BIG DATA ANALYTICS IN SUPPLY CHAIN MANAGEMENT: TRENDS AND RELATED RESEARCH

Ivan Varela Rozados
Supply Chain Research Centre, School of Management, Cranfield University
Cranfield, Bedford MK43 0AL, UK

Benny Tjahjono
Supply Chain Research Centre, School of Management, Cranfield University
Cranfield, Bedford MK43 0AL, UK, E-mail: b.tjahjono@cranfield.ac.uk

ABSTRACT
Big Data Analytics offers vast prospects in today's business transformation. Whilst big data have remarkably captured the attentions of both practitioners and researchers especially in the financial services and marketing sectors, there is a myriad of premises that big data analytics can play even more crucial roles in Supply Chain Management (SCM). This paper therefore intends to explore these premises. The investigation ranges from the fundamentals of big data analytics, its taxonomy and the level of maturity of big data analytics solutions in each of them, to implementation issues and best practices. Finally, some examples of advanced analytics applications will also be presented as a way of unveiling some of the relatively unexplored territories in big data analytics research.

Keywords: Analytics, Big Data, Business Transformation, Data Science, Predictive Analytics, Supply Chain Management.

1. INTRODUCTION
Major business players who embrace Big Data as a new paradigm are seemingly offered endless promises of business transformation and operational efficiency improvements. In Supply Chain Management (SCM) in particular, some examples have captured the attention of both practitioners and researchers, hitting the headlines of recent news. Amazon uses Big Data to monitor, track and secure 1.5 billion items in its inventory that are laying around 200 fulfilment centres around the world, and then relies on predictive analytics for its 'anticipatory shipping' to predict when a customer will purchase a product, and pre-ship it to a depot close to the final destination (Ritson, 2014). Wal-Mart handles more than a million customer transactions each hour (Sanders, 2014), imports information into databases to contain more than 2.5 petabytes and asked their suppliers to tag shipments with radio frequency identification (RFID) systems (Feng et al., 2014) that can generate 100 to 1000 times the data of conventional bar code systems. UPS deployment of telematics in their freight segment helped in their global redesign of logistical networks (Davenport and Patil, 2012).

SCM organisations are inundated with data, so much that McAfee and Brynjolfsson (2012) reported "business collect more data than they know what to do with". This is apparently true in firms that are considered a benchmark for warehouse data management, marketing or transportation. Nonetheless, the reality reveals that these cases are not just anecdotes of success; they are the face of a change where failure to adapt could mean irrelevance. Hopkins et al. (2010) reported from a Sloan Management Review survey that analytics' top performers outpace industry peers performance up to three times.

While most organisations have high expectations from Big Data Analytics (BDA) in their supply chain, the actual use is limited and many firms struggle to unveil its business value (Pearson et al., 2014). In the pursuit of a change to that situation and a willingness to guide the SCM practice to capitalise BDA, the overall aim of this research is to close the knowledge gap between data science and Supply Chain Management domain, linking the data, technology and functional knowledge in BDA applications across procurement, transportation, warehouse operations and marketing. Specifically, this paper will (1) redefine, by research on previous scientific work, what BDA means in the context of Supply Chain Management, and how it differs and has evolved from previous analytics technologies; (2) develop taxonomy of Big Data within SCM that identifies and classifies the different sources and types of data arising in modern supply chains and (3) suggest some applications of BDA and show the potential high value this technology offers to solve complex SCM challenges.

2. RELATED WORK AND KNOWLEDGE GAPS

BDA in SCM is a heterogeneous topic as it builds upon cross-disciplinary work from various areas. Business challenges rarely show up in the appearance of a perfect data problem (Provost and Fawcett, 2013), and even when data are abundant, practitioners have difficulties to incorporate it into their complex decision making that adds business value (Shah et al., 2012). Hazen et al. (2014) described the field as “new and emergent”. Barratt et al. (2014) recognised the need for searching more practical implications of BDA in SCM, and they manifested their intention to attract research projects about BDA for the Council of Supply Chain Management Professionals (CSCMP) 2014 annual conference. Sanders (2014) published the first book combining both SCM theory and Big Data, Big Data Driven Supply Chain Management that provides great insight in the managerial implications of implementing BDA.

The most cited ‘call for research’ in BDA came from Waller and Fawcett (2013), who highlighted the importance that conducting scientific research in the area where SCM intersects with Big Data and advanced analytics techniques from Operational Research domain could illuminate a “myriad of new opportunities” for both practitioners and academia. They attributed the lack of publications or applications of data science, predictive analytics, and Big Data in the context of SCM, to not fully address the conceptual requirements in integrating domain knowledge with quantitative skills. From the abovementioned evidence, a clear knowledge gap has been identified, and with the intention to bridge the gap, this research has set off.

