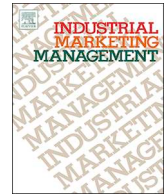




ELSEVIER

Contents lists available at ScienceDirect

Industrial Marketing Management

journal homepage: www.elsevier.com/locate/indmarman

Research paper

B2B analytics in the airline market: Harnessing the power of consumer big data

Christopher P. Holland^{a,e}, Sabrina C. Thornton^{b,*}, Peter Naudé^{c,d}^a School of Business and Economics, Loughborough University, Loughborough LE11 3TU, UK^b Huddersfield Business School, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK^c Manchester Metropolitan University Business School, Oxford Road, Manchester M15 6BH, UK^d Sydney Business School, University of Sydney, NSW, Australia^e Münster University, Leonardo Campus 11, 48149 Münster, Germany

ARTICLE INFO

Keywords:

Online market analysis
Big data
B2B analytics
Online panel data
Online performance measures
B2B networks

ABSTRACT

This paper utilizes market-level data to explore the relative performance of individual companies amongst defined competitors. We show the potential of using consumer clickstream data, an important type of big data, to create a new set of B2B analytical frameworks. In the markets where complex interactions between competitors, search intermediaries and consumers create a network, B2B relationships can be inferred from consumer search patterns, and can then be modeled to gauge the online performance. A commercial dataset from ComScore's US panel of one million users is used to illustrate a new approach to measure and evaluate the online performance of competitors in the US airline market. The methodology and associated performance framework demonstrate the potential for new forms of market intelligence based on the visualization of market networks, online performance calculated from matrix algorithms, the measurement of the impact of search intermediaries, and the identification of latent relationships. This research makes theoretical and empirical contributions to the debate on the use of big data for B2B market analytics. B2B managers can use this approach to extend their network horizon from an egocentric to a network view of competition and map out their competitive landscape from the perspective of the customer.

1. Introduction

Business-to-business (B2B) analytics is relatively undeveloped compared to business-to-consumer (B2C) analytics (Wedel & Kannan, 2016). There are some interesting challenges and opportunities to develop novel approaches to improving the situation for practitioners in B2B markets (Lilien, 2016). B2B and B2C markets often have different characteristics, and this is reflected in the development of analytics and big data strategies. However, the emergence of new business models and ecommerce businesses (Zott & Amit, 2010) raises the question of whether this binary divide between B2B and B2C markets in the marketing discipline is actually relevant in a digitally connected world.

New business models often involve networks of organizations and consumers, organized around platforms (Eisenmann, Parker, & Alstyne, 2006). For example, in many ecommerce markets, including airlines, financial services, energy, grocery and telecommunications, search intermediaries incorporate networks of business organizations that are used by consumers, which are termed 'platforms' (Eisenmann et al.,

2006). In these cases, complex interplays between competitors, search intermediaries and consumers, characterized by a business-to-business-to-consumer (B2B2C) network structure, cannot be modeled as simple supply chains (Kumar, Lahiri, & Dogan, 2017). For example, in the airline industry a simple B2B or B2C delineation is insufficient because it fails to capture the complexity of the interactions between search intermediaries, airlines and individual consumers. In this case it is necessary to map out the customer journey between airline websites and search intermediaries as a network of search paths. This means that to understand the wider competitive landscape, business organizations must look beyond their immediate customers and suppliers, and extend their network horizon (Holmen, Aune, & Pedersen, 2013). By extending the network to include consumers (or end users) and competitors, a company gains a much richer and more realistic view of market competition and is able to evaluate its own performance relative to the market.

To measure performance in a competitive context using analytics, one needs to consider the market-level view rather than a single

* Corresponding author.

E-mail addresses: c.p.holland@lboro.ac.uk (C.P. Holland), s.thornton@hud.ac.uk (S.C. Thornton), p.naude@mmu.ac.uk (P. Naudé).<https://doi.org/10.1016/j.indmarman.2019.11.002>

Received 20 June 2018; Received in revised form 7 October 2019; Accepted 4 November 2019

0019-8501/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

organization, because strategic performance is fundamentally about business performance relative to competitors (Porter, 1980). As the market activities are increasingly being transformed into network forms (Achrol & Kotler, 1999; Gummesson, 2002; Thorelli, 1986; Thornton, Henneberg, Leischnig, & Naudé, 2019), a market level view lends itself to a network approach, and this is the approach taken in this paper. In this new networked environment, focusing on an individual organization and its relationships, or limiting the scope of analysis to just business relationships is inadequate (Thornton, Henneberg, & Naudé, 2013, 2014). Instead, a market network approach is necessary to create meaningful B2B analytics.

In ecommerce markets where online consumer search is extensive, online panels can be used to map out market networks of competitors and search intermediaries based on consumer search patterns (e.g., Holland, Jacobs, & Klein, 2016). We follow this general approach and develop a novel methodology to B2B analytics that is based on consumer search patterns. In doing so, we show how online panel data, a new type of data, can be used to generate visual interpretations of the networks between firms, and develop a set of algorithms that exploit this data to measure the online performance of a firm relative to its competitors. In ecommerce markets where business competitors and search intermediaries are all vying for the search attention of consumers, consumer search patterns that connect a price comparison engine and a competitor are a proxy for a B2B relationship. Similarly, search flows between two competitors tell us about the nature and intensity of the competitive relationship between two companies based on the search activities of their customers. In these types of ecommerce markets, understanding how consumers spend their attention in terms of search flows between business organizations' websites is an important performance indicator. Consumer search patterns can therefore be used to infer both the existence of B2B relationships and also the performance of competitors and search intermediaries.

There is considerable scope for developing universal, rigorous, analytical approaches to measuring firm performance in ecommerce markets in the literature (Chen, Chiang, & Storey, 2012; Law, Qi, & Buhalis, 2010) as well as in practice (Hieronimus & Kullmann, 2013). Previous research into online performance analytics tended to use a company's own web server data, and this yields important insights into path analyses and sales conversions (Montgomery, Li, Srinivasan, & Liechty, 2004). However, it does not inform managers of online performance *relative to competitors*, which is a basic requirement of strategic analysis (Porter, 1980). With new sources of big data, in particular online panel data and search data (e.g., Callegaro et al., 2014; Göritz, Reinhold, & Batinic, 2002), it is now possible to evaluate search behavior and online performance at a *market-level* (as opposed to organization-level) to consider search behavior across a set of websites that together constitute an online market, and to use this data as the basis to evaluate performance relative to competitors (e.g., Holland et al., 2016).

This paper therefore addresses the research question of how consumer search patterns can be analyzed to evaluate the market network composed of competitors and search intermediaries. It answers the call to measure online performance in order to inform online marketing strategies (Wind, 2008) by developing a new, externally focused analytical approach that is complementary to existing web analytics. This includes identifying the scope and therefore the boundary of the network that shows the positions of the key competitors within the competitive landscape (Koka & Prescott, 2002); the ability of an organization to attract visitors from its competitors and search intermediaries; and to use this data to inform the discussion regarding the performance of an organization *relative to its competitors*. That is, to assess online performance in a market context and to provide managers with an analytical framework that gives them a market-level view of competitors and search intermediaries.

Derived from the research question, this study has three main objectives. These are (1) to show the complexity and hierarchical nature

of the analysis of big data problems, which starts from online activity that generates clickstream data and then progresses through successive stages culminating in strategic interpretation of the results; (2) to develop a methodology for identifying B2B relationships that are inferred from consumer search patterns; and (3) to define, measure and evaluate an *online performance framework* using search metrics derived from the novel interpretation of online panel data.

This study makes two contributions to the literature. First, it offers new insights into the growing research area in big data and B2B analytics (Chen et al., 2012) by providing a network visualization of a defined market of interconnected key players. The crucial point here is that B2B analytics at the level of the market using big data from all of the competitors and search intermediaries within a market is fundamentally different from the use of web analytics that analyzes a single website using a company's own web server data (Chen et al., 2012). Second, it answers the call of Law et al. (2010) that research into business performance suffers from a lack of universal applications and approaches to measuring and evaluating the online performance of business websites in a holistic manner. This study offers a diagnostic tool that allows ecommerce managers to examine their competitive position and strategic marketing planning through a holistic view of the networked market. It also offers an analytical framework through which they can assess and benchmark their own performance against the market.

In the next section the literature on big data and analytics is reviewed, and the use of online panel data as an important source of big data is discussed. Building on a synthesis of strategic concepts and new ideas from the big data literature, a new online performance framework is proposed based on a novel set of theoretical constructs, which is then applied to the US airline market. The commercial and managerial implications of the results are described, the synthesized framework is presented, and future research opportunities and limitations based on this new approach are outlined.

2. Organization-level versus market-level analysis of online performance

In a B2B2C network structure, it is clearly important to understand the interplay and dynamics between competitors, search intermediaries and consumers (Kumar et al., 2017). This view is supported by the following quote from the consultancy firm PwC (cited in Borders, Johnston, & Rigdon, 2001):

“On the Internet, consumers will interface with a value network that is made up of a coalition of trading partners linked together around a common goal – connecting to the consumer.”

The importance of competitor analysis in online marketing has also been identified by McKinsey and Google who acknowledged the relatively immature nature of online analytics at the market level (Hieronimus & Kullmann, 2013).

“...there is no standardized, widely accepted approach to measuring a company's online-marketing performance.”

There is no universally recognized and applied online performance definition in the literature. Depending on the context and the focus of the research, most studies define online performance as measures related to or leading to the eventual purchase, such as customer engagement (Kuo & Chuang, 2016), return visits (Plaza, 2011), visit times (Pakkala, Presser, & Christensen, 2012), conversation rates (Wilson, 2010) and actual purchases (Martens, Provost, Clark, & Junqué de Fortuny, 2016).

