

An Analytic Infrastructure for Harvesting Big Data to Enhance Supply Chain Performance

Dr Yuanzhu Zhan^{a*}

Lecturer in Operations Management

^{a*} Management School,

University of Liverpool, United Kingdom

Address: Management School, Chatham Street,

Liverpool. L69 7ZH

E-mail: yuanzhu.zhan@liverpool.ac.uk

Tel: +44 (0)151 794 9853

Dr Kim Hua Tan^b

Professor of Operations and Innovation Management

^b Nottingham University Business School,

The University of Nottingham, United Kingdom

Address: Si Yuan Building, Jubilee Campus, Wollaton Road,

Nottingham. NG8 1BB

E-mail: Kim.Tan@nottingham.ac.uk

Tel: +44 (0)115 846 7749

^{a} = Corresponding Author*

An Analytic Infrastructure for Harvesting Big Data to Enhance Supply Chain Performance

Abstract

Big data has already received a tremendous amount of attention from managers in every industry, policy and decision makers in governments, and researchers in many different areas. However, the current big data analytics have conspicuous limitations, especially when dealing with information silos. In this paper, we synthesise existing researches on big data analytics and propose an integrated infrastructure for breaking down the information silos, in order to enhance supply chain performance. The analytic infrastructure effectively leverages rich big data sources (i.e. databases, social media, mobile and sensor data) and quantifies the related information using various big data analytics. The information generated can be used to identify a required competence set (which refers to a collection of skills and knowledge used for specific problem solving) and to provide roadmaps to firms and managers in generating actionable supply chain strategies, facilitating collaboration between departments, and generating fact-based operational decisions. We showcase the usefulness of the analytic infrastructure by conducting a case study in a world-leading company that produces sports equipment. The results indicate that it enabled managers: (a) to integrate information silos in big data analytics to serve as inputs for new product ideas; (b) to capture and interrelate different competence sets to provide an integrated perspective of the firm's operations capabilities; and (c) to generate a visual decision path that facilitated decision making regarding how to expand competence sets to support new product development.

Keywords: Decision support systems; big data; analytic infrastructure; competence set; deduction graph

1. INTRODUCTION

Big data is attracting considerable attention worldwide. For companies, taking advantage of valuable knowledge that can be drawn from big data is becoming the basis to compete in today's rapidly changing business environment (Tambe, 2014; Kache and Seuring, 2017; Kunc and O'Brien, 2018). Researchers indicate that many business advantages can be achieved through harvesting big data, including better customer services, higher operational efficiency, better informed strategic direction, the identification of new markets and customers, and suggestions for new services and products (Opresnik and Taisch, 2015; Swaminathan, 2017). In particular, Sanders (2014) states that big data analytics has given rise to intelligent supply chains and it can help to improve supply chain management in many ways. Hazen et al. (2016) support this argument by pointing out that without big data, a competitive level of excellence

across a global supply chain cannot be accomplished. Retailers and manufacturers capture data all along their supply chains. It includes data collected from different sources such as RFID tags, GPS locations, Member Card and Point of Sale (POS), data emitted by social media feeds and equipment sensors (Gandomi and Haider, 2015; Choi et al., 2017; Swaminathan, 2017). Analysing vast amounts of data from a wide array of sources can help firms to understand their customers better, reduce costs and better control their supply chain uncertainties (Demirkan and Delen, 2013; Sanders and Ganeshan, 2015; Vidgen et al., 2017). For example, Walmart is an early adopter of information-driven supply chains and improves its operational efficiency by leveraging big data analytics. As a result, it informs its everyday decision making from sources across its whole supply chains - from automatic purchase to inventory tracking and customer fulfilment (Waller and Fawcett, 2013; Tan and Zhan, 2017; Fisher and Raman, 2017). Moreover, the analysis of big data may have a huge operational and strategic impact on supply chain management, and therefore allow firms that adopt it to achieve competitive advantage by improving effectiveness and efficiency, enhancing decision making, increasing the value-added services and producing higher-quality products (Coussement et al., 2015; Pape, 2016; Kunc and O'Brien, 2018).

Currently, there is a variety of data analytics including business intelligence, machine learning algorithms, data mining techniques, and case-based reasoning can help firms to capture information from different data sources in order both to improve supply chain visibility and to identify customers' needs and their preferences (Hendricks and Singhal, 2014; Hindle and Vidgen, 2017; Vidgen et al., 2017). With diversified sources of data, such as social media, sensor networks, high-throughput instruments, and scientific experiments, data sets are increasing at an exponential rate (Tambe, 2014; Ferreira et al., 2015). However, the off-the-shelf techniques and technologies that we currently use to collect and analyse data do not work efficiently or satisfactorily (McAfee and Brynjolfsson, 2012; Abrahams et al., 2014; Müller et al., 2016). Moreover, most of the surge in the use of data analytics for supply chain management draws on 'information silos'. Large amounts of valuable data lurk in information silos in different divisions, specialised agencies and regional offices (Wang and Alexander, 2015; Müller et al., 2016). According to Borgman (2015), the great value of big data generates from the use of integrated data sources and the collection of actionable decisions. Although many data analytics are available to help firms to create much more information, they tend to be unfocused and hence inefficient. According to Tan et al. (2015), a lot of time and effort is required to re-organise the data captured and to determine the useful information generated. Thus, firms require a way to link and structure various streams from information silos in order to create a coherent picture that can help to deal with a specific problem.

The objective of this study is to develop an analytical infrastructure: (a) to help firms to expand their competence set by capturing values from big data; (b) to dismantle information silos and to connect

islands of information; and (c) to create a comprehensive interrelationship network in relation to a particular issue to allow better fact-based supply chain decision making on the part of managers. We showcase the analytic infrastructure through a case study of a leading international Chinese sports equipment manufacturer; the analytic infrastructure helped the company to identify the best product development processes and to determine supply chain strategies to support collaboration and decision making. The empirical results demonstrate that the proposed analytic infrastructure offers a method for quantifying and eliciting the connections essential to harvest valuable information from various data sources. Application of this analytic infrastructure offers a comprehensive perspective of department interaction and offers valuable business insights that help make critical decisions.

A previous study by Tan et al. (2015) has proposed the earliest extant knowledge on the implementation of big data and competence set analysis by suggesting a framework to enhance supply chain innovation. This research extends Tan et al.'s (2015) big data framework by recasting and augmenting the conceptual basis of the data analytics and competence set analysis. In particular, Tan et al. (2015) have not provided a systematic overview of the data analytics that is focused on dealing with different kinds of data sources. As a result, it is not evident which data analytics were used as input for the competence set analysis. Also, it is impossible to determine which data analytics can be applied to the different steps of the framework. Moreover, Tan et al. (2015) fail to unambiguously describe the processes for conducting the big data analytics infrastructure, as well as the procedure applied to determine the appropriate competence sets through different data analytics. This causes doubts about the validity of the two-step based analytic framework, as noted by Tan et al. (2015) themselves. Furthermore, the model proposed by Tan et al. (2015) only investigate the issue of two decision makers. The integrating of competence sets among multiple supply chain decision makers has not yet been studied. Therefore, as a significant advantage over the Tan et al.'s (2015) model, the analytic infrastructure developed in this study can involve multiple decision makers in the competence set network, thereby making the infrastructure more applicable in managing complex supply chain situations in real business.

This study makes the following contributions related to the most recent research in big data and supply chain management. Firstly, the developed analytic infrastructure can support firms in harvesting big data to expand their competence sets and to enhance their supply chain performance. It provides integrated support for tapping different data analytics. It captures and interrelates information from big data and provides an integrated view of a firm's capabilities, for strategy analysis and better decision making. Hence, managers are able to derive optimal strategies for collaborating and new product development in terms of expanding their competence set. Application of the analytic infrastructure not only improves supply chain performance but also enhances firms' collaboration and innovation. Secondly,

this analytic infrastructure allows managers to create a visual knowledge path from information harvested from big data. The approach allows managers to link together various information silos generated from data analytics to create a coherent picture of their operational capabilities. As a result, the network of competence sets describes strategic options, different departments' options and routines available to all decision makers. Thirdly, the analytic infrastructure, based on competence set analysis, can be used by researchers as a generic procedural guideline to the development of alternative frameworks for constructs of their interest. Finally, it usefully extends the big data framework proposed by Tan et al. (2015), and elaborates more on different data analytics and demonstrates how the big data is integrated with the competence set analysis. This research aims to stimulate future theoretical and empirical study in this area. The results reveal the promise of this integrated infrastructure, which we believe is worth further pursue by scholars interested in operations management and big data.

The rest of the study is outlined as follows. The next section introduces competence set analysis and discusses its values in supply chain management. In section 3, we review the literature on competence set models and analytics for different data sources. We then present in section 4 an integrated analytic infrastructure that considers a step-by-step information-sorting process for the harvesting of big data. In section 5, we verify and showcase the usefulness of the analytic infrastructure by conducting a company case study. The outcomes gathered from the case is discussed in section 6, and section 7 concludes the paper by outlining the contributions and limitations of the present study, and indicating directions for future research.

2. COMPETENCE SET ANALYSIS AND ITS VALUES IN SUPPLY CHAIN

The concept of competence set was first introduced by Yu and Zhang (1990; 1992). For each important decision problem, there is a competence set consisting of ideas, knowledge, information and skills for its satisfactory solution. If decision makers have acquired and mastered the relevant competence set, they can quickly make the decision confidently. Otherwise, decision makers may need to expand their existing competence set (Li and Yu, 1994). For example, why can some large firms easily attract targeted consumers in certain markets? Because these firms have acquired and thoroughly mastered the competence set required to determine who their consumers are, as well as their preferences and purchasing habits, and so can act quickly to address their needs. For some small firms, their initial competence set might be very small and incomplete and this explains why they are unable to attract as many consumers as the large firm does. The competence set can be acquired through experience, learning

and consultants. It can include information on, for instance, how to detect their potential consumers, judge their behaviours or preferences, identify new fashion trends, and respond appropriately (Chen, 2002).

In terms of competence set analysis (as shown in Figure 1), for each decision problem, E , there exists a competence set that has been acquired by decision makers (DM), denoted $Sk(E)$, and a truly needed competence set, denoted $Tr(E)$. When $Sk(E)$ covers $Tr(E)$, optimal decisions can be quickly made with confidence (Case A). Otherwise shown as Case B, if the truly needed competence set (Tr) is not completely acquired by decision makers' competence set (Sk), the decision makers cannot make the decision quickly with confidence. Therefore, it is necessary for decision makers to acquire the needed competence set (Tr) through learning, consultants or cooperation with others (Junwen, 1999). Competence set expansion indicates the process of incorporating the competence sets of others. For instance, Case C explains competence set expansion which represents a way of finding proper learning sources and sequences of acquiring needed competence sets $Tr(E)$ from existing competence sets $SK(E)$ to deal with the decision-making problem. Thus, $Tr(E)$ can be arrived at from a sequence of $Sk(E)$ to deal with the decision-making problem (Li and Zhang, 1990). In particular, this study aims to support supply chain managers to harvest big data to make better fact-based decisions based on competence set analysis. The important functions of big data are to identify the relevant information of competence set by applying different analytics, and then support decision makers to effectively obtain their needed competence set from the acquired competence set, thus allowing the decision makers to confidently make optional decisions towards a particular problem.