3. RESEARCH METHODOLOGY

Gimenez (2005) argued that conducting research in SCM through the application of multiple methods assures that variances are trait-related and not method-related, as well as the fact that each methodology is more appropriate for the development of a particular stage of the research. In order to build a definition of BDA and its associated list of themes, the first part of the research was about understanding BDA in its own terms. Like most of the areas close to Big Data, BDA meaning is mainly what people have made of it. The *systematic literature review* transformed a broad spectrum of documentation first into a delimited set of themes, and then into synthesised extracted data. The analysis of the themes structure resulted in a somewhat exhaustive description of its features, specifically in the SCM context and produced a solid base of knowledge and substantive justification on which to build subsequent phases of the research.

The inclusion of the *case study* in this work was to maintain practicality at the core. Case studies investigated simultaneous BDA examples, typically in emerging practices, thus being a successful way of including the latest trends detected in the industry. Both business cases from the literature as well as those reported through semi-structured interviews with consultants at a major consulting company in the UK were used. The combined systematic literature review and case studies was used to create a toolset that is based on academic sources as well as practical experience and that was helpful and useful to use.

4. SYSTEMATIC REVIEW ON BIG DATA ANALYTICS

The identified lack of results from previous peer-review published work brought numerous questions to a field not yet formally covered. Closing some of this research gaps drove the following review question: "*What are the definition and the thematic domains of BDA in SCM context, and how they apply to Big Data sources in modern supply chains*?" This question should help ensure a comprehensive review, but it would not necessarily lead to the direct research findings.

The search strategy was developed by first identifying the relevant data sources. An extensive selection of databases was selected as a way of having access to a diverse range of publications (e.g. journal articles, conference proceedings, dissertations, theses, books, magazine articles, newspaper articles and trade journals). Databases such as EBSCO, Emerald and Scopus were searched. This process was complemented with an Internet search to retrieve additional materials, e.g. white papers. Keywords identified were directly associated with BDA (e.g. Big Data Analytics, Big Data, analytics, advanced analytics, predictive analytics, data science, Supply Chain Analytics, etc.). These keywords were then combined with terms such as "Supply Chain Management" or "SCM" in order to ensure their relevance to this study. Depending on the database, the search field for the strings was also adopted (Title, Abstract, Keywords, etc.).

The search process was iterative in nature. Once the articles were collected, the abstracts and keywords were used as a preliminary filter, and those articles not relevant to the review were removed from the list. There were 129 items proposed for review in their full content, including journal articles, books and other reports. The journals reviewed are for instance Big Data, International Journal of Logistics Management, Harvard Business Review, Journal of Business Logistics, Supply Chain Management Review, Supply Chain Quarterly etc.

By carefully revising each item, a collection of 11 themes that aggregated research contents was built. Then each document was indexed with a score in all the 11 themes depending on its incidence (whether that piece of work had a topic focus, detailed discussion or reference to the BDA theme, as well as a clear context in SCM). The documents whose total index was low, i.e. those who only refer to one of the topics to a low degree or have very little to do with SCM, were removed. This criterion was cross-validated, so items not contextualised in SCM but are otherwise contributors to important concepts in BDA, were still included. This process concluded in 85 papers, and subsequent cross-checking of references increased the list to 87.

4.1 Information flows in SCM: An Extended Supply Chain

Supply Chain Management is defined by Christopher (2011) as the management, across and within a network of upstream and downstream organisations, of both relationships and flows of material, information and resources. For centuries, information of the goods that were stored and shipped was transported with the goods themselves in the form of physical documents, but actual supply chains have little resemblance with that. Our interest in the extended supply chain considers a model where technologies, such as BDA, synchronise SCM by driving a separate flow of information (Edwards et al., 2001) that enables organisations to capture, process, analyse, store and exchange data about their operations (Smith et al., 2007).

An extended supply chain is a multi-echelon system that connects organisations allowing collaboration and integration, as competition between supply chains is perceived to be more intense than individual firms (Antai and Olson, 2013). The long list of IT systems that have been used for this purpose before included Electronic Data Interchange (EDI), Vendor Managed Inventory (VMI), Efficient Consumer Response (ECR), Collaborative Planning Forecasting and Replenishment (CPFR), Collaborative Planning System (CPS), Sales Force Automation (SFA), Point Of Sale data (POS) or Customer Service Manager (CSM) (Barrat and Oke, 2007). Amongst the phases of the SCM information flow (capture, process, analyse, store and exchange), BDA specifically focus on the analysis. Tools that facilitate analysis of SCM data are englobed in the "Analytics" domain.