Chen et al. (2012) showed that there are two broad analytical approaches to analyze online firm performance using big data: (1) website analytics of a single organization; and (2) network analytics of a set of websites such as a market of competitors. Website analytics provides valuable marketing information about online performance for a *single*

Table 1

Two approaches to online performance evaluation.

| | Authors | Analytical context | Independent variable | Dependent variable | Analytical technique |
|------------------------------|-----------------------|---|--|--|---|
| Organization-level Analytics | Wilson (2010) | Specialized medical instruments B2B website | <ul style="list-style-type: none"> Product index Product specifications Shopping area Shopping cart Purchasing | Conversion rate | Descriptive statistics |
| | Plaza (2011) | Not-for-profit tourism website | <ul style="list-style-type: none"> Direct visits Referring site visit Search engine visit | Return visit | Regression analysis Time series analysis |
| | Pakkala et al. (2012) | Three Scandinavian food composition websites | <ul style="list-style-type: none"> Content Layout | Return visit Visit time | Descriptive statistics |
| | Kuo and Chuang (2016) | Educational gamification website | <ul style="list-style-type: none"> Bounce rate Pages per session Average session duration Returning visitors (%) | User engagement | Descriptive statistics |
| Market-level Analytics | Johnson et al. (2004) | Key sites in three product categories (books, CDs and Air travel) | <ul style="list-style-type: none"> Depth of search Dynamics of search Activity of search | Search propensity (leading to purchase) | Markov chain Monte Carlo simulation |
| | Zhang et al. (2006) | Key sites in three product categories (music, computer hardware and air travel) | <ul style="list-style-type: none"> Broadband Internet adoption Price Search depth Household income Household education Household age | Search pattern (leading to purchase) | Regression Analysis |
| | Holland et al. (2016) | Airline websites | <ul style="list-style-type: none"> Airline 'direct' search only OTA search only Airline / OTA interaction | Propensity to increase consideration set | Descriptive statistics |
| | Martens et al. (2016) | Banking websites | <ul style="list-style-type: none"> Consumer transaction behaviors | Actual purchase | Predictive analysis |

organization, e.g., to identify the propensity to search and to buy, the sources of visitors and sales, and the performance of online campaigns (Plaza, 2011). Network analytics measures online performance relative to competitors and requires market-level data such as online panel data, i.e., data from a set of competitors. Online panel data is extremely powerful because it contains very detailed, granular information about search behavior across multiple websites at scale. Consumers will typically search for a product across several related competitor websites and search intermediaries (Holland et al., 2016). This online search trajectory is captured with online panels, and the related nature of an individual's search behavior is recorded (Anderl, Becker, Wangenheim, & Schumann, 2014; Edelman, 2010). Commercial examples of online panels include ComScore, Alexa and GfK.

Following the categorization provided by Chen et al. (2012), Table 1 identifies studies that exemplify the examination of online performance at the level of the organization and at the level of the market. Note that there are many more examples of organization-level analytics, because the software is freely available and almost pervasive in commercial organizations. In contrast, the use of online panel data to evaluate online performance at the market level is far less developed. Our research aims to address this lack of universal applications in order to develop and understand the potential of market-level online analytics.

The *organization-level analytics approach* examines users' online behavior, such as purchase conversion rate (Wilson, 2010), number of return visits (Plaza, 2011), visit time (Pakkala et al., 2012), and user engagement (Kuo & Chuang, 2016). The purpose of these studies is to understand how website users' behavioral patterns inform their decision-making. The *market-level analytics approach* using online panel data at a very large scale provides different types of insights into online performance. Here, the focus shifts from a focal company to a market with a defined set of competitors and related companies, such as search intermediaries. Online panel research has investigated the size of the consideration set (Johnson et al., 2004; Zhang et al., 2006), propensity to buy (Martens et al. (2016) and interaction effects between airline and search intermediaries (Holland et al. (2016)). The market-level research demonstrates that there is significant scope and potential to use structured big data to explore online performance relative to competitors.

In summary, the evaluation of online performance in the academic literature is predominately focused on a single, focal organization. The limitation of a focal company analysis is that it cannot provide an assessment of performance against competition. There are only nascent examples of online competitor analysis by leading consultancy firms using surrogate measures of online sales, such as search behavior, to evaluate a set of competitors, e.g., McKinsey's digital marketing practice (Dörner, Galante, & Kauter, 2013).

3. Big data: Definition, characteristics and utilization

Big data has been defined by its key characteristics of Volume, Velocity and Variety (Goes, 2014), to which Veracity and Value have also been added (George, Haas, & Pentland, 2014; Goes, 2014; Opresnik & Taisch, 2015). The concept of 'big data' signifies the newly available streams of large volumes of data that are generated by activities including commercial transactions, the Internet of Things, and the online activities of people tracked by their mobile devices, tablets and computers (Chang, Kauffman, & Kwon, 2014; Gandomi & Haider, 2015; Goes, 2014). Growth in the volume of data has been accompanied by a rapid rise in computer processing power (Kambatla, Kollias, Kumar, & Grama, 2014) that makes it technically possible to sift, structure, analyze and model very large data sets in a timely and cost effective manner in order to improve firm performance (Järvinen & Karjaluo, 2015; Wamba et al., 2017; Wang, Kung, Wang, & Cegielski, 2017). The combination of cheap or freely available computing power, new sources of big data (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2014) and innovations in analytical methods (Ketter, Peters, Collins, & Gupta, 2015) has led to transformational capabilities in modeling problems and perhaps even in the method by which social science is developed in what has been termed 'computational social science' (Chang et al., 2014).

3.1. Online clickstream panel data as a type of big data

The collection of clickstream data directly from a panel of online users creates exciting new research opportunities (Bucklin et al., 2002;

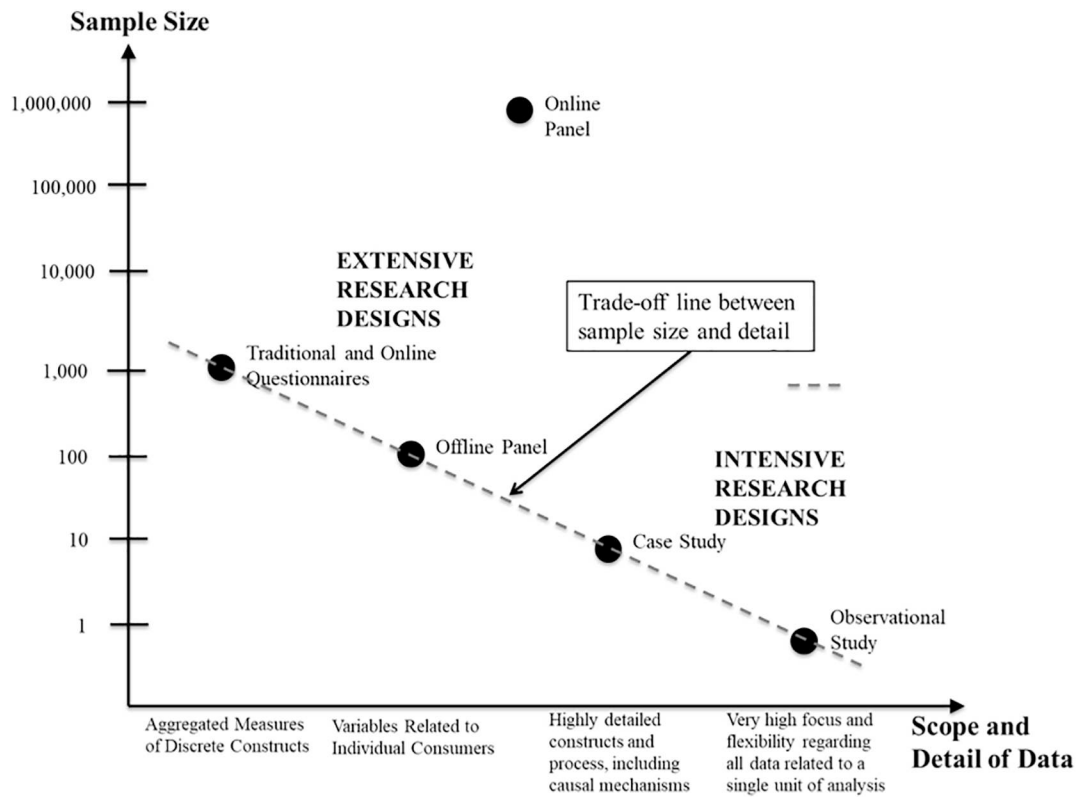


Fig. 1. Research designs plotted on a sample size – scope & detail of data map.