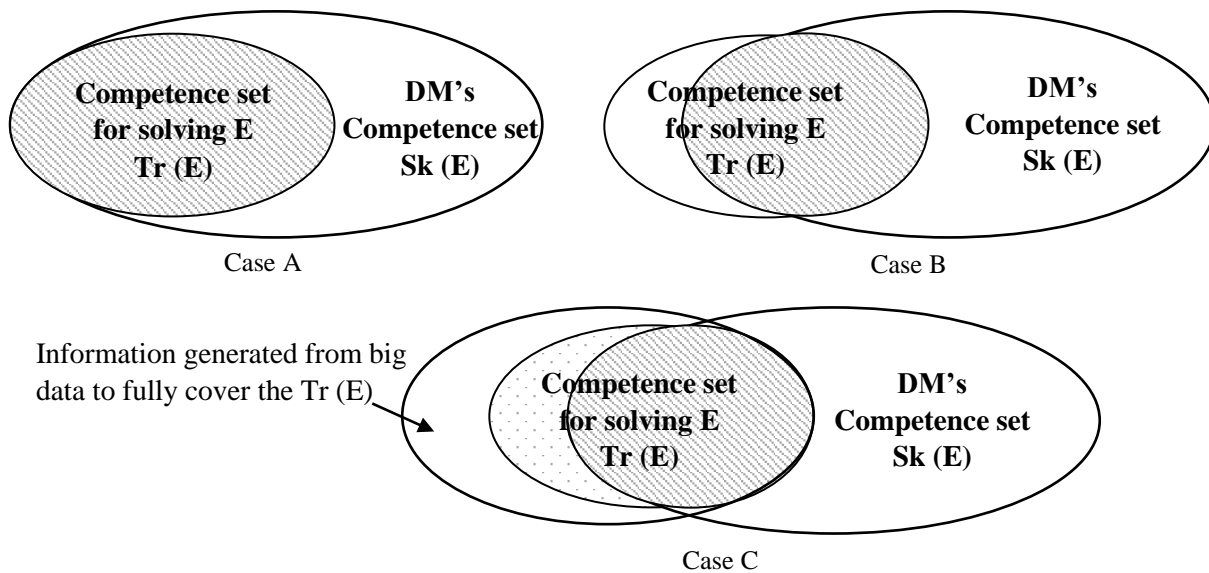


Figure 1: Competence set expansion process

Many researchers point out that competence set analysis is vital to improve firms' collaboration strategy, innovation capabilities, operational efficiency as well as supply chain performance (Schmenner and Vastag, 2006; Chen, 2002; Mishra and Shah, 2009; Li et al., 2014; Tan et al., 2015). According to decision consultants, the challenge of supply chain success rests in how to make decisions by expanding their competence set efficiently (Yu and Zhang, 1990; Kunc and Morecroft, 2010). Thus, competence set expansion plays an essential role in enhancing today's operations as well as improving supply chain performance. For example, most managers find the main challenge in operations and supply chain management is to provide products/services to generate increased revenues and achieve a high level of customer satisfaction. To overcome this challenge, it requires each participant in a supply chain – including distributors, retailers, manufacturers and suppliers – to offer the chain with their own competence sets (Greco et al., 2010). For example, the main competence set of distributors is associated with delivery, distributions and the planning of all related logistics tasks (Sander, 2014); the retailers' competence set is to manage the upstream supplier exchanges and downstream customer demands effectively (Wang and Alexander, 2015; Fisher and Raman, 2017); the competence set of manufacturers is related to offer the form utility of products by transforming components and raw materials into goods that are beneficial to end consumers (Sander, 2014; Tan et al., 2015); the suppliers' competence set is focused on providing essential inputs to the production process of supply chain manufacturers (Wang et al., 2016). Moreover, Yu and Chen (2010) states that competence set expansion plays an important role in today's changing business environment. It can integrate a wide variety of information to support managers in making better decisions. In this way, it can dismantle information silos effectively and connect these islands of information to create a comprehensive interrelationship network in the solving of a specific problem (Yu and Chen, 2002; Tan et al., 2015).

3. RESEARCH BACKGROUND

In big data analytics, the fact that information tends to reside in silos represents a common problem for many organisations (Lavalle et al., 2011; Wong, 2012; Rangarajan et al., 2018). However, when proper access controls are in place and data can be shared, companies are able to increase productivity and make stronger business decisions. Sanders (2014) points out that leading-edge companies are redesigning internal processes, such as improving collaboration and breaking down silos, in order to utilise new information capabilities to their fullest extent. In this research, we focus our attention on how to support firms in enhancing their supply chain performance by harvesting big data, and subsequently, how to dismantle information silos and provide managers with better strategies for decision making, by

implementing competence set analysis. We start with a review of the literature on different models of competence set analysis. We then turn to the big data literature and review different sources of data, analytics, and values that can be captured by analytics to increase supply chain performance.

3.1 Models of Competence Set Analysis

Over the years, researchers have proposed several models for managing competence set expansion process. Particularly, Yu and Zhang (1992) addressed the issue by introducing the ‘minimum spanning tree’ concept to optimally managing the competence set expansion; Li and Yu (1994) proposed a mathematical method for competence set expansion based on deduction graph; Shi and Yu (1996) presented an approach for competence set expansion with asymmetric acquiring costs in a cyclic causal network; Li (1999) further improved his deduction graph model by incorporating situations with multiple decision makers and cyclical situations; Li et al. (2000) introduced the concept of the stage expansion process (SEP) and applied mathematical programming methods to determine an optional ordering expansion with the minimal cost SEP; Chen (2002) provided a new ‘forest’ learning approach to expand the competence set, which was capable of addressing consumer decision problems. Compared to other competence set expansion methods, Li’s (1999) deduction graph model can treat both group decision making and cyclic causal network problems; it is, therefore, more practical than other models in generating supply chain solutions and so is more promising in this research for achieving competence set expansion and improving operations decision making processes. The deduction graph is also an efficient method of operations research (Li et al., 2000; Yu and Chen, 2010; Tan et al., 2015). It not only generates networks which illustrate the process of competence set expansion vividly but also provides more alternatives for managers dealing with a particular problem (Tan et al., 2015). Many optimisation programs can be applied to obtain optimal strategies to acquire the required competence set. By doing so, it embeds insights to drive decisions more understandable and actionable.

In implementing the deduction graph, the critical function of big data is to allow the required information to be determined regarding a specific problem (e.g., learning costs, required competence set and internal skills); the decision maker’s current competence set can then be expanded accordingly, thus allowing right decisions to be made. For example, consider a manufacturing firm consisting of sub-firms with particular competence sets. Big data can be applied to support each sub-firm to determine its best-selling products and the manufacturing capabilities required to produce these products. Big data can also provide more options to each sub-firm to acquire a new capability (e.g., new skill or equipment), either by learning it or purchasing it from other sub-firms. In this way, firms are able to produce new products

either by themselves or incorporating with others. Moreover, all sub-firms can effectively collaborate by exchanging information generated from big data to maximise their benefits.

3.2 Data Sources and Analytics

Big data will not be helpful unless we have methods to make sense out of it (Davenport and Patil, 2012; Louridas and Ebert, 2013; Zhan et al., 2016; Kunc and O'Brien, 2018). Big data analytics refers to the use of predictive analytics or other advanced methods to extract knowledge or information from data (Müller et al., 2016; Choi et al., 2017). In this study, big data refers to the massive amounts of information that can be drawn from the supply chain (a) whose source is autonomous and heterogeneous, (b) whose dimension is diverse, (c) whose format and size is beyond the capacity of conventional tools for effective management (e.g., collect, store, exploit and visualise), and (d) whose relationship is dynamic, complex and evolving. This study identifies three key aspects of extracting values from different data sources to leverage opportunities.

3.2.1 Extracting values from structured databases. Big data relies on various data capture and extraction techniques (Mortenson et al., 2015; Vidgen et al., 2017). It has its roots in the longstanding database management field (Watson and Wixom, 2007; Sanders, 2014; Zhan et al., 2017). Most of these data are structured, captured and stored in commercial relational database management through companies' various legacy systems (Mortenson et al., 2015). It can be used to analyse and visualise a variety of performance metrics, enhance business performance management, develop better predictive modelling and conduct different statistical analyses (Ferreira et al., 2015); it can also be adopted in various business applications (Matthias et al., 2017). A significant advantage for managers is that they can leverage different aspects of company data, including sales, human resources, marketing, manufacturing data, and customer service, and gain insight into different aspects of how the business is doing (Tambe, 2014; Pape, 2016). For example, operational data such as levels of inventory, company turnovers and customer purchasing activities are captured to provide managers with a window on a considerable amount of business transactions that can be used to support their day-by-day decision making.

Structured data analysis is based on data warehousing and data management (Turban et al., 2008; Singh and Singh, 2017). The commonly used analytical techniques in these systems are grounded mainly in data mining and statistical analyses (Chen et al., 2012). For example, dashboards and scorecards can be applied to visualise a variety of performance metrics in business performance management. In addition, several data mining techniques and statistical analyses can be applied for data analysis, including data manipulation and segmentation, clustering, anomaly detection, association rules, classification, and predictive modelling in world-wide business applications (Gandomi and Haider, 2015; Oztekin et al.,

2017). The majority of these analytical and data mining techniques have been incorporated into the leading BI Platforms developed by leading IT vendors, such as SQL Server Analysis and Reporting Services by Microsoft, IBM's Cognos/SPSS and Oracle's CRM-OD Business Intelligence (Louridas, 2013; Witten et al., 2016).

3.2.2 Extracting values from web and social media data. With advanced data analytics and computing power, open data - from both private and public data sources - is offering a new dimension to data analytics and making rise to data-driven innovation (Wong, 2012; Sander and Ganeshan, 2015; Matthias et al., 2017). According to McKinsey (2013), open data have the potential to generate more than \$3 trillion in additional value annually across different fields. Notably, social media is a growing source of wide-ranging information on customer experiences and preferences (Lusch et al., 2010; Moreno and Terwiesch, 2014; Chan et al., 2017). Although it is not completely open along the cost dimension, it is relatively liquid and shareable at an unprecedented level (Chae, 2015). For example, the world's most innovative Internet companies such as Yahoo, eBay, Google and Twitter continue to lead the development of social media platforms, cloud sourcing and web analytics (Grover et al., 2018). Unlike traditional database captured from various operational systems, the data that collected from the web and social media (e.g. e-communities) mostly are unstructured but often contain a great deal of customer behavioural information and opinions (Chan et al., 2015; Swaminathan, 2017). Sources can include search and user logs, news and web contents, feedback and comments, transaction records and contents. For example, the emergence of customer-generated data on various web blogs, forums, newsgroups, crowd-sourcing systems, and social media platforms offers opportunities for firms to gain a better understanding about their customers from different business constituents, including investors, social media and employees (Coussement et al., 2015).

The analysis of web and social media data requires the integration of scalable and mature techniques in text and web mining (e.g. opinion mining, question-answering, topic identification, and information extraction), spatial-temporal analysis and social network analysis with different data sources (Coussement et al., 2015; Grover et al., 2018). In particular, the data generated from web pages can be collected manually, by RSS feeds or applying more advanced methods such as application programming interfaces (APIs), HTML parsing and data streaming (Kalampokis et al., 2013). Additionally, researchers are adopting machine learning algorithms and text analytics to web and social media data that involve personal movements, conversations and activities in an attempt to improve business performance and social life (Abbasi et al., 2018; Di Minin et al., 2018). For example, Abbasi et al. (2018) proposed a language-action perspective (LAP) based text analytics framework for collecting and analysing sense-making in social media. Today, many firms make their data banks on users and usage patterns available

through their APIs in different disciplines such as weather, maps, traffic and so on. Through collecting the data, it can offer companies with improved supply chain efficiency, enhanced traffic and road conditions, and advanced environmental monitoring (Ferreira et al., 2015; Grover et al., 2018). However, while social media APIs make data publicly available, they are not open in the sense of providing full and unlimited access to the entire databases of companies such as Facebook and Twitter. Some popular database engines have developed specialised performance optimised APIs for bulk-loading data (e.g., Oracle GoldenGate and OpenRefine) rather than using standard structured query language (Chaudhuri et al., 2011; Lomborg and Bechmann, 2014; Wang et al., 2016).

While a traditional data warehouse based on relational database management systems (RDBMS) can work as a repository for the structured data, its lack of ability to handle unstructured data prevents it from performing as a unified data source for big data analytics (Zhang et al., 2014; Müller et al., 2016; Kache and Seuring, 2017). With the exponential growth of the data being generated, it is becoming more and more costly for organisations to store these data in the traditional data warehouse (Chaudhuri et al., 2011). In addition, due to humongous data being captured, the performance also suffers and requires heavily invest in the hardware configurations (Konrad et al., 2017). Therefore, the concepts such as data lake and non-relational database are emerging as popular ways to manage and develop the next generation of systems to overcoming new big data challenges (Ramakrishnan et al., 2017). These systems are typically developed to manage large and quickly arriving volumes of unstructured data (e.g., sensor and social media data) from which further insights are acquired. For instance, the data lakes can hold a massive amount of raw data in its native format, and the data stored in the lake becomes accessible as soon as it is created (Ramakrishnan et al., 2017). However, it can be complicated and time-consuming for companies to integrate data lakes with other parts of the technology architecture, develop appropriate rules, identify the capabilities required to adopt data lakes and realise significant business values from them (Hagstroem et al., 2017).