4.2 Advanced analytics

Advanced analytics is defined as the scientific process of transforming data into insight for making better decisions. As a formal discipline, advanced analytics have grown under the Operational Research domain. There are some fields that have considerable overlap with analytics, and also different accepted classifications for the types of analytics (Chae et al., 2014). Lustig et al. (2010) proposed a classification of advanced analytics in three main sub-types.

4.2.1 Descriptive analytics

These are the data analysis made to describe a past business situation in a way that trends, patterns and exceptions become apparent. The first level of analytics explores what has occurred as a way to gain insight for better approaching the future, usually trying to answer the question of "what happened". Some of the techniques that are included in this group, as detailed in Zeng et al. (2011), include:

- Standard reporting and dashboards: Off-the-shelf packages, executing queries internally implemented.
- Ad-hoc reporting: Queries customised by the final user on the interface of the package.
- Query drilldown (OLAP): A first level of data mining that allows obtaining complex information from databases by aggregating multidimensional structures such as information cubes, where the data can be interrogated from different variables perspective.
- Alerts: Developed on any of the previously cited groups by aggregating a rule-based mechanism that generates a "lead" to the user when a certain variable of interest or other measures cross a baseline value.
- Visualisation: Data into visual forms in order to enhance facts and patterns that may not be easy or feasible at all to identify in other formats.

4.2.2 Predictive analytics

Predictive analytics (PA) analyses real time and historical data to make predictions in the form of probabilities about future events. They encompass technology able to learn from data (Siegel, 2013), based on the machine learning techniques and other computational algorithms of data mining. Predictive analytics are typically algorithmic-based techniques that include (but are not limited to):

- Time series methods and advanced forecasting, vastly used in SCM for marketing measures such as predicting sales or safety stocks. Models have evolved from basic ones, e.g. Holt-Winters to ARIMA or ARMA.
- Supervised learning, which includes Regression (linear and logistic), statistical algorithms such as Discriminant Analysis, k-NN, Naïve Bayes (NB) and Bayes Networks (BN);

Decision trees, CART and Random Forests that use a hierarchical sequential structure; Kernel methods: Support Vector Machines (SVM, LS-SVM) and Neural networks/multi-layer perceptron

- Clustering, the most extended unsupervised learning technique that includes hierarchical, k-means and density based models.
- Dimensionality reduction, such as t-distributed stochastic neighbour embedding.

4.2.3 Prescriptive analytics

Prescriptive analytics use predictions based on data to inform and suggest proposed sets of actions that can serve to take advantage or to avoid on a particular outcome. They also include the study of addressing variability on the expected outcomes by what/if scenario analysis or game theory. Prescriptive analytics are mainly associated with optimisation and simulation, and have special relevance in contexts of uncertainty (i.e. where deterministic algorithms are infeasible) relying on stochastic computational programming of random variables (e.g. Monte Carlo).

4.3 Definitions of BDA in SCM

BDA is the union of two disciplines intrinsically linked: Big Data and advanced analytics. Formally there is no single definition adopted for the term Big Data, a buzzword not yet attributed to any particular author, and that even shows some fight between its claimers (Lohr, 2013) but on a review by Ward and Barker (2013), Laney (2001) proposed a magnitude data framework that explained an explosion in data based on the “3 Vs”:

- *Volume*: The volume of the Big Data datasets becomes a more relevant factor as it is beyond the capacity of traditional database management. For example, Intel considers that organisations creating approximately 300 terabytes of data weekly are in the group of Big Data volume generators.
- *Velocity*: Data is now created at higher speed than ever. According to IBM, “every day 2.5 quintillion bytes of data are created, so much that 90% of the data in the world today has been created in the last two years alone”. Velocity is also referred to as the transmission of data moving from batch processing to real time operation.
- *Variety*: Big Data can be in many different formats. Until now, structured data was the normal standard for data storage in most organisations, using relational databases managed by languages such as SQL. Now semi-structured data like XML and mostly unstructured data in any type that has not table fields could include digital information not “tagged” such as video, free form text or images.

Manyika et al. (2011) reflected their vision of Big Data as “the next frontier for innovation, competition, and productivity”. Their definition of Big Data is associated with high computer power requirements: “Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse”. The application of advanced analytics in SCM derived in the appearance of Supply Chain Analytics, a subset of technologies part of the extended supply chain and the precedent of what BDA is considered today in SCM. Early Supply Chain Analytics resembled OLAP tools that support multidimensional analysis of data from transactional databases, allowing for summarisation, consolidation and multi-perspective data view, enabling to measure, monitor, forecast and manage data on SCM business processes (Smith, 2000).

