



WORKING PAPER

ITLS-WP-18-19

**Creation of unstructured big data from
customer service: The case of parcel
shipping companies on Twitter**

By

**Jyotirmoyee Bhattacharjya¹, Adrian Bachman
Ellison¹, Vincent Pang², and Arda Gezdur¹**

¹Institute of Transport and Logistics Studies (ITLS),
The University of Sydney Business School, Australia

²The University of New South Wales

October 2018

ISSN 1832-570X

**INSTITUTE of TRANSPORT and
LOGISTICS STUDIES**

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-18-19

TITLE: **Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter**

ABSTRACT: **Purpose** - Customer service provision is a growing phenomenon on social media and parcel shipping companies have been among the most prominent adopters. This has coincided with greater interest in the development of analysis techniques for unstructured big data from social media platforms, such as the micro-blogging platform, Twitter. Given the growing use of dedicated customer service accounts on Twitter, this paper investigates the effectiveness with which parcel shipping companies use the platform.

Design/methodology/approach – This paper demonstrates the use of a combination of tools for retrieving, processing and analysing large volumes of customer service related conversations generated between parcel shipping companies and their customers in Australia, United Kingdom and the United States. Extant studies using data from Twitter tend to focus on the contributions of individual entities and are unable to capture the insights provided by a holistic examination of the interactions.

Findings – This study identifies the key issues that trigger customer contact with parcel shipping companies on Twitter. It identifies similarities and differences in the approaches that these companies bring to customer engagement and identifies opportunities for using the medium more effectively.

Originality/value – The development of consumer-centric supply chains and relevant theories require researchers and practitioners to have the ability to include insights from growing quantities of unstructured data gathered from consumer engagement. This study makes a methodological contribution by demonstrating the use of a set of tools to gather insight from a large volume of conversations on a social media platform.

KEY WORDS: *unstructured big data, big social data, Twitter, parcel shipping, media richness, customer service, global logistics, logistics services, retail logistics*

AUTHORS: *Bhattacharjya, Ellison, Pang, and Gezdur*

CONTACT:

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES
(H73)

The Australian Key Centre in Transport and Logistics
Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1824

E-mail: business.itlsinfo@sydney.edu.au

Internet: <http://sydney.edu.au/business/itls>

DATE:

October 2018

Introduction

The speed of generation of large volumes of data from business operations and the wide variety of sources for such data (e.g., customer transactions, social media updates, GPS signals from mobile phones) have led to the recognition of both opportunities and challenges for managerial decision makers (Goes, 2014; McAfee and Brynjolfsson, 2012). Big data analytics involves working with high volume and high velocity data from a range of different sources (Richey et al., 2016) to increase operational visibility, identify performance variabilities, tailor products and services to market segments, support business model innovation, and achieve competitive advantage in dynamic business environments (McAfee and Brynjolfsson, 2012; Wamba et al., 2015; Wang et al., 2016). Big data can be of different types: structured, semi-structured and unstructured (Gandomi and Haider, 2015; Sagiroglu and Sinanc, 2013). While much of the extant literature focuses on structured data, the volume of unstructured data around the world is growing considerably more rapidly than structured data (Schneider, 2016).

Social media platforms including blogs, microblogging sites, wikis, social networking and video sharing sites and virtual worlds have played a significant role in the growth of high volume and high velocity unstructured data in a variety of contexts (Kaplan and Haenlein, 2010). The rapid adoption of such platforms by end-consumers have led managers to recognise new engagement and promotion opportunities for their businesses (Kiron et al., 2012). Companies have increasingly started connecting with customers using Facebook, Twitter and blogs, and the provision of links between company websites and social media sites has become part of the norm (Culnan et al., 2010; Gunton and Davis, 2012). This phenomenon is driven by the potential for creating virtual customer communities that can support branding, product development and sales (e.g., Jansen et al. 2009).