3.2.3 Extracting values from mobile and sensor data. Mobile and sensor data is the next new rich data source (Ferreira et al., 2015; Kache and Seuring, 2017). According to Chen and Zhang (2014), mobile and sensor data can generate large amounts of values in many fields, and this information can be used to make supply chain management more efficient and effective. Mobile and sensor-based content are mostly unstructured, mainly coming in data streams and the form of query processing on mobile devices and sensor-based resources (Whitaker et al., 2007; Witten et al., 2016; Grover et al., 2018). Analytics including person-centred analysis, context-relevant analysis, location-aware analysis, human-computer interaction and mobile visualisation can be conducted to offer organisations unique advantages if the data can be used to inform actionable decisions, or at least to generate insight (Chen et al., 2012; Gandomi and

Haider, 2015; Kunc and O'Brien, 2018). Chan et al. (2015) suggest that mobile sensing supports the acquisition of more in-depth and abundant data sets relating to the consumer experience, which can potentially reveal a greater diversity of patterns of consumer behaviour as well as the changing dynamics of visiting over time. For example, in order to improve supply chain efficiency, cut costs and lessen its environmental impact, UPS, the international courier company, applies sensors on its vehicles to monitor mileage, miles per gallon, speed, engine health and number of stops (Chae, 2015; Davenport and Harris, 2017). This helps the company reduce fuel consumption, harmful emissions and idling time.

More important, rather than looking at data to understand what happened in the past, firms require to consider regarding continuous flows and processes (Davenport et al., 2012; Lomborg and Bechmann, 2014; Konrad et al., 2017). In particular, streaming data is becoming increasingly important for extracting values from sensor and mobile data as it can enable firms to analyse the data during a specific event to improve their outcomes (Huda et al., 2018). Today, the increased velocity and volume of data in production settings indicates that firms will require to build continuous processes for capturing, analysing and interpreting data (Davenport et al., 2012; Larson and Chang, 2016). The insights from these practices can be integrated with manufacturing processes and applications to allow continuous processing (Wang et al., 2016). Although stocks of data located in data warehouses can continue to be useful for establishing and improving the analytical models applied on big data, they require to process continuing data flows precisely and efficiently (Davenport et al., 2017). Leading tools such as Google Analytics, App Clix, Converseon, Power BI, Birst, NodeXL network graphs and IBM Watson can be used for pre-processing, capturing, analysing and visualising sensor and mobile flows to some extent (Chen and Zhang, 2014; Gandomi and Haider, 2015; Huda et al., 2018).

Although various analytic techniques are identified to support firms in extracting values from different sources of data, there are inherent limitations to these data analytics (Chen and Zhang, 2014; Tan et al., 2015; Vidgen et al., 2017). Most of the time, the information generated resides in silos, and this can make it extremely hard for managers to gain values from big data in real applications, especially for a comprehensive decision making purpose (Sanders, 2015). For example, if managers have identified a new product, how can those managers then determine the related information that will be required to develop the product? Therefore, what company managers require is a method that serves different data sources as inputs can support the making of more informed and comprehensive strategic decisions.

4. DEVELOPMENT OF AN INTEGRATED ANALYTIC INFRASTRUCTURE

Synthesising the prior literature in competence set analysis and big data analytic techniques, we extend the knowledge of the framework proposed by Tan et al. (2015) and incorporate competence set analysis with different data analytics to yield an integrated analytic infrastructure that can link quantitative information generated from different data analytics to the underlining deduction graph, described in the supply chain content, to support a firm's operational decisions.

An analytic infrastructure is a combination of the services, systems, applications and utilities that are used for either preparing data for scoring, modelling, estimating and validating models, or related activities (Grossman, 2009). Research into very recent industry experience suggests that an integrated analytic infrastructure, which incorporates various data analytics and competence set analysis, has significant potential benefits for manufacturers of physical products, yielding surprisingly and dramatically positive results. Also, it provides structured guidance to practitioners and researchers on how to combine different data analytics with competence set analysis to support companies in breaking down information silos, improving collaboration and decision making (Oztekin, 2017). Referring to Figure 2, the proposed analytic infrastructure consists of five main stages:

- Stage One: Data capture and management
- Stage Two: Data cleaning and integration
- Stage Three: Data analytics
- Stage Four: Competence set analysis - deduction graph
- Stage Five: Information interpretation and decision making

The primary focus of this research is on Stages 3, 4 and 5. It illustrates a step-by-step information sorting process, and the application of competence set analysis to integrate different data analytics in Stage 3 and 4 can create a path for competence set expansion processes.

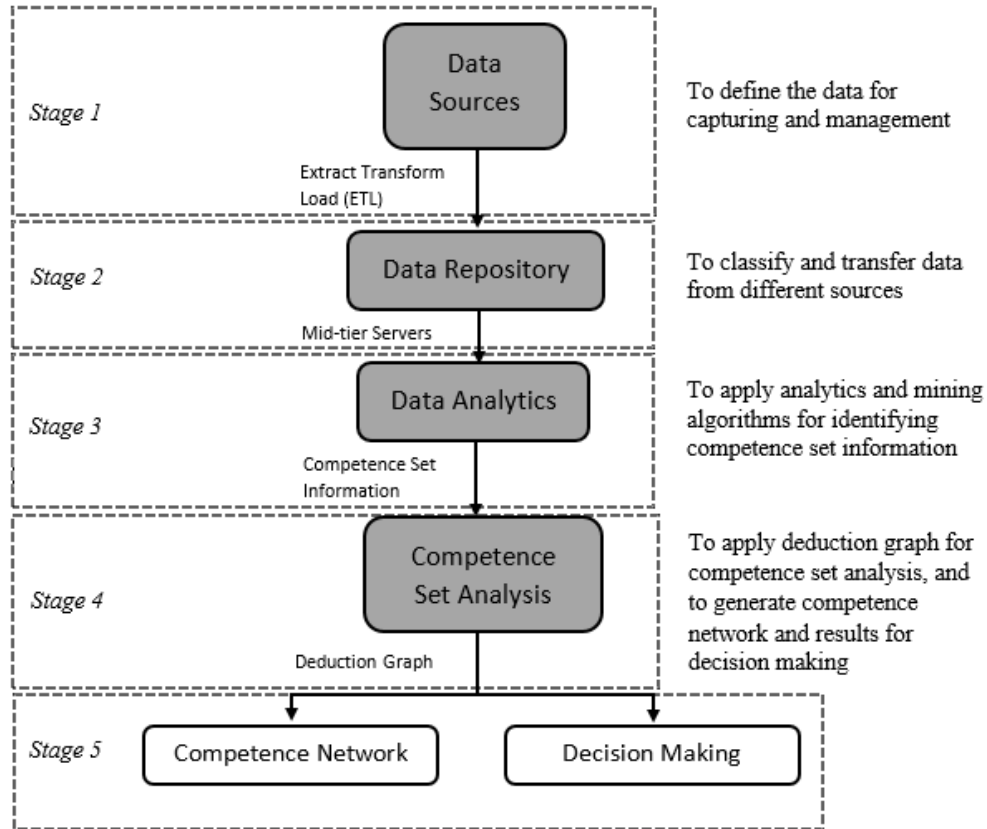


Figure 2: The integrated analytic infrastructure

4.1 Stage One: Data Capture and Management

In companies, much valuable information is captured and processed at a high cost, but most of it is ignored finally (Davenport and Harris, 2017). In this stage, it is essential for organisations to determine the data sources that will be used to create as much value as possible. It is also important to develop a particular department for expert information management; in part, this will be to meet the data storage requirements in the data capture and management, with the establishment of experimental databases, as well as storage arrays for large output files, and large-scale scientific computations (Tan et al., 2015; He et al., 2016). The exact requirements will depend on different organisations' needs and problems.

The information over which company data analytics are performed frequently originates from various data sources – typically from different operational databases across divisions within the organisation, and third parties (Davenport, 2012; Chae, 2015; Hindle and Vidgen, 2017). Moreover, big data management often require to be conducted incrementally as new information arrives such as the latest update of the operational databases (Borgman, 2015). This provides scalable and efficient data

loading and refreshes capabilities imperative for company decision making (Chen and Chiang, 2012). There are various back-end techniques for pre-processing the data for analytics and referred to as extract transform load (ETL) tools (Chen and Zhang, 2014; Louridas and Ebert, 2013). Popular ETL tools such as IBM DataStage, Oracle Warehouse Builder, Microsoft SQL Server Integration and SAS Data Management can be applied to provide great advantages in different stages of ETL process starting from data collection to data transformation and data warehousing (Vassiliadis, 2009; Witten et al., 2016).

4.2 Stage Two: Data Cleaning and Integration

Diverse information sources often contain data of varying quality, applying incompatible formats, codes, and representations which have to be reconciled (Wang and Alexander, 2015). Thus, in Stage Two a number of data pre-processing techniques, including data integration, reduction and transformation are applied to correct inconsistencies and remove noise (McAfee and Brnjolfsson, 2012; Witten et al., 2016). For example, text tagging and annotation can be used as a technique toward cleaning and integrate unstructured data from different sources like email, text files, web pages, and social media (Sukumaran and Sureka, 2016).

After that, the data will be stored in a back-end data repository that is managed by different servers for subsequent analysis (Gandomi and Haider, 2015; He et al., 2016; Bicevska and Oditis, 2017). One of the most popular databases in widespread use is relational database management systems (RDBMS) (Vassiliadis, 2009; Stieglitz et al., 2018; Wang et al., 2018). However, as the volume of digital data is growing rapidly, there is increasing demand to develop low-cost databases that can support much more extensive data volume than traditionally managed by RDBMSs (Huda et al., 2018). Therefore, engines based on non-relational (NoSQL) data warehouse are being developed such as MongoDB and Cassandra for analysing more completed data sources and query logs (e.g., web documents and online query logs) (Vassiliadis, 2009; Bicevska and Oditis, 2017). In the meanwhile, data warehouse servers are complemented by a set of mid-tier servers such as online analytic processing (OLAP) server, reporting server and enterprise search engine that can provide specialised functionality for different purposes of data analytics (Chaudhuri et al., 2011; Huda et al., 2018). For example, the star schema is the simplest style of data model used in most relational OLAP services to handle simple queries, while some other models such as snowflake schema can be used to deal with multiple levels of relationships and multiple related tables. Nonetheless, firms now are facing more and more emerged data repository options such as data lakes, cloud, enterprise data hub, hybrid storage, HTAP, and so on. The main difference between big data solutions and traditional data management solutions is that the former has a unique ability to analyse semi- or unstructured data. In particular, semi- and unstructured data cannot easily but stored in a

traditional database or fit into predefined tables (Tan et al., 2015). Most importantly, the ability to analyse semi- and unstructured data plays a vital role in the success of big data applications in businesses since 90% of the data is unstructured (IBM, 2013). In the meanwhile, the massive size and high dimensionality of data could lead to some unique challenges, such as data cost and storage, data integration, real-time analysis, data governance and organisational resistance (Ramakrishnan et al., 2017; Yichuan et al., 2018).

4.3 Stage Three: Data Analytics

Big data has a dynamic, diverse and unreliable nature, thus for the mining and analysis it needs to be integrated and accessible (Ferreira et al., 2015; Grover et al., 2018). On the one hand, statistical machine learning, often based on powerful algorithms and well-grounded mathematical models, such as ensemble models, reinforcement learning techniques and Bayesian networks, have been adopted to everyday operational logs, social media, and text mining analytics (Louridas and Ebert, 2013; Witten et al., 2016). On the other hand, some up-to-date analytic technologies leverage unique data characteristics, from data digging to spatial mining and sequential mining for mobile data, sensor data and high-speed data streams (Borgman, 2015; Chan et al., 2017). Increasingly, there is a need to support data analytics in near real-time, that is, identify the key information and make strategic decisions in an automated fashion. Sepcialised engines such as Complex Event Processing (CEP) has built to support such scenarios (Chaudhuri et al., 2011). In addition, some open source architectures have provided significant traction for data analytics such as Apache Kafka for real-time data management, Hive for data summarisation and ad hoc querying, Spark and HBase for distributed data storage, Chukwa for large volume log collection and analysis, and Mahout for data machine learning algorithms (Chaudhuri et al., 2011; Gandomi and Haider, 2015).

