)	Manyika, 2011	567	0.538	0.087
188		Chen, 2012	550	0.513	0.042
767		McAfee, 2012	482	0.493	0.025
231		Cohen, 1990	457	0.474	0.016
1179		Wernerfelt, 1984	415	0.480	0.016
633	6. Information Technology (IT) & Supply Chain (SC) (106)	Kohli, 2008	316	0.465	0.005
1103		Trkman, 2010	314	0.464	0.009
844		Nunnally, 1994	252	0.450	0.001
1147		Wade, 2004	234	0.448	0.003
412		Galbraith, 1974	218	0.451	0.004
182	7. Adoption & Integration (94)	Chatterjee, 2002	126	0.426	0.001
480		Hambrick, 1988	126	0.431	0.007
691		Liang, 2007	126	0.426	0.001
262		Davenport, 1998	106	0.444	0.009
572		Jansen Jjp, 2005	104	0.427	0.000
1146	8. Corporate Social Responsibility (CSR) (55)	Waddock, 1997	109	0.409	0.003
447		Graves, 1994	82	0.383	0.002
858		Orlitzky, 2003	75	0.399	0.002
1011		Sharfman, 1996	75	0.368	0.001
961		Russo, 1997	73	0.402	0.002
407	9. Sports Analytics (28)	Gabbett, 2012	18	0.246	0.001
409		Gabbett, 2014	18	0.246	0.001
601		Kempton, 2013	18	0.246	0.001
602		Kempton, 2015	18	0.246	0.001
1035		Sirotic, 2011	18	0.246	0.001
108	10. Brain–Computer Interfaces (BCI) (11)	Blankertz, 2010	19	0.232	0.000
375		Farwell, 1988	19	0.232	0.000
477		Halder, 2011	19	0.232	0.000
481		Hammer, 2012	19	0.232	0.000
625		Kleih, 2011	19	0.232	0.000

1986; Devaraj and Kohli, 2003; Tippins and Sohi, 2003), or measurement issues (Podsakoff *et al.*, 2003; Santhanam and Hartono, 2003).

Second, this first cluster is closely connected to several other clusters, which cover more specialized topics related to BDA. For instance, there is a separate cluster focusing on how IT and business intelligence and analytics add value to organizations (Elbashir, Collier and Davern, 2008; Fairbank *et al.*, 2006; Kohli and Grover,

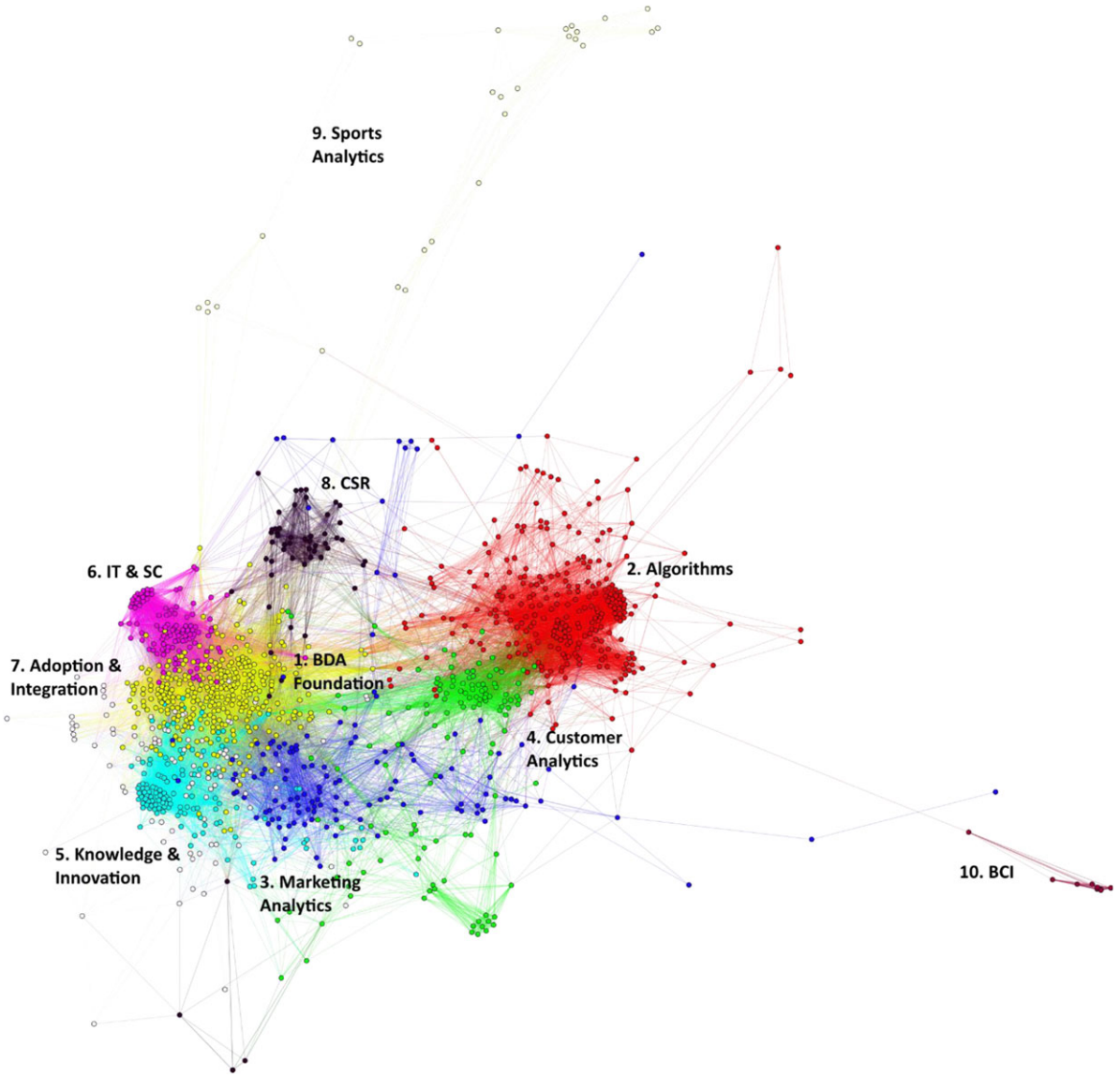

Figure 2. The co-citation network with 1252 secondary papers and ten clusters
Note: Different shades are used to indicate the cluster to which a secondary paper has been assigned. The clusters represent closely related papers, which share thematic similarities.
[Colour figure can be viewed at wileyonlinelibrary.com]

2008) particularly in improving supply chain management (Dehning, Richardson and Zmud, 2007; Hendricks, Singhal and Stratman, 2007; Kannan and Tan, 2005; Stadtler, 2005; Trkman *et al.*, 2010). Here too, the resource-based view seems a central theory (Newbert, 2007; Wade and Hulland, 2004). Another example is cluster five ($N = 116$), which we dubbed Knowledge and Innovation. Although it includes some seminal publications in the general BDA debate (e.g. Hsinchun, Chiang and Storey, 2012; Manyika *et al.*, 2011; McAfee and Brynjolfsson, 2012) – evidenced by their high weighted degree and closeness centrality in the network (Table 2) – the majority of its publications focused specifically on how organizations create, transfer and manage knowledge, innovation and learning (e.g. Cohen and Levinthal, 1990; Grant, 1996; Kogut and Zander, 1992; Nonaka and Takeuchi, 1995; Zander and Kogut, 1995). Owing to a lack of space, details on the Marketing Analytics and Adoption and Integration clusters can be found in

Appendix S1 and in the Supporting information, Appendices S1–S4.

Third, the cluster containing publications on statistics and machine learning algorithms was far removed from the above central clusters. Statistical innovations – such as the bagging of multiple predictors (Breiman, 1996) or decision tree and random forest algorithms (Breiman, 2001; Breiman *et al.*, 1984) – have only been fully leveraged by the customer analytics cluster ($N = 124$). Here, scholars have used advanced algorithms and predictive designs to try and predict customers' loyalty, retention and purchasing behaviours (e.g. Buckinx and Van den Poel, 2005; Larivière and Van den Poel, 2005; Verbeke *et al.*, 2011). All other large clusters seemed to draw on the algorithms cluster to a lesser extent.

For a fifth insight, we refer to the existence of cluster eight ($N = 55$) on the relationship between ethics, corporate social responsibility and firm performance. Most of its core publications (e.g. Berman *et al.*, 1999; Graves and Waddock, 1994; Russo and Fouts, 1997) show the (mutually) positive relationships between ethical and green business policies and their performance (for an exception, Hillman and Keim, 2001), as reverberated by the meta-analysis in this cluster (see Orlitzky, Schmidt and Rynes, 2003). Other papers consider the strengths and weaknesses of measuring corporate social responsibility with the social ratings of Kinder, Lydenberg, Domini Research & Analytics (e.g. Berman *et al.*, 1999; Chatterji, Levine and Toffel, 2009; Sharfman, 1996). Nevertheless, this CSR cluster remains somewhat dislocated from the main network.

Sixth and final, two small clusters were found: one on big data analytics in sport ($N = 28$) and one on brain–computer interfaces ($N = 11$). The publication dates of their main papers suggest that they are relatively emerging fields (see Figure 2) and these clusters also appeared only marginally connected to the rest of the network.

Overall, Study 1 provided insights into the intellectual structure of the BDA and performance debate. The most important cluster involves the main debate on the implications of BDA for organizational performance and seems closely knit with a cluster on BDA from IT and Supply Chain perspectives. The methodological cluster dealing with big data algorithms is, surprisingly, situated in the periphery (Figure 2) and linked to the rest of the network predominantly through Customer Analytics research.

Study 2: Algorithmic historiography

The development of a field over time can be displayed by ordering the most important publications in a field in the sequence in which they appeared, along with the citation relations between these publications (Garfield, 2004; van Eck and Waltman, 2014a). Such an evolutionary visualization of a field illustrates the history of science and scholarship and has been referred to as an algorithmic historiography (Garfield, 2001, 2004). Like other bibliometric methods, a historiography considers the relationships between various primary papers. However, the direction rather than the weight of this relationship is of importance as relationships are binary – a primary paper either does or does not cite a second primary paper. As the changes in the citation rate of key papers of a field inform how basic concepts within and the perception of the paradigm as a whole have changed over time, the resulting historiography helps the understanding of paradigms (Garfield, Pudovkin and Istomin, 2003).