The focus on better business process has led some authors such as Grimes (2000) to identify Supply Chain Analytics as a business process reengineering enabler. Marabotti (2003) added the fact that the analytics information must be presented and extracted in a way that supported the final user. The evolution of Business Intelligence (BI) enabled wider possibilities of data integration, and Supply Chain Analytics targeted enhanced visibility across the whole supply chain (Sahay and Ranjan, 2008). Also, processing velocity made the use of data mining intelligent methods to extract more complex patterns much more accessible, as well as to update information in real time, so the patterns responded not only to past but to current business situations.

Pearson (2011b) made a shift in the definition referring to the fact that the purpose of the analysis should be "forward-looking", and also assessing the impact on "prospective" decisions. O'Dwyer and Renner (2011) synthesised this shift, already evolving to the term Advanced Supply Chain Analytics, describing a new paradigm where models have to be proactive to data instead of reactive. Waller and Fawcett (2013), reaffirmed the need for including domain knowledge in the use of analytics. Sanders (2014) offered a generic definition of BDA without specifically tailoring it for SCM. The evolution of definitions of Analytics in SCM is summarised in Table 1.

Table 1. Definitions of analytics in Supply Chain Management

Author	Definition of SCM Analytics
Smith (2000)	"Supply chain analytics is the process by which individuals, organizational units, and companies leverage supply chain information through the ability to measure, monitor, forecast and manage supply chain related business process."
Marabotti (2003)	"Supply chain analytics is the process of extracting and presenting supply chain information to provide measurement, monitoring, forecasting and management of the chain.
Sahay and Ranjan (2008)	"Supply chain analytics provides a broad view of an entire supply chain to reveal full product and component. Supply chain analytics provides a single view across supply chain and includes pre-packaged KPI, analytics."
Pearson (2011b)	"Supply Chain Analytics is […] using quantitative methods to derive forward-looking insights from data in order to gain a deeper understanding of what is happening upstream and downstream, being as a result able to assess the operational impacts of prospective supply chain decisions."
O'Dwyer and Renner (2011)	"Advanced supply chain analytics represents an operational shift away from management models built on responding to data. Advanced supply chain analytics can help supply chain professionals analyze increasingly larger sets of data using proven analytical and mathematical techniques".
Waller and Fawcett (2013)	"SCM data science is the application of quantitative and qualitative methods from a variety of disciplines in combination with SCM theory to solve relevant SCM problems and predict outcomes, taking into account data quality and availability issues."
Sanders (2014)	"Analytics is applying math and statistics to these large quantities of data. […] big data without analytics is just lots of data, Analytics without big data is simply mathematical and statistical tools and applications."

So far, the concept of Supply Chain Analytics does not appear to cover the interaction with Big Data technologies until very recently. This situation is identified as a lag between the emergence of new BDA technologies and their accepted use in SCM. BDA is the natural evolution of data analysis in SCM. The lack of previous attempts to conceptualise this phenomenon has led us to propose the following definition that converged the general concepts above and closed the research question of the systematic review.

Finding 1: SCM Big Data Analytics is the process of applying advanced analytics techniques in combination with SCM theory to datasets whose volume, velocity or variety require information technology tools from the Big Data technology stack; leveraging supply chain professionals with the ability to continually sense and respond to SCM relevant problems by providing accurate and timely business insights.

5. BIG DATA DRIVEN SUPPLY CHAIN MANAGEMENT

In this section, current trends in the generation of Big Data in SCM are analysed. Our understanding of the supply chain revolves around four main activities: *buy*, *sell*, *move* and *store*; associated with four main SCM levers: *procurement*, *marketing*, *transportation* and *warehouse* operations. The identified data sources that may be considered for decision-making purposes in each of that SCM levers are classified in the *taxonomy* according to their features in the 3 Vs framework.

5.1. SCM Big Data and the 3 Vs

A full identification of data sources used in the business cases and guidelines/methods for successful implementation obtained from the systematic review produced a list of 52 mainstream sources of Big Data across the supply chain. Each of the sources was reported in one or more of the SCM four levers, with a level of incidence from 0 (does not appear in that lever) to 4 (core for processes at that lever). In the same way, each data source was classified according to its reported volume and velocity in a 0-4 scale. Variety was described in a 3-level classification: Structured, Semi-Structured or Unstructured. Although these three subcategories are statistically dependent in the scores of a given data source, in order to facilitate analysis of some patterns of interest later discussed, they are reported separately.

