Management Science and Big Data: A Text Mining Meta-Analysis Study

Arben Asllani
University of Tennessee – Chattanooga

D. Michael Long
University of Tennessee – Chattanooga

This study explores the impact of Big Data on the major research fields of management science. Text-mining and exploratory factor analysis are applied on articles published in the Management Science journal from 1998-2017 to measure changes in general trends, models, and methodologies. The findings show a significant shift in the underlying research themes in management science over this period. However, despite the Big Data buzz, there is no significant increase over the study period in Big Data categories. Although, the results did show a recent significant correlation between specific themes and Big Data models and techniques.

INTRODUCTION

This paper offers an empirical investigation on the impact that big data has had on Management Science (MSc) in today's dynamic business environment. The business of any science is to develop new understandings of past, present, or newly identified natural phenomena (Jackson, 1996). In the case of MSc, these phenomena have dramatically changed as business organizations have entered into the era of big data. For example, as digital storage has become less and less expensive, it is now economical for companies like Walmart to continuously collect and store vast amounts of information and data. Every hour Wal-Mart handles more than a million customer transactions and generates over 2.5 petabytes of data. Successful organizations can evaluate and synthesize this data and convert it into meaningful business information to use for better decision-making. For example, some web statistics suggest that by better integrating big data, healthcare organizations may save as much as $300 billion per year, thus reducing annual costs by almost $1,000 for every man, woman, and child in the US (Wikibon, 2012). For a typical Fortune 1000 company, a mere 10% increase in data accessibility can result in more than $65 million in additional net income (McCafferty, 2014). Unsurprisingly, MSc techniques have become the weapon of choice for many successful managers because it allows them to work with large amounts of data and make sense of all the information collected.

Davenport (2013) coined the term prescriptive analytics to describe models that "involve large-scale testing and optimization and are a means of embedding analytics into key processes and employee behaviors" (Davenport, 2013, p. 70). The 44th Annual Decision Sciences Institute Meeting in 2013 recognized the changes introduced in the business world by big data and challenged its members to rediscover their decision analytics roots. With big data and new technologies in the spotlight, traditional

MSc models are beginning to exhibit new characteristics and become more complex. This has forced scholars and practitioners to develop better techniques and acquire new problem-solving tools.

Consider how the classic traveling salesman problem (TSP), a true representative of traditional MSc, has evolved into its modern version of an unmanned aerial vehicle (UAV). TSP was first identified by mathematicians Hamilton and Kirkman in the 1800s (Biggs et al., 1986) and later formulated in 1930 by Lawler et al. (1991). TSP has been one of the most intensively studied problems in MSc and computational mathematics. In the problem, one is presented with a list of cities, and the distances between each pair of adjacent cities. The goal is to find the shortest possible route that stops at each city exactly once and ultimately returns to its origin. The solution to the puzzle can be a very complex NP-hard problem, especially when the number of cities is large. Therefore, as in many other MSc techniques, when the number of decision variables increases, the decision maker needs to employ alternative techniques such as the brute force attack, or the nearest neighbor algorithm. Although the brute force attack will generate an optimal solution, it is an inefficient algorithm; whereas the nearest neighbor algorithm is efficient, but may not yield an optimal solution.

The modern version of TSP is the unmanned aerial vehicle (UAV), i.e., an aircraft without a human pilot on board. The challenge for any UAV is similar to a TSP: fly over several destinations and return home via the shortest possible flying distance. However, the modern version of TSP faces new challenges. The traditional TSP assumes that the decision maker knows in advance the number of cities to visit and the distance between each pair of adjacent cities. However, for UAVs the information is constantly updated via position and movement sensors, or by remote human controllers. As such, using LP or the brute force attack algorithm becomes impractical for UAVs' scheduling problems. Thus, the nearest neighbor algorithm becomes the method of choice. A comparison between the TSP and UAV problem provides an intuitive explanation of the changes in the world of analytics and MSc, where sensors and other components of the Internet of Things produce a wide variety of data and information in high volumes and at high speeds.

Thus, this paper offers an empirical investigation of the influence big data may have on MSc in today's dynamic business environment. The next section briefly discusses Big Data and its potential impact on management science. This is followed with a discussion on the methodology and results of the text-mining and exploratory factor analysis approach. The paper closes with the main findings and conclusions.

THE IMPACT OF BIG DATA ON MANAGEMENT SCIENCE

By definition, big data is a combination of: (1) structured in-house operational databases; as well as, (2) external databases that contain data that is automatically captured, but often unstructured, from social media networks, web server logs, banking transactions, webpage content, financial market data, etc. Big data is characterized in this paper by the three Vs: volume, variety, and velocity (Laney, 2001; Dumbill, 2012). There are other sources that recommend four or five Vs to describe big data. However, volume, variety, and velocity are good representative characteristics of big data. Other Vs simply describe data in general.

Volume

Today, organizations use internal and external databases to generate and store large amounts of transactional data. High volumes of transactions are captured in structured and unstructured records, and then combined into de-normalized data warehouses. De-normalized data are intentionally redundant, and thus yield high volumes of data. Statistically, more data points should improve the accuracy of the input variables in MSc models. To automatically capture and process those input variables, MSc models use known extract-transform-load (ETL) processes which permit the live stream of input data by querying transactional records. The automatic capturing and processing of input data allow practitioners to design optimization models embedded within business processes (LaValle et al., 2011), and to periodically adjust input parameters to produce dynamic optimal solutions.