We conducted the historiography in CitNetExplorer (van Eck and Waltman, 2014a) on the earlier described full sample of primary papers. CitNetExplorer is a software tool for visualizing and analysing citation networks of scientific publications. It is especially useful for analysing the development of a research field over time, as it shows how publications build on each other. CitNetExplorer reduces this full citation network in two ways. First, it identifies the core publications through the concept of k-cores (Seidman, 1983), where publications are considered core when they have a certain minimum number of ingoing or outgoing citation relations with other core publications. Van Eck and Waltman (2014a) consider publications 'core' if they have citation relations with at least ten other core publications, whereas Garfield, Pudovkin and Istomin (2003) propose limiting the core publications to approximately 5% of the total number of publications. For our data set, such settings resulted in a network including fewer than 21 publications, quite incomprehensive. We, therefore, decided to expand this set iteratively – balancing the network's comprehensiveness and interpretability – which resulted in an optimal

network of 50 core publications (approximately 15% of the total number of publications).

Second, CitNetExplorer performed a so-called transitive reduction of the citation network. Here, the program distinguishes essential from non-essential citation relations in the network, and only the essential relations are retained (van Eck and Waltman, 2014a). Citation relations are classified as essential if there are no other pathways (i.e. relations) connecting two publications. Removing all non-essential relations minimizes the edges in the network while ensuring that all previously connected publications still have a pathway connecting them. CitNetExplorer draws the resulting network by, on the vertical axis, the publication year and, on the horizontal axis, the closeness between publications (see van Eck *et al.* (2010) for a more technical explanation).

Results

The results of the historiography are presented in Figure 3. Although the 50 core publications formed six clusters, Figure 3 clearly demonstrates that the BDA–performance research field has two main evolutionary streams. The first stream is rooted in statistics and algorithms and their application to financial/customer topics. The seminal paper by Altman (1968) is the first root publication of this stream. Other root papers come from a more statistical perspective (e.g. classification and regression trees, bagging, random forests) (Breiman, 1996, 2001; Breiman *et al.*, 1984). About forty years later, several publications in Expert Systems with Applications followed, examining predictive analytics applications within finance, such as a credit risk scoring (e.g. Twala, 2010; Wang *et al.*, 2011). Other contemporary papers build mostly on the statistical perspective and cover predictive analytics focused on customer behaviour (e.g. Ballings and Poel, 2012). Generally speaking, the left side of Figure 3 relates to the development of new statistical methods and applications within the fields of financial and customer analytics.

Second, a more management and strategically oriented stream evolved on the right side of Figure 3. Although the first paper has a statistical perspective, covering structural equation modelling (Fornell and Larcker, 1981), other root papers in this second stream discuss the resource-based view (Barney, 1991), the dynamic capabilities of organizations (Wernerfelt, 1984) and a knowledge-based theory of organizations (Barney, 1991; Grant, 1996; Wernerfelt, 1984). This foundation has resulted in two main themes in contemporary papers within the stream. On the one hand, there is a general discussion regarding how BDA influences organizational performance, and specifically the performance of several management functions (e.g. supply chain, human resource management) (LaValle *et al.*, 2011; Trkman *et al.*, 2010). On the other hand, there are papers discussing the general topics related to business intelligence in this second stream (Fosso Wamba *et al.*, 2015; Hsinchun, Chiang and Storey, 2012). These publications review how BDA and business intelligence would – theoretically and empirically – influence organizational performance. Yet, this stream does not include advanced analytical applications or empirical investigations.

An interesting final deduction that we can make from Figure 3 is that the above two evolutionary streams have only recently been connected. The responsible papers cover customer event history (Ballings and Poel, 2012) and the ways in which big data may form a competitive advantage for organizations (Manyika *et al.*, 2011).

Study 1 elucidated the intellectual structure of the field, and Study 2 adds to this by providing an overview of its historical evolution. Some findings of this second study align with those of the first: the large gap between the methodological and theoretical discussions surrounding BDA is visible in both Figures 1 and 2. Similarly, the paper (Ballings and Poel, 2012) linking the two evolutionary streams in Figure 3 studied customer event history, whereas the Customer Analytics cluster bridged the algorithms with the rest of the BDA network in Study 1.

Study 3: Bibliographic coupling

Bibliographic coupling examines the extent to which documents cite the same secondary documents. This implies that the primary, citing document rather than the cited, secondary documents is the focus of analysis (Vogel and Güttel, 2013). The general assumption is that the more the bibliographies of two documents overlap, the stronger their connection is.

Bibliographic coupling is different from other bibliometric methods as it does not derive the

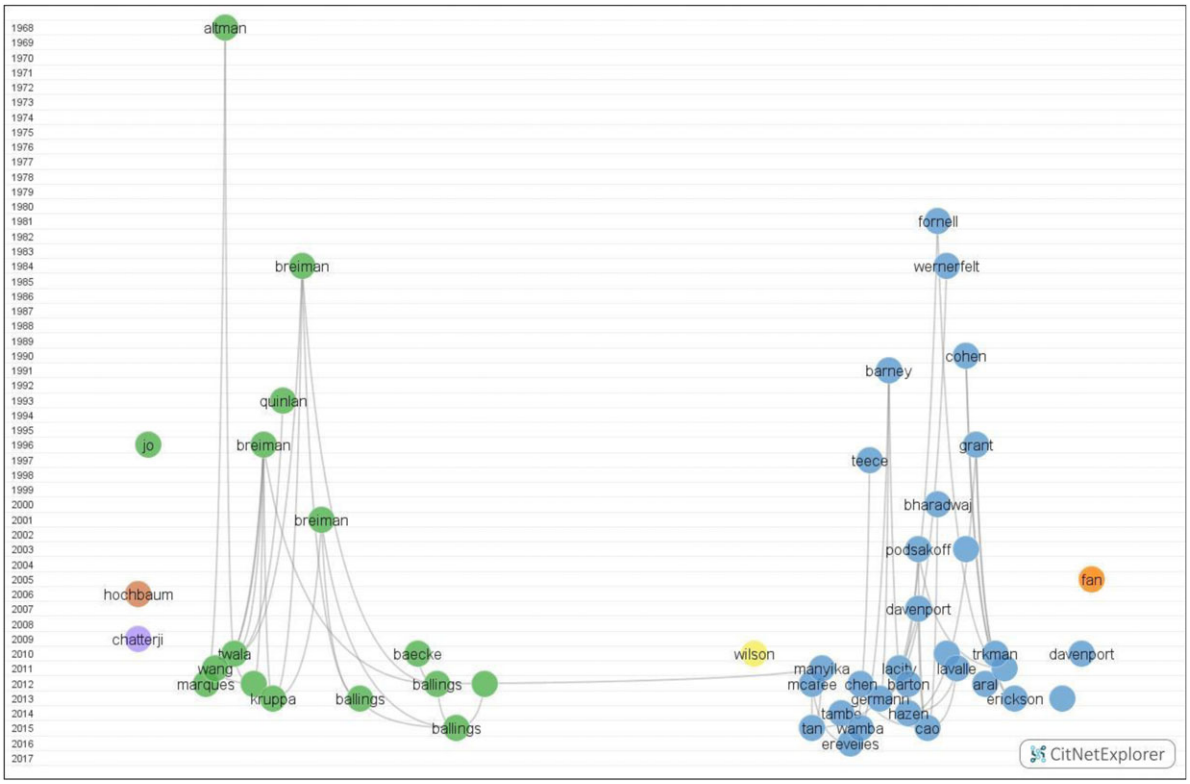

Figure 3. Citation network of the evolution of the BDA–performance debate
Note: Curved lines are used to indicate citation relations. Different shades represent the cluster to which primary papers have been assigned. Clusters represent closely related papers, sharing thematic similarities.
[Colour figure can be viewed at wileyonlinelibrary.com]

importance of papers within a scholarly community from their citation count or relations (Verbeek *et al.*, 2002). This prevents an (over)emphasis on mainstream documents that may be popular but insignificant to a fields' intellectual development. Moreover, because it relies on the references within documents, the results of bibliographic coupling are more stable over time because reference lists do not change over time (in contrast to citation counts and relations). All this makes coupling particularly suitable for detecting current trends and future priorities, as these are commonly covered in the more recent publications, which inherently are not the most cited.

Although we intended to use the retrieved data set of 324 primary papers in the bibliographic coupling, only 211 of these primary documents (65.12%) were interconnected in the same network. The other papers had completely unconnected reference lists and were thus automatically removed by VOSviewer (van Eck and Waltman, 2014b). The normalized network data of the included papers were loaded into Gephi (Bastian, Heymann and Jacomy, 2009), and visualized with a force-directed layout (Hu, 2005).

Results

The 211 primary documents in the bibliographic coupling network formed eight clusters. Table 3 provides an overview of the clusters and the most important papers (by weighted degree) per cluster. Based on the full text of their most important papers, we named the clusters (1) Risk and Customer Predictions, (2) Strategic BDA, (3) Information and Knowledge Management, (4) Text and Genetic Algorithms, (5) CSR, (6) Clustering, (7) Sports Analytics, (8) BCI.