Figure 1 shows the average volume and velocity versus the variety of the data sources in a model such as $E(Y \mid X) = f(X, \beta)$ with Y=0.5$(Volume+Velocity)$ and X=$Variety$.

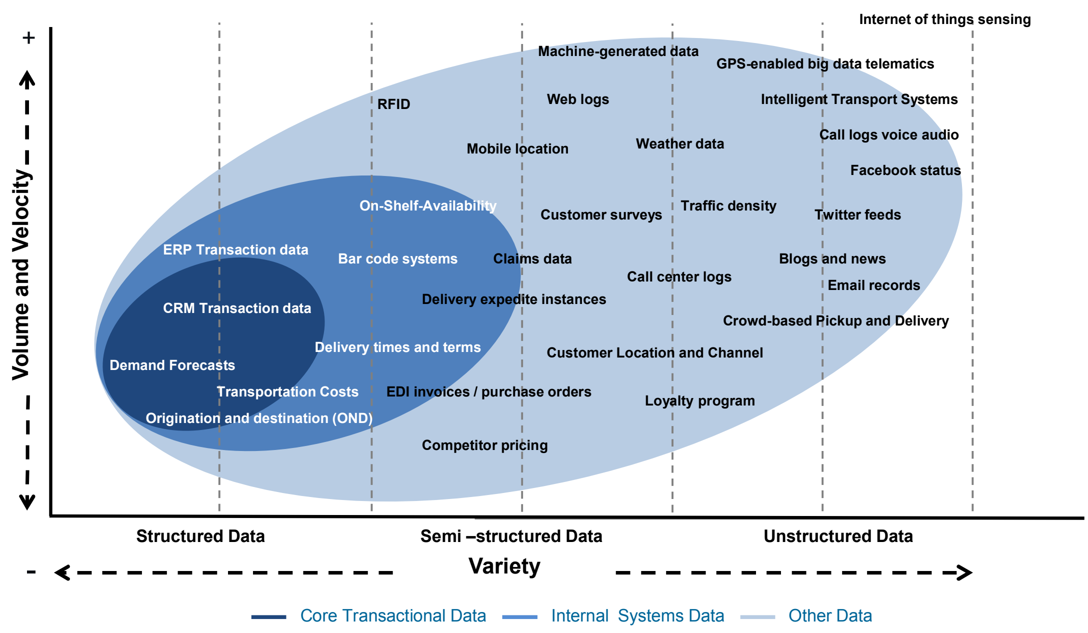

Figure 1. SCM Data Volume and Velocity vs. Variety

Each of the three shaded areas includes data sources that fall between core transactional data, internal systems data or others, respectively. The frontier of all three areas has a much wider horizon when moving along the variety of formats (horizontally) than on the other two dimensions (vertically). If the model E above is a linear regression, all parameters in vector β are strictly positive. In practical means, that fact relates to a positive correlation between larger volumes and velocity of information in unstructured formats. This proposition is supported by many practitioners and academia, and although there is no previous conclusive quantitative analysis, it is considered as rule of thumb that 80% of usable business information is unstructured (Roberts, 2010).

Validation of that trend in SCM has clear implications in the approach to data management for BDA. Although transactional data in relational databases from different systems such as ERP, CRM or SRM remain as the core of internal information and have relative high volumes, they are relatively a small fraction of the total data sources available for use (8 out of 52 in the taxonomy).

Following an observation of high concentration of points at the top right, most of the customer interface data platforms are in this high volume/unstructured region: social media, online surveys or mobile location devices. Email data is another example. Massively employed nowadays as the first communication and information tool, email is rarely used for analysis, when it certainly provides unstructured feedback about experiences with clients or suppliers (Ordenes et al., 2014).

Finding 2: SCM Big Data sources are commonly generated in unstructured formats that are difficult to analyse with traditional IT tools. Whilst data management has focused on expanding velocity and volume capabilities for transactional data, the number of core transactional data sources is relatively small. There is an asymmetry in SCM data sources between the relatively smaller variations of volume and speed versus the larger ones in data variety, and a positive correlation between the unstructured formats and high volume/velocity.

5.2 Four levers in the Big Data Driven Supply Chain

BDA can work across all SCM levers, conveying information from one area to another but the aggregation requires accuracy, timeliness, consistency and completeness (Hazen et al., 2014). For instance, marketing captures and tracks demand through Point of Sale (PoS) data, transportation creates records from GPS transponders, RFID data identifies stored goods and electronic data interchange sends automatic buying orders.