Variety

Variety refers to the mix of different data formats that are derived from different sources. Variety is an important dimension of big data that is usually considered an additional challenge when implementing optimization models. Since MSc models require the input data to be uniform, adding the ETL layer between the data sources and optimization models can mitigate variety-related issues. The transform component of the ETL layer can then be used to convert data into the format required by the optimization models.

Velocity

Velocity refers to the ever-changing nature of input data. Today, data and information are generated and flow into optimization models at a far greater rate. Trends such as mobile computing, online sales, smartphones, and social media networks is producing more new data than ever before. For example, traditional household meters are being replaced with digital smart meters that allow electric power companies to read usage every 15 minutes instead of once every month. Such fast-moving data offer new opportunities for real-time business intelligence.

Big Data and Modeling

Volume, variety and velocity describes the availability of big data that allow organizations to explore, formulate, and solve previously unsolvable problems. However, in the era of big data, successful implementation of optimization models requires decision scientists to not only store and process large amounts of data and information, but to also modify their problem-solving methods to better accommodate big data. Today, technologies such as cloud computing and distributed file systems have dramatically increased the ability of businesses to store and process information. Those technologies also offer large, dynamic distributed platforms for organizations to process input parameters and solve large-scale models. However, current big data algorithms (e.g., MapReduce), which run in distributed files systems (e.g., Hadoop) are embarrassingly parallel computational algorithms. This is a term used in computer programming to describe problems that can be divided into many parallel tasks with little effort (Herlihy & Shavit, 2012). As such, big data platforms which engage multiple clusters are not suited to run advanced MSc models. For example, a linear programming (LP) model traditionally requires decision makers to optimize a given goal under a predefined set of constraints. Finding a feasible solution within that construct requires input variables be in a single cluster. Accordingly, in the era of Big Data, genetic and other evolutionary algorithms are used far more successfully to solve LP models than the traditional simplex methods.

The implementation of MSc models also requires data scientists to consider a trade-off between practical but less-than-optimal solutions, versus optimal but more complex and delayed solutions. That is, an approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem (Sashihara, 2012). Thus, it is necessary that management science continuously adapts in order to solve ever changing real-life problems. Successful companies such as Amazon and Google are at the forefront of incorporating big data for big optimizations. While Amazon consistently strives to reduce the delivery times of its orders (Levy, 2011), Google has paved the way for big data optimization strategies that use search engine optimization to capture, process, and then produce insights from large datasets (Sashihara, 2012).

As business analytics models continue to be widely implemented in the business world, the challenge of converting big data into optimal decisions and actions remains. To meet that challenge, practitioners and scholars of MSc need to modify their approach in implementing traditional MSc tools and techniques to better accommodate new business models affected by high volume, wide variety, and high velocity of data. Some of these changes are described using the acronym SMART (***S***treaming data; The ***M***AD approach; ***A***utomatic decision making processes; ***R***eal-time operational intelligence; and ***T***raditional tools and techniques). SMART management science models must process high volumes of streaming data as business transactions generate them, while the MAD approach allows the models to attract data from multiple and heterogeneous sources while replacing missing values. At the same time, SMART MSc models are also directly connected with operational databases that enables decision support systems to dynamically respond to ever-changing data input and provide automatic solutions free of human interaction with the system. In this sense, generating real-time actionable operational business intelligence is the new goal of MSc models. However, even though SMART elements are necessary, they do not mark sufficient changes in MSc in the era of big data. It is doubtful modern MSc will fully replace the traditional values in the field since the new models are rooted in the tools and technology developed over the last century. Thus, further empirical research is needed to better understand the metamorphosis of MSc.

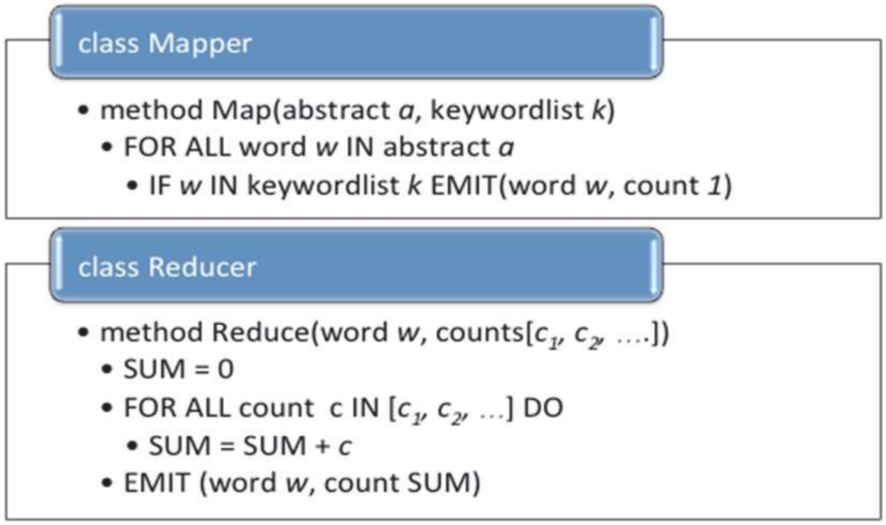
METHODOLOGY AND RESULTS

This study uses a text-mining methodology and exploratory factor analysis to better understand changes in research themes and the impact of Big Data on the field of management science. Article abstracts, keywords, and full text corpus are collected from the Management Science journal during the 1998-2017 period. As previously mentioned, the analysis consists of two parts: (1) exploring the major research themes of management science and their transformation; and (2) investigating the infusion of Big Data concepts, tools, and methodologies in research.