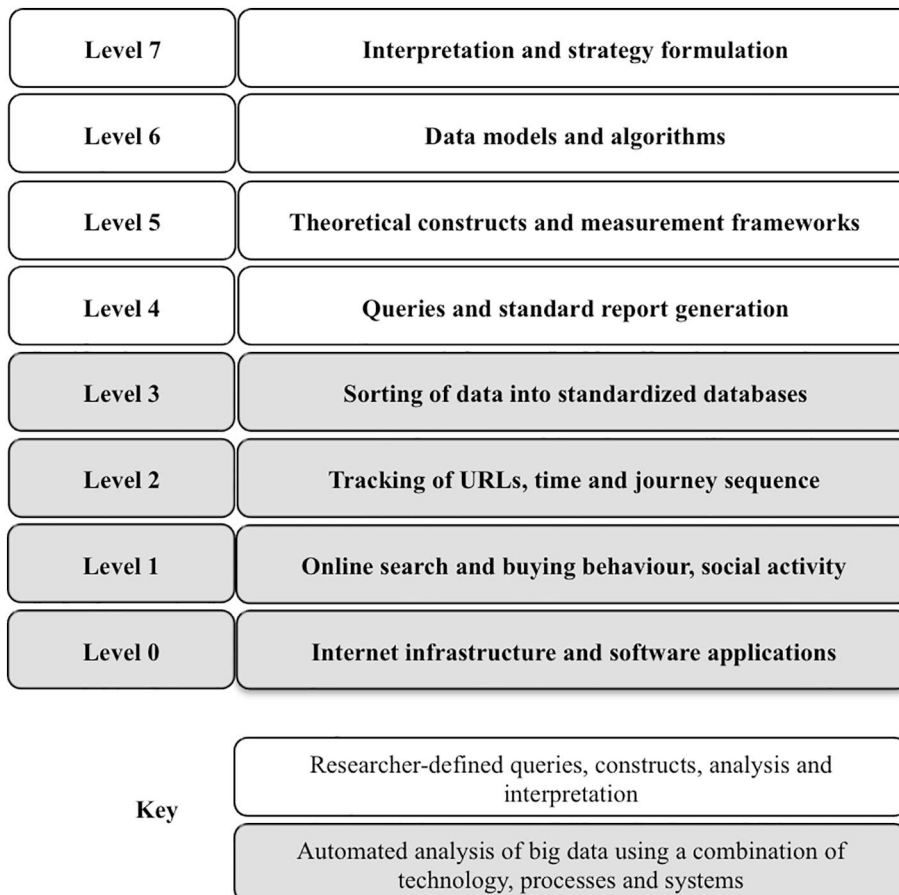


Fig. 2. A hierarchical model of big data analysis and interpretation.

Bucklin & Sismeiro, 2009). It is an important type of big data (Gandomi & Haider, 2015) that fulfills the criteria of high volume, velocity and variety. Importantly, it also has a high level of veracity, arising from the careful panel composition and technical innovations involved in managing the inherently complex set of unstructured clickstream data, and then transforming this raw data into meaningful reports that make sense and can be interpreted by managers to measure and evaluate marketing activities such as advertising effectiveness (Wixom, Ross, Beath, & Miller, 2013) and so to create value (George et al., 2014).

In addition to the 5Vs, the concept of *granularity* is important in understanding the nature of big data as distinct from other types of data such as that generated from survey questionnaires or self-reported sources from offline panels (George et al., 2014). Granularity is concerned with the level of detail captured within the data at the level of the individual. To illustrate the differences between online panels and traditional data sources, the trade-off line between detail and sample size is shown in Fig. 1. Online panel data breaks the trade-off line because the data capture process is automated and uses advanced software on consumer devices to measure online behavior in a direct, accurate and comprehensive manner.

Fig. 1 is an illustrative diagram that shows how online panels are different to other research data collection methods in terms of the relationship between sample size and detail of the data that is collected. The sample size as an order of magnitude is mapped on the vertical axis and the scope and level of detail in the data on the horizontal axis. The general pattern with different research designs is that as the level of detail increases, the sample size decreases, and this is the trade-off that researchers make between sample size that may give more confidence about generalizability and empirical detail, which potentially generates richer insights. By inspection, online panels are two or three orders of magnitude larger than a questionnaire-based study, and crucially the data contains granular information, e.g. search trajectory between websites, which is crucial for determining analytics data based on structural properties of the network (George et al., 2014).

3.2. A hierarchical model of online panel data

A hierarchical model of online panel data as a type of big data is shown in Fig. 2, which depicts the transformation of big data from its inception through to strategy formulation.

The diagram conveys the importance of the transformation of raw, unstructured clickstream data that is captured from members of online panels (Levels 0-2) into structured and standardized databases (Level 3). Levels 0-3 are the domains of the company that manages the online panel and this requires significant computing power and technical expertise. As researchers we took the standardized data and extracted a specific set of market-focused data on the US airline market (Level 4). The new B2B analytical framework and theoretical constructs (Levels 5 and 6) are described in the methodology section, and level 7 is the interpretation and evaluation of the empirical results.

The hierarchical framework allows us to position previous research into market-level analytics. For example, the advanced McKinsey and Google project that utilized an online marketing excellence model (Dörner et al., 2013; Hieronimus & Kullmann, 2013) through the usage of Google's software and processes to manage levels 1-3 and to present a standard reporting interface (level 4). However the data itself was mainly descriptive (level 5) and expressed as a set of key performance indicators, i.e., no examples of more sophisticated predictive and explanatory modeling was attempted, which is characteristic of level 6 and a pre-requisite of achieving high-level insights and interpretation of the data to inform strategic decision-making (Agarwal & Dhar, 2014; Saboo, Kumar, & Park, 2016). The focus of a user-generated big data study (Tirunillai & Tellis, 2014) was on interpreting level 2 data, which in that case was the complex opinions, moods and evaluations of a large sample social media content to create simple meaning and measurements of brand performance (level 5). In a very advanced study of a

significantly sized dataset, for example in Martens et al. (2016), predictive modeling of purchasing (levels 5 and 6) was improved by exploiting big data that was analyzed at the level of individual consumers. However, this particular study did not explore the strategic implications of the results.

3.3. Interpretation of big data

With commercial online panel data, the marketing challenges are managerial and interpretive, i.e., how to make sense of standardized online reports to generate new insights and meaning in terms of online performance based on large-scale patterns of consumer search behavior. A recent article has argued that there are four key areas in which big data analytics has a role to play in marketing: customer relationship management, optimization of the marketing mix, personalization of the marketing mix, and privacy and security (Wedel & Kannan, 2016). In this paper we argue that there is a fifth approach that can be taken, namely to look at online performance of competitors relative to each other. Law et al. (2010) argue that research in the field of online analytics suffers from a lack of universal applications and approaches to measuring and evaluating the online performance. New applications are therefore needed to assess more precisely consumers' browsing and purchasing behaviors on a continuous basis (Wind, 2008).

4. Research methodology and online panel data


The US airline market was chosen because of its economic size and the very high use of online search to find and book airline tickets. The US is the largest airline market worldwide (Pearce, 2014); in 2017 there were over 742 million domestic passengers and 223 million international passengers (US Department of Transportation, 2017). In 2016, 51 million US consumers visited airline websites (ComScore, 2016).

Online panels work in fundamentally the same manner as pre-Internet market research panels, e.g., Goodhardt, Ehrenberg, and Chatfield (1984), and Goodhardt and Ehrenberg (1967). Individual members are recruited to the panel and their Internet behavior is tracked through the automatic capture of their clickstream data (Bucklin et al., 2002; Bucklin & Sismeiro, 2009). The key differences between online and offline panels are the size of the panels, the accuracy and granularity of the data, and the ability to track the sequence of online behavior, which gives related information about separate websites (Göritz et al., 2002). Online panels have been used in consumer marketing (Bucklin et al., 2002; Bucklin & Sismeiro, 2009), economics (Baltagi, 2013) and psychology research (Göritz, 2007).

By using this data source, our study identifies and develops three new applications. The first is to visualize the market network of airlines and Online Travel Agents (OTAs) across a network of inter-connected websites. This exploits the 'relatedness' aspect of panel data, which is an important granular feature of big data (Göritz et al., 2002). This application yields important insights into the scale and structure of the market network that extends the network horizon of managers beyond their immediate economic partners. The second is concerned with the understanding of the performance of an airline's website *in relation to other airlines* and in particular to measure the performance of individual airlines in terms of their ability to attract visitors from competitor airline companies. The third market-level performance metric is concerned with the relationships with OTAs. Most airlines utilize OTAs to gain access to a wider reach of customers (Koo, Mantin, & O'Connor, 2011). The OTAs are intermediaries and have gained popularity in the marketplace due to their value proposition as an information provider, which offers a platform for consumers to compare the attributes of multiple flights, hotels and other travel items (Bhargava & Choudhary, 2004; Dubé & Renaghan, 2000).

Table 2

Excerpt from source/loss report for American Airlines.



United States
Home and Work
March 2014
American Airlines

5/9/2014

©2014

| | Unique Visitors (000) | American Airlines visitor flows with airlines and OTAs | |
|----------------------------|-----------------------|--|------------|
| | | Source (000) | Loss (000) |
| American Airlines | 4,676 | | |
| Expedia Inc | | 146 | 267 |
| US Airways Group, Inc. | | 231 | 209 |
| Kayak.com Network | | 195 | 200 |
| United Airlines | | 131 | 157 |
| Delta Airlines | | 105 | 134 |
| Priceline.com Incorporated | | 138 | 118 |
| Orbitz Worldwide | | 27 | 118 |
| TripAdvisor Inc. | | 70 | 92 |
| Travelocity | | 44 | 76 |
| Fareportal Media Group | | 75 | 67 |
| Southwest Airlines Co. | | 136 | 61 |

Source: ComScore, covering the month of March 2014.

4.1. Online panel data source

Online panel data from ComScore was used in this research (www.comscore.com). ComScore was a pioneer in online panels and its worldwide panel has approximately two million members, with one million in the United States. It is a leading provider of digital intelligence to advertisers. It collected approximately 14 petabytes of online data in 2013, measuring a wide variety of online consumer behavior including websites visited, time spent per website, search terms used, and customer search journeys and buying behavior (Ferguson, 2014; Wixom et al., 2013). In this paper, data from the US airline market is used to illustrate online market-level performance analytics. The clickstream data is organized into a database from which a variety of standard reports can be generated. As an example, an excerpt from a source/loss report is shown in Table 2, with all data being based on a one-month American sample in March 2014. Referring back to Fig. 2, in our study, the shaded elements of the model, Levels 0 – 3, are covered by ComScore's proprietary technology. Levels 4 – 7 are the new ideas proposed in this paper that interrogate the ComScore results from a strategic perspective by defining a market network of search intermediaries, competitors and consumers, and implementing a novel analytics framework based on our definition of theoretical constructs to define a multi-dimensional online performance model. The source/loss report is at the lowest of those levels, being a report generated from the data that has been sorted into a standardized database.