There are three large clusters in the network. In the first, largest cluster ($N = 74$), several papers sought to predict the risk of credit applicants (Abellán and Castellano, 2017; Florez-Lopez and

Table 3. Statistics of the clusters and papers in the bibliographic coupling network

ID	Cluster (*N*)	First author, year	Weighted degree	Closeness	Betweenness
165	10. Risk & Customer Predictions (74)	Twala, 2010	150	0.420	0.020
62		Florez-Lopez, 2015	141	0.461	0.031
175		Twala, 2009	139	0.417	0.017
64		Ballings, 2015	96	0.431	0.026
129		Ballings, 2012	94	0.448	0.032
13	20. Strategic Big Data and Analytics (56)	Ren, 2017	201	0.441	0.008
102		Chae, 2014	193	0.451	0.018
23		Wamba, 2017	177	0.470	0.044
24		Akter, 2016	170	0.477	0.060
149		Coltman, 2011	147	0.417	0.013
15	30. Knowledge & Information (40)	Rothberg, 2017	118	0.454	0.032
111		Erickson, 2013	64	0.385	0.006
57		Jarvinen, 2015	41	0.387	0.012
191		Cross, 2006	40	0.391	0.023
205		Osborn, 1998	29	0.385	0.010
65	40. Text & Genetic Algorithms (19)	Van de Kauter, 2015	17	0.359	0.016
88		Lau, 2014	14	0.385	0.038
52		Nguyen, 2015	9	0.319	0.001
85		Kim, 2014	9	0.297	0.001
150		Esfahanipour, 2011	8	0.297	0.007
35	50. Corporate Social Responsibility (CSR) (10)	Lucas, 2016	55	0.400	0.014
130		Nandy, 2012	46	0.384	0.027
124		Boesso, 2013	45	0.324	0.001
178		Chatterji, 2009	43	0.320	0.000
60		Kang, 2015	41	0.341	0.002
116	60. Clustering (6)	Song, 2013	9	0.340	0.022
107		Chen, 2013	8	0.307	0.001
193		Hochbaum, 2006	7	0.327	0.001
71		Ghazarian, 2015	2	0.286	0.000
75		Munivrana, 2012	1	0.254	0.000
33	70. Sport Analytics (4)	Hogarth, 2016	7	0.207	0.010
48		Kempton, 2016	7	0.261	0.019
12		Woods, 2017	5	0.349	0.028
31		Wilkerson, 2016	1	0.172	0.000
121	80. Brain–Computer Interfaces (BCI) (2)	Halder, 2013	11	0.265	0.010
89		Hammer, 2014	10	0.210	0.000

Ramon-Jeronimo, 2015; Twala, 2010; Wang *et al.*, 2011), others predicted customer churn/retention risks (Ballings and Poel, 2012; Moeyersoms and Martens, 2015; Morales and Wang, 2010), whereas more niche topics are also included, for instance, social media usage predictions (Ballings and Van den Poel, 2015). Papers in the second cluster ($N = 56$) examined what organizational characteristics affect firm performance in the era of BDA (Akter *et al.*, 2016; Ji-fan Ren *et al.*, 2017; Wamba *et al.*, 2017) and how BDA improved decision-making and value creation in organizations (Cao, Duan and Li, 2015; Chae, Olson and Sheu, 2014; Chae *et al.*, 2014; Chen, Preston and Swink, 2015; Coltman, Devinney and Midgley, 2011). A closely connected third cluster ($N = 40$) focused on how knowledge and information can be strategically developed, managed and leveraged in organizations (e.g. Erickson and Rothberg, 2013), and the role of BDA therein (Rothberg and Erickson, 2017; Tsui *et al.*, 2014; Wang *et al.*, 2013).

Next, five smaller clusters were identified. Cluster four ($N = 19$) examined how text analytics and sentiment analysis of social media data can, for instance, predict stock markets (Kim and Kim, 2014; Nguyen, Shirai and Velcin, 2015; Van de Kauter, Breesch and Hoste, 2015) and crime (Gerber, 2014), or optimize product design and marketing strategies (Lau, Li and Liao, 2014). Moreover, it includes research on genetic algorithms predicting stock markets (Esfahanipour and Mousavi, 2011) and optimizing production lines (Balakrishnan,

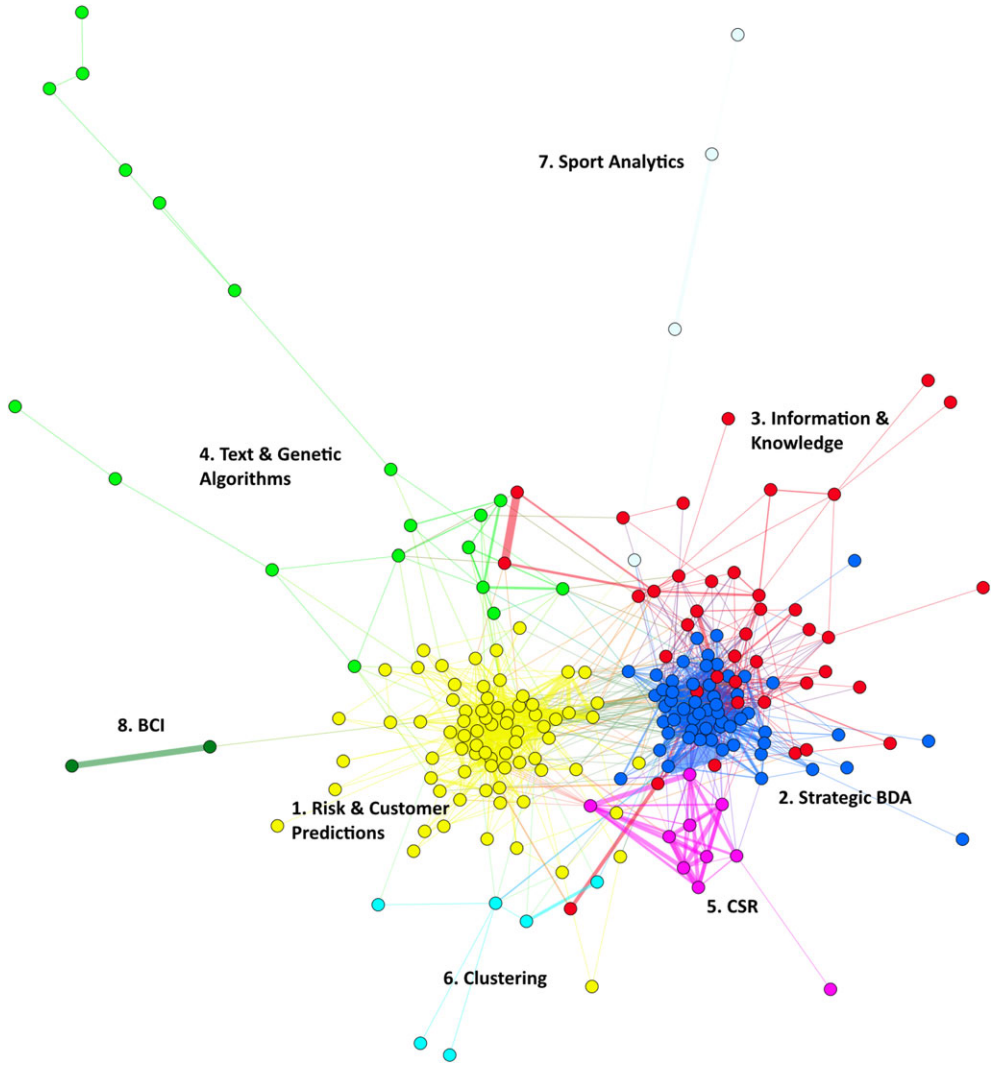

Figure 4. The bibliographic coupling network with 211 papers and eight clusters
Note: Line strength reflects bibliometric overlap. Different shades represent the cluster to which primary papers have been assigned. Clusters represent closely related papers, sharing thematic similarities.
[Colour figure can be viewed at wileyonlinelibrary.com]

Gupta and Jacob, 2006). Cluster five ($N = 10$) examined corporate social responsibility with the ratings of Kinder, Lyndenberg, Domini Research and Analytics (e.g. Lucas and Noordewier, 2016; Nandy and Lodh, 2012). Cluster six ($N = 6$) examined how clusters can be identified and ranked in order to improve recommendation engines and other business processes (e.g. Chen, Cheng and Hsu, 2013; Song *et al.*, 2013). Studies in cluster seven used big data analytics in sports to analyse the evolution of gameplay in Australian football (Woods, Robertson and Collier, 2017), the relationship between practice and injury in American football (Wilkerson *et al.*, 2016), and the possession value (Kempton, Kennedy and Coutts, 2016) and match demands in rugby football (Hogarth, Burkett and McKean, 2016). Finally, the two studies in cluster eight used machine learning to predict the performance of brain–computer interfaces (Halder *et al.*, 2013; Hammer *et al.*, 2014).

A final deduction of Figure 4 is that the reference lists of the more strategic research streams were closely interrelated (Strategic BDA, Information and Knowledge and CSR) whereas the other, more

technical and operational streams are dispersed across the network.

While Studies 1 and 2 looked at the intellectual roots and the historical evolution of the BDA–performance debate, the purpose of Study 3 was to look ahead, at the future of the debate. Figure 4 again centres the Customer Analytics cluster – which also proved to be an important bridge in the networks of Figure 2 and 2. In this future outlook, the cluster appears to move even closer to the Strategic BDA cluster as well as the overall centre of the BDA debate. Similar to the co-citation analysis (Figure 2), clusters relating to new technological and methodological advances (e.g. brain–computer interfaces, text analysis, genetic algorithms) seem to arise at the periphery of Figure 4. In terms of important publications in the future of the debate, Figure 4 puts forward Ji-fan Ren *et al.* (2017) and Wamba *et al.* (2017), both in the Strategic BDA cluster, and published in the *International Journal of Production Research* and the *Journal of Business Research*, respectively. Both studies examine the effect of BDA in relation to dynamic capabilities.

Discussion

This paper reviews the literature on the relationship between big data, analytics (BDA) and the performance in and of organizations with three bibliometric methods (co-citation analysis, algorithmic historiography and bibliographic coupling). The results provide insight into the intellectual structure and the past and future evolution of research linking BDA to organizational performance. The number of academic publications on the topic is rising quickly. We identified ten clusters of research on which studies that link BDA to organizational performance build: including a large BDA foundation cluster, some closely intertwined fields (IT and supply chain research, innovation research and research on marketing analytics), and more peripheral scholarly communities (algorithmic research, customer analytics research, corporate social responsibility research). We uncovered that, historically, BDA research has evolved in two large, but isolated research streams, but cross-disciplinary bridges have formed during the past decade. Regarding the future evolution, we identified strong research clusters focused on financial risk management, customer relationship management and strategic management considering BDA.