Marketing has transformed customer knowledge into an agile system that sends large amount of information flowing upstream in the chain (Jüttner et al., 2010). Intimacy with customers can be achieved through increasingly more sophisticated methods of analysing customer data, and at this lever, data sources that include social media, mobile apps, or loyalty programmes can be found; all of them are the enablers for the *sentiment analysis*. Similarly, recording omnichannel sales information can be facilitated by the electronic and cloud PoS, and by machine generated data that record transactions. Butner (2008) stated that customer inputs need to be better aligned to SCM systems, and that supply chain managers have a tendency to focus more on their suppliers than their customers, but for our interest, he also reflected that technology has made it more feasible than ever to access and understand customer data, as Big Data enables sensing of social behaviour (Shmueli et al., 2014).

Procurement deals with the relationships at the upstream supply chain. Data complexities on this side might arise from globalised purchasing strategies with thousands of transactions. In this lever, a strong connection with internal finance reporting led to adopt measures on spend visibility data, to achieve granular levels on aggregated procurement patterns. Nevertheless, according to Ainsworth (2014), data on external expenditure, which can be more than 50% of a company's cost, are "often backward looking, often inconsistently categorised and not integrated with internal costs". A subgroup of data that is still to be fully integrated and appears in the taxonomy as semi-structured are the business documents (purchase orders, shipping notices, invoices) sent through the EDI. Still et al. (2011) concluded that the procurement needs to activate the data sources not only for spending data management process, but also for the entire procurement function.

Warehouse management (particularly inventory management) has been radically changed by modern identification systems after successful introduction of RFID. Within this group, the largest clusters of data are related to an automated sensing capability, especially as the Internet of Things and extended sensors, connectivity and intelligence to material handling and packaging systems applications evolved. Position sensors for on-shelf availability share space with traditionally SKU levels and BOMs.

Transportation analysis applying Operational Research models has been widely used for location, network design or vehicle routing using origin and destination (OND), logistics network topology or transportation costs as "static" data, as described by Crainic and Laporte (1997). New alternatives to manage and coordinate in real time using operational data rely on mobile and direct sensing over shipments that are integrated into in-transit inventory, estimated lead times based on traffic conditions, weather variables, real time marginal cost for different channels, intelligent transportation systems or crowd-based delivery networks among sources of Big Data. A detailed analysis of the 3 Vs in transportation data revealed to be the lever with proportionally higher speeds in data transition.

5.3 Data integration for BDA in SCM

Figure 2 shows a Kamada-Kawai network, the distance forces between the 52 data sources and each of the four SCM levers. Those data sources that are linked to only one lever appear in the periphery of the visualisation, whereas those who are equally associated with the whole supply chain appear in the core (e.g. 34-Machine generated data, reveals association with the 4 levers, whereas 30-Invoice data reveals association only with Procurement and Transportation).