The source/loss report is specific to each individual airline and OTA. It gives the size of the focal website measured by unique visitors, and the source and loss of its traffic from/to other websites measured by the number of visits by consumers. Note that an individual consumer could have multiple visits to one website and therefore arrive from different websites, and depart to different websites. The source/loss is therefore reported in visits rather than unique visitors. In this case, the set of websites considered are airlines and OTAs. The data excerpt shows that American Airlines had 4.676 million unique visitors in March 2014. Taking one other website, Expedia generated 146,000 visits to the American Airlines website. Looking at the loss column, American generated 267,000 visits to the Expedia website, shown here as a loss from the American Airlines website to Expedia. The source and loss data for

all the airlines and OTAs are used to create a market network view of consumer search flows. The number of unique visitors within a one-month time-period defines the size of each node in the network. Structural network information is therefore generated, which is the foundation for evaluating online performance.

This type of data cannot be captured by an individual airline, because it would only have access to its own web server data. Clickstream data from online panels can therefore be used to establish the trajectory of online users across travel-related websites. The related nature of the search data enables researchers to measure patterns across websites. Online panel data has some characteristics of intensive research methods, while also being an example of an extensive research method because of the size of the sample (Sayer, 1992). The combination of extensive and intensive characteristics is unusual, because traditional research methods lose detail and relatedness as the extensiveness of the research is increased. With online panel data, an increase in the scale of the sample does not affect the level of granular detail or relatedness of the data.

4.2. Panel recruitment and management

Panel recruitment is done through a variety of online and offline recruitment incentives (Callegaro et al., 2014). Offline recruitment is used to overcome possible issues with online recruitment bias. Incentives to join the panel include the use of free software, e.g., email virus protection and a tree-planting program. The range of incentives offered is very wide in order to give the panel a diverse psychographic appeal. Affiliate marketing techniques are used to recruit panel members, where ComScore works with other organizations in the same way that online retailers use affiliate marketing to acquire new customers (Duffy, 2005). A potential member is told that they are joining an online panel where their behavior will be tracked, and that their individual identity is guarded and remains private. Demographic data is only used at an aggregate level and can be used to segment the panel.

One of the issues with panels is the true identity of individual users (Bucklin & Sismeiro, 2009), in this case the user of a computer. To overcome the problem of multiple users of a single machine within an individual household or place of work, ComScore uses a patented technology, called 'User Demographic Reporting', which identifies the user based on characteristic patterns in their keystrokes and mouse movement patterns.

The data collection process entails a hierarchical model. It originates and moves from clickstream data, which is codified into a flexible database of online usage patterns, including demographic profile, search trajectories, buying behavior on specific pages and time spent per page. The reports are linked to ComScore's own categorization of the websites into specific industry sectors. Standard management reports are then generated on demand from the database, and this research utilized and built on these reports (Level 4, Fig. 2) to develop a new B2B analytics to assess online performance (Level 5-7).

4.3. The use of online panel data

The main advantages of ongoing measurements using an online panel are accuracy, market network scope, measurement of the scale and direction of relationships, identification of latent relationships (Mariotti & Delbridge, 2012), measurement of changes over time and the ability to conduct cross-sector and international analyses in a cost effective manner. The approach is particularly valuable when online consumer search activity plays an important role in the customer journey, which is true in the airline market where over 50% of bookings are made online (Expedia, 2015) and inform the search stage of the customer journey, regardless of where the final purchase is made (Xiang, Magnini, & Fesenmaier, 2015). A disadvantage of this method is that online panels require significant effort to set up and maintain in order to ensure that their composition reflects the general population.

In practical terms, this means that researchers are more likely to gain interesting and relevant data by working in partnership with existing commercial panels rather than setting up their own initiatives.

The Hawthorne effect is well recognized in all forms of social science research (Lewis-Beck, Bryman, & Liao, 2003), but its effects in online panels are very small or negligible because of the non-intrusive nature of the tracking software and the long-term membership of the panel. Even if users are conscious of the software initially, over time it will become the norm (Callegaro et al., 2014; Göritz et al., 2002). For further detailed discussion of methodological issues including panel recruitment, composition and validity of the data the interested reader is referred to two core texts in this area (i.e., Batinic, Reips, Bosnjak, & Werner, 2002; Callegaro et al., 2014).

5. Assessing online performance in the US airline travel network

The analysis of the data concerns the business relationships between the major US airline companies and OTAs based on consumer traffic flows between their websites. The largest five American airline companies were analyzed, which collectively account for over 70% market share of domestic air travel (US Department of Transportation, 2019). The largest OTAs were sampled based on their economic importance, and online size measured by the number of unique visitors reported by ComScore. The two largest OTAs account for gross bookings of \$50 billion each (Expedia, 2015; Priceline, 2015) and play a crucial role in the online search process (Xiang et al., 2015). Collectively, these airline and OTA websites account for the majority of online search behavior in the airline market.

5.1. Queries and report generation

The source/loss data for each website is combined to create a network view of online activity. Note that the term OTA here also refers to websites such as Kayak that does price comparisons only and does not have a booking functionality. The online visitor matrix is shown in Table 3.

5.2. Visualization of the online market

Each of the major airlines and OTAs are shown in each row and column in Table 3. The visitor data is presented as ‘from website A to website B’ where A is in the row, and B is in the column. For example, the intersection of United Airlines in the second row and SouthWest in the first column, the figure of 197 means that 197,000 visitors went from United Airlines’ website to SouthWest. The sum of the rows is therefore the total losses from an individual website to other websites, and the sum of the columns is the total gains. A visual representation of the network is shown in Fig. 3, which is based on standard graph techniques and software using the matrix data from Table 3 (Borgatti,

Everett, & Freeman, 2002; Freeman, 2004). Note that for the purpose of visualization, the ‘from – to’ aspect of the data has been used, representing in effect the top half of the matrix in Table 3.

The picture shows the complexity of traffic flows and starts to uncover interesting differences in terms of the scale of individual business relationships from which online performance metrics can be calculated. The existence of a relationship between airlines and OTAs is inferred from the search network, in what could be termed a pseudo market-network (Martens et al., 2016). Marketing managers of an individual airline will realize that they are related to other organizations in their market and will typically have detailed information about specific regions of this network. For example, American Airlines will have very detailed information about the number of referrals from Expedia, because this will be governed by contractual arrangements. However, individual airlines are unlikely to have visibility of the network as a whole, and to be able to conduct competitor analysis based on the scale of individual relationships. It is also possible to focus on specific categories of business organization, e.g., OTAs and airlines as separate groups. The network diagrams for these groups are shown in Figs. 4 and 5.

It can be seen from Fig. 4 that consumers use multiple OTAs, and the high volume of visitors flowing from one travel agent to another demonstrates the importance of these online flows. This result is not entirely intuitive because one could expect that consumers would choose a preferred agent and then rely on that agent to give them comprehensive market coverage in terms of potential flight options. However, this is clearly not the case in practice. Relationships between OTAs may not be fully apparent to OTA marketing managers, whose natural focus would be towards the airlines for business development. In this instance, latent relationships can be identified through the extent of the visitor flow between two parties (Mariotti & Delbridge, 2012). The network diagram also yields interesting and commercially valuable information about the size and network position of competing OTAs.

The traffic flows between airlines and OTAs can be used to generate measures of online performance that takes into account market size. Whilst the use of visualization techniques do not in themselves contain any additional or new information to the source/loss matrix, they present an intuitive perspective of the network data that is easier to interpret than a matrix. Visualization makes certain properties of the network very clear by inspection only and without recourse to formal social network analysis, for example the size of the network, its members, the boundary and its structure. In particular, the level of connectedness between nodes (Anderson, Håkansson, & Johanson, 1994) has implications for assessing the state of the business relationships, even between competitors (Bengtsson & Kock, 2000). This visual description of the data can then be used to guide more specific analytical questions that are answered through the use of numerical analysis of the matrix. The analytical possibilities for competitor profiling are explored in the next section.