Due to the success achieved by different information technologies and analytics (Chae, 2015; Chen et al., 2013; Swaminathan, 2017), big data can be applied to support competence set analysis and help managers generate lots of useful information, relating to the company's needed competence set, existing competence set, internal skills, learning costs information, as well as the relevant skills for a specific issue. For example, competitive analytics and intelligence reports for each competitor can be generated on a daily, weekly, or monthly basis to provide the firms with information regarding unknown skills required for supply chain innovation or development of a new product (Panagiotidou and Tagaras, 2010). Moreover, data transparency and various mining techniques make it easier to access different competence set information from competitors (Zobel et al., 2006; Li et al., 2014; Davenport and Harris, 2017). Notably, by comparing competitors and customer comments about a specific competing product, a company could identify its needed competence set information more precisely, responsive to consumer

demand and thereby increase its sales potential (He et al., 2017). Furthermore, by analysing the operational data generated from day to day operations of the business, companies can achieve a better understanding of their patterns, trends and problems reflected in operational content (Hindle and Vidgen, 2017). Therefore, companies can make more informed identification of its current existing competence set and learning costs that can potentially improve their product performance and can help achieve better business outcomes. The outcome of this stage is information regarding different competence sets generated by data analytics.

4.4 Stage Four: Competence Set Analysis – Deduction Graph

We derive the deduction graph in Stage Four. It is an optimisation model based on cooperation with other firms, specifically in relation to competence set analysis; the model vividly illustrates how a competence set can be expanded (Yu and Zhang, 1992; Li et al., 2000). The application of deduction graph model consists of two main forms of competence set information. Firstly, existing competence set (SK), is consisting of skills, information, resources and knowledge that have been acquired by the decision maker. For example, it can refer to relevant equipment or skilled labour acquired by the firm for manufacturing a specific new product. It is followed by the second form, required competence set (Tr) which consisting of skills, information, resources and knowledge that are truly needed to solve a specific problem (e.g., the new product development may require some additional equipment and skilled labour which hasn't been acquired by the company). Therefore, the objective of deduction graph model is to obtain the optimal path for competence set expansion from the existing competence set (SK) to the required competence set (Tr) (Chen, 2002; Schmenner and Vastag, 2006). Further description of the deduction mathematic model in details, is available at Yu (1990, 1992), Li (1999) and Tan et al. (2015).

This method is particularly useful for illustrating the competence set expanding process vividly and optimising the cooperation among multiple decision makers. In comparison to other competence set analysis models, deduction graph model is a more effective method for treating both the multiple decision making problems and the cyclic causal network problem. In this stage, the information of existing competence set, needed competence set, learning costs, and the relevant skills can be acquired from Stage Three through applying data analytics (Tan et al., 2015). Therefore, by applying these competence set information as inputs to the deduction graph, a unique mathematic model can be developed. Thus, managers can dismantle information silos, visualise the expansion process, and apply tools (e.g. LINGO, MATLAB) to obtain optimal decision making regarding a specific problem.

4.5 Stage Five: Information Interpretation and Decision Making

In this final stage, following the analytics and competence set analysis part of the process comes the interpretation of the outcomes. This stage is expected to generate most of the value of the infrastructure, by offering insight, supporting decision making or information in the form of patterns. A valid interpretation requires the user to comprehend the results, incorporated by the deduction graph. It not only generates learning networks which illustrate the competence set expansion process vividly but also provides more alternative ways to managers towards a particular problem. Also, it embedded insights to drive decisions and deliver values more understandable and actionable. In particular, a competence set network will be generated that provides managers to see different options to collaborate with others and achieve their objectives. The optimisation programme could be applied to support companies to identify the optimal solution. Besides, the network visualisation offers alternative paths to achieve a set of objectives.

Besides, driven by the goal of improving programmer productivity, there have been recent practices to build analytics that can be automated to generate solutions in terms of specific business problems (Davenport et al., 2012; Ramakrishnan et al., 2017; Grover et al., 2018). So far, automating data capture, management, retrieval, analysis, and decision making can make sound judgements in specific areas (e.g., self-driving cars) (Chaudhuri et al., 2011; Witten et al., 2016; Bicevska and Oditis, 2017). However, decision automation falls short for handling most significant business situations, which tend to have idiosyncratic and unique objectives, contexts, and approaches (Lavalley et al., 2011; Ramakrishnan et al., 2017). According to Davenport and Harris (2017), many decision automation applications were merely solutions looking for problems, contributing little to improve organisations' business performance. Over time, automation tools may reduce the demand for technical-only experts, but the need for data scientists who are equipped with analytical skills for decision making and acquired deep knowledge of the business context will keep growing (Davenport et al., 2012; Witten et al., 2016). Therefore, to succeed with analytics, firms need to find ways to integrate the automated decision-making applications into their strategies and processes to achieve competitive advantage (Chaudhuri et al., 2011; Grover et al., 2018).

5. EVALUATIONS AND FINDINGS

In order to validate the analytic infrastructure, we chose to conduct an in-depth case study of a firm that has implemented a new product development project, and empirically demonstrate that the analytic

infrastructure is indeed useful for firms seeking to dismantle their information silos and to expand their competence set, both of which serve to improve decision making.

5.1 Case Introduction

A company based case study is conducted to illustrate an application of the analytic infrastructure. Company LLY is a leading international Chinese producer of athletics goods – shoes and sporting equipment. It has over 6000 employees and annual revenue of about 1.3 billion US dollars in 2016. The company has three main manufacturing facilities in China, noted here as A, B and C, respectively. The CEO championed the case study, and 12 managers took part in the testing process. These participants were three senior managers from factories A, B and C, and nine managers from different departments (information, marketing, R&D, production and finance). After implementing the analytic infrastructure for 2 months, a panel of the 12 managers was set up. The panel was asked to comment on the analytic infrastructure regarding its feasibility for dismantling information silos in big data analytics and creating a comprehensive interrelationship network to deliver better fact-based supply chain decisions. The researchers played the facilitator roles, i.e. providing clarifications and guidance when needed.

Company LLY supplies table tennis bats and badminton rackets directly into UK market. The consolidated profit from that market has increased steadily year on year. To further improve its business performance and expand its market penetration in the sports equipment market, Company LLY decided to expand its competence set to manufacture new products - golf sports equipment. By using analytics techniques, Company LLY was able to capture and analyse the relevant data. The approach, however, seemed likely to lead to the generation of an ‘island’ of information on the products, whereby management would not be able to use it to identify a comprehensive set of potential ideas or preferences for developing the new product. In short, the company was not confident that their current processes to extract understanding from various data sources were appropriate and sufficient to support them in operational decision making. Thus, the analytic infrastructure was applied to support Company LLY in making effective use of the ‘island’ of understanding extracted from different data sources as an input to the development of new products. Especially, Company LLY was keen to investigate how to extract the value from big data to achieve competence set expansion (i.e., to strengthen new product development). The overall journey through the five stages of the integrated analytic infrastructure is described below.

5.2 Stage One: Data Capture and Management

Company LLY had developed an information department to manage and analyse data captured from different sources to enhance its day-by-day supply chain performance. The main objective of the

department is to integrate the information generated from different data sources, for new product development and competence set expansion. The data is captured from five main areas: (a) *Existing customers' information and preferences*. The company collected customer information via its official website. From September 2003 to January 2016, 582,346 fully registered customers had given their name, age, gender, profession, income, address, contact details, communication records, financial income and transaction history; (b) *Product information*. The company stores specific information on more than 200 products. The information is collected and stored in their SAP systems, according to types of documents such as manuals, product descriptions and technical requirements. The company also captures the latest information on new products from its competitors via data mining, text and web analytics; (c) *Videos and photos of available products*. The company has built a video and images data set since November 2012. Up to January 2016, the data set comprised sampling frames from more than 18,000 relevant videos and 273,000 images. The data set is organised in different categories, and every member of staff can upload the video and images they find useful. The company is now focusing on analytics to extract structured insights out of these unstructured data; (d) *Social media (i.e. forums, Facebook, Tweets)*. This provides clues on existing products and potential new product ideas. The data source is chiefly user postings from the company's officially supported three e-communities for its current and new product discussion. The data set crawled as of January 2016 comprised a total of 10,746,331 discussion threads. Threads fewer than three postings or containing less than 30 words were excluded from further analysis as they were identified to contain insufficient information for describing and confirming a new or old product; (e) *Mobile and sensor data recording of the location of customers at specific time intervals*. The LLY Company develops mobile applications based on the IOS and Android systems and uses mobile sensing to record the location of its loyal customers at specific time intervals. The sensors capture more than 2.3 data points every day for each customer on average, for more than 147,000 active registered customers.

5.3 Stage Two: Data Cleaning and Integration

At this stage, different data analytics were applied, such as some open source technologies like Hadoop and Spark for extraction, indexing, storing, and analysis of data from many operational sources; cutting edge web services and programming models (e.g., XML, J2E, Flash, HTML) for crowd-sourcing, mobile visualisation and mobile social networking; opinion mining for understanding, extracting, and examining the opinions expressed in different social networking comments, web sources, sensor networks, and so on.

The LLY Company has established data centres of cloud computing specialising in managing and storing big data and applied it to support new product projects, to minimise the product development risk and achieve more efficient operations and management. These centres focus on research of big data

technology including Big Data Management and Visualisation, Distributed Computing Algorithm, Flow Calculation and Real-Time Decision Making. In particular, the company has grasped the core technology of big data such as HBase, Hadoop and Spark SQL. Meanwhile, in order to improve easy-to-use and complex query ability of its big data platform, the proposal of the cluster and multilevel index were designed, which successfully made an extension and seamless connection with the currently existing SQL-based RDBMS.

5.4 Stage Three: Data Analytics

In this stage, LLY Company identified a significant amount of information generated from various data sets regarding its new product development project. In particular, analysis of existing big data indicated that seven different types of golf clubs would have a significant impact on customer satisfaction and have great potential for long-term benefits. The golf clubs were: 460cc 10.5° driver, small cavity back 3-iron, #4 hybrids, 56° wedge, Sand wedge, Standard head Standard length Putter, and 5-wood. Also, the company identified all the competence set information required for manufacturing the seven different golf clubs; these competence sets are here denoted $a, b, c, d, e, f, g, h, i, j, k, l, m$, and n (Table 1). For example, set a is sophisticated CAD design skill; set b is X-Ray metalwood checking; set c is Billet forging technology, and so on.

Table 1: Competence set information for new golf sports equipment

| Competence Set | Manufacturing Capability | Competence Set | Manufacturing Capability |
|----------------|------------------------------------------|----------------|----------------------------------|
| a | Sophisticated CAD design skill | b | X-ray metal-wood checking |
| c | Billet forging technology | d | Stamping technology |
| e | Quality assurance process | f | Polishing and sandblasting skill |
| g | Wood manufacturing | h | Chrome plating skill |
| i | Interchangeable hosel system | j | CNC milling machine |
| k | Physical vapour deposition coating skill | l | Die casting skill |
| m | Metal analysers | n | Investment casting skill |

Specifically, the production of each golf club requires different manufacturing capabilities (i.e., competence sets). Table 2 shows the competence set information required to manufacture each of the seven specific golf clubs. For example, to produce 460cc 10.5° driver will require Billet forging technology (c), Stamping technology (d), Quality assurance process (e), Wood manufacturing (g) and Chrome plating skill (h).