This study is conducted using data from the microblogging site, Twitter, which has over 320 million individual and organizational users around the world (Statista, 2017). The platform is used for information dissemination by government agencies (Waters and Williams, 2011), institutions of higher education (Linvill et al., 2012) and not-for-profit organisations (Waters and Jamal, 2011). It has served as an important communication tool during natural disasters (Drake,

2013; Papadopoulos et al., 2017). Businesses and professionals use the platform for a range of different purposes including the introduction of new products (Abney et al., 2017), sales improvement (Miller, 2009), responses to service failures (Fan and Niu, 2016), and the reporting of breaking news (Vis, 2012). The platform enables e-retailers to engage with a globally distributed consumer base (Bhattacharjya et al., 2016). Since parcel shipping companies play an important role in the supply chains of these e-retailers, the big data generated from the interactions of these carriers with end consumers on Twitter could be useful in generating service quality related insights for both partners.

Literature review

Big Data in Business Operations and Logistics

Recent evidence from research on big data analytics in operations and supply chain management suggests that while shortage of relevant skills and managerial inertia continue to pose challenges, big data analytics can provide significant benefits to businesses and create lasting competitive advantage (Arya et al., 2017; Dutta and Bose, 2015; Matthias et al., 2017; Richey et al., 2016). Big data sources may be harnessed for predictive analytics (Ilie-Zudor et al., 2015; Waller and Fawcett, 2013) to provide insights for a number of supply chain management challenges including predictive facility layout optimisation, identification of qualitative and quantitative criteria related to suppliers and preferred supplier and carrier selection (Lamba and Singh, 2017).

Unstructured Big Data from Social Media

The fastest growing type of big data around the world is unstructured and available in text, audio, video and image formats (Gandomi and Haider, 2015). Social media platforms host large volumes of unstructured big data created from the consumer engagement and brand management activities of many companies (Aral et al., 2013) and have the potential for producing significant insights for both practice and research in supply chain management (Chae, 2015). Many companies are increasingly using the microblogging platform, Twitter, for providing customer

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

service (Gunton and Davis, 2012). E-retailers and e-commerce companies have established customer service accounts on Twitter to address the concerns of their customers (Bhattacharjya et al., 2016; Bhattacharjya and Ellison, 2015). Parcel shipping companies which play important roles in the supply chains of e-retailers have also recognised the need for setting up accounts on Twitter to provide customer service.

Extracting insights from unstructured big data from social media requires techniques beyond those that are easily applicable to structured data. Fan and Gordon (2014) highlight three important aspects of analytics in relation to such data sources. In the first stage, data is captured from relevant social media sources and archived and relevant information is extracted (e.g., Rieder 2013). The capture of large volumes of data generated on Twitter has its share of challenges because of limitations of the search functions offered by the platform (Boyd and Crawford, 2012). Given the difficulties of accessing archival data, most research using Twitter has focused on events in the present or immediate past such as elections (Freelon and Karpf, 2014; Tumasjan et al., 2011) and natural disasters (e.g., Ashktorab et al. 2014; Lee & Chien 2013). The limited research in relation to interactions between companies and customers has generally involved the extraction of customer tweets (e.g., Jansen et al. 2009; Liau & Tan 2014; Misopoulos et al. 2014) or company tweets (e.g., Swani et al. 2014) but not both. The next major challenge involves making sense of these large volumes of data using text mining techniques (e.g., Fan et al. 2006). The third and final stage in analytics for unstructured big data involves the synthesis of actionable findings for different decision makers and is dependent on the quality of insights gained from text mining. While software packages used for analysing unstructured data have become very efficient over the years, the process of extracting intelligence from unstructured data beyond existing sentiment analysis approaches remains a labour-intensive process. This paper addresses the challenges of simultaneous gathering of tweets from companies and their customers and the subsequent analysis of both to provide insights into the customer service interactions of parcel shipping companies.

Media richness theory and social media

Media richness theory posits that organisations process information to reduce equivocality and uncertainty (Daft and Lengel, 1986) and selection of media by senior management needs to be aligned with purpose for successful communication. Media channels can be organised in a hierarchy based on four criteria that determine the facilitation of understanding (and therefore media richness): feedback, multiple cues, language variety and personal focus (Daft et al., 1987; Lengel and Daft, 1988). Firstly, instant feedback implies quicker exchange of clarifications. Secondly, depending on the medium, multiple cues, including voice inflections, words, numbers and symbols may be used in a given message. Thirdly, natural language can present a broader range of meanings while numbers convey precision. Finally, a message can be conveyed more fully when personalised and infused with emotion. Based on these criteria, Daft et al. (1987) considered face-to-face interaction to be the richest medium of communication followed by telephonic exchange and then documents.