Part A: The Changing Nature of Management Science

First, a MapReduce program is applied to identify the frequency of over 7000 management science keywords in a set of over 2900 abstracts from all 239 issues published in the Management Science journal from 1998-2017. In this stage, due to the complexity and size of the text, abstracts are used as the text corpus. Using abstracts is a form of stratified sampling and is "almost always more representative than non-stratified" sampling (Biber, 1993). One hundred percent of the abstracts over the 1998-2017 study period are used to ensure better variability of language. The traditional MapReduce algorithm (Dean & Ghemawat, 2008) is modified here to count the frequency of the words that exists in both the abstract corpus and keyword list as defined by the authors of the published articles. The pseudo code of the MapReduce program is shown in Figure 1. The MapReduce algorithm is executed in a Hadoop cluster with four nodes and produced just over 4000 highly frequented keywords. Due to space limitations, Figure 2 shows only a sample of the keywords and Table 1 lists the top 30 most frequent keywords appearing in the abstract corpus during the period 1998-2017.

FIGURE 1
MODIFIED MAP-REDUCE PROGRAM TO INCLUDE KEYWORDS



The exploratory factor analysis that follows requires a much smaller number of measurable variables. Therefore, keywords are limited to those that were repeated more than 200 times as shown in Figure 2. The frequency cutoff of 200 represents the keywords that lie above the bend in the "elbow" in Figure 2. This reduced the list from 4000 words down to a more manageable 113 keywords. The 113 top keywords were further processed to combine synonyms, and other obvious groupings. For example, words like *knowledge*, *data*, and *information* are combined into just "*information*"; and stems of words like *finance*, *financial*, and *financing* are represented by the stem "*financ*∗". As a result, the top 113 keywords are narrowed down to 87 keywords that are available for factor analysis.

FIGURE 2
GRAPHICAL REPRESENTATION OF KEYWORD FREQUENCIES

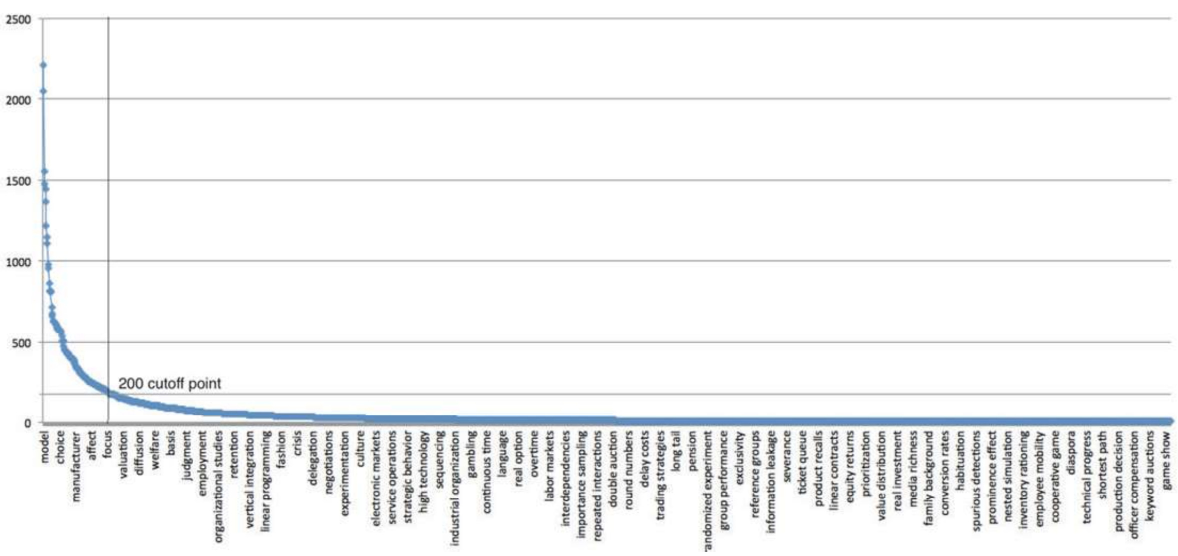


TABLE 1
TOP 30 FREQUENT WORDS IN MANAGEMENT SCIENCE (1998-2017)

Keyword	Frequency	Keyword	Frequency
Model	2209	Strategy	717
Information	2052	Theory	672
Market	1554	Behavior	662
Management	1475	Prices	629
Risk	1445	Production	626
Performance	1363	Process	624
Optimal	1216	Markets	624
Cost	1146	Policy	611
Demand	1113	Design	604
Quality	980	Knowledge	594
Value	954	Competition	587
Inventory	863	Technology	579
Pricing	814	Capacity	577
Models	812	Choice	575
Service	809	Uncertainty	571