Table 3

Online visitor matrix for US airline market, source ComScore, US panel of approximately 1 million, Units (000s).

| | SouthWest | United Airlines | Delta | American | US Airways | Expedia | Priceline | Travelocity | Orbitz | Kayak | Tripadvisor | Fareportal | Losses: |
|-----------------|-----------|-----------------|-------|----------|------------|---------|-----------|-------------|--------|-------|-------------|------------|---------|
| SouthWest | | 213 | 211 | 136 | 88 | 196 | 177 | 98 | 157 | 138 | 35 | 181 | 1630 |
| United Airlines | 197 | | 261 | 131 | 134 | 129 | 105 | 61 | 112 | 146 | 31 | 49 | 1356 |
| Delta | 103 | 229 | | 105 | 56 | 133 | 155 | 111 | 147 | 120 | 47 | 69 | 1275 |
| American | 61 | 157 | 134 | | 209 | 267 | 118 | 76 | 118 | 200 | 92 | 67 | 1499 |
| US Airways | 99 | 96 | 82 | 231 | | 57 | 65 | 29 | 64 | 129 | 0 | 14 | 866 |
| Expedia | 296 | 191 | 255 | 146 | 88 | | 1872 | 1134 | 873 | 621 | 1708 | 794 | 7978 |
| Priceline | 191 | 120 | 187 | 138 | 43 | 1372 | | 923 | 709 | 2886 | 1332 | 902 | 8803 |
| Travelocity | 157 | 31 | 102 | 44 | 42 | 556 | 880 | | 523 | 520 | 705 | 584 | 4144 |
| Orbitz | 88 | 180 | 128 | 27 | 31 | 654 | 512 | 546 | | 426 | 399 | 636 | 3627 |
| Kayak | 122 | 194 | 311 | 195 | 120 | 343 | 3397 | 373 | 411 | | 380 | 518 | 6364 |
| Tripadvisor | 65 | 119 | 65 | 70 | 0 | 1394 | 884 | 1153 | 588 | 358 | | 559 | 5255 |
| Fareportal | 335 | 105 | 76 | 75 | 50 | 756 | 806 | 455 | 636 | 413 | 495 | | 4202 |
| Sources: | 1714 | 1635 | 1812 | 1298 | 861 | 5857 | 8971 | 4959 | 4338 | 5957 | 5224 | 4373 | |

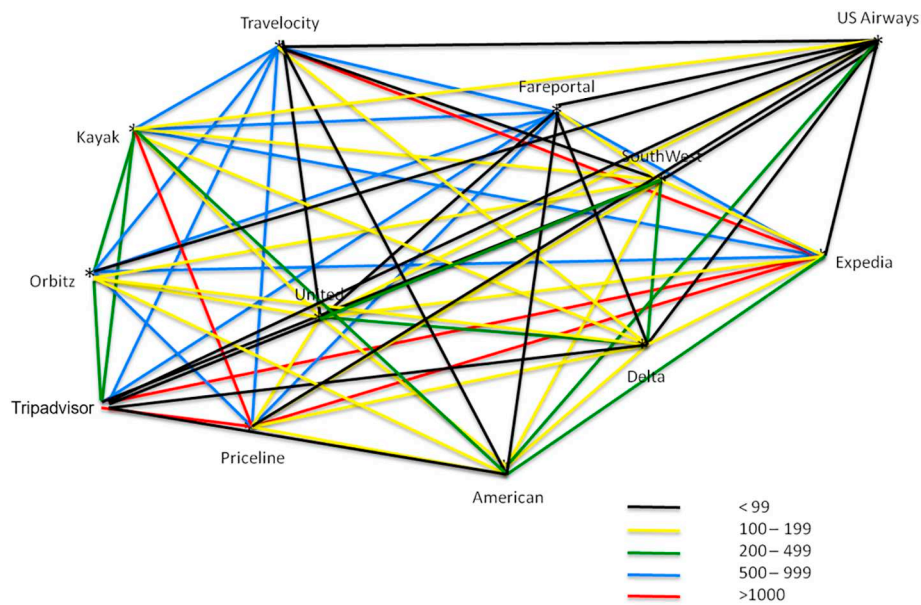


Fig. 3. Network diagram of the airline travel market, units 000s.

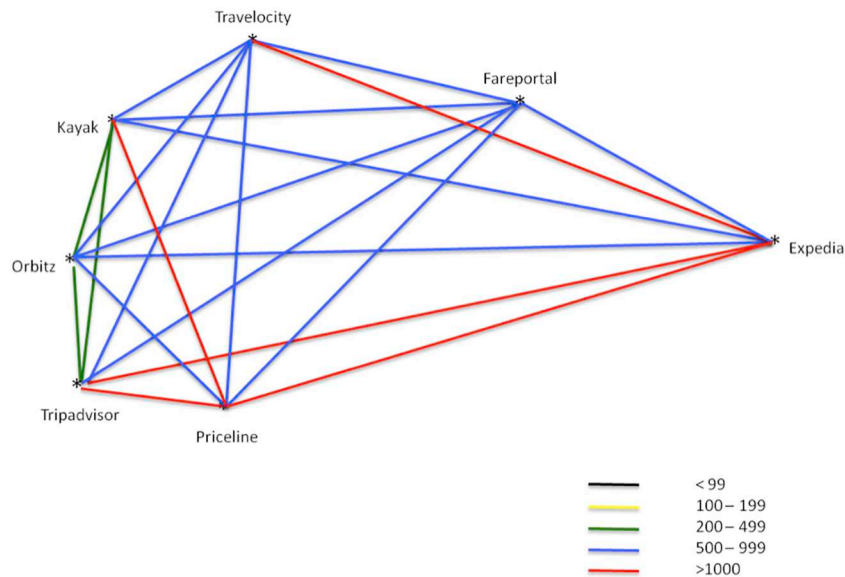


Fig. 4. Network diagram of US Online Travel Agents (OTAs), units 000s.

5.3. Theoretical constructs and measurement framework

The sources of traffic to airline websites from both other airlines and OTAs are strategically important because these aspects of online search trajectories generate visitors, which in turn generate sales. The online visitor matrix in Table 3 can be used to calculate sources of traffic to airline websites from both other airlines and the OTAs, which in turn can be used to calculate the performance of an airline to generate traffic from other airlines and OTAs respectively.

The simple measure of ‘traffic’ would not be a good indicator of online performance per sé, because it includes the effect of market size. To develop a metric that is meaningful for competitive comparisons, the effect of market size is removed by using the measure *share of traffic divided by passenger market share*, which is how advertising companies calculate advertising indices (Jones, 1989). An index of 1.0 is interpreted to mean that the airline’s online performance matches its market share. If it is greater than 1.0, then its online performance is greater

than the market average, and less than 1.0 means that it is underperforming. This approach has not previously been demonstrated in academic or consultancy frameworks of online performance. For example, McKinsey used the simple measure, volume of search activity, to identify the growth of a new entrant and did not attempt to carry out more sophisticated analyses (Dörner et al., 2013).

5.4. Data models and algorithms

The description above can be generalized into mathematical notation. The first measure, identified as ‘Algorithm I’ is concerned with ‘airline online performance’ and indicates the ability of an individual airline to attract traffic from other airlines. An airline only matrix, M , with n airline competitors (in this case $n=5$), is a sub-set of the matrix shown in Table 3. The entry in $M(i,j)$ is the flow of traffic from the airline in row i to the airline in column j . From this we derive the following definitions:

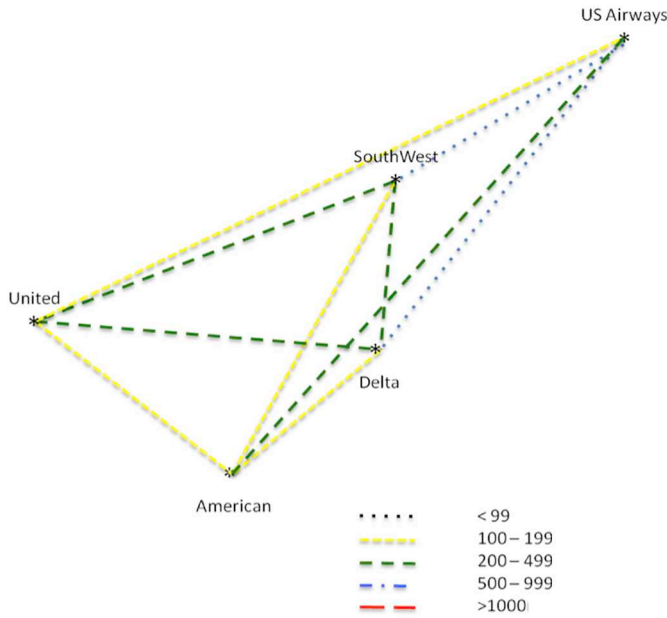


Fig. 5. Network diagram of US airlines, units: 000s.

Algorithm I.

1. A, competitor airline visitors into an individual airline x equals $\sum_{i=1}^{i=n} M_{i,x}$
2. B, the sum of all competitor airline visitors to all airlines equals $\sum_{j=1}^{j=n} \sum_{i=1}^{i=n} M_{i,j}$
3. C, the share of competitor airline visitors for airline x equals $\frac{A}{B}$
4. Passenger market share, D, for airline x in a given time period t, say one year, is: $D = \frac{\text{Passengers for airline } x}{\text{sum of passengers for all airlines}}$
5. The airline online performance E for an individual airline x for attracting traffic from other airlines, taking out the effect of market size is defined by: $E = \frac{C}{D}$

The algorithm described above can also be applied to calculate the performance of the business relationships for an airline with OTAs, expressed as an online performance index that is independent of market size. A more focused performance index, which looks at individual relationships with other airlines, is identified as ‘Algorithm II’ and analyzes the detailed composition of E.

Algorithm II.

1. The flow of traffic, a, from airline y to airline x is given by $a = M(y,x)$
2. The sum of all traffic, b, flowing from airline y to other airlines is given by: $b = \sum_{j=1}^{j=n} M_{y,j}$
3. The share of traffic, c, of website x of all of website y’s traffic is given by: $c = \frac{a}{b}$
4. The share of passengers, d, for airline x in a given time period t, say one year, is: $d = \frac{\text{Passengers for airline } x}{\text{sum of passengers for all airlines}}$
The performance of the relationship, e, of website x with website y, relative to competitors and taking out the effect of market size, is defined by: $e = \frac{c}{d}$

5.5. Interpretation of results and strategy formulation

The online performance results for gaining traffic from other airlines and OTAs are shown in Table 4. Note that the approach outlined is similar to density measures in social network analysis, but in this case

the definitions defined here are focused on performance and utilize the concept of ‘share of traffic’, and also take out the effect of market size, which is relevant in a competitive context when the purpose is to evaluate performance relative to competitors.