Main findings

Our bibliometric review provides one of the first overarching overviews of the perspectives that have been taken in exploring the BDA–performance linkage. Similar to other reviews, we found that BDA applications are already being considered, developed, implemented and adding value in the management of customers, information, innovation, technology and supply chains (Fosso Wamba *et al.*, 2015; Grover and Kar, 2017). Moreover, we found similar key topics, including machine learning, business intelligence, text analytics and social media data (Grover and Kar, 2017). Our results also cover four of the six BDA debates found by Günther *et al.* (2017), related to algorithms, organizational capabilities, innovation and strategy, and corporate social responsibility. While the number of scientific publications in our reviewed sample was considerably larger than prior reviews, our focus was narrower (i.e. performance in organizations). Potentially, as a result, our review does not replicate the big data research streams in healthcare, education and public management/government included in previous work (Fosso Wamba *et al.*, 2015; Grover and Kar, 2017; Sheng, Amankwah-Amoah and Wang, 2017), or the other two BDA debates found by Günther *et al.* (2017) – the inductive–deductive debate and the modes of big data access.

Dispersed research and theory

Our review provides several new insights. First, while it seems without doubt that there is increased attention for BDA in management research (Figure 1), our analyses suggest that research on the strategic management in the era of BDA and research on actual implementations and operationalizations of BDA are still two different worlds. The related academic communities and their discourse are quite dispersed. For instance, the co-citation network (Figure 2), which displays the intellectual foundation, demonstrates a strong divide between the core BDA research stream and the clusters developing and implementing predictive algorithms. Similarly, the historiography (Figure 3) and the coupling network (Figure 4), underline the weak overlap

in the shared knowledge and discourse between research covering strategical issues in BDA research (e.g. value, management, ethics) and research covering operational implementations (e.g. predictive analytics, text analytics, clustering). Relatedly, we could not even include over a third of the primary documents in the bibliographic coupling analysis, because they lacked bibliographic connections to any other document in the network. This is a worrying development as it suggests that a vast amount of information and knowledge is not diffused in the greater scientific community, meaning scholars and practitioners could overlook best practices or novel algorithms.

This dispersion could have affected the theoretical foundation of the field. The most frequently cited theoretical perspective in our sample was the resource-based view. Yet, seeing BDA as an organizational resource or capability leading to improved performance seems quite fitting from an IT or general management perspective (Mikalef *et al.*, 2018), but potentially less relevant when considering other functional management perspectives. For example, from a marketing, risk or customer management perspective, having solid behavioural theories that drive what data is collected for micro-level predictions is potentially more value-adding. Hence, in such fields a large variety of other theories was used to ground the BDA–performance relationship, including for instance echelons theory in the Marketing Analytics cluster (cf. Germann, Lilien and Rangaswamy, 2013) and institutional theory in the Adoption & Integration cluster (cf. Liang *et al.*, 2007).

We believe that improved cross-disciplinary collaborations might improve the diversity of perspectives and ultimately lead to better theoretical understanding of the full BDA–performance link. For instance, while many sampled IT papers draw on the resource-based view, we did not encounter behavioural psychology theories to help unravel the role of intangible resources (e.g. culture, knowledge). Potentially, IT scholars could draw on research on marketing, organizational behaviour or human resource management for such insights. Fortunately, knowledge sharing and cross-disciplinary collaboration seems to be occurring at an increasing pace. Our historiography (Figure 3) demonstrates that the first bridges between the two main research streams have recently been established, building on Ballings and Poel (2012) and Manyika *et al.* (2011), which we regard as a promising first step.

Differing levels of maturity

Second, our studies suggest that the various management functions in organizations are at different stages of BDA maturity. The use of BDA seems established in relation to financial risk and customer relationship management, where predictive modelling and the more advanced statistical algorithms are already widely applied, researched and discussed. Figures 2 and 4 suggest that developments within marketing, supply chain and IT are on their way as well. However, particularly in the latter two domains, research is focused mostly on the high-level strategic impact of BDA (Chen, Preston and Swink, 2015; Germann, Lilien and Rangaswamy, 2013; Trainor *et al.*, 2014; Trkman *et al.*, 2010) rather than actual applications or individual-level predictions within these functional domains (for some exceptions see Ballings *et al.*, 2015; Chi *et al.*, 2007; Esfahanipour and Mousavi, 2011). This shows that there is a divide between fields taking micro- vs. macro-level approaches to exploring the value of BDA for organizational performance.

Several management functions seem to be trailing behind, at least in terms of academic discourse on the value of BDA. For instance, although studies mention the rise of BDA and algorithmic intelligence in the HR field (e.g. LaValle *et al.*, 2011), little research has been done. Arguably, this is undesirable: HR missing the big data bandwagon may imply a loss for organizations and cause harm for employees, whose interests could consequently be overlooked in BDA initiatives (Angrave *et al.*, 2016; Liang and Liu, 2018). Similarly, we did not encounter studies on the use of BDA in legal, procurement, M&A, health and safety, public administration or facility management. On the one hand, this could mean that some functions (e.g. IT, marketing) are more mature in leveraging the value of BDA than others. On the other hand, it could be that researchers in some fields (e.g. legal, human resources) do not have readily available (big) data to theorize or test the implications of BDA for performance.

Moreover, in contrast earlier studies (Fosso Wamba *et al.*, 2015; Grover and Kar, 2017; Sheng, Amankwah-Amoah and Wang, 2017), we did not encounter studies exploring BDA applications

in the public sector specifically. Nevertheless, there is a lot of potential impact for predictive analytics and data-driven strategies in these settings (cf. Reinmoeller and Ansari, 2016; Sheng, Amankwah-Amoah and Wang, 2017). For example, Sheng, Amankwah-Amoah and Wang (2017), in their review study, have found that public services and administration can benefit from big data for e-voting and e-government (e.g. with cloud computing). It could be that our research setup (e.g. used keywords) caused the public sector to be underrepresented in our sample.

Alternatively, the above differences in levels of maturity could be due to other geographical, sectoral or domain-level differences in the value and/or applicability of BDA. For instance, the General Data Protection Regulation (GDPR, 2016) in Europe makes the gathering and use of personal data of individuals significantly more challenging for organizations. This could (have) cause(d) differences in the speed of development of BDA applications in Europe compared with, for instance, the Americas or Asia. Similarly, such legislation may cause differences between functional domains that mainly process personal data (e.g. marketing, customer relationship management, human resources management) vs. those that rely more strongly on non-personal data (e.g. finance, supply chain, IT). Finally, you could expect differences on a sectoral level, where sectors that work with more and more sensitive personal data (e.g. healthcare) could be hindered in their development of BDA applications. Other geographical, sectoral or domain-level differences in BDA development may include the technological capabilities of the workforce, or the perceived ethicality of using predictive profiling in specific settings. More research attention is needed on the primary causes of such differences and their implications.

Ethics and corporate social responsibility

A third insight is the cluster on the corporate social responsibility that arose in both the co-citation and bibliographic coupling networks. Although the core publications in these clusters did consider the effect of (perceived) corporate social responsibility on organizational performance, they had little to do with BDA (e.g. Chatterji, Levine and Toffel, 2009; Lucas and Noordewier, 2016; Waddock and Graves, 1997). On the one hand, the proprietary nature of social and environmental ratings such as those of Kinder, Lyndenberg, Domini Research and Analytics (currently MSCI) did not allow us to assess accurately whether they truly use 'big' data. On the other hand, the studies in these CSR clusters did not employ the more advanced predictive algorithms, but instead relied on traditional linear and logistic regression methods. We had hoped to find studies demonstrating how organizations may deal with ethics and privacy concerns when deriving business value through BDA, or how organizations may use BDA to solve costly environmental issues, such as pollution or energy waste. The lack of such studies in this review is striking and worrying, and we urge scholars to pay more focused attention to this topic.

Limitations

This study faces several limitations, of which we discuss three below. A first limitation involves our search strategy. Although we reached out to nearly fifty experts in the field, only ten responded with keywords for our search. Their responses were internally consistent and had high face validity (e.g. big data, machine learning, deep learning, data science, analytics, artificial intelligence), but may have had a strong influence on our results. For instance, one could question whether the more distant clusters (e.g. brain–computer interfaces) belong in a review on BDA and performance in organizations. Alternatively, our search strategy may have caused an underrepresentation of specific data sources (e.g. wearables, sensors), algorithms (e.g. long-short-term memory networks) or sectors (e.g. healthcare, government).

Second, the interpretation of the results – the networks and the clusters – was limited to our human capabilities in terms of text and information processing. In line with our BDA topic, future studies could extend our current analysis with a more data-driven approach. For instance, text-mining algorithms such as latent Dirichlet allocation (Blei, Ng and Jordan, 2003) could be used to identify the state-of-the-art topics in BDA research. Additionally, meta-analytical review approaches could help future researchers to quantify the added value of BDA for organizational performance, to test the effectiveness of different BDA applications and strategies, or to compare the potential geographical, sectoral or functional differences in BDA impact.

A third and final limitation is that we had to apply certain thresholds in order to process the data. Here, we followed the established guidelines (Eck and Waltman, 2014a; Garfield, Pudovkin and Istomin, 2003), and we compared different settings in order to test the robustness of analyses. Nevertheless, we acknowledge that these thresholds may have introduced bias in the otherwise relatively objective bibliometric methods.

Future research directions

Apart from its limitations, this current review extends our knowledge of how BDA influence the management and performance in and of organizations. Based on our results, we propose four overall directions advancing the BDA–performance debate.