Table 2: Competence set information required by golf clubs

| | <i>a</i> | <i>b</i> | <i>c</i> | <i>d</i> | <i>e</i> | <i>f</i> | <i>g</i> | <i>h</i> | <i>i</i> | <i>j</i> | <i>k</i> | <i>l</i> | <i>m</i> | <i>n</i> |
|--------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 460cc 10.5° driver | | | √ | √ | √ | | √ | √ | | | | | | |
| 3-iron | | | √ | | | √ | √ | | | | | | | |
| #4 hybrids | | | | | | √ | √ | | | | | | | |
| 56° wedge | | | | | √ | | | | | | | √ | | √ |
| Sand wedge | | | √ | | | | | | | | | | √ | √ |
| Standard head Standard length Putter | | | | √ | | | | | √ | | √ | | | |
| 5-wood | | √ | | | √ | | | | | √ | | | | |

(“√” means required)

Having identified the competence set information required for manufacturing each club, the company managers were asked to identify the existing manufacturing capabilities (i.e., competence sets) in each of their department A, B and C. The manufacturing capabilities of department A (S_A) and department B (S_B) were in fact, the same: *a*, Sophisticated CAD design skill; *b*, X-Ray metal-wood checking; and *d*, Stamping technology. The manufacturing capabilities of department C (S_C) differed somewhat: *a*, Sophisticated CAD design skill; *b*, X-Ray metal-wood checking; *c*, Billet forging technology; and *e*, Quality assurance process.

A quick analysis indicates that the departments do not have all the manufacturing capabilities required to manufacture the seven newly identified golf clubs. Therefore, the departments need to expand their own existing competence sets or acquire some competence sets among from each other. Notably, the selling price for competence sets in each department is estimated in Table 3 (i.e., the selling price for competence *c* in department C is 1.44 units, and 1.8 and 6 units for competence *d* in Departments A and B, respectively). In particular, the unit refers to the expending cost of acquiring a competence set that takes into account the resources, time, labour, energy, fund, and so on.

Table 3: The selling price for competence sets

| Department | <i>c</i> | <i>d</i> | <i>e</i> |
|------------|----------|----------|----------|
| A | | 1.8 | |
| B | | 6 | |
| C | 1.44 | | 8.4 |

Besides, the expending cost for department A of expanding its competence set (henceforth ‘expanding cost’) is illustrated in Table 4a, for department B in Table 4b, and for department C in Table 4c. Also, there are compound nodes (e.g., $a \wedge b$ and $e \wedge f$). In order to manufacture the newly identified golf

clubs, the required competence set will be acquired by learning from existing competence set or through purchasing from other departments specifically.

Table 4a: Competence expanding cost for department A

| | <i>c</i> | <i>e</i> | <i>f</i> | <i>g</i> | <i>h</i> |
|------------------------------|----------|----------|----------|----------|----------|
| <i>a</i> | 2.4 | | 3.6 | 4.8 | |
| <i>b</i> | 2.4 | | | | |
| <i>c</i> | | 3.6 | | | |
| <i>d</i> | | 4.8 | 2.16 | | |
| <i>e</i> | 1.2 | | | | 4.8 |
| <i>f</i> | | | | 2.4 | |
| <i>g</i> | | | 1.2 | | 2.4 |
| <i>h</i> | | 2.4 | | 1.2 | |
| <i>a</i> ^{<i>b</i>} | 1.7 | | 3 | 4.2 | |
| <i>e</i> ^{<i>f</i>} | | | | 1.8 | 1.8 |

Table 4b: Competence expanding cost for department B

| | <i>c</i> | <i>e</i> | <i>l</i> | <i>m</i> | <i>n</i> |
|------------------------------|----------|----------|----------|----------|----------|
| <i>a</i> | | | 2.4 | | |
| <i>b</i> | | 6 | | | |
| <i>c</i> | | | | 6 | |
| <i>d</i> | | | | | 3.6 |
| <i>m</i> | 1.2 | | | | |
| <i>a</i> ^{<i>b</i>} | | 4.8 | 1.8 | | |
| <i>a</i> ^{<i>c</i>} | | | 2.16 | 1.92 | |

Table 4c: Competence expanding cost for department C

| | <i>d</i> | <i>i</i> | <i>j</i> | <i>k</i> |
|------------------------------|----------|----------|----------|----------|
| <i>a</i> | 2.4 | 3.6 | 4.8 | |
| <i>b</i> | 2.4 | 3.6 | | |
| <i>c</i> | | | 4.8 | |
| <i>e</i> | 2.4 | | | 2.4 |
| <i>i</i> | | | | 1.8 |
| <i>k</i> | | 1.2 | | |
| <i>a</i> ^{<i>b</i>} | 2.4 | 3 | 3.6 | |
| <i>e</i> ^{<i>i</i>} | 1.8 | | | 1.2 |

From the analysis, the three departments should focus on manufacturing different golf club sets. The overall cost of acquiring the necessary combination of competence sets indicates that department A is

better for 460cc 10.5° driver, 3-iron and #4 hybrids. Department B is more suitable to produce Standard head standard length putter and 5-wood production. Department C should be responsible for 56° wedge and Sand wedge manufacturing. Table 5 illustrates the products to be manufactured in each department. We denoted 460cc 10.5° driver, 3-iron and #4 hybrids as X1, X2, X3 respectively, Standard head standard length putter and 5-wood production as Y1 and Y2, 56° wedge and Sand wedge as Z1 and Z2.

Table 5: Products to be manufactured in departments

| | |
|---------------------|--------------------------------------|
| Department A | |
| X1 | 460cc 10.5° driver |
| X2 | 3-iron |
| X3 | #4 hybrids |
| Department B | |
| Y1 | Standard head standard length putter |
| Y2 | 5-wood |
| Department C | |
| Z1 | 56° wedge |
| Z2 | Sand wedge |

The possible revenue for manufacturing different product sets is specified in Table 6 (the figures presented were estimated by the panel through their information system models). For example, if department A manufactures golf club X1, department B manufactures golf club Y1 and department C manufactures golf club Z1, the possible revenue in the following year earned by the three departments are 19.8, 20.7 and 10.8 respectively. However, it assumes that all three departments are willing to communicate and collaborate to maximise their benefits.

Table 6: Revenues for different golf club sets

| Z1 | X1 | X2 | X3 |
|-----------|------------------|------------------|------------------|
| Y1 | 19.8, 20.7, 10.8 | 14.4, 18, 14.4 | 12.6, 16.2, 9 |
| Y2 | 14.4, 14.4, 14.4 | 12.6, 16.2, 17.1 | 10.8, 16.2, 19 |
| Z2 | X1 | X2 | X3 |
| Y1 | 21.6, 10.8, 10.8 | 14.4, 13.5, 14.4 | 14.4, 12.6, 13.5 |
| Y2 | 14.4, 13.5, 16.2 | 12.6, 11.7, 18 | 11.7, 12.6, 18.9 |

Although various information was identified to support Company LLY in extracting many understandings of their customers' preferences and market potentials, the departmental and senior managers have not been able to use the various bits of analysed information to make informed decisions. Instead of generating vast amounts of information through applying different data analytics, Company

LLY requires a method to integrate information bits in data analytics and create a coherent network that can provide managers with understanding and actionable decisions.

5.5 Stage Four: Competence Set Analysis – Deduction Graph

The main focus of this stage was on providing an approach for incorporating the isolated information generated and to develop a comprehensive understanding of potential new product decisions. In stages 2 and 3 outlined in previous sub-section, the information generated remains within silos, managers are unable to use this information for making integrated decisions. This is also one of the critical contributions of this research.

Based on the collected competence set information (i.e. required competence set information (Tr), existing competence set information (Sk), internal and relevant competence set (I), as well as revenues for different product sets and learning costs), the problem encountered by Company LLY can be established as a mathematical operations research model by applying the deduction graph. Then, several optimisation programs (e.g., LINGO, MATLAB) can be applied to support managers in identifying optimal solutions according to the equation:

$$\text{Maximize } \sum_{l=1}^n \omega_l \left[\sum_{k1=1}^{q(1)} \dots \sum_{kn=1}^{q(n)} \theta_{k1}^1 \theta_{k2}^2, \dots \theta_{kn}^n \cdot \text{Rev}(\kappa_l | \kappa_1 \dots, \kappa_n) - \text{Cost}_l(S_l, T_l) \right]$$

The mathematic constraints and properties of the objective equation are illustrated and explained in Appendix 8. As a result, the outcomes can be applied to improve managerial decision making process by solving the information silos effectively and supporting managers in making better judgements.

5.6 Stage Five: Information Interpretation and Decision Making

By applying the deduction graph, a competence set network can be generated to illustrate the expansion process of competence sets to manufacture the new golf clubs. In this case, it considers a cyclical situation and involves compound nodes (i.e., compound competence set information) in the analysis. Figure 3 illustrates the competence set network for departments A, B and C to produce different products based on each department's existing competence sets.

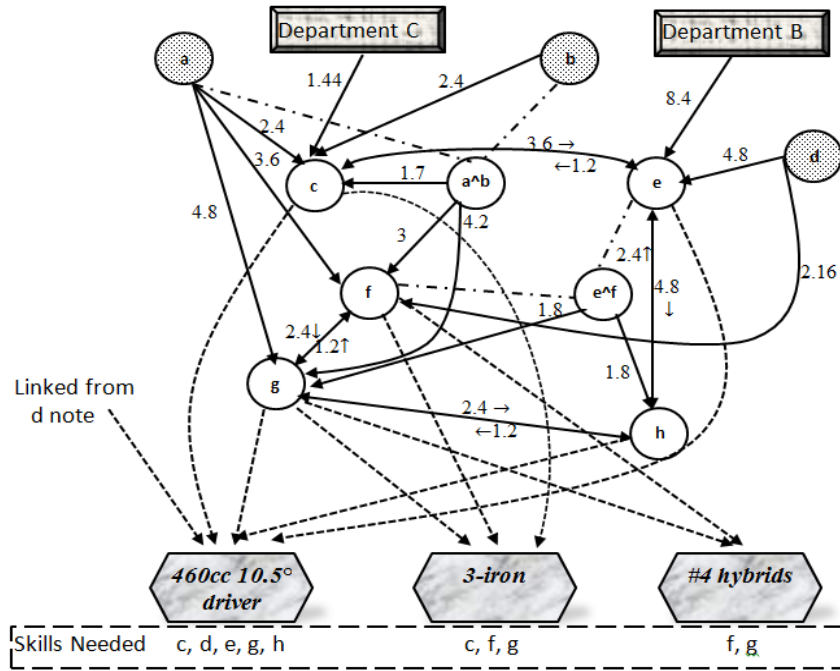


Figure 3a: Competence set network of department A (existing competence sets are *a*, *b* and *d*)

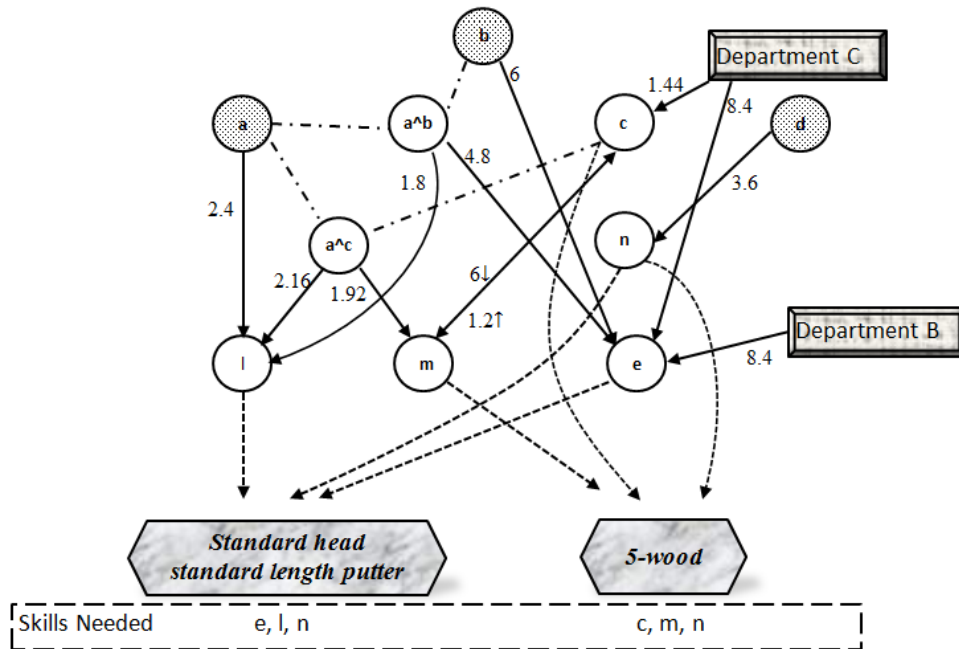


Figure 3b: Competence set network of department B (existing competence sets are *a*, *b*, and *d*)