Table 4 shows the performance of the relationships between the individual airlines and (a) the other airlines as a group of websites, and (b) the OTAs as a group of websites. The measures ‘airline online performance’ and ‘OTA online performance’ are based on the inflows of traffic from these two groups of websites for each airline. For example, SouthWest attracts 460 thousand visitors from the other airlines, which gives it a share of airline visitors of 16%. Relative to its market size, this gives it a score of 0.53, i.e., 16% share of inflows of traffic from airline websites divided by its market share. In contrast, United Airlines has a score of 1.51, which means that the inflow of traffic from other airlines is much higher, relative to its market share. Indeed, in absolute terms, United Airlines has the highest inflow of airline-generated traffic. The OTA online performance for SouthWest is near parity, i.e., 1.0. This is despite its strategy of deliberately avoiding the price comparison engines as a distinctive part of its marketing strategy. That is, even though SouthWest is not listed on the comparison websites, its share of inflows from OTAs matches its market share, which means that it is performing at the market average. In order to illustrate the types of analyses possible, more details of the relative performance of individual relationships between the airlines and OTAs for two exemplar airlines, SouthWest and United Airlines, are shown in Table 5. These two were chosen simply because they represent the lowest and the highest scores on Airline online performance (column E in Table 4 above).

SouthWest’s online performance of 0.53 in Table 4 is analyzed in more detail in Table 5, which shows the performance of individual relationships with all of the other airlines range from 0.37 with American to 0.93 with United Airlines. From Table 4, SouthWest’s OTA online performance is 0.97. Table 5 shows the composition of this average performance and shows a large variance of scores with individual OTAs, which identifies a potentially fruitful area for further market research, possibly related to the segmentation of customers using different OTAs.

United Airlines’ online performance is 1.51 in Table 4, and the detailed composition of this average score are shown in Table 5, where it attracts a higher share of traffic from every other airline, relative to its market share. This tells us that United Airlines is more likely to be included in customers’ consideration sets when a customer conducts direct search with a competitor airline than one would expect from its market share position. United also scores highly with its average OTA online performance of 1.36, which includes a wide range of individual scores from 0.52 with Travelocity to 2.52 with Orbitz. Possible explanations for the variation in the average scores and the individual scores with other airlines and OTAs include the success of online marketing, the value of an airline’s offers, its commercial relationships with OTAs, and its segmentation strategy. However these ideas would require further analysis and synthesis of the big data results shown here with other market information such as company market research, qualitative customer research and passenger sales data, in order to be confident in the interpretation of the results.

The big data results shown here represent a new form of data analysis that would not be possible to derive using existing methods of market research and on its own can generate some useful new ideas and insights. From the perspective of SouthWest, the results suggest that it is not as closely integrated into the airline travel network as the legacy airlines. However, it is successful in generating visitors, and must therefore be compensating for this relative network weakness in other areas such as television, paid search and online advertising, which is typical of a direct distribution strategy of low-cost airlines (Wensveen & Leick, 2009). This brief discussion is not meant to be an exhaustive analysis, but to illustrate some possible analytical approaches that can be used with market network information derived from panel data.

Table 4
Online performance based on Algorithm I.

| Airline | Passenger market share (D) | Competitor airline generated visitors (000s) (A) | Share of competitor airline visitors (C) | Airline online performance (E) | OTA generated visitors (000s) | Share of OTA visitors | OTA online performance |
|---------------|----------------------------|--|--|--------------------------------|-------------------------------|-----------------------|------------------------|
| Southwest | 29% | 460 | 16% | 0.53 | 1,254 | 29% | 0.97 |
| Delta | 25% | 688 | 23% | 0.93 | 1,124 | 26% | 1.02 |
| American | 17% | 603 | 21% | 1.20 | 695 | 16% | 0.93 |
| United | 16% | 695 | 24% | 1.51 | 940 | 21% | 1.36 |
| US | 13% | 487 | 17% | 1.32 | 374 | 8% | 0.68 |
| TOTALS | 100% | 2,933 (B) | 100% | | 4,387 | 100% | |

N.B. % figures rounded to nearest 1 per cent. Performance ratios calculated using 2 decimal places in the original numbers. The letters A to E refer to Algorithm I defined in paragraph 5.4

6. Discussion and conclusions

This paper has demonstrated a new approach that measures online performance using market-level data in a rigorous and structured manner, which can be replicated and applied in different market contexts. A synthesis of the online performance framework is shown in Fig. 6. Starting with clickstream data, the network structure can be visualized and the results are shown in Figs. 3–5. The visitor matrix in Table 3 and the passenger data are the input data for Algorithms I and II, which generate online performance indices for each airline.

This is the first time that this type of approach has been applied to infer the strength of business relationships between competitors, and between competitors and search intermediaries, using online panel data. The starting point is to utilize online panel data to derive visualizations of the network structure, showing the extent of interactions between the airlines and OTAs. The methodology then expands on the simple visualization to develop formal and mathematical definition of new theoretical constructs to measure performance based on network diagrams derived from a market-level source/loss matrix. The paper has therefore contributed to the growing research area in network analytics (Chen et al., 2012). Our market network level research is distinct from existing organization-level research that adopts a focal company perspective (e.g., Kuo & Chuang, 2016; Pakkala et al., 2012; Plaza, 2011; Wilson, 2010). In this paper, *no one* company is at the heart of the analysis, because a market-wide, network perspective is taken. In the field of market-level analytics, the comparisons made between different competitors do not provide a holistic market view of the competitive landscape (e.g., Holland et al., 2016; Johnson et al., 2004; Martens et al., 2016; Zhang et al., 2006) as we demonstrate in our new analytical approach and applications. This helps to identify the linkages that exist in the consumers' minds, the implications of which might not be readily transparent to the managers concerned.

The advantage of online panel data over methods such as market research projects with customers or survey data is that the size of the panel makes it possible to develop reliable and objective measures of

the wider market network beyond an individual airline. In this way it is possible to derive performance measures based on search behavior, and to gain a descriptive, visual analysis of the network as a whole, which makes it possible to see how different airlines and OTAs are related to each other and to measure precisely the strength of these relationships in an objective manner. The methodology described here could be applied to larger market networks that include smaller competitors and the results would enable managers to see beyond their immediate economic partners and extend their network horizon to gain a broader understanding of the competitive landscape (Holmen & Pedersen, 2003). This application answers the call of Law et al. (2010) that research suffers from a lack of universal applications and approaches to measure and evaluate the online performance of travel websites.

Methodologically, we add to the growing literature on how big data can be used to deepen our understanding of emerging business models (Amankwah-Amoah, 2016; Cachia, Compañó, & Da Costa, 2007), and specifically shed new light on how consumer big data can be transformed and analyzed to inform B2B marketing decision. The paper has shown that online panel data is a new source of big data that can be used to design a new research methodology to measure and track the online behavior of very large number of consumers in the online channel. The comparison of online panel data with traditional research methods in Figure 1 clearly shows that online panel data breaks the trade-off between scale and detail in a comparison of research methods.

The hierarchical model of big data shown in Fig. 2 is a contribution to the big data discussion, in that it demonstrates that the analysis and interpretation of raw big data such as clickstream data requires multiple levels of analysis in order to interpret it, and that there are significant managerial problems to handle big data in addition to the technical problems associated with managing it. There are important software and technical challenges in levels 0 to 3 that require significant investment into software, technology platforms and technical staff. In levels 4 to 7, the challenges are more concerned with market definition, development of meaningful theoretical constructs, data models, algorithms and interpretation to develop insights that can be used to inform

Table 5
Detailed online performance based on Algorithm II.

| | SouthWest share of visitors from individual websites | SouthWest online performance | United Airlines share of visitors from individual websites | United Airlines online performance |
|-----------------------------|--|------------------------------|--|------------------------------------|
| Airlines | SouthWest | - | 33% | 2.09 |
| | United Airlines | 27% | - | - |
| | Delta | 21% | 46% | 2.95 |
| | American | 11% | 28% | 1.78 |
| | US Airways | 19% | 19% | 1.20 |
| Online Travel Agents (OTAs) | Expedia | 30% | 20% | 1.24 |
| | Priceline | 28% | 18% | 1.12 |
| | Travelocity | 42% | 8% | 0.52 |
| | Orbitz | 19% | 40% | 2.52 |
| | Kayak | 13% | 21% | 1.31 |
| | Tripadvisor | 20% | 37% | 2.37 |
| | Fareportal | 52% | 16% | 1.04 |