Cross-functional bridges

First, we demonstrated that the cross-functional adoption and application of BDA is scarce, but imminent. Scholars have noted that, for a long time, management researchers have been focused on traditional methodology (e.g. general linear models), thereby not realizing the full potential of the 'big' data collected through modern technology (e.g. social media, wearables, sensors, video, audio) (e.g. social media, wearables, sensors, video, audio; Angrave *et al.*, 2016; van der Laken *et al.*, 2018; Yarkoni and Westfall, 2017). Fortunately, our algorithmic historiography demonstrates that the first bridges between the management and statistical research innovation communities have been made (Figure 3). Future scholars and practitioners should jump on the bandwagon and seek cross-functional collaborations, where domain experts within managerial functions team up with experts in statistics and machine learning domains in order to test academic theories and deploy relevant business applications simultaneously. Preliminary empirical evidence from fields such as operations and IT management shows that a combination of management and statistical perspectives can add great value to firm performance (cf. Wamba *et al.*, 2017). One direction would be to apply advanced statistical methods to leverage value from big data in underexplored management functions. For instance, HR data may be used to predict the hiring success of applications, the effectiveness of training courses, or the number of workplaces needed (Marler and Boudreau, 2017).

Great potential lies in cross-disciplinary knowledge exchange. Here, mainstream clusters such as Strategic Big Data and Analytics could learn from collaborations with scholars in the peripheral clusters. For instance, scholars in the Sports Analytics domain already leverage data from wearables and sensors for scientific and practical purposes. From a management perspective, wearables can be used to explore the communication patterns in organizations with the aim of improving knowledge sharing, or to monitor employees' health in order to improve their well-being (e.g. Wenzel and Van Quaquebeke, 2018).

Big data analytics and ethics

In applied BDA research, ethical considerations are essential (Boyd and Crawford, 2012; Herschel and Miori, 2017). Hence, we were surprised that no cluster or studies in our results specifically focused on ethical perspectives related to BDA or the ethical issues related to predictive analytics particularly. It goes without saying that all researchers should make sure that the privacy and the interests of their study subjects are protected, but ethicality is even more important when dealing with sensitive 'big' data, such as continuous audiovisual, biometric, behavioural or geolocation monitoring. Particularly when it comes to predictive analytics, scholars and practitioners should take additional care in preventing the creation of self-fulfilling prophecies or the incorporation of human bias into decision-making algorithms (Herschel and Miori, 2017). Additionally, BDA is often seen as objective and accurate (Boyd and Crawford, 2012). Complex and inaccurate data or predictions can create a false sense of authority, whereby organizational decisions based on them appear objective and indisputable. We call for future research examining to what extent the above issues occur in organizations, how they are currently handled, and what best practices can be implemented to prevent them from happening. In practice, continuously exploring and testing both the financial and ethical implications of analytical initiatives would allow organizations to establish their long-term survival more firmly.

New research methods

We provide a first and novel review approach for the BDA–performance debate. Yet, different

review methods can be used to shed light further on the debate. One such method is text analysis or text mining (Kobayashi *et al.*, 2018). For example, text mining can be used to explore abstracts or whole papers to reveal new facts, trends or constructs deriving from patterns and relationship in the text. The style of writing the papers may differ from function to function, which can, for example, suggest that certain writing styles are more frequent in one function over the other (e.g. Thorpe *et al.*, 2018) and hinder the dissemination of findings (e.g. methodological advancements clusters vs. mainstream management clusters). Our second suggestion is to use temporal networks that can inform the evolution and the future trends at the same time. In such networks, nodes can interact via a sequence of temporary events. For example, temporal networks can be applied on the secondary papers, and the temporal closeness centrality (Pan and Saramäki, 2011) – which measure how quickly all other nodes can be reached from a given node – can be used to show the intellectual evolution and possible future trends.

Future direction by theoretical advancements

Finally, we suggest that scholars exploring the BDA–performance relationship should explore a more diverse range of theoretical perspectives. The current repertoire is based predominantly on the resource-based view (Barney, 1991). Based on the content of strategic BDA cluster in Study 3, we suggest two ways for potential expansion. First, strategic management theories can help to explain the fit between BDA and organizational strategy. One such framework is Porter's value chain (Porter, 1980). This framework displays the set of activities that an organization can carry out to generate value for its customers (e.g. inbound logistics, operations). Here, BDA can provide better information for the decision-making process in such activities. For example, in the inbound logistics part of the framework, BDA can analyse historical data to provide support for a just-in-time approach to receiving, storing and distributing inputs internally. This can further enhance the value for the end customers: for example, end products can be delivered to the customer sooner and cheaper, resulting in increased organizational performance.

Second, the usage and efficiency of BDA can be related to the organizational culture and climate in place. Big data and analytics needs to be in line not only with the organizations' strategy, but also with its culture (Gupta and George, 2016). While BDA may be implemented to stimulate a data-driven culture, managerial decisions on various hierarchical levels will often still be based mainly on the experience and intuition of decision-makers (McAfee *et al.*, 2012). Hence, a change in individual mindsets and organizational culture is necessary to achieve a more data-driven, objective and impactful decision-making. Various management and behavioural theories can help BDA research address these topics. For example, contextual and multi-level theories (e.g. Johns, 2006; Kozlowski and Klein, 2000) are used to observe, predict and change behaviours considering the stimuli provided by the context. We argue that data-driven culture comes through strategic alignment between strategy, human resource management and culture (Buller and McEvoy, 2012; Ogbonna and Harris, 1998). For instance, organizations might design their HR systems (e.g. selection, training, rewards) to stimulate individual BDA usage and acceptance (cf. Ostroff and Bowen, 2016) or to increase their employees' human capital, which, in turn, might make them more proficient with the BDA tools (Mikalef *et al.*, 2018; Rasmussen and Ulrich, 2015).

References

Abellán, J. and J. G. Castellano (2017). 'A comparative study on base classifiers in ensemble methods for credit scoring', *Expert Systems with Applications*, **73**, pp. 1–10.

Akter, S., S. F. Wamba, A. Gunasekaran, R. Dubey and S. J. Childe (2016). 'How to improve firm performance using big data analytics capability and business strategy alignment?', *International Journal of Production Economics*, **182**, pp. 113–131.

Altman, E. I. (1968). 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *Journal of Finance*, **23**, pp. 589–609.

Angrave, D., A. Charlwood, I. Kirkpatrick, M. Lawrence and M. Stuart (2016). 'HR and analytics: why HR is set to fail the big data challenge', *Human Resource Management Journal*, **26**, pp. 1–11.

Balakrishnan, P. S., R. Gupta and V. S. Jacob (2006). 'An investigation of mating and population maintenance strategies in hybrid genetic heuristics for product line designs', *Computers & Operations Research*, **33**, pp. 639–659.

Ballings, M. and D. V. D. Poel (2012). 'Customer event history for churn prediction: how long is long enough?', *Expert Systems with Applications*, **39**, pp. 13517–13522.

Ballings, M. and D. Van den Poel (2015). 'CRM in social media: predicting increases in Facebook usage frequency', *European Journal of Operational Research*, **244**, pp. 248–260.

Ballings, M., D. Van den Poel, N. Hespeels and R. Gryp (2015). 'Evaluating multiple classifiers for stock price