Although the theory preceded the advent of the digital age, it has seen useful application in a number of more recent contexts including e-learning (Sun and Cheng, 2007), communications of employee benefits (Freitag and Picherit-Duthler, 2008), listening skills for accounting practitioners (Stone and Lightbody, 2012) and the selection and use of email, short message service (SMS) and Facebook Wall postings by university students (Park et al., 2012). In this paper, the theory is applied to gain better insight into the nature of communications between customer service staff at parcel shipping companies and their customers on Twitter.

Methodology

Background

The purpose of the study was to investigate how different parcel shipping companies utilise the microblogging platform Twitter for customer service interactions. The platform allows users (individuals or organisations) to post microblogs referred to as ‘tweets’. Each tweet is limited to 140 characters and can be accompanied by a link or a photo or a video (support.twitter.com). A

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

user may include a hashtag (i.e., the symbol ‘#’) before a keyword in the tweet if he/she wishes the tweet to be found easily by other users on the platform. Messages addressed to other users generally include the screen name of the intended user with the ‘@’ symbol included in front of the screen name. Most tweets are visible to the public but the platform also provides support for sending direct messages (DMs) which are only visible to the intended recipient. Direct messages can only be sent to users following a given user account. Companies were found to follow individual customers to allow these customers to send DMs regarding specific issues.

Sampling and unit of analysis

In selecting the cases, the study adopted a theoretical sampling approach (Barratt and Choi, 2007; Eisenhardt, 1989; Glaser and Strauss, 1967). The aim of the study was to examine the Twitter activities of parcel shipping companies based in Australia, United Kingdom and the United States. The leading parcel shipping companies in these countries as well as their competitors were identified using the Hoover’s Company Profiles database. The Twitter site was then checked to confirm that each relevant Twitter account provided a link to the corresponding company website and had been verified by Twitter. Only companies with verified Twitter accounts are included in the study as this ensures the authenticity of the accounts. A summary of the companies and their Twitter accounts is provided in Table 1. As shown in the table, a parcel shipping company may choose to have one or more accounts on Twitter. The study involves seven companies and 12 Twitter accounts. The data was collected between December 1, 2014 and July 10, 2015 and includes a total of 706,582 tweets.

Table 1: Overview of the parcel shipping companies

Parcel shipping firms	Sector		Geographical reach		Type of Twitter account		Number of tweets		Number of Conversations
	Public	Private	Domestic	International	Single all-purpose account	Dedicated customer service account	Parcel shipping firms	Customers	
PS1	X		X		X		48,454	43,059	27,610
PS2		X	X	X	X		7,736	3,857	3,071
PS3		X	X	X		X	7,421	2,990	2,244
PS4		X	X	X		X	58,011	25,309	24,684
PS5		X	X	X	X		91,030	68,016	37,745
PS6		X	X	X		X	148,606	147,806	1,697
PS7	X		X			X	29,926	24,361	16,009

Data collection

The premise of the study was to examine the large volume of data generated from interactions on Twitter between parcel shipping companies and their customers. The data collection was designed to retrieve tweets from or received by these companies while maintaining the inherent relationships between tweets. Existing free and commercially available tools have been designed to retrieve tweets individually without regard to these interactions. Therefore, it was necessary to develop a new tool to retrieve the relevant tweets while maintaining the relationships between tweets in a conversation. The data collection process works by using the companies' Twitter timelines as seeds for a recursive algorithm that descends each conversation in reverse chronological order until the original tweet is retrieved. In many cases a conversation may include several branches involving a different (but partially overlapping) group of companies/people.