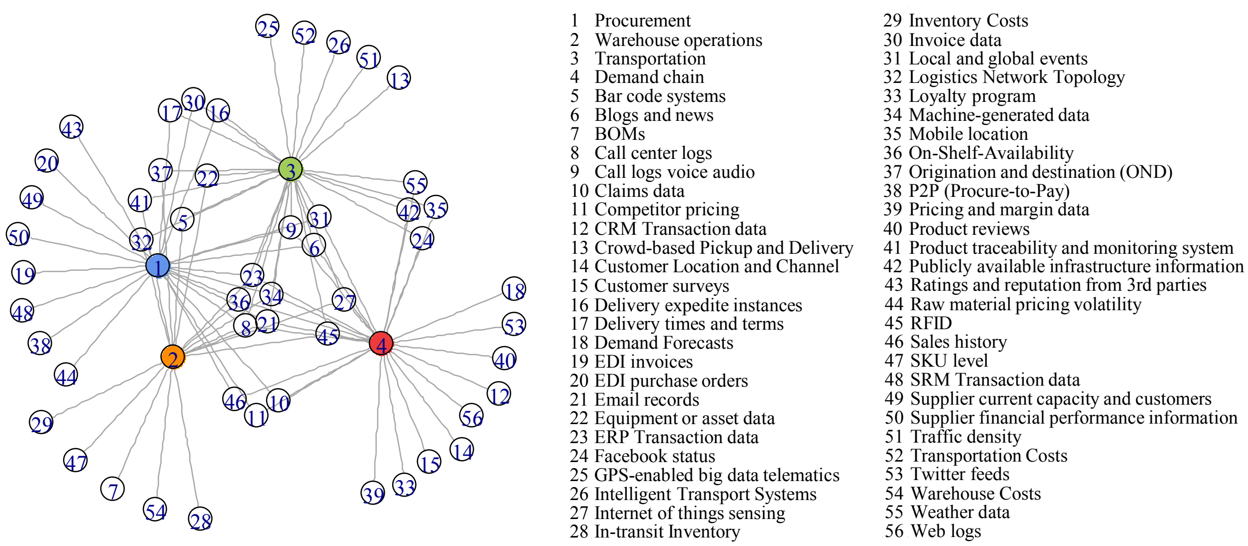

Figure 2. Kamada-Kawai Network of the identified Big Data sources across SCM

Most data sources appearing in the periphery, with a high level of symmetry, suggest large data sets with incidence only on one of the four SCM functions, or at least with a much stronger association with one of the four, rather than being utilised across the whole SCM enterprise. There are a number of data sources that can be grouped together, e.g. location information (Clusters 55, 42, 35, 24) between marketing and transportation, or data from shipment orders (Clusters 17, 30, 16) between procurement and transportation; but most sources are hosted by a single domain.

In a more favourable scenario, not necessarily all sources would need to share the same degree of incidence across the whole SCM, as certainly all kinds of data have a particular area where they are more useful, but the prevalence of many datasets anchored to only one area (52%), the fact that in most of that cases the area is the same as where the data was first generated, and the belief that among then a large fraction would add value in other areas, makes us suspect of systemic data silos.

This generates a barrier for BDA implementation due to the emphasised importance of aggregated layers of data from multiple sources in order to enhance the predictive capabilities of such models. As an example, a procurement department that has managerial incentives to apply a BDA model that monitors raw material pricing (44) in order to predict the best moment to buy at a low in the market. If they do not include in their model in-transit inventory (28) or inventory costs (29) associated with the final products using the raw materials (which is information hosted at the warehouse operations lever in our model and not at procurement), then obtaining more raw materials when there is enough final products in stock, even at low prices, could be suboptimal for the company by creating higher inventories and pushing costs downstream.

Research by Dell'Anno and Dukatz (2014) found out that leveraging many different data sources unlock value by fostering data connections and gaining actionable insights quickly. Addressing key challenges on "movement, processing and interactivity" of the data would help organisations achieving the modern data supply chain. Daugherty et al. (2014) reported that only 1 out of 5 organisations integrate their data across the enterprise.. In order to improve this situation they presented a model of data intelligent transportation throughout the organisation that could help breaking down data silos, usually built and owned by a single department, and enable data to flow freely for the benefit of the whole organisation.

Finding 3: SCM Big Data are made up of large information silos distributed among business functions and external sources, largely not interconnected, and therefore do not provide an end-to-end visibility of SCM. As a basis for BDA models generating accurate insights valuable to the organisation as a whole, and not only to single processes or sub-functions, most organisations must strive to make disparate data sources accessible by aggregating their data into a single point of access.

6. SOME APPLICATIONS OF BIG DATA ANALYTICS IN SCM

This section intends to provide some assistance to practitioners to understand where they could begin to incorporate Big Data Analytics across their supply chains, allowing them to potentially solve complex problems relevant for SCM. Table 2 briefly summarises some practical applications on how BDA can transform particular areas of SCM[1].

[1] A fuller list can be provided upon request

Table 2. Some examples of practical applications of BDA in SCM

SCM lever	Functional problem	Type of data	BDA proposed solution	BDA techniques
Marketing	Sentiment analysis of demand new trends	Blogs and news, feeds, ratings and reputation from 3rd parties, web logs, loyalty programs, call centres records, customer surveys	1. Create lexicons from training datasets that identify key terms that relate to the demand of a product. 2. Integrate all data sources that relate to a product into a unified text corpus. 3. Use supervised learning algorithms to predict sentiment scores of the corpus' term document matrix based on training datasets.	Natural language processing Text mining with R tm package: (Corpus, term-document matrix) Logistic regression, random forests, CART, Naïve Bayes, k-NN;
Procurement	Informing supplier negotiations	SRM Transaction data, Supplier current capacity & top customers, supplier financial performance information	1. Capture performance requirements for procurement contracts (SLA or other quality measures). 2. Require or publicly capture data regarding previous transactions of the supplier with other third parties in similar characteristics (delivery locations, lead times).	Suitable supervised learning algorithms, expert systems modelling
Warehouse Operations	Warranty Analytics	Internet of things sensing, user demographics, historical asset usage data	1. Aggregate multiple sensing sources on real time with reports on monitored assets together with user demographics. 2. Aggregate patterns in user and usage clusters in order to generate multidimensional segmentations.	t-distributed stochastic neighbour embedding (t-SNE)
Transportation	Real time route optimisation	Traffic density, weather conditions, transport systems constraints, intelligent transport systems, GPS-enabled Big Data telematics	1. In order to address time variability for deliveries in predefined networks, model the delivery network and update it with current position of delivery units. 2. New requirements for delivery are entered in the system. Taking into account all network availability factors, from each delivery unit a spatial regression predicts time/cost of serving a delivery to other point of the network.	Spatial regression modelling