Once the most frequent keywords are identified, the frequency index f_{ij} of each key word i in article j is calculated as follows:

$$F_{yt} = \frac{\sum_{i=1}^{|T_t|} f_{iy}}{|T_t|} \tag{1}$$

where F_{ij} is the frequency of keyword i in article j and T_j is the total number of words in article j. To calculate F_{ij} and T_j the MapReduce algorithm is ran against each full article with the new keyword list as an input to the program. In order to manage the complexity and size of the data, full text articles are only collected for years 1998, 2002, 2007, 2012, and 2017. A total of 730 articles are collected, which serve as the basis for the exploratory factor analysis. Figure 3 shows the overall model parameters for the exploratory factor analysis. A Kaiser-Meyer-Olkin (KMO) value of 0.667 indicates the data is suitable for factor analysis (Cerny & Kaiser, 1977). In addition, Bartlett's test of sphericity is used to test the hypothesis that the variables are unrelated and therefore unsuitable for structure detection. The small value of significance (p< 0.001) indicates that a factor analysis is useful with this data (Snedecor & Cochran, 1989).

FIGURE 3
MODEL VALIDITY FOR FACTOR ANALYSIS

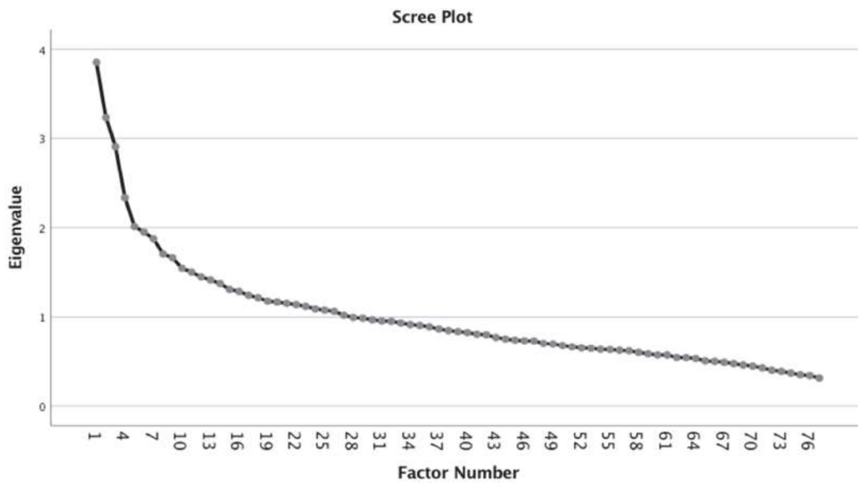
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.667
Bartlett's Test of Sphericity	Approx. Chi-Square	8179.298
	df	2926
	Sig.	.000

The scree plot in Figure 4 is used to help determine the number of factors for principal component analysis. The number of factors chosen was 6, because visually one can see a significant drop in

Eigenvalues through the first six factors. Beyond this point the remaining factors belong to the flatter area of the diagram.

FIGURE 4
IDENTIFYING THE NUMBER OF FACTORS



The factor-loading matrix shown in Figure 5 displays the six factors. The output is sorted in descending order and the keywords are filtered where the factor loading is greater than 0.35. Based on the set of keywords for each factor, it appears that the following six major research themes have dominated the discourse in management science over the last two decades: Supply Chain Management, Financial Modeling, Mathematical Programming, Learning Organizations, Market Equilibrium, and Incentive Contracting.

FIGURE 5
MAJOR THEMES IN MANAGEMENT SCIENCE-LOADING FACTOR > 0.35

Rotated Factor Matrix

Factor	1	2	3	4	5	6
Themes	**Supply Chain Management**	**Financial Modeling**	**Mathematical Programming**	**Learning Organization**	**Market Equilibrium**	**Incentive_Contracting**
demand	0.647					
inventory	0.561					
supply_chain	0.442					
retailers	0.39					
return		0.715				
portfolio		0.484				
financial		0.462				
volatility		0.424				
investment		0.412				
programming			-0.524			
optimal			-0.482			
algorithm			-0.369			
linear			-0.356			
utility				-0.427		
organization				0.389		
theory				-0.376		
learning				0.351		
pricing					0.518	
equilibrium					0.426	
market					0.423	
profits					0.388	
incentive						0.524
contract						0.424
effort						0.381

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization
Rotation converged in 22 iterations.

Supply Chain Management

The first factor identified is *Supply Chain Management.* The keywords for this factor use words like *demand*, *inventory*, *supply chain*, and *retailers*. This theme started in the late 1980s (Adam & Swamidass, 1989; Flynn, Sakakibara, Schroeder, & Bates, 1990) and was significantly impacted by advances in technology, Internet of Things and smart devices, social media, robotics and Big Data (Melnyk, Flyn, & Awaysheh, 2017).