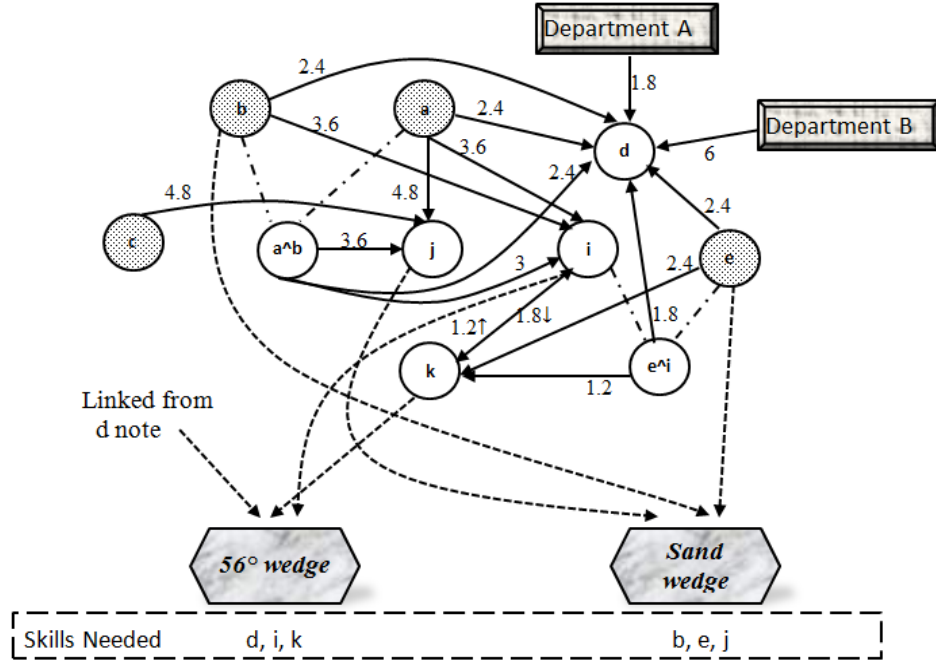


Figure 3c: Competence set network of department C (existing competence sets are a , b , c and e)

Each node indicates a competence set information. An arc means there is a relationship between two different competence sets; for instance, $a \rightarrow c$ refers to competence set c can be learned from competence set a . In terms of m and d , it shows no arc between them, indicating that it is impossible to get m from d or get d from m (Figure 3b). The figure next to the arc indicates the cost of acquiring the competence sets. Also, the compound node such as $a \wedge b$ and $a \wedge c$ can be applied only when the decomposed competence sets are acquired. In order to manufacture the new golf clubs, the required competence sets will be acquired by learning from existing competence sets or by getting from different company departments directly. For instance, in Figure 3a, competence set c can be acquired from competence set a , b and $a \wedge b$ at the cost of 2.4, 2.4 and 1.7. Also, competence set c can be acquired from department C at the cost of 1.44. Moreover, the learning sequence such as $a \rightarrow f \rightarrow g \rightarrow h$ indicates the learning process starts from a , then f , then g , then h from g . The final objective of the competence set network is to identify the optimal expansion sequence with the highest potential profit.

Additionally, by setting the objective equations and constraints in Stage Four, we can acquire the optimised solution for manufacturing the new golf clubs. According to Table 7, it shows that department A should produce #4 hybrids (X3) while department B should produce 5-wood (Y2) and department C should produce 56° wedge (Z1). As for department A, competence set f is acquired from d , and competence set g is acquired from f . For department B, competence set c is purchased from department C,

competence set n is acquired from d , competence set m is acquired from a^c . For department C, competence set d is purchased from department A, competence set k is acquired from skill e , and competence set i is acquired from skill k . The expanding deduction graph is shown in Figures 4a, 4b and 4c.

Table 7: Solutions of the new golf club sets problem

| | A | B | C |
|-------------------------|-----------------|-------------|----------------|
| Product | #4 hybrids (X3) | 5-wood (Y2) | 56° wedge (Z1) |
| Required competence set | f, g | c, m, n | d, i, k |
| Total cost | 4.56 | 6.96 | 5.4 |
| Revenue | 10.8 | 16.2 | 19 |
| Profit | 6.24 | 9.24 | 13.6 |

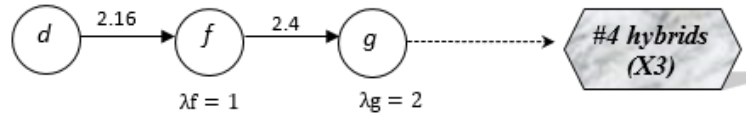


Figure 4a: Deduction graph of department A

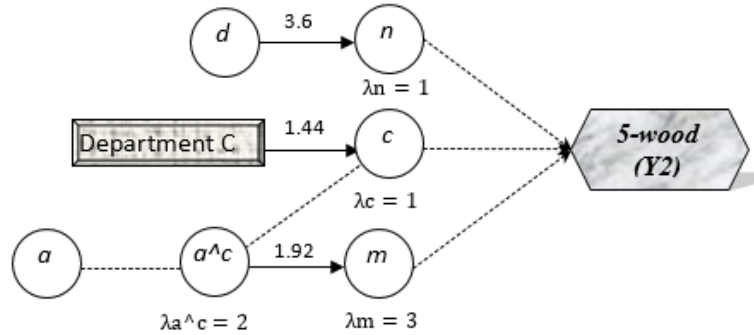


Figure 4b: Deduction graph of department B

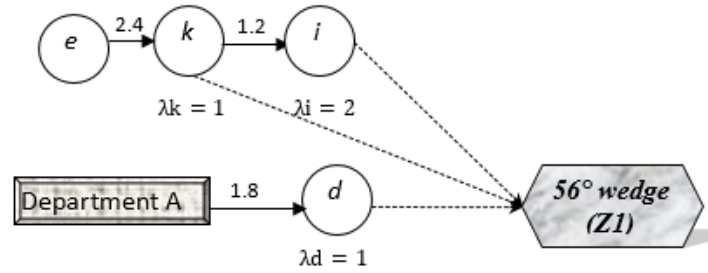


Figure 4c: Deduction graph of department C

6. DISCUSSION

The case study helps to identify the pros and cons of the proposed analytic infrastructure. A final session was conducted as part of the case study in order to obtain feedback from the company's panel (the three senior managers and nine managers from different departments). The applicability of the developed analytic infrastructure was evaluated based on the criteria of feasibility, usability, and utility (Platts, 1994; Tan and Matthews, 2009). Overall, the managers reported a reasonable level of satisfaction with regard to the quality of the results. The senior managers were confident that the derived paths of competence sets would help LLY to better understand its competence set boundary and identify strategies for new product development. They also stated that LLY would be happy to apply the analytic infrastructure again to support other big data analysis.

Moreover, the proposed analytic infrastructure shows to be a systematic method to support managers in making operational decisions, as well as the expansion of competence sets and the integration of otherwise isolated information. The network generated provides clear and easy-to-understand indications about the interrelationships of competence sets. Moreover, it can help managers to link and structure various streams of information to develop a comprehensive map of a particular issue – so that greater insight into the problem being analysed could be acquired. The information manager pointed out that *'most of the acquired data is worth next to nothing; it is only worth something when I put the data together with other data. The hybrid competence set and deduction graph approach helps us to put those data into a context that can create values, i.e. help us to identify new product ideas and understand our operations better'*. In this way, the competence set network developed can provide a clear view regarding the different expansion processes to managers, as well as suggesting the optimal expansion process for operational strategies and new product development. A manager from the R&D department further commented it as *'a roadmap that provides many alternative ways to arrive at the*

destination'. The proposed integrated analytic infrastructure and case results helped to identify the challenges (regarding either conventional techniques or the latest analytics) for future projects that aim to capture potential values from big data.

6.1 Implications for Research

This study has implications for both research and practice. In terms of implications for research, the analytic infrastructure demonstrates a fairly comprehensive analysis of competence sets that can be based on big data analytics embedded in the vast amount of data sources. The required information relating to these competence sets can then be applied as inputs for operations strategies and decisions making. When applying the analytic infrastructure, it is important to note that not all required information pointed out in our infrastructure has to be adopted. In this study, we did not fully include the information captured from current available videos and images as we found that adding such information would greatly increase the difficulty of interpreting our research outcomes (since the current video and image analytics are still limited and hard to quantify). This study has extended the earliest extant knowledge on the implementation of big data and competence set analysis proposed by Tan et al. (2015), and complemented the framework by recasting and augmenting the conceptual basis of the data analytics and competence set analysis. This results indicate that competence set analysis can be utilised to analyse large amounts of isolated information that is produced from the analytics of different big data sources, and achieve better decision makings. Moreover, this study opens avenues for implementation of data analytics and competence set analysis in the OM research field (Chan et al., 2015). While competence set analysis is not a new method (see e.g., Yu and Zhang (1992); Li (1999)) it was impossible to utilise such method easily in the past. This research proposes a solution to this issue while applying different data analytics for competence set analysis. The incorporation is novel and of particular significance today, because of the enhancement in computational capacity derived from the advanced of information technology and growing amounts of data from different sources. Furthermore, this study identified three main directions for big data research (i.e., operational database, social media, and mobile and sensor network) to explain how firms can extract values from different sources to enhance their supply chain performance. In addition, this research can be identified as the first step toward action research in the area of big data analytics and competence set analysis. The findings and future research can be incorporated into a comprehensive predictive infrastructure to offer a solution for the sorting and classification of big data.

6.2 Implications for Practice

This study also has offered implications for practice. First of all, as companies keep tapping into big data for value creation, this research provides them with a systematic approach to engage in such efforts. In

particular, the approach integrates isolated information generated from data analytics to provide companies with a coherent picture of a specific operational problem and thereby enhances their decision making. The analytic infrastructure provides companies with concrete ideas on how to link ‘islands’ of information, extracted values from different data sources, with leading analytics and tools. Since our proposed analytic infrastructure needs people with expertise in big data analytics, companies should acquire people who equip the relevant skills in data manipulation, data mining, value extraction, and advanced modelling. Secondly, the competence set network generated here can be applied in practice to support collaboration, facilitate information identification, and reduce production costs. It illustrates the interrelationship of different competence sets visually so that managers can gain a clear view of their operations problems. By generating the competence set network, managers can have more choices regarding how to expand their company’s competence set with others to support their operational strategies and decision making. Moreover, the information generated from the analytic infrastructure can be applied to get a normalised evaluation of value-and-return, which can be used as a basis for comparing competence set networks, either between organisations (e.g. own firm and partner firms) or between products (or classes of product), from the same or other partners, which may be a comparison of a new product with an earlier one. Also, this information can be adopted to distribute production resources to improvement efforts, or to enhance collaboration with partners, and can be used to identify areas for potential enhancement. For example, by means of reconstructing analytic system based on the proposed infrastructure and establishing product innovation and competence set network based on data analytic technologies, LLY Company enhanced the competitiveness of products and satisfaction of customers. Since September 2015, when LLY Company involved in golf sports equipment, it has gained great progress in concise time and achieved good results that annual sales reached a hundred million and market share reached over 6.8 per cent, thus becoming one of the largest golf sports equipment manufacturers in the Chinese market. Finally, when applying the analytic infrastructure, companies should be creative, since the analytic infrastructure is comprehensive, but not exhaustive, and is extensible. While we have shown that certain types of competence sets are the most significant in supporting operations decision making, companies should be careful to establish dynamic big data analytics for their particular situation (such as new product development and competence set expansion) as specific outcomes will vary by product, geography and time. The proposed analytic infrastructure and case study conducted here can be seen as a springboard to expedite and guide such efforts.