7. CONCLUDING REMARKS

We concur with Waller and Fawcett (2013) who (more or less) argue that previous research had not yet properly closed the gap between supply chain functional knowledge, supply chain data and BDA techniques which was the reason to present this paper bottom-up, inferring the strategic benefits of BDA from the understanding of the data sources present in the supply chain, and from the application of BDA models to specific problems in SCM. Some of the practical applications proposed a disruptive shift for certain SCM activities that require a holistic change in the strategy. However, in other cases, BDA offers substantial efficiency improvements to existing processes with minor modifications, apart from the fact of understanding problems both functionally in SCM terms, and analytically in BDA terms.

We argue that in order to succeed in Big Data, we need to consider the data no longer as an information asset but as a strategic asset. By doing so, organisations in SCM could realise the economic value inherent in the data and the potential to capitalise it when combined with BDA through revenue generating activities. Some evidence presented here demonstrated that BDA is in its early stages in the supply chain, but the incoming steps will show the potential of BDA through more specific applications in SCM.

8. REFERENCES[2]

Antai, I and Olson, H., (2013). Interaction: a new focus for supply chain vs supply chain competition, *International Journal of Physical Distribution & Logistics Management* 43 (7), pp.511- 528

Barratt, M. and Oke, A., (2007). Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective, *Journal of Operations Management* 25 (6), pp.1217-1233

Chae, B., Sheu, C., Yang, C. and Olson, D., (2014). The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective, *Decision Support Systems* 59 (1), pp. 119-126

Christopher, M. (2011). *Logistics & supply chain management*, 4th Ed, FT Prentice Hall, NY

Edwards, P., Peters, M. and Sharman, G., (2001). The Effectiveness of Information Systems in Supporting the Extended Supply Chain, *Journal of Business Logistics* 22 (1), 1-27

Grimes, S. (2000). Here today, gone tomorrow, *Intelligent Enterprise* 3 (9), pp. 42-48

Laney, D., (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*, Gartner

Lustig, I., Dietrich, B., Johnson, C. and Dziekan, C., (2010). *The Analytics Journey*, Institute for Operations Research and the Management Sciences

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity, *Big Data: The Next Frontier for Innovation*, Competition & Productivity, pp. 1-143

Marabotti, D. (2003). Build supplier metrics, build better product, *Quality* 42 (2), pp. 40-43.

O'Dwyer, J. and Renner, R. (2011). The Promise of Advanced Supply Chain Analytics, *Supply Chain Management Review* 15(1), pp. 32-37

Pearson, M. (2011b). Predictive Analytics: Looking forward to better supply chain decisions, *Logistics Management* 50 (9), pp. 22

Sahay, B.S. and Ranjan, J. (2008). Real time business intelligence in supply chain analytics, *Information Management & Computer Security* 16 (1), pp. 28-48

Sanders, N. R. (2014). *Big Data Driven Supply Chain Management: A Framework for Implementing Analytics and Turning Information into Intelligence*, 1st Ed, Pearson, NJ

Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*, Wiley Publishing

Smith, M. (2000). The visible supply chain, *Intelligent Enterprise* 3 (16), pp. 44-50

Smith, G. E., Watson, K. J., Baker, W. H. and Pokorski, J. A., (2007). A critical balance: collaboration and security in the IT-enabled supply chain, *International Journal of Production Research* 45 (11), pp. 2595-2613

Waller, M. A. and Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management, *Journal of Business Logistics* 34(2), pp. 77-84

Ward and Barker (2013). School of Computer Science, University of St Andrews, *Undefined By Data: A Survey of Big Data Definitions UK*

Watson, H. J. (2013a). The Business Case for Analytics, *BizEd* 12(3), pp. 49-54

Zeng, X., Lin., D and Xu, Q. (2011). Query Performance Tuning in Supply Chain Analytics, *4th Int Conf on Computational Sciences and Optimization (CSO)*, pp. 327

[2] Due to space limitation, other reference items can be provided upon request