direction prediction', *Expert Systems with Applications*, **42**, pp. 7046–7056.
Bar-Ilan, J. (2008). 'Which h-index? – A comparison of WoS, Scopus and Google Scholar', *Scientometrics*, **74**, pp. 257–271.
Barney, J. (1991). 'Firm resources and sustained competitive advantage', *Journal of Management*, **17**, pp. 99–120.
Baron, R. M. and D. A. Kenny (1986). 'The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations', *Journal of Personality and Social Psychology*, **51**, pp. 1173–1182.
Barton, D. and D. Court (2012). 'Making advanced analytics work for you', *Harvard Business Review*, **90**, pp. 78–83.
Bastian, M., S. Heymann and M. Jacomy (2009). 'Gephi: an open source software for exploring and manipulating networks'. Paper presented at the 3rd International AAAI Conference on Weblogs and Social Media, San Jose, May 17–20, 2009.
Batistič, S., M. Černe and B. Vogel (2017). 'Just how multilevel is leadership research? A document co-citation analysis 1980–2013 on leadership constructs and outcomes', *Leadership Quarterly*, **28**, pp. 86–103.
Berman, S. L., A. C. Wicks, S. Kotha and T. M. Jones (1999). 'Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance', *Academy of Management Journal*, **42**, pp. 488–506.
Bharadwaj, A. S. (2000). 'A resource-based perspective on information technology capability and firm performance: an empirical investigation', *MIS Quarterly*, **24**, pp. 169–196.
Blankertz, B., M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, … and K. R. Mueller (2010). 'The Berlin brain–computer interface: non-medical uses of BCI technology', *Frontiers in neuroscience*, **4**, 198.
Blei, D. M., A. Y. Ng and M. I. Jordan (2003). 'Latent Dirichlet allocation', *Journal of Machine Learning Research*, **3**, pp. 993–1022.
Blondel, V. D., J.-L. Guillaume, R. Lambiotte and E. Lefebvre (2008). 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, **2008**, P10008.
Boesso, G., K. Kumar and G. Michelon (2013). 'Descriptive, instrumental and strategic approaches to corporate social responsibility: Do they drive the financial performance of companies differently?', *Accounting, Auditing & Accountability Journal*, **26**, pp. 399–422.
Bowman, C. and V. Ambrosini (2003). 'How the resource-based and the dynamic capability views of the firm inform corporate-level strategy', *British Journal of Management*, **14**, pp. 289–303.
Boyd, D. and K. Crawford (2012). 'Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, **15**, pp. 662–679.
Brandon-Jones, E., B. Squire, C. W. Autry and K. J. Petersen (2014). 'A contingent resource-based perspective of supply chain resilience and robustness', *Journal of Supply Chain Management*, **50**, pp. 55–73.
Breiman, L. (1996). 'Bagging predictors', *Machine Learning*, **24**, pp. 123–140.
Breiman, L. (2001). 'Random forests', *Machine Learning*, **45**, pp. 5–32.
Breiman, L., J. Friedman, C. J. Stone and R. A. Olshen (1984). *Classification and Regression Trees*. Boca Raton, CA: CRC Press.
Brynjiolfsson, E., L. Hill and H. H. Kim (2011). 'Strength in numbers: how does data-driven decision-making affect firm peformance'. MIT Sloan Working Paper, Cambridge, MA.
Buckinx, W. and D. Van den Poel (2005). 'Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting', *European Journal of Operational Research*, **164**, pp. 252–268.
Buller, P. F. and G. M. McEvoy (2012). 'Strategy, human resource management and performance: sharpening line of sight', *Human Resource Management Review*, **22**, pp. 43–56.
Cao, G., Y. Duan and G. Li (2015). 'Linking business analytics to decision making effectiveness: a path model analysis', *IEEE Transactions on Engineering Management*, **62**, pp. 384–395.
Chae, B., D. Olson and C. Sheu (2014). 'The impact of supply chain analytics on operational performance: a resource-based view', *International Journal of Production Research*, **52**, pp. 4695–4710.
Chae, B. K., C. Yang, D. Olson and C. Sheu (2014). 'The impact of advanced analytics and data accuracy on operational performance: a contingent resource based theory (RBT) perspective', *Decision Support Systems*, **59**, pp. 119–126.
Chatterjee, D., R. Grewal and V. Sambamurthy (2002). 'Shaping up for e-commerce: institutional enablers of the organizational assimilation of web technologies', *MIS quarterly*, 65–89.
Chatterji, A. K., D. I. Levine and M. W. Toffel (2009). 'How well do social ratings actually measure corporate social responsibility?', *Journal of Economics & Management Strategy*, **18**, pp. 125–169.
Chen, D. Q., D. S. Preston and M. Swink (2015). 'How the use of big data analytics affects value creation in supply chain management', *Journal of Management Information Systems*, **32**, pp. 4–39.
Chen, Y. L., L. C. Cheng and W. Y. Hsu (2013). 'A new approach to the group ranking problem: finding consensus ordered segments from users' preference data', *Decision Sciences*, **44**, pp. 1091–1119.
Chen, H., R. H. Chiang and V. C. Storey (2012). 'Business intelligence and analytics: from big data to big impact', *MIS quarterly*, 1165–1188.
Chi, H.-M., O. K. Ersoy, H. Moskowitz and J. Ward (2007). 'Modeling and optimizing a vendor managed replenishment system using machine learning and genetic algorithms', *European Journal of Operational Research*, **180**, pp. 174–193.
Cohen, W. M. and D. A. Levinthal (1990). 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, pp. 128–152.
Coltman, T., T. M. Devinney and D. F. Midgley (2011). 'Customer relationship management and firm performance', *Journal of Information Technology*, **26**, pp. 205–219.
Cross, R., T. Laseter, A. Parker and G. Velasquez (2006). 'Using social network analysis to improve communities of practice', *California Management Review*, **49**, pp. 32–60.
Davenport, T. H. and L. Prusak (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.
Davenport, T. H. (2006). 'Competing on analytics', *Harvard Business Review*, **84**, pp. 98–107.
Davenport, T. H. and J. G. Harris (2007). *Competing on Analytics: The New Science of Winning*. Boston, MA: Harvard Business School Publishing.
Davenport, T. H., P. Barth and R. Bean (2012). 'How big data is different', *MIT Sloan Management Review*, **54**, pp. 43–46.

Dehning, B., V. J. Richardson and R. W. Zmud (2007). 'The financial performance effects of IT-based supply chain management systems in manufacturing firms', *Journal of Operations Management*, **25**, pp. 806–824.

DeLong, E. R., D. M. DeLong and D. L. Clarke-Pearson (1988). 'Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach', *Biometrics*, 837–845.

Devaraj, S. and R. Kohli (2003). 'Performance impacts of information technology: is actual usage the missing link?', *Management Science*, **49**, pp. 273–289.

Easterby-Smith, M., M. A. Lyles and M. A. Peteraf (2009). 'Dynamic capabilities: current debates and future directions', *British Journal of Management*, **20**, pp. S1–S8.

Elbashir, M. Z., P. A. Collier and M. J. Davern (2008). 'Measuring the effects of business intelligence systems: the relationship between business process and organizational performance', *International Journal of Accounting Information Systems*, **9**, pp. 135–153.

Erickson, G. S. and H. N. Rothberg (2013). 'A strategic approach to knowledge development and protection', *Service Industries Journal*, **33**, pp. 1402–1416.

Esfahanipour, A. and S. Mousavi (2011). 'A genetic programming model to generate risk-adjusted technical trading rules in stock markets', *Expert Systems with Applications*, **38**, pp. 8438–8445.

Fairbank, J. F., G. J. Labianca, H. K. Steensma and R. Metters (2006). 'Information processing design choices, strategy, and risk management performance', *Journal of Management Information Systems*, **23**, pp. 293–319.

Farwell, L. A. and E. Donchin (1988). 'Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials', *Electroencephalography and clinical Neurophysiology*, **70**, pp. 510–523.

Florez-Lopez, R. and J. M. Ramon-Jeronimo (2015). 'Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal', *Expert Systems with Applications*, **42**, pp. 5737–5753.

Fornell, C. and D. F. Larcker (1981). 'Evaluating structural equation models with unobservable variables and measurement error', *Journal of Marketing Research*, **18**, pp. 39–50.

Fosso Wamba, S., S. Akter, A. Edwards, G. Chopin and D. Gnanzou (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.

Gabbett, T. J., D. G. Jenkins and B. Abernethy (2012). 'Physical demands of professional rugby league training and competition using microtechnology', *Journal of Science and Medicine in Sport*, **15**, pp. 80–86.

Gabbett, T. J. (2014). 'Effects of physical, technical, and tactical factors on final ladder position in semiprofessional rugby league', *International journal of sports physiology and performance*, **9**, pp. 680–688.

Galbraith, J. R. (1974). 'Organization design: An information processing view', *Interfaces*, **4**, pp. 28–36.

Garfield, E. (1979). 'Is citation analysis a legitimate evaluation tool?', *Scientometrics*, **1**, pp. 359–375.

Garfield, E. (2001). 'From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography: a citationist's tribute to Belver C. Griffith'. Presented at Drexel University, Philadelphia, PA, 27 November.

Garfield, E. (2004). 'Historiographic mapping of knowledge domains literature', *Journal of Information Science*, **30**, pp. 119–145.

Garfield, E., A. I. Pudovkin and V. S. Istomin (2003). 'Why do we need algorithmic historiography?', *Journal of the American Society for Information Science and Technology*, **54**, pp. 400–412.

GDPR (2016). 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46', *Official Journal of the European Union (OJ)*, **59**, pp. 1–88.

George, G., M. R. Haas and A. Pentland (2014). 'Big data and management', *Academy of Management Journal*, **57**, pp. 321–326.

Gerber, M. S. (2014). 'Predicting crime using Twitter and kernel density estimation', *Decision Support Systems*, **61**, pp. 115–125.

Germann, F., G. L. Lilien and A. Rangaswamy (2013). 'Performance implications of deploying marketing analytics', *International Journal of Research in Marketing*, **30**, pp. 114–128.

Ghazarian, S. and M. A. Nematbakhsh (2015). 'Enhancing memory-based collaborative filtering for group recommender systems', *Expert systems with applications*, **42**, pp. 3801–3812.

Grant, R. M. (1996). 'Toward a knowledge-based theory of the firm', *Strategic Management Journal*, **17**, pp. 109–122.

Graves, S. B. and S. A. Waddock (1994). 'Institutional owners and corporate social performance', *Academy of Management Journal*, **37**, pp. 1034–1046.

Grover, P. and A. K. Kar (2017). 'Big data analytics: a review on theoretical contributions and tools used in literature', *Global Journal of Flexible Systems Management*, **18**, pp. 203–229.

Gunasekaran, A., T. Papadopoulos, R. Dubey, S. F. Wamba, S. J. Childe, B. Hazen and S. Akter (2017). 'Big data and predictive analytics for supply chain and organizational performance', *Journal of Business Research*, **70**, pp. 308–317.

Günther, W. A., M. H. Rezazade Mehrizi, M. Huysman and F. Feldberg (2017). 'Debating big data: a literature review on realizing value from big data', *Journal of Strategic Information Systems*, **26**, pp. 191–209.

Gupta, M. and J. F. George (2016). 'Toward the development of a big data analytics capability', *Information & Management*, **53**, pp. 1049–1064.

Hair, J. F., C. M. Ringle and M. Sarstedt (2011). 'PLS-SEM: indeed a silver bullet', *Journal of Marketing Theory and Practice*, **19**, pp. 139–151.

Halder, S., D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, … and A. Kübler (2011). 'Neural mechanisms of brain–computer interface control', *Neuroimage*, **55**, pp. 1779–1790.

Halder, S., B. Varkuti, M. Bogdan, A. Kübler, W. Rosenstiel, R. Sitaram and N. Birbaumer (2013). 'Prediction of brain–computer interface aptitude from individual brain structure', *Frontiers in Human Neuroscience*, **7**, p. 105.

Hambrick, D. C. and P. A. Mason (1984). 'Upper echelons: The organization as a reflection of its top managers', *Academy of management review*, **9**, pp. 193–206.

Hammer, E. M., S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, … and A. Kübler (2012). 'Psychological predictors of SMR-BCI performance', *Biological psychology*, **89**, pp. 80–86.

Hammer, E. M., T. Kaufmann, S. C. Kleih, B. Blankertz and A. Kübler (2014). 'Visuo-motor coordination ability predicts performance with brain–computer interfaces controlled by modulation of sensorimotor rhythms (SMR)', *Frontiers in Human Neuroscience*, **8**, p. 574.

Hendricks, K. B., V. R. Singhal and J. K. Stratman (2007). 'The impact of enterprise systems on corporate performance: a study of ERP, SCM, and CRM system implementations', *Journal of Operations Management*, **25**, pp. 65–82.

Hanley, J. A. and B. J. McNeil (1982). 'The meaning and use of the area under a receiver operating characteristic (ROC) curve', *Radiology*, **143**, pp. 29–36.