Financial Modeling

Another important theme derived from the list of keywords is *Financial Modeling*. The keywords with the highest factor loading in this category include *return, portfolio, financial, volatility,* and *investment*. Financial management and modeling have been present in management science research for years. One issue that Financial Modeling emphasizes is the growing importance and use of sophisticated financial models in practice (Zenios & William, 1992). The text mining analysis also shows a strong presence of financial modeling on several aspects, such as optimization, product development, supply chain investments, etc.

Mathematical Programming

The third factor in our analysis is *Mathematical Programming*. This factor is represented in the analysis with keywords such as *programming, optimal, algorithm*, and *linear.* However, research in these areas started much earlier with the work of (Dantzig, 1955), and continues to remain relevant in today's world of Big Data business (Asllani, 2014).

Learning Organizations

The fourth factor showed a significant loading of keywords like *utility, organization, theory,* and *learning*. Factor analysis indicates a strong correlation between these keywords, although further research is needed to show a causal association within this group. For now, this factor is identified as the *Learning Organization* theme. Research on learning organizations examines a system's ability to adjust and improve in response to ever changing market conditions. Good examples in Learning Organizations include studies on shared learning (Paul, 1990); social learning (Li, Xiao, Xue, & Olivia, 2016); and open innovation (Parker & Van Alstyne, 2017).

Market Equilibrium

The study of *Market Equilibrium* has been another central theme in the field of management science over the years and is the fifth factor in this study. Some notable research in this area include (Biais, et al, 2013; Subramanian, 2013; Draganska and Jain, 2004). The *Market Equilibrium* theme is represented in the analysis using a significant group of keywords such as *pricing, equilibrium, market,* and *profits*.

Incentive Contracting

The last factor is *Incentive Contracting* and is represented by such keywords as *incentive, contract,* and *effort*. The keywords in this theme are mainly related to coordinating contracts and providing incentives among players of different decentralized systems. For example, Krishnan, Kapuscinski, & Butz (2010) investigated the impact of lowering downstream inventories of a manufacturer's supply chain, and the incentives to encourage the retailer to exert more effort to increase the sales of the manufacturer's products. Similar research includes incentive contracts for portfolio managers (Cohen & Starks, 1988); risk incentives on performance-based contracts (Grinblatt & Titman, 1989); incentive contracts on compensation (Golec, 1993); and forecast-based or linear contracts (Chen, Lai, & Xiao, 2015).

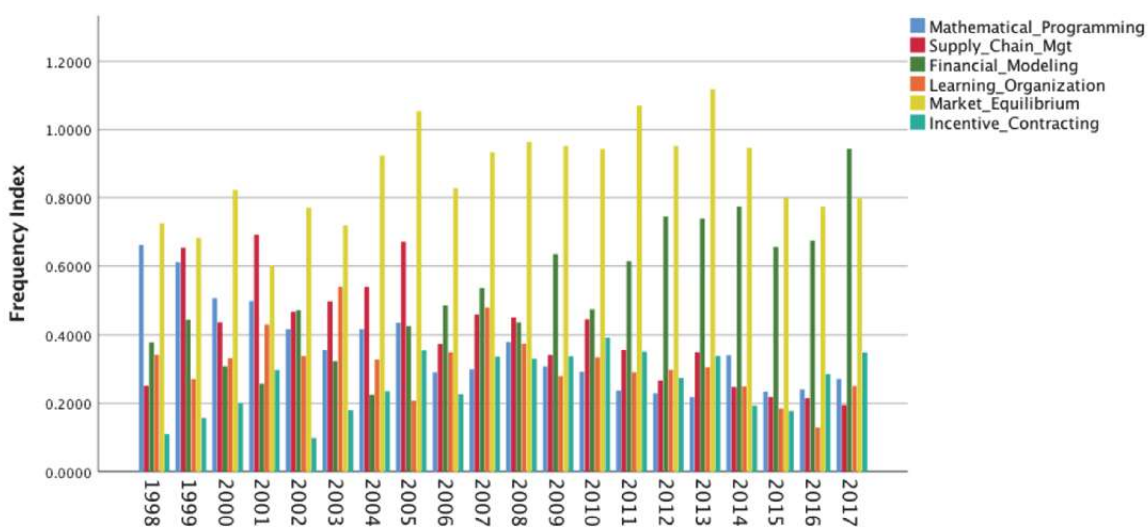
Next, the average frequency index is calculated for each theme during a given year as shown below:

$$F_{yt} = \frac{\sum_{i=1}^{|T_t|} f_{iy}}{|T_t|} \qquad (2)$$

where:
t - represents the theme number (t = 1, 2, …, 6)
y- represents the year (y = 1998, 1999, …, 2017)
F_{yt} - is frequency of theme *t* in year *y*
T_t - the set of keywords in theme t.
$|T_t|$-the cardinality of set T_t, that is, the number of keywords in each set T_t
f_{iy} - is frequency of keyword *i* in year *y*

Figure 6 represents the frequency of each research theme during the 1998-2017 period. As shown, the text mining analysis indicates that the *Market Equilibrium* theme has dominated the discourse of the Management Science journal. *Mathematical Programming* was also at the center of the research discourse during the initial years of our analysis period. However, there was a gradual decrease of this theme over the last twenty years. Similarly, *Supply Chain Management* theme has decreased over the years, especially after 2005. *Financial Modeling* research appears to exhibit the opposite behavior. It starts rather slowly in the late 1990s and early 2000s, but has an ever-increasing presence over the last 10 years. The *Incentive Contracting* research theme is comparatively weak during the entire 1998-2017 period.