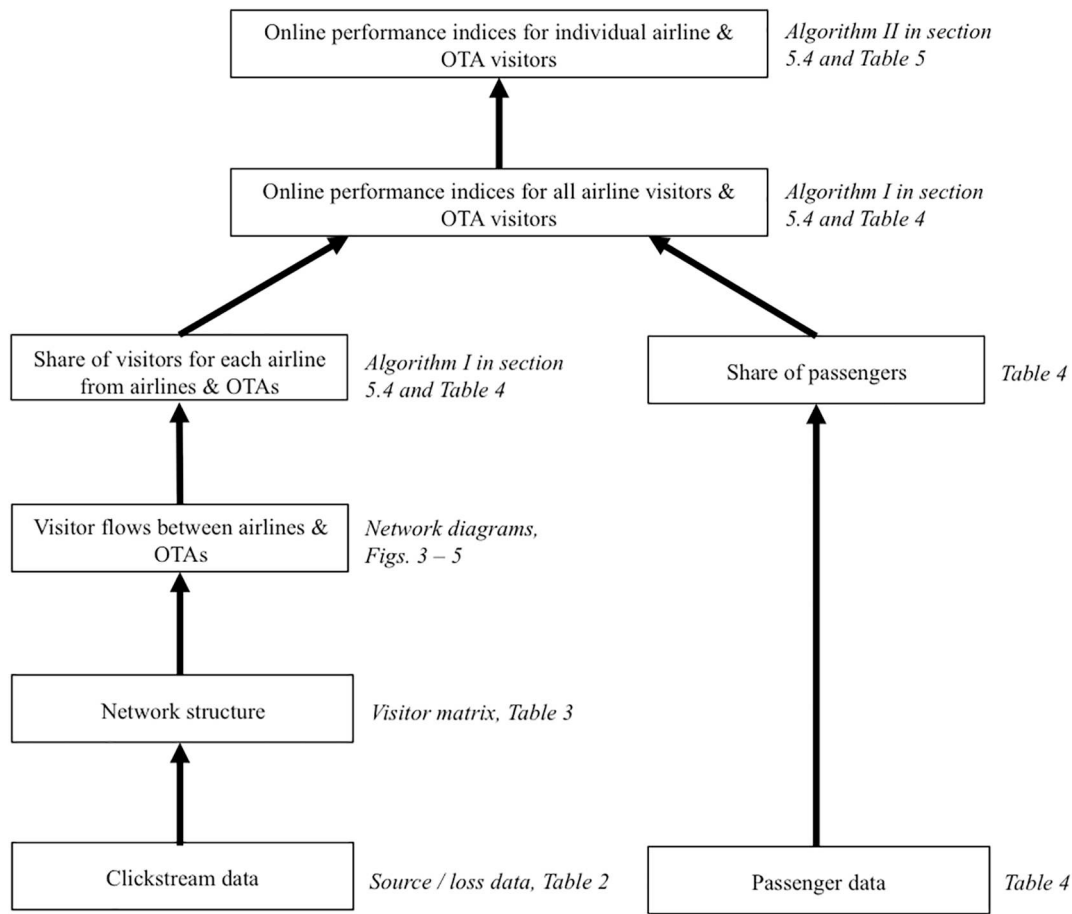


Fig. 6. Online performance framework based on a synthesis of the theoretical constructs.

strategy formulation. All of these levels of analysis are crucial in order to reach insights that have strategic value for managers involved in monitoring online performance within a market and to inform the allocation of online marketing resources. The model also shows that the managerial analysis and interpretation of big data is quite distinct from the technical challenges of handling raw data such as clickstream data. The analysis and results are presented using the structure of the hierarchical model of big data and show that consumers define their own networks, and the scale and importance of relationships between different organizations can be measured in a consistent and accurate manner.

The managerial implications of this approach are that companies will be able to map out their competitive landscape from the perspective of the customer, and thereby gain different insights about the market, particularly regarding structural information that is not possible using web server software. There are clear applications for this approach in market segmentation, advertising, competitor analysis and evaluation of relationships with other airlines and OTAs. The use of big data in the form of online panel data creates new analytical possibilities for online performance and competitor profiling in a market context.

Four types of marketing intelligence can be identified and offered to B2B marketing decision-making. (1) A network perspective of the data gives a view of the overall market, including the size, network position of individual airlines and OTAs, and structural information of the network. (2) The online performance of all airlines against each other and against the overall market can be measured in an objective and dynamic manner. (3) Potential new relationships based on consumer activity with competitors and intermediaries can be identified and the commercial potential of these possible partnerships estimated. (4) In fast changing markets, or in a situation when the market is in a state of

flux, online activity provides current information and can track the scale and growth of new trends such as the growth of a new competitor, the impact of a change in advertising or a new offer, or the impact of mergers and acquisitions on search behavior.

An important point about the formal definition of the performance metrics described and explained in the methodology and illustrated with the US airline market is that this type of approach can be automated to provide ongoing, longitudinal analyses of performance and market changes, which apply to all of the four types of market intelligence described above (Sheth & Parvatiyar, 2002; Wilson, 2010). The concepts are also equally applicable in any ecommerce market, characterized by a B2B2C market structure, where there is significant online activity and where search intermediaries form an important part of how the market functions.

The approach that we have developed has broader implications for managers in the B2B community generally. We have shown how a new type of big data, from an online panel, can be used to generate an understanding of a B2B network of important players, and also of their online performance relative to a predefined set of competitors or other counterparts in an industry. This overcomes the limitation of taking only a focal firm's perspective and provides a way to be much more ambitious regarding the formal analysis of much larger networks. B2B managers can therefore adopt this approach to create new possibilities for understanding behavior and strategies that lie beyond what is visible to the firm. It opens up two distinct possibilities for B2B managers. The first is to undertake the analysis at regular intervals, thereby providing a track record of current performance relative to competitors, something that cannot be done using traditional market research methods. This, in turn, facilitates the identification of particular competitors or other players that might require a particular strategic

response. Depending on the diagnostics of the relationship strength of a player and its network position in relation to a focal firm, relationship initiation, development and termination routines can be applied in keeping a strong relationship portfolio (Forkmann, Henneberg, Naudé, & Mitrega, 2016; Mitrega, Forkmann, Ramos, & Henneberg, 2012). These routines can be exercised through conscious networking, such as strengthening existing business relationships and introducing new relationships that can bring about new opportunities. Empirical evidence shows that the ability to sense the network and seize new opportunities has a positive impact on relationship portfolio management and subsequently enhances firm performance (Thornton, Henneberg, & Naudé, 2015). For instance, the identification of an increasingly important player that is not currently connected to a focal firm could motivate an initiation of a formal relationship, e.g., strategic partner, in order to capture an emerging market opportunity.

7. Limitations and future research directions

Online panel data is a relatively new source of research information and one that has been under-utilized by academics. As with all types of panel research, the size of the panel places limitations on what is possible regarding more focused analyses that circumscribe the data sets, e.g., to analyze the interaction and customer search behavior between two companies means that the research sample only includes those members of the online panel that are users of both of these websites. If the focus becomes very narrow, then this may cause the problem of minority sampling. Having said that, in the case of the data source used here, ComScore has a global panel of two million users, with one million in the US and very large panels in Europe and China. However, even with such large online panels, the application of the research design is inherently more reliable when examining the online activities of larger organizations, and it is more difficult to track the online users of Small and Medium sized Enterprises (SMEs).

The new concepts and analytical approaches that have been introduced in this paper can be applied and extended in a variety of ways: (1) to other markets in which on-line behavior forms a large part of the buying decision, (2) internationally, e.g., in order to compare international markets, and (3) to evaluate markets in cross-sector studies to develop more general theories in areas such as online search, use of decision tools such as OTAs and price comparison engines and online buying behavior.

The future potential of this approach will be achieved by exploiting the unique features of online panel data and also by combining it with other sources of market data such as sales data and strategic insights gained from working closely with managers from individual companies within a network. The most exciting developments are likely to be in research that exploits the ability to make international comparisons, conduct cross-sector research, utilize the demographic information of users and compare behavior across desktop and mobile devices. These new approaches will exploit the inherent flexibility of a database constructed from big data that would otherwise not be economically feasible for academic or commercial research organizations using traditional data sources. There is also the promise of being able to develop much more robust theory that is based on large-scale empirical evidence that can be tested and refined using natural experimental designs.

References

Achrol, R. S., & Kotler, P. (1999). Marketing in the network economy. *Journal of Marketing*, 63(4), 146–163.

Agarwal, R., & Dhar, V. (2014). Editorial - big data, data science, and analytics: The opportunity and challenge for is research. *Information Systems Research*, 25(3), 443–448.

Amankwah-Amoah, J. (2016). Emerging economies, emerging challenges: Mobilising and capturing value from big data. *Technological Forecasting and Social Change*, 110167–110174.

Anderl, E., Becker, I., Wangenheim, F. V., & Schumann, J. H. (2014). Mapping the customer journey: A graph-based framework for online attribution modeling. Available at SSRN: <https://ssrn.com/abstract=2343077>.

Anderson, J. C., Håkansson, H., & Johanson, J. (1994). Dyadic business relationships within a business network context. *Journal of Marketing*, 58(4), 1–15.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2014). Transformational issues of big data and analytics in networked business. *MIS Quarterly*, 38(2), 629–631.

Baltagi, B. H. (2013). *Econometric analysis of panel data* (5th). Chichester: Wiley.

Batinic, B., Reips, U.-D., Bosnjak, M., & Werner, A. (2002). *Online social sciences*. Seattle: Hogrefe & Huber Publishers.

Bengtsson, M., & Kock, S. (2000). Coopetition in business networks— to cooperate and compete simultaneously. *Industrial Marketing Management*, 29(5), 411–426.

Bhargava, H. K., & Choudhary, V. (2004). Economics of an information intermediary with aggregation benefits. *Information Systems Research*, 15(1), 22–36.

Borders, A. L., Johnston, W. J., & Rigdon, E. E. (2001). Beyond the dyad: Electronic commerce and network perspectives in industrial marketing management. *Industrial Marketing Management*, 30(2), 199–205.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for windows: Software for social network analysis*.

Bucklin, R. E., Lattin, J. M., Ansari, A., Gupta, S., Bell, D., Coupey, E., & Steckel, J. (2002). Choice and the internet: From clickstream to research stream. *Marketing Letters*, 13(3), 245–258.

Bucklin, R. E., & Sismeiro, C. (2009). Click here for interact insight: Advances in click-stream data analysis in marketing. *Journal of Interactive Marketing*, 23(1), 35–48.

Cachia, R., Compañó, R., & Da Costa, O. (2007). Grasping the potential of online social networks for foresight. *Technological Forecasting and Social Change*, 74(8), 1179–1203.

Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). *Online panel research: A data quality perspective*. Chichester: John Wiley & Sons.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.

ComScore (2016). *Key measures report*.