Herschel, R. and V. M. Miori (2017). 'Ethics & big data', *Technology in Society*, **49**, pp. 31–36.

Hillman, A. J. and G. D. Keim (2001). 'Shareholder value, stakeholder management, and social issues: what's the bottom line?', *Strategic Management Journal*, **22**, pp. 125–139.

Hochbaum, D. S. and A. Levin (2006). 'Methodologies and algorithms for group-rankings decision', *Management Science*, **52**, pp. 1394–1408.

Hogarth, L. W., B. J. Burkett and M. R. McKean (2016). 'Match demands of professional rugby football codes: a review from 2008 to 2015', *International journal of Sports Science & Coaching*, **11**, pp. 451–463.

Hsinchun, C., R. H. L. Chiang and V. C. Storey (2012). 'Business intelligence and analytics: from big data to big impact', *MIS Quarterly*, **36**, pp. 1165–1188.

Hu, Y. (2005). 'Efficient, high-quality force-directed graph drawing', *Mathematica Journal*, **10**, pp. 37–71.

Jacso, P. (2008). 'Google Scholar revisited', *Online Information Review*, **32**, pp. 102–114.

Jansen, J. J., F. A. Van Den Bosch and H. W. Volberda (2005). 'Managing potential and realized absorptive capacity: how do organizational antecedents matter?', *Academy of management journal*, **48**, pp. 999–1015.

Järvinen, J. and H. Karjaluoto (2015). 'The use of Web analytics for digital marketing performance measurement', *Industrial Marketing Management*, **50**, 117–127.

Ji-fan Ren, S., S. Fosso Wamba, S. Akter, R. Dubey and S. J. Childe (2017). 'Modelling quality dynamics, business value and firm performance in a big data analytics environment', *International Journal of Production Research*, **55**, pp. 5011–5026.

Johns, G. (2006). 'The essential impact of context on organizational behavior', *Academy of Management Review*, **31**, pp. 386–408.

Kang, J. (2015). 'Effectiveness of the KLD social ratings as a measure of workforce diversity and corporate governance', *Business & Society*, **54**, pp. 599–631.

Kannan, V. R. and K. C. Tan (2005). 'Just in time, total quality management, and supply chain management: understanding their linkages and impact on business performance', *Omega*, **33**, pp. 153–162.

Kempton, T., A. C. Sirotic, M. Cameron and A. J. Coutts (2013). 'Match-related fatigue reduces physical and technical performance during elite rugby league match-play: a case study', *Journal of Sports Sciences*, **31**, 1770–1780.

Kempton, T., A. C. Sirotic, E. Rampinini and A. J. Coutts (2015). 'Metabolic power demands of rugby league match play', *International Journal of Sports Physiology and Performance*, **10**, pp. 23–28.

Kempton, T., N. Kennedy and A. J. Coutts (2016). 'The expected value of possession in professional rugby league match-play', *Journal of sports sciences*, **34**, pp. 645–650.

Kim, S.-H. and D. Kim (2014). 'Investor sentiment from internet message postings and the predictability of stock returns', *Journal of Economic Behavior & Organization*, **107**, pp. 708–729.

Kleih, S. C., T. Kaufmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, … and A. Kuebler (2011). Out of the frying pan into the fire—the P300-based BCI faces real-world challenges. In *Progress in brain research* (Vol. **194**, pp. 27–46). Elsevier.

Kobayashi, V. B., S. T. Mol, H. A. Berkers, G. Kismihók and D. N. Den Hartog (2018). 'Text mining in organizational research', *Organizational Research Methods*, **21**, pp. 733–765.

Kogut, B. and U. Zander (1992). 'Knowledge of the firm, combinative capabilities, and the replication of technology', *Organization Science*, **3**, pp. 383–397.

Kohli, R. and V. Grover (2008). 'Business value of IT: an essay on expanding research directions to keep up with the times', *Journal of the Association for Information Systems*, **9**, pp. 23–28, 30–34, 36–39.

Kozlowski, S. W. J. and K. J. Klein (2000). 'A multilevel approach to theory and research in organizations: contextual, temporal, and emergent processes'. In K. J. Klein and S. W. J. Kozlowski (eds), *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*, pp. 3–90. San Francisco, CA: Jossey-Bass.

Larivière, B. and D. Van den Poel (2005). 'Predicting customer retention and profitability by using random forests and regression forests techniques', *Expert Systems with Applications*, **29**, pp. 472–484.

Lau, R. Y., C. Li and S. S. Liao (2014). 'Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis', *Decision Support Systems*, **65**, pp. 80–94.

LaValle, S., E. Lesser, R. Shockley, M. S. Hopkins and N. Kruschwitz (2011). 'Big data, analytics and the path from insights to value', *MIT Sloan Management Review*, **52**, pp. 21–32.

Liang, H., N. Saraf, Q. Hu and Y. Xue (2007). 'Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management', *MIS Quarterly*, **31**, pp. 59–87.

Liang, T.-P. and Y.-H. Liu (2018). 'Research landscape of business intelligence and big data analytics: a bibliometrics study', *Expert Systems with Applications*.

Liu, Y. (2014). 'Big data and predictive business analytics', *Journal of Business Forecasting*, **33**, pp. 40–42.

Lucas, M. T. and T. G. Noordewier (2016). 'Environmental management practices and firm financial performance: the moderating effect of industry pollution-related factors', *International Journal of Production Economics*, **175**, pp. 24–34.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers (2011). 'Big data: the next frontier for innovation, competition, and productivity'. Technical Report, McKinsey Global Institute.

Marler, J. H. and J. W. Boudreau (2017). 'An evidence-based review of HR analytics', *International Journal of Human Resource Management*, **28**, pp. 3–26.

McAfee, A. and E. Brynjolfsson (2012). 'Big data: the management revolution', *Harvard Business Review*, **90**, pp. 60–68.

McCain, K. W. (1990). 'Mapping authors in intellectual space: a technical overview', *Journal of the American Society for Information Science (1986–1998)*, **41**, pp. 433–443.

Melville, N., K. Kraemer and V. Gurbaxani (2004). 'Information technology and organizational performance: an integrative model of IT business value', *MIS Quarterly*, **28**, pp. 283–322.

Mikalef, P. and A. Pateli (2017). 'Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: findings from PLS-SEM and fsQCA', *Journal of Business Research*, **70**, pp. 1–16.

Mikalef, P., I. O. Pappas, J. Krogstie and M. Giannakos (2018). 'Big data analytics capabilities: a systematic literature review and research agenda', *Information Systems and e-Business Management*, **16**, pp. 547–578.

Michaelidou, N., N. T. Siamagka and G. Christodoulides (2011). 'Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands', *Industrial marketing management*, **40**, pp. 1153–1159.

Mithas, S., N. Ramasubbu and V. Sambamurthy (2011). 'How information management capability influences firm performance', *MIS Quarterly*, **35**, pp. 237–256.

Moeyersoms, J. and D. Martens (2015). 'Including high-cardinality attributes in predictive models: a case study in churn prediction in the energy sector', *Decision Support Systems*, **72**, pp. 72–81.

Morales, D. R. and J. Wang (2010). 'Forecasting cancellation rates for services booking revenue management using data mining', *European Journal of Operational Research*, **202**, pp. 554–562.

Munivrana, G., G. Furjan-Mandić and M. Kondrič (2015). 'Determining the structure and evaluating the role of technical-tactical elements in basic table tennis playing systems', *International journal of sports science & coaching*, **10**, pp. 111–132.

Nandy, M. and S. Lodh (2012). 'Do banks value the eco-friendliness of firms in their corporate lending decision? Some empirical evidence', *International Review of Financial Analysis*, **25**, pp. 83–93.

Newbert, S. L. (2007). 'Empirical research on the resource-based view of the firm: an assessment and suggestions for future research', *Strategic Management Journal*, **28**, pp. 121–146.

Newman, M. E. J. (2004). 'Fast algorithm for detecting community structure in networks', *Physical Review E*, **69**, p. 066133.

Nguyen, T. H., K. Shirai and J. Velcin (2015). 'Sentiment analysis on social media for stock movement prediction', *Expert Systems with Applications*, **42**, pp. 9603–9611.

Nonaka, I. and H. T. Takeuchi (1995). *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York, NY: Oxford University Press.

Nooy, W., A. Mrvar and V. Batagelj (2011). *Exploratory Social Network Analysis with Pajek*. New York, NY: Cambridge University Press.

Nunnally, J. C. and I. H. Bernstein (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ogbonna, E. and L. C. Harris (1998). 'Managing organizational culture: compliance or Genuine Change?', *British Journal of Management*, **9**, pp. 273–288.

Orlitzky, M., F. L. Schmidt and S. L. Rynes (2003). 'Corporate social and financial performance: a meta-analysis', *Organization Studies*, **24**, pp. 403–441.

Osborn, C. S. (1998). 'Systems for sustainable organizations: emergent strategies, interactive controls and semi-formal information', *Journal of Management Studies*, **35**, pp. 481–509.

Ostroff, C. and D. E. Bowen (2016). 'Reflections on the 2014 Decade Award: is there strength in the construct of HR system strength?', *Academy of Management Review*, **41**, pp. 196–214.

Pan, R. K. and J. Saramäki (2011). 'Path lengths, correlations, and centrality in temporal networks', *Physical Review E*, **84**, 016105.

Pasadeos, Y., J. Phelps and B.-H. Kim (1998). 'Disciplinary impact of advertising scholars: temporal comparisons of influential authors, works and research networks', *Journal of Advertising*, **27**, pp. 53–70.

Pauwels, K., T. Ambler, B. H. Clark, P. LaPointe, D. Reibstein, B. Skiera, … and T. Wiesel (2009). 'Dashboards as a service: why, what, how, and what research is needed?', *Journal of Service Research*, **12**, pp. 175–189.

Podsakoff, P. M., S. B. MacKenzie, J. Y. Lee and N. P. Podsakoff (2003). 'Common method biases in behavioral research: a critical review of the literature and recommended remedies', *Journal of Applied Psychology*, **88**, pp. 879–903.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York, NY: Free Press.