7. CONCLUSION

In today's competitive changing environment, managers need to harvest values from different data sources to enhance their supply chain performance. However, most such captured data – is of value only when it is put together with other data in a specific context. According to Borgman (2015), information silos vary widely across disciplines, and the biggest challenge in many cases is to dismantle information silos and to incorporate these islands of information. In this way, firms can utilise these new information capabilities to their fullest extent (Sander, 2014). Therefore, there is a need for an integrated analytic infrastructure to aid managers in improving their operations performance and dismantling information silos generated from big data analytics.

Our primary goal of this research is to investigate how, and to what extent, competence set analysis can be applied to overcome information silos in big data analytics and improve operational decision making. The research is motivated by practice and extend the framework proposed by Tan et al. (2015) to guide the better practice of big data analytics and competence set analysis in operations management. The analytic infrastructure provides a practical approach to systematically structure information generated from big data analytics to useful advance demand information for collaboration and decision making. In practice, firms should develop information management systems and use the infrastructure to take advantage of various computer science and data analytic techniques, such as predictive modelling and artificial intelligence, to better harvest the value of advance demand information from various data sources, and to enhance the accuracy and reliable of the prediction in decision making.

The research findings must be interpreted cautiously given the limitations of this study: one of the most significant assumptions is that decision makers can freely share information and is willing to sell and purchase their competence set at prescribed prices. However, in reality information sharing is not easy, incentives to do so are minimal, and data practices vary widely across disciplines. Hence, managers may not be willing to share information freely. Moreover, we only used the data available from the LLY Company to generate the information required in competence set analysis, which can lead to biases for our empirical study. Therefore, firms should understand their goals and invest in big data to collect vast amounts of information from different sources about their customers. Furthermore, the proposed analytic infrastructure may or may not generalise to other industries, such as services, general retail, and hospitality. In the meanwhile, there are a number of interesting directions for future research that build upon the findings of this study. First of all, the learning costs of competence set are essential to the analysis, as every little change could result in different outputs. Nonetheless, there are no systematic methods to define them automatically. The information on interrelationships and costs can be improved

when the attributes of competence sets and their interrelationships are considered. Hence, future research should be carried out to develop a systematic approach (e.g., via explicit programming or machine learning algorithms) to measure these costs and interrelationships precisely and automatically. Secondly, mathematical programming and algorithms could be developed to simplify the deduction graph computation and to achieve decision automation. Thirdly, further research is required to test how the analytic infrastructure will play out in different companies and industries, and to identify which associations generalise across times, industries and locations. Finally, although our analytic infrastructure can be used to generate a comprehensive decision support system to facilitate operational collaboration and decision making, all the findings herein rely on the data from a particular industrial firm with limited data captured. Therefore, we hope our study stimulates more research in this significant, practice-driven and big data analytic area.

Acknowledgements: We thank the Editors of this Special Issue, as well as the anonymous referees, for their valuable suggestions to improve the paper. Thanks also to the informants of the case company, who contributed their valuable thoughts and time to this research. This research was partially supported by grants from the University of Liverpool Early Career Researchers & Returners Fund, and UK Royal Society International Exchanges 2015 Cost Share (China), IE150652.

8. APPENDIX:

Consider a problem related to n decision makers, the strategy set for the i th decision maker is:

$$\theta_i = \{\theta_1^i, \theta_2^i, \theta_3^i \dots, \theta_{q(i)}^i\},$$

$$\text{And } \theta_k^i = 0 \text{ or } 1, \sum_{k=1}^{q(i)} \theta_k^i = 1. \quad (1)$$

The revenue can be denoted as $Rev(\kappa_i | \kappa_1 \dots, \kappa_n)$, representing the revenue that can be obtained if the DM_i chose strategy k_i ; For example, $Rev(x1 | x1, y1, z1) = 19.8$ and $Rev(x1 | x1, y1, z2) = 21.6$. The cost can be denoted as $Cost_i(S_i, T_i)$. It indicates the minimum cost for the efficient deduction graph for DM_i .

Given a competence set b in a graph, the arc relating c from b is defined by $r(b, c)$; the corresponding cost of acquiring c from b is defined by $c(b, c)$. Denote $Bf(b)$ as the set of competences immediately before b in the graph and $Af(b)$ as the set of competences immediately after in the graph.

If the strategy set k has been chosen by i th decision maker, the all relevant skills to achieve set k need to be within the deduction graph. This can be expressed as:

$$[Bf(k)] \theta_k^i \leq \sum_{j \in Bf(k)} u_j \quad (2)$$

Since the variables μ_i , $\nu(i, j)$ are binary variables, and therefore suggest λ_i be integers.

$$[Af(i)]u_i \geq \sum_{j \in Af(i)} \nu(i, j), \text{ if } i \in S, \text{ (this is for the starting node)} \quad (3)$$

$$u_i \leq \sum_{j \in Bf(i)} \nu(j, i) + \sum_{1 \in Bf(k)} \nu(DM_\nu, i), \text{ (if } i \notin S \text{ and } i \text{ is not compound node)} \quad (4)$$

The $\nu(DM_\nu, i)$ is used for purchasing skills from other DM . If the skill is learned itself, then $\nu(DM_\nu, i) = 0$. If the skill is purchased from others, then $\nu(DM_\nu, i) = 1$.

$$([Af(i)] + [Bf(i)])u_i \geq \sum_{j \in Bf(i)} \nu(j, i) + \sum_{j \in Af(i)} \nu(i, j), \text{ if } i \notin S$$

$$\lambda_i = 0, \text{ if } i \in S \quad (5)$$

$$\lambda_i - \lambda_j + m\nu(i, j) \leq m - 1, \text{ (} j \text{ is not the compound node)}$$

$$u_i \leq \lambda_i \leq mu_i \quad (6)$$

The constriction (5), (6) are for solving the cyclic graph; the m represents the single nodes number and λ_i should be integer variables.

$$Bf(i)u_i \leq \sum_{it \in Bf(k)} u_{it}, \text{ } i \text{ is compound node} \quad (7)$$

$$\lambda_i + m(Bf(i) - \sum_{it \in Bf(k)} u_{it}) \geq \lambda_{it}, \text{ } i \text{ is compound node} \quad (8)$$

In order to achieve the whole group profit maximisation, decision makers have to choose an optimised solution of strategies. This can be built as the following linear mixed 0-1 optimisation model.

$$\text{Maximize } \sum_{t=1}^n \omega_t \left[\sum_{k1=1}^{q(1)} \dots \sum_{kn=1}^{q(n)} \theta_{k1}^1 \theta_{k2}^2, \dots \theta_{kn}^n \cdot Rev(\kappa_t | \kappa_1, \dots, \kappa_n) - Cost_t(S_t, T_t) \right] \quad (9)$$

The constraints are as follows:

$$\text{If maximize } k1, k2, \dots, kn, \text{ There } \theta_{k1}, \theta_{k2}, \dots, \theta_{kn} \leq \theta_{ki}; \quad (10)$$

Subject to: (1)-(10), the weights benefit is denoted as ω_t . And it should be noted that the cost of the expanding has to be less or equal than the budget.

$$Cost_t(S_t, T_t) \leq Budget_t \text{ (} t=1, \dots, n \text{)}$$

Based on the above formulas, the case can be built as a mathematical model. A portion of this project mainly related to department A, is demonstrated below, where: $\omega_A = \omega_B = \omega_C = 1$, and $Budget_A = Budget_B = Budget_C = 7.2$.

$$\text{Maximize } 19.8\theta_{x1,y1,z1} + 21.6\theta_{x1,y1,z2} + \dots + 11.7\theta_{x3,y2,z2} - cost_A(S_A, T_A) \\ - cost_B(S_B, T_B) - cost_C(S_C, T_C)$$

$$\text{Subject to: } \theta_{x1,y1,z1} \leq \theta_{x1}, \theta_{x1,y1,z1} \leq \theta_{y1}, \theta_{x1,y1,z1} \leq \theta_{z1}$$

$$\begin{aligned}
Cost_A(S_A, T_A) = & 3.6v(a, f) + 2.4v(a, c) + 4.8v(a, g) + 2.4v(b, c) + 1.7v(a^{\wedge}b, c) + 2.16v(d, f) \\
& + 4.8v(d, e) + 2.4v(f, g) + 2.4v(g, h) + 4.8v(e, h) + 1.8v(e^{\wedge}f, g) + 1.2v(g, f) \\
& + 2.4v(h, e) + 1.7v(a^{\wedge}b, c) + 3v(a^{\wedge}b, f) + 4.2v(a^{\wedge}b, g) + 1.8v(e^{\wedge}f, g) \\
& + 1.8v(e^{\wedge}f, h) + 1.2v(h, g) + 1.44v("C", c) + 8.4v("C", e)
\end{aligned}$$

$$5\theta_{x1} \leq u_c + u_d + u_e + u_g + u_h$$

$$3\theta_{x2} \leq u_c + u_f + u_g$$

$$2\theta_{x3} \leq u_f + u_g$$

$$6u_c \geq v(a, c) + v(a^{\wedge}b, c) + v(b, c) + v(c, e) + v(e, c) + v("C", c)$$

$$6u_e \geq v(c, e) + v(d, e) + v(e, h) + v(e, c) + v(h, e) + v("C", e)$$

$$5u_f \geq v(a, f) + v(a^{\wedge}b, f) + v(d, f) + v(f, g) + v(g, f)$$

...

$$\lambda_a - \lambda_c + 5v(a, c) \leq 4$$

$$\lambda_a - \lambda_f + 5v(a, f) \leq 4$$

$$\lambda_a - \lambda_g + 5v(a, g) \leq 4$$

...

$$u_c \leq \lambda_c \leq 5u_c$$

$$u_f \leq \lambda_f \leq 5u_f$$

$$u_g \leq \lambda_g \leq 5u_g$$

...

$$2u_{e^{\wedge}f} \leq u_e + u_f$$

$$2u_{a^{\wedge}b} \leq u_a + u_b$$

$$\lambda_{a^{\wedge}b} + 5(2 - u_a - u_b) \geq \lambda_a$$

$$\lambda_{a^{\wedge}b} + 5(2 - u_a - u_b) \geq \lambda_b$$

$$\lambda_{e^{\wedge}f} + 5(2 - u_e - u_f) \geq \lambda_e$$

$$\lambda_{e^{\wedge}f} + 5(2 - u_e - u_f) \geq \lambda_f$$

$$u_a = u_b = u_d = 1, \lambda_a = \lambda_b = \lambda_d = \lambda(\text{Department } C) = \lambda(\text{Department } B) = 0$$

$$Cost_A(S_A, T_A) \leq 7.2, Cost_B(S_B, T_B) \leq 7.2, Cost_C(S_C, T_C) \leq 7.2,$$

$\theta_{x1}, \theta_{x2}, \theta_{x3}, \theta_{y1}, \theta_{y2}, \theta_{z1}, \theta_{z2}, u_i, v(i, j)$ are 0 – 1 variables, λ_i are integers, $\lambda_i \leq 5$.