FIGURE 6
MAJOR THEMES OF MANAGEMENT SCIENCE



Part B: Infusion of Big Data Concepts in Management Science

This part of the analysis uses the Linguistic Inquiry and Word Count (LIWC) program, a text analysis program that counts words in psychologically meaningful categories. LIWC has the "ability to detect meaning in a wide variety of experimental settings" (Tausczik & Pennebaker, 2010) by using psychological categories that show attention focus, emotion, social relationships, thinking styles, and individual differences. In addition, the LIWC software program allows a researcher to create a unique dictionary to investigate the presence of a specific category or a set of categories in a text file or series of text documents. Based on the earlier discussion on Big Data and the six themes from the analysis in Part A, a dictionary file is created and shown in Table 2. The dictionary and full text of all 730 total articles published in the years 1998, 2002, 2007, 2012, and 2017 are uploaded into the LIWC program and processed. An analysis of variance (ANOVA) on the results generated by LIWC is performed using SPSS software.

The ANOVA in Figure 7 tests for differences in the means of the frequency indices for each year under investigation. The results show significant changes in the means of the frequencies for all six management science research themes. This suggests that the major themes in management science have changed significantly across the 5-year periods. It is also interesting to note that *Market Equilibrium* dominated all of the management science themes across all periods, except in 2017, where *Financial Modeling* dominated.

More importantly, the discussion of *Big Data* driven categories (*Data*, *Big Data*, *SMART*) do not show any significant change over the last two decades despite of all the buzz this category has picked up in trade magazines, professional conferences, and white paper publications. One possible explanation for this is that over the years MSc has maintained a relatively constant focus on data, so there is no change to be observed. As shown later in the paper, Big Data and Data categories have a strong positive correlation of 99.9%. This suggests that "big data" is just another way of saying "data". Thus, the buzzword Big Data may simply be "white noise" that provides no additional information on the discourse in management science.

To further investigate the relationship between the themes in management science and Big Data, Pearson correlations are calculated between these categories. These results are shown in Figure 8. A significant correlation of 0.138 is found between *Supply Chain Management* and *SMART* model approaches. Although not shown, the correlation between themes is also ran for each year. A significant and stronger correlation between *Supply Chain Management* and SMART is only found in 2017 at □ = 0.566. There is also a significant correlation between *Mathematical Programing* and *SMART* at □ = 0.208 for 2017. Another interesting finding is that although the *Supply Chain Management* theme is significantly correlated with the *SMART* model category, it is not significantly related to categories such as *Data* and *Big Data*. This indicates that the focus of the research is not *Data* or *Big Data* per se, but rather on models and technologies around *Big Data.* Also, the very strong and positive correlation (*.999*) between *Data* and *Big Data* shows that management science research makes virtually no distinction between these two categories, at least in the context of the 3Vs driven-definition of *Big Data,* as used in this analysis. Thus, the term "big data" may just be another way of saying "data".

TABLE 2
CLASSIFICATION OF KEYWORDS INTO CATEGORIES

Keywords	Categories								
	Supply Chain Mgt	Financial Modeling	Math Program	Learning Org.	Market Eq.	Incentive Contracts	Data	Big Data	SMART
demand*	x								
inventory*	x								
supply chain	x								
retailers*	x								
return*		x							
portfolio*		x							
financ*		x							
volatil*		x							
invest*		x							
programm*			x						
optim*			x						
algorithm*			x						
linear*			x						
utilit*				x					
organization*				x					
theory				x					
learn*				x					
pric*					x				
equilibrium*					x				
market*					x				
profit*					x				
incentive*						x			
contract*						x			
effort*						x			
volume							x	x	
velocity							x	x	
variety							x	x	
veracity							x		
integrity							x		
stream*									x
magnet*									x
agil*									x
depth									x
automatic*									x
self learning									x
real time									x

Other notable correlations include significantly negative correlations between Financial Modeling and Mathematical Programming and Learning Organization. This gives statistical weight to the trend previously observed in Figure 6, i.e., Financial Modeling is increasing while the other two themes are decreasing. This indicates there was a significant shift toward Financial Modeling research and a significant shift away from Mathematical Programing and Supply Chain over the study period.

FIGURE 7
COMPARISON OF MEANS FOR 1998-2002-2007-2012-2017

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Supply_Chain_Mgt	Between Groups	4.093	4	1.023	5.958	.000
	Within Groups	124.527	725	.172		
	Total	128.620	729			
Financial_Modeling	Between Groups	18.160	4	4.540	10.273	.000
	Within Groups	320.402	725	.442		
	Total	338.562	729			
Mathematical_Programming	Between Groups	4.294	4	1.074	12.296	.000
	Within Groups	63.300	725	.087		
	Total	67.594	729			
Learning_Organization	Between Groups	2.072	4	.518	4.699	.001
	Within Groups	79.938	725	.110		
	Total	82.011	729			
Market_Equilibrium	Between Groups	5.751	4	1.438	2.851	.023
	Within Groups	365.656	725	.504		
	Total	371.407	729			
Incentive_Contracting	Between Groups	2.041	4	.510	3.719	.005
	Within Groups	99.487	725	.137		
	Total	101.528	729			
Data	Between Groups	.016	4	.004	.605	.659
	Within Groups	4.818	725	.007		
	Total	4.835	729			
Big_Data	Between Groups	.016	4	.004	.600	.663
	Within Groups	4.805	725	.007		
	Total	4.821	729			
SMART	Between Groups	.010	4	.003	1.874	.113
	Within Groups	1.003	725	.001		
	Total	1.014	729			