Mining the tweets

The data was then imported into SAS Enterprise Miner for text mining purposes. A parsing functionality (SAS, 2012) was used to identify terms and their occurrences (frequency) for each

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

Twitter account. This step also included stemming and identification of multiword terms and synonyms in each data file. The stemming process recognised the presence of different tenses of the same verb as well as singular and plural forms of the same noun. Multiword terms, i.e., terms composed of more than one word, e.g., ‘Tracking Number’, ‘Reference Number’, ‘Phone Number’ and ‘First Class’ were also identified as part of the parsing process.

The parsed data was then filtered to separate the tweets of a company from those of its customers. The text filtering process also allowed for the correction of spellings and the manual creation of synonyms. Some multiword terms were identified to be synonyms of each other.

This was followed by text topic analysis which associated each tweet with a collection of terms that characterise the key theme. This step also assisted in the identification of terms that were not of significant value in the analysis (e.g. a company’s username included in a customer’s tweet), additional multiword terms and synonyms (e.g., “Track No”, “Tracking Number”, and “Tracking #”).

For each of the Twitter accounts included in the study, the analysis was iterative. Once the topics and corresponding terms had been examined, terms with similar meaning were substituted by a synonym before the steps from parsing through to topic analysis could be iterated. For example, in cases where the customer service officers included their names in their tweets, the names were treated as a single entity to avoid these names showing up as significant terms for any of the topics identified for the relevant companies. The initials of customer service officers and preceding characters (e.g., -, ^) were substituted by a single term (e.g., <PS7CSOfficer>, <PS1CSOfficer>. Topic maps were generated for customer service officers as shown in the Figure 1 for PS4’s customer service officers.