Dörner, K., Galante, N., & Kauter, C. *Radar for the e-commerce market. About McKinsey Capabilities*. (2013). Available at: http://www.mckinsey.com/~media/mckinsey/dotcom/client_service/retail/articles/perspectives-winter-2013/9_radar_for_the_e-commerce_market_vf.aspx [Accessed October 2019].

Dubé, L., & Renaghan, L. M. (2000). Marketing your hotel to and through intermediaries. *Cornell Hospitality Quarterly*, 41(1), 73.

Duffy, D. L. (2005). Affiliate marketing and its impact on e-commerce. *Journal of Consumer Marketing*, 22(3), 161–163.

Edelman, D. C. (2010). Branding in the digital age. *Harvard business review*, 88(12), 62–69.

Eisenmann, T., Parker, G., & Alstyne, M. W. V. (2006). Strategies for two-sided markets. *Harvard Business Review*, 84(10), 92–101.

Expedia (2015). *Annual report 2015*. Expedia Inc. Available at: <http://www.sec.gov/Archives/edgar/data/1324424/000119312515035706/d838066d10k.htm> [Accessed October 2019].

Ferguson, R. B. (2014). Comscore: The art and science of big data, from the inside. *Big Idea: Competing With Data & Analytics [Online]*. Available from: <https://sloanreview.mit.edu/article/comscore-the-art-and-science-of-big-data-from-the-inside/> [Accessed October 2019].

Forkmann, S., Henneberg, S. C., Naudé, P., & Mitrega, M. (2016). Supplier relationship management capability: A qualification and extension. *Industrial Marketing Management*, 57, 185–200.

Freeman, L. C. (2004). *The development of social network analysis. A study in the sociology of science*. Vancouver: Empirical Press.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35, 137–144.

George, G., Haas, M. R., & Pentland, A. (2014). From the editors big data and management. *Academy of Management Journal*, 57(2), 321–326.

Goes, P. B. (2014). Editor's comments: Big data and is research. *Mis Quarterly*, 38(3), iii–viii.

Goodhardt, G. J., & Ehrenberg, A. S. C. (1967). Conditional trend analysis: A breakdown by initial purchasing level. *Journal of Marketing Research*, 4(2), 155–161.

Goodhardt, G. J., Ehrenberg, A. S. C., & Chatfield, C. (1984). The dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)*, 147(5), 621–655.

Göritz, A. S. (2007). Using online panels in psychological research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.). *The oxford handbook of internet psychology* (pp. 473–485). Oxford: Oxford University Press.

Göritz, A. S., Reinhold, N., & Batinic, B. (2002). Online panels. In B. Batinic, U.-D. Reips, M. Bosnjak, & A. Werner (Eds.). *Online social sciences* (pp. 27–47). Seattle: Hogrefe & Huber Publishers.

Gummeson, E. (2002). Relationship marketing and a new economy: It's time for de-programming. *Journal of Services Marketing*, 16(7), 585–589.

Hieronimus, F., & Kullmann, M. (2013). *Is your online marketing any good? Perspectives on Retail and Consumer Goods*. McKinsey & Company. Available at: http://www.mckinsey.com/client_service/retail/latest_thinking/perspectives_second_edition/~media/8a29b8969df44bed8409ca9236c39b66.aspx [Accessed October 2019].

Holland, C. P., Jacobs, J. A., & Klein, S. (2016). The role and impact of comparison websites on the consumer search process in the us and german airline markets. *Information Technology & Tourism*, 16(1), 127–148.

- Holmen, E., Aune, T. B., & Pedersen, A.-C. (2013). Network pictures for managing key supplier relationships. *Industrial Marketing Management*, 42(2), 139–151.
- Holmen, E., & Pedersen, A.-C. (2003). Strategizing through analyzing and influencing the network horizon. *Industrial Marketing Management*, 32(5), 409–418.
- Järvinen, J., & Karjalainen, H. (2015). The use of web analytics for digital marketing performance measurement. *Industrial Marketing Management*, 50, 117–127.
- Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science*, 50(3), 299–308.
- Jones, J. P. (1989). Ad spending: Maintaining market share. *Harvard Business Review*, 68(1), 38–42.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573.
- Ketter, W., Peters, M., Collins, J., & Gupta, A. (2015). Competitive benchmarking: An is research approach to address wicked problems with big data and analytics. *MIS Quarterly*, 40(4), 1057–1080.
- Koka, B. R., & Prescott, J. E. (2002). Strategic alliances as social capital: A multi-dimensional view. *Strategic Management Journal*, 23(9), 795–816.
- Koo, B., Mantin, B., & O'Connor, P. (2011). Online distribution of airline tickets: Should airlines adopt a single or a multi-channel approach? *Tourism Management*, 32(1), 69–74.
- Kumar, V., Lahiri, A., & Dogan, O. B. (2017). A strategic framework for a profitable business model in the sharing economy. *Industrial Marketing Management*, 69, 147–160.
- Kuo, M.-S., & Chuang, T.-Y. (2016). How gamification motivates visits and engagement for online academic dissemination—an empirical study. *Computers in Human Behavior*, 55, 16–27.
- Law, R., Qi, S., & Buhalis, D. (2010). Progress in tourism management: A review of website evaluation in tourism research. *Tourism Management*, 31(3), 297–313.
- Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (2003). *The sage encyclopedia of social science research methods*. Thousand Oaks: Sage Publications.
- Lilien, G. L. (2016). The b2b knowledge gap. *International Journal of Research in Marketing*, 33(3), 543–556.
- Mariotti, F., & Delbridge, R. (2012). Overcoming network overload and redundancy in interorganizational networks: The roles of potential and latent ties. *Organization Science*, 23(2), 511–528.
- Martens, D., Provost, F., Clark, J., & Junqué de Fortuny, E. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS Quarterly*, 40(4), 869–888.
- Mitrega, M., Forkmann, S., Ramos, C., & Henneberg, S. C. (2012). Networking capability in business relationships—concept and scale development. *Industrial Marketing Management*, 41(5), 739–751.
- Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4), 579–595.
- Opresnik, D., & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics*, 165, 174–184.
- Pakkala, H., Presser, K., & Christensen, T. (2012). Using google analytics to measure visitor statistics: The case of food composition websites. *International Journal of Information Management*, 32(6), 504–512.
- Pearce, B. (2014). The shape of air travel markets over the next 20 years. *Global Airport Development Conference*.
- Plaza, B. (2011). Google analytics for measuring website performance. *Tourism Management*, 32(3), 477–481.
- Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York: The Free Press.
- Priceline (2015). *Annual report 2014*. Edgar Online. Available at: http://www.annualreports.com/HostedData/AnnualReportArchive/t/NASDAQ_PCLN_2014.pdf [Accessed October 2019].
- Saboo, A. R., Kumar, V., & Park, I. (2016). Using big data to model time-varying effects for marketing resource (re) allocation. *MIS Quarterly*, 40(4), 911–939.
- Sayer, A. (1992). *Method in social science: A realist approach* (2nd). London & New York: Routledge.
- Sheth, J. N., & Parvatiyar, A. (2002). Evolving relationship marketing into a discipline. *Journal of Relationship Marketing*, 1(1), 3–16.
- Thorelli, H. B. (1986). Networks: Between markets and hierarchies. *Strategic Management Journal*, 7(1), 37–51.
- Thornton, S. C., Henneberg, S. C., Leischnig, A., & Naudé, P. (2019). It's in the mix: How firms configure resource mobilization for new product success. *Journal of Product Innovation Management*, 36(4), 513–531.
- Thornton, S. C., Henneberg, S. C., & Naudé, P. (2013). Understanding types of organizational networking behaviors in the uk manufacturing sector. *Industrial Marketing Management*, 42(7), 1154–1166.
- Thornton, S. C., Henneberg, S. C., & Naudé, P. (2014). Conceptualizing and validating organizational networking as a second-order formative construct. *Industrial Marketing Management*, 43(6), 951–966.
- Thornton, S. C., Henneberg, S. C., & Naudé, P. (2015). An empirical investigation of network-oriented behaviors in business-to-business markets. *Industrial Marketing Management*, 49, 167–180.
- Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4), 463–479.
- US Department of Transportation (2019). *Airline domestic market share July 2018 - June 2019*. [online]. Bureau of Transportation Statistics. Available at: <http://www.transtats.bts.gov/> [Accessed October 2019].
- US Department of Transportation (2017). *2017 traffic data for U.S. Airlines and foreign airlines U.S. Flights*. [online]. Bureau of Transportation Statistics. Available at: <http://www.bts.gov/newsroom/2017-traffic-data-us-airlines-and-foreign-airlines-us-flights> [Accessed October 2019].
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, J. F., Ren, R. D., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.
- Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2017). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, 55(1), 64–79.
- Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.
- Wensveen, J. G., & Leick, R. (2009). The long-haul low-cost carrier: A unique business model. *Journal of Air Transport Management*, 15(3), 127–133.
- Wilson, R. D. (2010). Using clickstream data to enhance business-to-business web site performance. *Journal of Business & Industrial Marketing*, 25(3), 177–187.
- Wind, Y. J. (2008). A plan to invent the marketing we need today. *MIT Sloan Management Review*, 49(4), 21.
- Wixom, B. H., Ross, J. W., Beath, C. M., & Miller, C. A. (2013). *Capturing value from big data at comscore through platform, people, and perception*. MIT CISR.
- Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 22, 244–249.
- Zhang, J. J., Fang, X., & Sheng, O. R. L. (2006). Online consumer search depth: Theories and new findings. *Journal of Management Information Systems*, 23(3), 71–95.
- Zott, C., & Amit, R. (2010). Business model design: An activity system perspective. *Long Range Planning*, 43(2), 216–226.