FIGURE 8
CORRELATION BETWEEN THEMES AND BIG DATA

Correlations

		Supply_Chain_Mgt	Financial_Modeling	Mathematical_Programming	Learning_Organization	Market_Equilibrium	Incentive_Contracting	Data	Big_Data	SMART
Supply_Chain_Mgt	Pearson Correlation	1	-.130**	.185**	-.139**	.125**	.022	-.019	-.017	.138**
	Sig. (2-tailed)		.000	.000	.000	.001	.552	.617	.651	.000
	N	730	730	730	730	730	730	730	730	730
Financial_Modeling	Pearson Correlation	-.130**	1	-.117**	-.093*	.065	-.056	.004	.004	-.062
	Sig. (2-tailed)	.000		.002	.012	.078	.132	.919	.908	.096
	N	730	730	730	730	730	730	730	730	730
Mathematical_Programming	Pearson Correlation	.185**	-.117**	1	-.101**	-.064	.072	-.062	-.059	.042
	Sig. (2-tailed)	.000	.002		.006	.083	.051	.095	.110	.260
	N	730	730	730	730	730	730	730	730	730
Learning_Organization	Pearson Correlation	-.139**	-.093*	-.101**	1	-.103**	-.013	.024	.025	.005
	Sig. (2-tailed)	.000	.012	.006		.005	.728	.510	.499	.901
	N	730	730	730	730	730	730	730	730	730
Market_Equilibrium	Pearson Correlation	.125**	.065	-.064	-.103**	1	.018	.020	.022	-.022
	Sig. (2-tailed)	.001	.078	.083	.005		.619	.584	.555	.554
	N	730	730	730	730	730	730	730	730	730
Incentive_Contracting	Pearson Correlation	.022	-.056	.072	-.013	.018	1	-.051	-.051	-.025
	Sig. (2-tailed)	.552	.132	.051	.728	.619		.171	.172	.492
	N	730	730	730	730	730	730	730	730	730
Data	Pearson Correlation	-.019	.004	-.062	.024	.020	-.051	1	.999**	.023
	Sig. (2-tailed)	.617	.919	.095	.510	.584	.171		.000	.533
	N	730	730	730	730	730	730	730	730	730
Big_Data	Pearson Correlation	-.017	.004	-.059	.025	.022	-.051	.999**	1	.023
	Sig. (2-tailed)	.651	.908	.110	.499	.555	.172	.000		.538
	N	730	730	730	730	730	730	730	730	730
SMART	Pearson Correlation	.138**	-.062	.042	.005	-.022	-.025	.023	.023	1
	Sig. (2-tailed)	.000	.096	.260	.901	.554	.492	.533	.538	
	N	730	730	730	730	730	730	730	730	730

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

CONCLUSIONS

This paper offers a text mining approach to examine the themes of management science research over the last two decades. In addition, the study also explores how management science has responded to the latest buzz of Big Data models, tools, and technologies. A unique approach is offered using a combination of text mining analysis (i.e., Big Data programming MapReduce), a LIWC linguistic inquiry software program, traditional factor analysis, ANOVA, and correlation analysis. The demonstrated methodology can be used to investigate the text based discourse of any science in attempts to better understand past or present trends in research themes. The application of this methodology is motivated by the need to validate recent claims that management science is significantly changing into "prescriptive analytics", and its models are becoming more complex to better represent data and Big Data driven methodologies.

In addition to the six research themes derived from text mining and exploratory factor analysis, three more categories are created: Data, Big Data, and SMART models and technologies, where each category is represented by a specific set of keywords. Over 7000 keywords, 2900 article abstracts, and 730 full text articles from the Management Science journal are applied in a two-stage analysis. The first stage explored the major research themes of management science and their transformation; and the second stage investigated the infusion of Big Data concepts, tools, and methodologies.

It was found that the Market Equilibrium theme was relatively constant over the time period and dominated the research in the Management Science journal. Interestingly, the study shows that while Mathematical Programming and Supply Chain Management themes were important during the initial years, these themes grew weaker in later years. Contrarily, the Financial Modeling research theme appears weaker at the beginning of the study period, but has grown to show a more central and stronger presence over the last 10 years.

Curiously, Data, Big Data, and SMART model categories showed no significant shift over the last 20 years. One reason for the lack of “change” around *Big Data* driven categories may be there is no change to be observed. *Big Data* and *Data* are found to have a strong positive correlation of 99.9%. This suggests that “big data” is just another way of saying “data”. Thus, the term Big Data may simply be “white noise” that can provide no additional information on the discourse in management science.