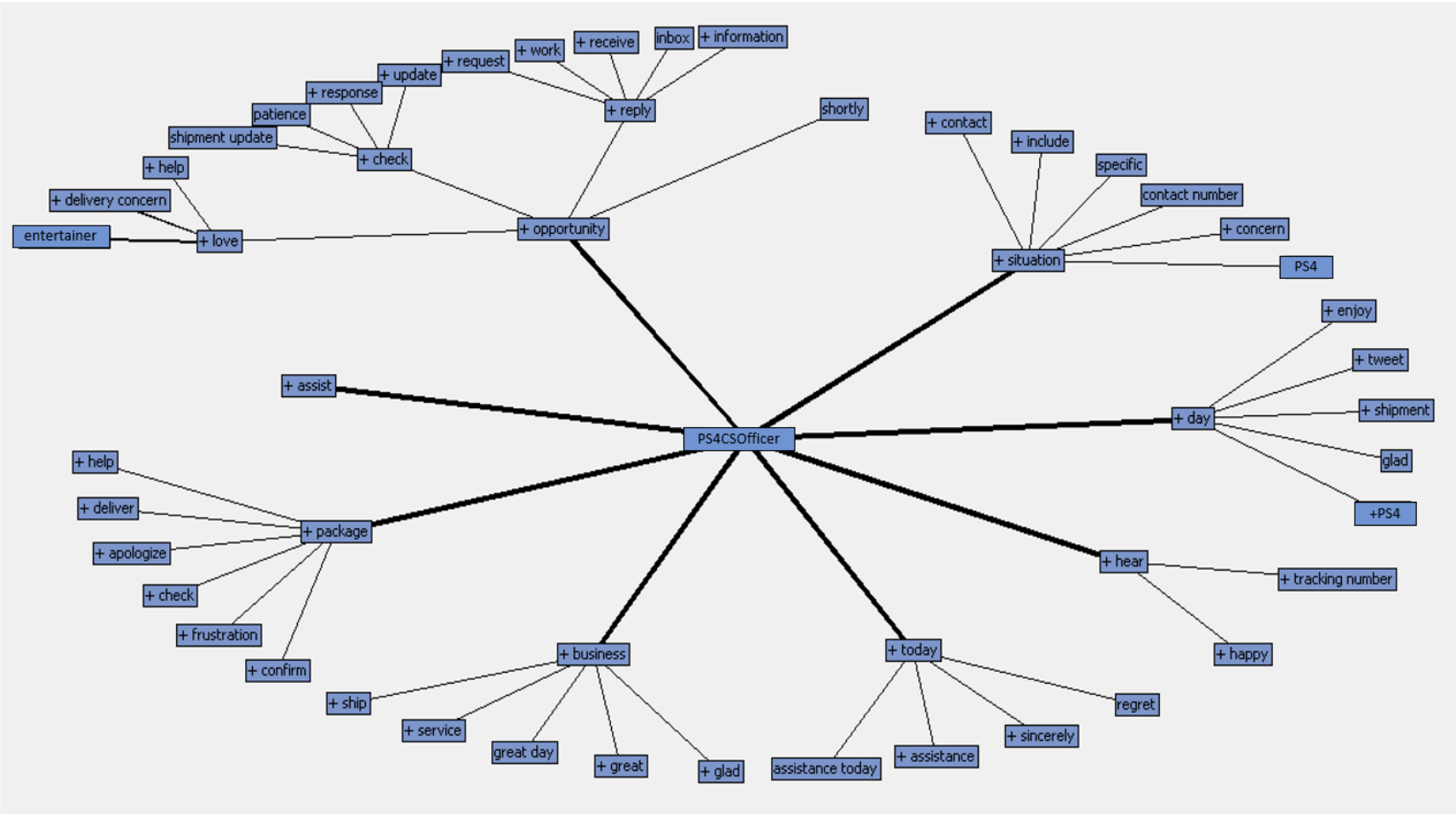


Figure 1: An example of a topic map for a customer service officer

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

For each Twitter account, iterations were performed till there were no significant changes to the terms associated with each topic. The dictionary and the lists of multiword terms and synonyms were augmented after the completion of analysis of each Twitter account. These were used as the starting point of analysis for the next account. Once the sequential analysis was completed, all the data files were reanalysed using the same dictionary, multiword terms and synonyms to ensure consistency.

Analysis of response delays and channel diversion

While the text mining process allowed the tweets from customers and companies to be analysed separately, understanding the extent to which the engagement capabilities of the platform are fully utilised from a media richness theory perspective (Daft et al., 1987) required a holistic consideration of the conversations. Each conversation consisted of two or more tweets with at least one tweet being posted by a customer and a response being posted by the corresponding company. The dataset was imported into the statistical package, R, and a script was created to determine the minimum, maximum and median response times for each parcel shipping company. Additionally, the frequencies with which customers were diverted to traditional channels or DM were also determined.

Findings

In the discussions below, the usernames of both the parcel shipping companies and the recipients (i.e., consumers) have been masked and only the first and last characters of tracking numbers are shown.

Characteristics of customer complaints

Tracking and delivery related issues were the most common reasons for which PS1 was contacted by customers via Twitter. Tracking enquiries included references to information provided by other parcel shipping companies in instances where PS1 was the last mile service provider. Reported delivery issues included delays, lost items and deliveries to wrong addresses as well as perceived incompetence in completing deliveries. Complaints about experiences in

physical stores were also posted on the platform. Hashtags such as #annoyed, #badservice, #fail and #poor were used to highlight negative experiences. Customers were also found to provide feedback on positive experiences but this was less frequent. PS2 received complaints involving similar issues. Some complaints involved references to other parcel delivery service providers in the chain.

PS3 and PS4 received queries regarding tracking and delivery issues as well as delays in customer service communications. Issues with the last mile service provider were also included amongst the complaints regarding delivery issues. Negative experiences with delivery occasionally led to tweets with embedded sarcasm and comparisons with its key competitors. Capitalizations and exclamation marks were used to convey emotion in relation to negative experiences. Hashtags were generally used to convey disappointment. While positive experiences with the delivery process and drivers were also reported, customers generally initiated contact while encountering problems.

Customers of PS5 deliveries also tweeted about a range of issues including delays, parcels delivered to wrong addresses and lack of care when leaving parcels at their intended destinations. Customers used a number of different hashtags including #badservice, #fail, #fuming, #frustration, #lateagain, #poorservice, #ripoff and #useless to express dissatisfaction and #greatcustomerservice, #thankyou and #welldone to express satisfaction.

Tracking related tweets directed at PS6 included both negative and positive feedback. Problems reported in relation to delivery included delays, damaged items, delivery driver incompetence, and packages left in insecure locations. Customer service communications drew both positive and negative feedback as customers experienced both prompt and delayed responses across different channels. Customers used a variety of hashtags in tweets about negative experiences. These include #badbusiness, #disappointing #fail, #PS4isbetter, #irresponsible, #NotHappy and #UselessPS6.