9. REFERENCES:

- Abbasi, A., Zhou, Y., Deng, S. and Zhang, P., 2018. TEXT ANALYTICS TO SUPPORT SENSE-MAKING IN SOCIAL MEDIA. *MIS Quarterly*, 42(2), pp.427-464.
- Abrahams, A. S., W. Fan, G. A. Wang, Z. Zhang, J. Jiao. 2015. An integrated text analytic framework for product defect discovery. *Production and Operations Management*. 24(6): 975-990.
- Bicevska, Z. and Oditis, I., 2017. Towards NoSQL-based data warehouse solutions. *Procedia Computer Science*, 104, pp.104-111.
- Borgman, C. L. 2015. *Big data, little data, no data: Scholarship in the networked world*. The MIT Press, Cambridge.
- Chae, B. 2015. Insights from hashtag#supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *Int.J.Production Economics*. 165: 247-259.
- Chan, H. K, X. Wang, E. Lacka, M. Zhang. 2015. A mixed-method approach to extracting the value of social media data. *Production and Operations Management*. doi/10.1111/poms.12390
- Chan, H.K., Lacka, E., Yee, R.W. and Lim, M.K., 2017. The role of social media data in operations and production management. *International Journal of Production Research*, 55(17), pp.5027-5036.
- Chaudhuri, S., U. Dayal, V. Narasayya, 2011. An Overview of Business Intelligence Technology. *Communications of the ACM*. 54(8): 89-98.
- Chen, C. L. P., C. Y. Zhang. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275: 314-347.
- Chen, H., R. H. L. Chiang, V. C. Storey. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4): 1165-1188.
- Chen, T. Y. 2002. Expanding Competence Sets for the Consumer Decision Problem. *European Journal of Operational Research*. 138(3): 622-648.
- Chen, Y.J., Tomlin, B. and Wang, Y., 2013. Coproduct technologies: Product line design and process innovation. *Management Science*, 59(12), pp.2772-2789.
- Choi, T.M., Wallace, S.W. and Wang, Y., 2017. Big Data Analytics in Operations Management. *Production and Operations Management*. <https://doi.org/10.1111/poms.12838>
- Coussement, K., D. F. Benoit, M. Antioco. 2015. A Bayesian approach for incorporating expert opinions into decision support systems: A case study of online consumer-satisfaction detection. *Decision Support Systems*. 79: 24-32.
- Davenport, T. and Harris, J., 2017. *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Press.
- Davenport, T. H., D. J. Patil. 2012. Data Scientist: The Sexiest Job of The 21st Century. *Harvard Business Review*. 90(10): 70-76.

- Demirkan, H., D. Delen 2013. Leveraging the capabilities of service-oriented decision support systems: Potting analytics and big data in cloud. *Decision Support Systems*. 55(1): 412-421.
- Di Minin, E., Fink, C., Tenkanen, H. and Hiippala, T., 2018. Machine learning for tracking illegal wildlife trade on social media. *Nature ecology & evolution*, 2(3), p.406.
- Ferreira, K.J., B. H. A. Lee, D. Simchi-Levi. 2015. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1), pp.69-88.
- Fisher, M. and Raman, A., 2017. Using data and big data in retailing. *Production and Operations Management*. <https://doi.org/10.1111/poms.12846>
- Gandomi, A., M. Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*. 35: 137-144.
- Greco, S., V. Mousseau, R. Słowiński. 2010. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*. 207(3): 1455-1470.
- Grossman, R. L. 2009. *What is analytic infrastructure and why should you care?*. ACM SIGKDD Explorations Newsletter. 11(1): 5-9.
- Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., 2018. Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, 35(2), pp.388-423.
- Guezguez, W., Amor, N.B., Mellouli, K. (2009). 'Qualitative possibilistic influence diagrams based on qualitative possibilistic utilities'. *European Journal of Operational Research*. 195 (1), 223-238.
- Hagstroem, M., Roggendorf, M., Saleh, T. and Sharma, J. 2017. *A smarter way to jump into data lakes*, McKinsey Global Institute, available at: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/a-smarter-way-to-jump-into-data-lakes> (accessed July 30, 2018)
- Hazen, B.T., Skipper, J.B., Boone, C.A. and Hill, R.R., 2016. Back in business: Operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, pp.1-11.
- He, W., Tian, X., Chen, Y. and Chong, D., 2016. Actionable social media competitive analytics for understanding customer experiences. *Journal of Computer Information Systems*, 56(2), pp.145-155.
- Hendricks, K.B., V.R. Singhal. 2014. The effect of demand–supply mismatches on firm risk. *Production and Operations Management*, 23(12), pp.2137-2151.
- Hindle, G.A. and Vidgen, R., 2017. Developing a business analytics methodology: a case study in the foodbank sector. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2017.06.031>
- Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K. and Muhamad, N., 2018. Big data emerging technology: insights into innovative environment for online learning resources. *International Journal of Emerging Technologies in Learning (iJET)*, 13(1), pp.23-36.
- Kache, F., S. Seuring. 2017. Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management. *International Journal of Operations & Production Management*. 37 (1): 10-36.

- Kalampokis, E., Tambouris, E. and Tarabanis, K., 2013. Understanding the predictive power of social media. *Internet Research*, 23(5), pp.544-559.
- Konrad, R.A., Trapp, A.C., Palmbach, T.M. and Blom, J.S., 2017. Overcoming human trafficking via operations research and analytics: Opportunities for methods, models, and applications. *European Journal of Operational Research*, 259(2), pp.733-745.
- Kucukyazici, B., V. Verter, N. E. Mayo. 2011. An analytical framework for designing community-based care for chronic diseases. *Production and Operations Management*. 20(3): 474-488.
- Kunc, M.H. and Morecroft, J.D., 2010. Managerial decision making and firm performance under a resource-based paradigm. *Strategic Management Journal*, 31(11), pp.1164-1182.
- Larson, D. and Chang, V., 2016. A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), pp.700-710.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. and Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), p.21.
- Li, H. L. 1999. Incorporating competence sets of decision makers by deduction graphs. *Operations Research*. 47(2): 209-220.
- Li, H. L., P. L. Yu. 1994. Optimal Competence Set Expansion Using Deduction Graphs. *Journal of Optimization Theory and Application*. 80(1): 75-91.
- Li, J.M., Chiang, C.I. and Yu, P.L., 2000. Optimal multiple stage expansion of competence set. *European Journal of Operational Research*, 120(3), pp.511-524.
- Li, T., S. Tong, H. Zhang. 2014. Transparency of information acquisition in a supply chain. *Manufacturing & Service Operations Management*, 16(3), pp.412-424.
- Lomborg, S. and Bechmann, A. (2014) Using APIs for Data Collection on Social Media, *The Information Society*, 30:4, 256-265, DOI: [10.1080/01972243.2014.915276](https://doi.org/10.1080/01972243.2014.915276)
- Louridas, P., C. Ebert. 2013. Embedded Analytics and Statistics for Big Data. *IEEE Software*. 30(6): 33-39.
- Lusch, R. F., Y. Liu, Y. Chen. 2010. The phase transition of markets and organisations: The new intelligence and entrepreneurial frontier. *IEEE Intelligent Systems*. 25(1): 71-75.
- Manyika, J., M. Chui, P. Groves, D. Farrell, S. V. Kuiken, E. A. Doshi. 2013. *Open data: Unlocking innovation and performance with liquid information*”, McKinsey Global Institute.
- Matthias, O., Fouweather, I., Gregory, I. and Vernon, A., 2017. Making sense of Big Data—can it transform operations management?. *International Journal of Operations & Production Management*, 37(1), pp.37-55.
- McAfee, A., E. Brnjolfsson. 2012. Big Data: the management revolution. *Harvard Business Review*. 90(10): 60-68.
- Mishra, A.A., R.Shah. 2009. In union lies strength: Collaborative competence in new product development and its performance effects. *Journal of Operations Management*. 27(4): 324-338.
- Moreno, A., C. Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research*, 25(4), pp.865-886.

- Mortenson, M.J., Doherty, N.F. and Robinson, S., 2015. Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3), pp.583-595.
- Müller, O., Junglas, I., vom Brocke, J. and Debortoli, S., 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), pp.289-302.
- Opresnik, D., M. Taisch. 2015. The value of big data in servitization, *Int. J. Production Economics*. 165:174-184.
- Oztekin, A., Al-Ebbini, L., Sevkibig datai, Z. and Delen, D., 2017. A Decision Analytic Approach to Predicting Quality of Life for Lung Transplant Recipients: A Hybrid Genetic Algorithms-based Methodology. *European Journal of Operational Research*.
- Panagiotidou, S., G. Tagaras. 2010. Statistical process control and condition-based maintenance: A meaningful relationship through data sharing. *Production and Operations Management*. 19(2): 156-171.
- Pape, T., 2016. Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research*, 252(2), pp.687-698.
- Ramakrishnan, R., Sridharan, B., Douceur, J.R., Kasturi, P., Krishnamachari-Sampath, B., Krishnamoorthy, K., Li, P., Manu, M., Michaylov, S., Ramos, R. and Sharman, N., 2017, May. Azure data lake store: a hyperscale distributed file service for big data analytics. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 51-63). ACM.
- Sanders, N. R., R. Ganeshan. 2015. Big Data in Supply Chain Management. *Production Operations Management*. 24(6): 1028-1029.
- Schmenner, R. W., G. Vastag. 2006. Revisiting the theory of production competence: Extensions and cross-validations. *Journal of Operations Management*. 24(6): 893-909.
- Shi, D. S., P. L. Yu. 1996. Optimal Expansion and Design of Competence Sets with Asymmetric Acquiring Costs. *Journal of Optimization Theory and Applications*. 88(3): 643-658.
- Singh, S. and Singh, J., 2017. Management of SME's Semi Structured Data Using Semantic Technique. In *Applied Big Data Analytics in Operations Management* (pp. 133-164). IGI Global.
- Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C., 2018. Social media analytics—Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39, pp.156-168.
- Swaminathan, J.M., 2017. Big data analytics for rapid, impactful, sustained and efficient (RISE) humanitarian operations. *Production and Operations Management*. <https://doi.org/10.1111/poms.12840>
- Tambe, P. 2014. Big data investment, skills, and firm value. *Management Science*. 60(6): 1452-1469.
- Tan, K. H., Y. Zhan, G. Ji, F. Ye, C. Chang. 2015. Harvesting big data to enhance supply chain innovation capabilities: an analytic infrastructure based on deduction graph, *International Journal of Production Economics*. 165: 223-233.
- Tan, K.H. and Zhan, Y., 2017. Improving new product development using big data: a case study of an electronics company. *R&D Management*, 47(4), pp.570-582.

- Turban, E., R. Sharada, J. E. Aronson, D. King. 2008. *Business intelligence: A managerial approach*. Pearson Prentice Hall, Boston.
- Vassiliadis, P., 2009. A survey of Extract–transform–Load technology. *International Journal of Data Warehousing and Mining (IJDWM)*, 5(3), pp.1-27.
- Vidgen, R., Shaw, S. and Grant, D.B., 2017. Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), pp.626-639.
- Waller, M. A. S. E. Fawcett. 2013. Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain, *Journal of Business Logistics*. 34(4): 249-252.
- Wang, G., Gunasekaran, A., Ngai, E.W. and Papadopoulos, T., 2016. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, pp.98-110.
- Wang, L. C. A. Alexander. 2015. Big data driven supply chain management and business administration. *American Journal of Economics and Business Administration*. 7(2): 60-67.
- Wang, Y., Kung, L. and Byrd, T.A., 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.
- Watson, H. J., B. H. Wixom. 2007. The current state of business intelligence, *IEEE Computer*, 40(9): 96-99.
- Whitaker, J., S. Mithas, M. S. Krishnan. 2007. A field study of RFID deployment and return expectations. *Production and Operations Management*. 16(5): 599-612.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wong, D. 2012. *Data is the Next Frontier, Analytics the New Tool: Five trends in big data and analytics, and their implications for innovation and organisations*, London: Big Innovation Centre.
- Yu, P. L., D Zhang. 1992. Optimal Expansion of Competence Sets and Decision Support. *Information Systems and Operational Research*. 30: 68-84.
- Yu, P. L., D. Zhang. 1990. A Foundation for Competence Set Analysis. *Mathematical Social Sciences*. 20: 251-299.
- Yu, P. L., D. Zhang. 1993. A marginal analysis for competence set expansion. *Journal of Optimization Theory and Applications*. 76(1): 87-109.
- Yu, P.L. and Chen, Y.C., 2010. Dynamic MCDM, habitual domains and competence set analysis for effective decision making in changeable spaces. In *Trends in multiple criteria decision analysis* (pp. 1-35). Springer, Boston, MA.
- Zhan, Y., Tan, K.H., Ji, G., Chung, L. and Tseng, M., 2017. A big data framework for facilitating product innovation processes. *Business Process Management Journal*, 23(3), pp.518-536.
- Zhan, Y., Tan, K.H., Li, Y. and Tse, Y.K., 2016. Unlocking the power of big data in new product development. *Annals of Operations Research*, pp.1-19.
- Zobel, C. W., D. F. Cook, C. T. Ragsdale. 2006. Data-driven classification using boundary observations. *Decision Sciences*. 37(2): 247-262.