Finally, the relationship between the six management science themes and Big Data is investigated. It was found that only the Supply Chain Management theme is correlated to SMART models and techniques. Supply Chain Management theme is also positively correlated to changes in other categories, such as Mathematical Programming and Market Equilibrium. This gives some indication that the focus of the research is not *Data* or *Big Data* per se, but on models and technologies around *Big Data*.

REFERENCES

Adam, E. E., & Swamidass, P. M. (1989). Assessing Operations Management from a Strategic Perspective. *Journal of Management*, *15*(2), 181-203.

Asllani, A. (2014). *Business Analytics with Management Science Models and Methods.* FT Press.

Biais, B., Bisière, C., & Pouget, S. (2013). Equilibrium discovery and preopening mechanisms in an experimental market. *Management Science*, *60*(3), 753-769.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, *8*(4), 243-257.

Biggs, N. L., Lloyd, E. K., & Wilson, R. J. (1986). *Graph Theory 1736-1936.* Oxford: Clarendon Press.

Cerny, C. A., & Kaiser , H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, *12*(1), 43-47.

Chen, F., Lai, G., & Xiao, W. (2015). Provision of Incentives for Information Acquisition: Forecast-Based Contracts vs. Menus of Linear Contracts. *Management Science*, *62*(7), 1899-1914.

Cohen, S. I., & Starks, L. T. (1988). Estimation risk and incentive contracts for portfolio managers. *Management Science*, *34*(9), 1067-1079.

Dantzig, G. B. (1955). Linear Programming Under Uncertainty. *Management Science*, *1*(3 and 4), 197-206.

Davenport, T. H. (2013, December). Analytics 3.0. *Harvard Business Review*, 65-72.

Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

Draganska, M., & Jain, D. (2004). A likelihood approach to estimating market equilibrium models. *Management Science*, *50*(5), 605-616.

Dumbill, E. (2012, January 11). *What is big data?* (O'Reilly Data) Retrieved May 14, 2014, from http://strata.oreilly.com/2012/01/what-is-big-data.html

Flynn, B. B., Sakakibara, S., Schroeder, R. G., Bates, K. A., & Flynn, E. J. (1990). Empirical Research Methods in Operations Management. *Journal of Operations Management*, *9*(2), 250-284.

Golec, J. H. (1993). The effects of incentive compensation contracts on the risk and return performance of commodity trading advisors. *Management Science*, *39*(11), 1396-1406.

Grinblatt, M., & Titman, S. (1989). Adverse risk incentives and the design of performance-based contracts. *Management Science*, *35*(7), 807-822.

Herlihy, M., & Shavit, N. (2012). *The Art of Multiprocessor Programming, Revised Reprint (revised ed.).* Elsevier.

Jackson, E. A. (1996). *The Second Metamorphosis of Science A Second View.* Retrieved March 12, 2017, from Santa Fe: http://samoa.santafe.edu/media/workingpapers/96-06-039.pdf

Krishnan, H., Kapuscinski, R., & Butz, D. A. (2010). Quick Response and Retailer Effort. *Management Science*, *56*(6), 962 - 977.

Laney, D. (2001, February 6). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011, September 8). *Big data, analytics and the path from insights to value.* Retrieved May 19, 2014, from INFORM: http://www.analytics-magazine.org/special-reports/260-big-data-analytics-and-the-path-from-insights-to-value.pdf

Lawler, E. L., Lenstra, J. K., Kan, A. R., & Shmoys, D. B. (1991). *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization.* John Wiley & Sons Inc.

Levy, S. (2011, Nov. 13). *Jeff Bezos Owns the Web in More Ways Than You Think.* Retrieved May 28, 2014, from Wired Magazine: http://www.wired.com/2011/11/ff_bezos/all/1

Li, Z., Xiao, F., Xue, B., & Olivia, R. L. (2016). Utility-based link recommendation for online social networks. *Management Science*, *63*(6), 1938-1952.

McCafferty, D. (2014, February 2). *Analytics/Big Data*. Retrieved March 20, 2017, from Surprising Statistics About Big Data: http://www.baselinemag.com/analytics-big-data/slideshows/surprising-statistics-about-big-data.html

Melnyk, S. A., Flyn, B. B., & Awaysheh, A. (2017). The best of times and the worst of times: empirical operations and supply chain management research. *International Journal of Production Research*, 1-28.
Parker, G., & Van Alstyne, M. (2017). Innovation, openness, and platform control. *Management Science , Article in Advance*, NA.
Paul, A. (1990). Shared learning. *Management Science , 36*(8), 938-957.
Sashihara, S. (2012, April 17). From Big Data to Big Optimization: What Every Executive Needs to Know. *The European Financial Review* , Online.
Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods, Eighth Edition.* Iowa State University Press.
Subramanian, A. (2013). Product market competition, managerial compensation, and firm size in market equilibrium. *Management Science , 59*(7), 1612-1630.
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology , 29*(1), 24-54.
Wikibon. (2012, August 1). Retrieved March 15, 2017, from A Comprehensive List of Big Data Statistics: http://wikibon.org/blog/big-data-statistics/
Zenios, S. A., & William, Z. T. (1992). Introduction to the Focused Issue on Financial Modeling. *Management Science , 38*(11), 1525-1528.