In relation to tracking, the two issues of concern for customers of PS7 included lack of updates and incorrect tracking information. Delivery related problems posted by customers included tweets about lost packages, packages delivered to wrong addresses, packages left without

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

knocking and stolen packages. Customer service communications were also a source of dissatisfaction as customers complained about the lack of prompt responses to both tweets and emails. Hashtags used to highlight negative experiences included #damagedpackage, #fail, #lost, #neveragain, #rude and #PS7fail. Table 2 presents examples of the complaints received by the parcel shipping companies.

Table 2: Examples of customer complaints received on Twitter by the parcel shipping companies

Companies to which the complaints are directed	Examples of customer complaints presented on Twitter
PS1	<p><i>@PS1 Can you please give me an update on item # C...E I am expecting from [country]? PS_other_1 tracking indicates it is now with you</i></p> <p><i>@PS1 trk #6...6 - signed by [wrong recipient]? Whoever that is. Incid#1...7. Was delivered to the wrong address, but signed</i></p> <p><i>Have been waiting @ home all day for @PS1 package to arrive. Just received email..it's been returned to warehouse What a joke!!</i></p> <p><i>@PS1 @friend's_username went to pick up a package on my behalf with proper ID from both of us. got harassed by staff. Not new for that store</i></p>
PS2	<p><i>Worst customer service @PS2, @PS_other_2 & @PS_other_3 - don't know where my parcel is, stop sending me from 1 place to another!!</i></p>
PS3	<p><i>@PS3 Look what you did to my shipment <sad face> This is our second email to claims and no one is answering! [image]</i></p> <p><i>PS3 tendered a package of mine to @PS7 Saturday and said it was supposed to arrive today. Of course, it didn't and no updates on PS7. #FAIL</i></p>
PS4	<p><i>Very interested to hear from the @PS4 guy tomorrow how he failed to find my giant office building. @PS6 never has a problem with it</i></p> <p><i>Hey @PS4, I know my house is hard to find but PS6, PS3, PS7, etc. can find it</i></p> <p><i>I order stuff and @PS4 and @PS6 give it to @PS7 to deliver? This is the second package to go missing. My iPod WAS NEVER FOUND. why!?</i></p>
PS5	<p><i>Got home to @PS5 card to say my parcel "left behind back gate" It's been raining & as gate locked it was thrown over! Lovely service</i></p>

PS6	<p><i>"@PS6 thanks for the help. Package was located and shipping information was updated. Guess it should be here today!</i></p> <p><i>Delivered 3 days late as couldn't find the address, shows up and it's broken! Still waiting email response to @PS6</i></p>
PS7	<p><i><abusive word> is the matter with you @PS7? Tracking says package delivered but it's not here. Been on hold waiting for help over 2 hrs! #PS7fail</i></p>

The reasons for which consumers were found to approach the parcel shipping companies on Twitter have been summarized in Table 3.

Table 3: A summary of conversation triggers requiring parcel shipping companies to respond on Twitter

Pre-delivery conversation triggers		Post-delivery conversation triggers		Pick-up/drop-off
Parcel shipping company's services	Parcel shipping partner's services	Delivery process	Physical item	
<p>Incorrect tracking information provided online</p> <p>Lack of online shipping information availability</p> <p>Lack of timely updates to tracking information</p> <p>Lack of response to tracking queries via other channels, e.g., email</p>	<p>Lack of timely updates to shipping information by last mile partners after handover of package</p> <p>Packages misplaced/lost during handover between parcel shipping companies</p>	<p>Deliveries to incorrect addresses</p> <p>Late deliveries</p> <p>Perceived driver incompetence</p> <p>Packages left in insecure and/or weather-exposed locations</p> <p>Lack of response to complaints via other channels</p> <p>Positive experience with speed of delivery</p>	<p>Damaged items</p> <p>Damaged packaging</p>	<p>Negative customer service experiences at physical outlets</p> <p>Positive customer service experiences at physical outlets</p>