Prinzie, A. and D. Van den Poel (2008). 'Random forests for multiclass classification: Random multinomial logit', *Expert systems with Applications*, **34**, pp. 1721–1732.

Rasmussen, T. and D. Ulrich (2015). 'Learning from practice: how HR analytics avoids being a management fad', *Organizational Dynamics*, **44**, pp. 236–242.

Reinmoeller, P. and S. Ansari (2016). 'The persistence of a stigmatized practice: a study of competitive intelligence', *British Journal of Management*, **27**, pp. 116–142.

Rothberg, H. N. and G. S. Erickson (2017). 'Big data systems: knowledge transfer or intelligence insights?', *Journal of Knowledge Management*, **21**, pp. 92–112.

Russo, M. V. and P. A. Fouts (1997). 'A resource-based perspective on corporate environmental performance and profitability', *Academy of Management Journal*, **40**, pp. 534–559.

Santhanam, R. and E. Hartono (2003). 'Issues in linking information technology capability to firm performance', *MIS Quarterly*, **27**, pp. 125–165.

Seidman, S. B. (1983). 'Network structure and minimum degree', *Social networks*, **5**, pp. 269–287.

Sharfman, M. (1996). 'The construct validity of the Kinder, Lydenberg & Domini social performance ratings data', *Journal of Business Ethics*, **15**, pp. 287–296.

Sheng, J., J. Amankwah-Amoah and X. Wang (2017). 'A multidisciplinary perspective of big data in management research', *International Journal of Production Economics*, **191**, pp. 97–112.

Sirotic, A. C., H. Knowles, C. Catterick and A. J. Coutts (2011). 'Positional match demands of professional rugby league competition', *The Journal of Strength & Conditioning Research*, **25**, pp. 3076–3087.

Sivarajah, U., M. M. Kamal, Z. Irani and V. Weerakkody (2017). 'Critical analysis of Big Data challenges and analytical methods', *Journal of Business Research*, **70**, pp. 263–286.

Small, H. (1973). 'Co-citation in the scientific literature: a new measure of the relationship between two documents', *Journal of the Association for Information Science and Technology*, **24**, pp. 265–269.

Small, H. (1999). 'Visualizing science by citation mapping', *Journal of the American Society for Information Science*, **50**, pp. 799–813.

Song, M., H. Yang, S. H. Siadat and M. Pechenizkiy (2013). 'A comparative study of dimensionality reduction techniques to enhance trace clustering performances', *Expert Systems with Applications*, **40**, pp. 3722–3737.

Stadtler, H. (2005). 'Supply chain management and advanced planning – basics, overview and challenges', *European Journal of Operational Research*, **163**, pp. 575–588.

Thorpe, A., R. Craig, D. Tourish, G. Hadikin and S. Batistic (2018). '"Environment" submissions in the UK's research excellence framework 2014', *British Journal of Management*, **29**, pp. 571–587.

Tippins, M. J. and R. S. Sohi (2003). 'IT competency and firm performance: is organizational learning a missing link?', *Strategic Management Journal*, **24**, pp. 745–761.

Trainor, K. J., J. M. Andzulis, A. Rapp and R. Agnihotri (2014). 'Social media technology usage and customer relationship performance: a capabilities-based examination of social CRM', *Journal of Business Research*, **67**, pp. 1201–1208.

Trkman, P., K. McCormack, M. P. V. de Oliveira and M. B. Ladeira (2010). 'The impact of business analytics on supply chain performance', *Decision Support Systems*, **49**, pp. 318–327.

Tsui, E., W. M. Wang, L. Cai, C. Cheung and W. Lee (2014). 'Knowledge-based extraction of intellectual capital-related information from unstructured data', *Expert Systems with Applications*, **41**, pp. 1315–1325.

Twala, B. (2009). 'An empirical comparison of techniques for handling incomplete data using decision trees', *Applied Artificial Intelligence*, **23**, pp. 373–405.

Twala, B. (2010). 'Multiple classifier application to credit risk assessment', *Expert Systems with Applications*, **37**, pp. 3326–3336.

Van de Kauter, M., D. Breesch and V. Hoste (2015). 'Fine-grained analysis of explicit and implicit sentiment in financial news articles', *Expert Systems with Applications*, **42**, pp. 4999–5010.

van der Laken, P., Z. Bakk, V. Giagkoulas, L. van Leeuwen and E. Bongenaar (2018). 'Expanding the methodological toolbox of HRM researchers: the added value of latent bathtub models and optimal matching analysis', *Human Resource Management*, **57**, pp. 751–760.

van Eck, N. J. and L. Waltman (2014a). 'CitNetExplorer: a new software tool for analyzing and visualizing citation networks', *Journal of Informetrics*, **8**, pp. 802–823.

van Eck, N. J. and L. Waltman (2014b). 'Visualizing bibliometric networks'. In Y. Ding, R. Rousseau and D. Wolfram (eds), *Measuring Scholarly Impact*, pp. 285–320. Cham: Springer.

van Eck, N. J., L. Waltman, R. Dekker and J. van den Berg (2010). 'A comparison of two techniques for bibliometric mapping: multidimensional scaling and VOS', *Journal of the American Society for Information Science and Technology*, **61**, pp. 2405–2416.

Van den Poel, D. and W. Buckinx (2005). 'Predicting online-purchasing behaviour', *European journal of operational research*, **166**, pp. 557–575.

Vargo, S. L. and R. F. Lusch (2004). 'Evolving to a new dominant logic for marketing', *Journal of marketing*, **68**, pp. 1–17.

Verbeek, A., K. Debackere, M. Luwel and E. Zimmermann (2002). 'Measuring progress and evolution in science and technology – I: the multiple uses of bibliometric indicators', *International Journal of Management Reviews*, **4**, pp. 179–211.

Verbeke, W., D. Martens, C. Mues and B. Baesens (2011). 'Building comprehensible customer churn prediction models with advanced rule induction techniques', *Expert Systems with Applications*, **38**, pp. 2354–2364.

Vogel, R. and W. H. Güttel (2013). 'The dynamic capability view in strategic management: a bibliometric review', *International Journal of Management Reviews*, **15**, pp. 426–446.

Waddock, S. A. and S. B. Graves (1997). 'The corporate social performance–financial performance link', *Strategic Management Journal*, **18**, pp. 303–319.

Wade, M. and J. Hulland (2004). 'Review: the resource-based view and information systems research: review, extension, and suggestions for future research', *MIS Quarterly*, **28**, pp. 107–142.

Waltman, L. and N. J. van Eck (2012). 'A new methodology for constructing a publication-level classification system of science', *Journal of the American Society for Information Science and Technology*, **63**, pp. 2378–2392.

Wamba, S. F., A. Gunasekaran, S. Akter, S. J.-f. Ren, R. Dubey and S. J. Childe (2017). 'Big data analytics and firm performance: effects of dynamic capabilities', *Journal of Business Research*, **70**, pp. 356–365.

Wang, G., J. Hao, J. Ma and H. Jiang (2011). 'A comparative assessment of ensemble learning for credit scoring', *Expert Systems with Applications*, **38**, pp. 223–230.

Wang, G. A., J. Jiao, A. S. Abrahams, W. Fan and Z. Zhang (2013). 'ExpertRank: a topic-aware expert finding algorithm for online knowledge communities', *Decision Support Systems*, **54**, pp. 1442–1451.

Webster Jr, F. E., A. J. Malter and S. Ganesan (2005). 'The decline and dispersion of marketing competence', *MIT Sloan Management Review*, **46**, pp. 35.

Wenzel, R. and N. Van Quaquebeke (2018). 'The double-edged sword of big data in organizational and management research: a review of opportunities and risks', *Organizational Research Methods*, **21**, pp. 548–591.

Wernerfelt, B. (1984). 'A resource-based view of the firm', *Strategic Management Journal*, **5**, pp. 171–180.

West, D. (2000). 'Neural network credit scoring models', *Computers and Operations Research*, **27**, pp. 1131–1152. https://doi.org/10.1016/S0305-0548(99)00149-5

Wetzels, M., G. Odekerken-Schröder and C. van Oppen (2009). 'Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration', *MIS Quarterly*, **33**, pp. 177–195.

Wilkerson, G. B., A. Gupta, J. R. Allen, C. M. Keith and M. A. Colston (2016). 'Utilization of practice session average inertial load to quantify college football injury risk', *Journal of Strength & Conditioning Research*, **30**, pp. 2369–2374.

Woods, C. T., S. Robertson and N. F. Collier (2017). 'Evolution of game-play in the Australian Football League from 2001 to 2015', *Journal of Sports Sciences*, **35**, pp. 1879–1887.

Yarkoni, T. and J. Westfall (2017). 'Choosing prediction over explanation in psychology: lessons from machine learning', *Perspectives on Psychological Science*, **12**, pp. 1100–1122.

Zander, U. and B. Kogut (1995). 'Knowledge and the speed of the transfer and imitation of organizational capabilities: an empirical test', *Organization Science*, **6**, pp. 76–92.

Zupic, I. and T. Čater (2015). 'Bibliometric methods in management and organization', *Organizational Research Methods*, **18**, pp. 429–472.

Saša Batistič is an assistant professor at the Department of Human Resource Studies at Tilburg University. He received his PhD from the University of Reading. His current research is focused on multi-level issues between HR systems, climates and employees' behaviours, leadership and organizational socialization. Saša's work has been published in journals such as *Leadership Quarterly*, *British Journal of Management*, *Human Resource Management Review* and *International Journal of Project Management.*

Paul van der Laken is a data scientist and applied management researcher. Paul conducted his PhD research on People Analytics and data-driven management at the Department of Human Resource Studies of Tilburg University. Since 2014, he has managed and executed data science, analytics and machine learning initiatives, mostly within the HR domain, at several national and multinational organizations. Paul's research has been published in journals such as *Human Resource Management* and *Human Resource Management Review*.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1
Appendix S2
Appendix S3
Appendix S4