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

Characteristics of responses from the companies

PS1 was found to respond to customers in one of four ways: offer assistance and request information, provide information and explanation, provide advice or redirection to another entity, and acknowledge feedback. The offer of assistance was generally accompanied by the request for a complaint number (for customers following up on Twitter regarding an existing complaint) or a tracking number (for customers unable to find information on expected deliveries). Where personal information (e.g., address, email address and phone number) was requested in addition to a tracking number or a complaint number and a follow up was required, the customers were generally requested to send messages via Twitter's DM service or email or call a customer service number. Provision of information and explanation was occasionally accomplished over consecutive tweets due to the platform's 140 character limit. In situations where packages could not be traced or had been returned to senders, customers were advised to contact the senders. Customer service officers were found to have added a personal touch to the responses by signing their names in their responses.

PS2 was found to take a more impersonal approach to customer service provision on the platform. Twitter's DM feature was suggested quite frequently when personal information was required. In tweets from PS3, customers needing assistance were frequently requested to email the company. Customer service officers included their initials in the tweets and also posted general greetings at the start and end of the working day.

Although PS4's customer service officials also signed their tweets, they were found to use more formulaic responses. Customers were primarily requested to email PS4 their tracking numbers together with their personal details. The customer service officers did not make use of sad or smiley face symbols. Customer service officers monitoring the PS5 account generally did not use their names or initials in their tweets. They provided advice and redirection to senders and other parcel delivery service providers including those based in other countries. Longer information provision responses were spread across two or more tweets.

Offers of assistance from PS6 were generally accompanied by requests to provide tracking and contact information by email or DM. Very little information was provided via the public Twitter

platform unless it was standard process related clarification. PS7 also directed consumers to use email or DM when offering assistance. The company provided information of a general nature over the public channel of Twitter. The advice provided generally included links for filing damage or loss claims or redirections to other parcel delivery service providers involved in the shipping and delivery chain. Table 4 presents examples of responses provided by the parcel shipping companies.

Table 4: Examples of responses and general information posted by parcel shipping companies

Companies	Examples of responses to customer queries/complaints and general information
PS1	<p><i>@CustomerA I'd be happy to assist you. Can you please share your tracking number with me and I'll take a look into this?</i></p> <p><i>@CustomerB Our records show your item is awaiting collection from [location], up until the scan advising this yesterday, your item... [followed by] @CustomerB ...was in transit. While items are in transit, they are inside a crate, on a truck and as such can't be scanned</i></p>
PS2	<p><i>CustomerC, we're sorry to hear this. Please DM us your waybill & contact number with full name, we can speak to Customer Service for you</i></p>
PS3	<p><i>@CustomerD To keep your information confidential, email your name, Twitter handle, tracking #, & phone # to [CS email address]</i></p> <p><i>Happy Friday!!!! I am on duty this morning. Let me know how I can help with your PS3 questions today!</i></p> <p><i>Hope everyone had a great Sunday! We will be back tomorrow at [time] to help with your PS3 questions!</i></p>
PS4	<p><i>@CustomerE Please email your name, twitter handle, delivery address, tracking & phone # to [email address]. Thanks</i></p>
PS5	<p><i>@CustomerF I'm sorry it's not arrived ... Signed For items aren't tracked in the network & are only updated once we've delivered ... [followed by] @CustomerF ... the item so we'd be unable to see what's caused the delay. I hope it arrives soon</i></p> <p><i>@CustomerG Hello, I'm sorry but the item is now being dealt with by PS7, please check for delivery details with them</i></p>
PS6	<p><i>"@CustomerH Apologies for the wait. Typically our drivers finish residential deliveries by 7 pm, but can run later due to volume.</i></p>

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

PS7	<p><i>@CustomerI Hi. We apologize for the delay. Can you please follow us and DM us your name, tracking #, and address? Thx.</i></p> <p><i>“@CustomerJ We provide last mile delivery for both PS4 and PS6 for packages in rural areas or areas they don't have as much coverage</i></p>
-----	---

Table 5 summarizes the characteristics of responses from parcel shipping companies engaging in conversations with their customers on Twitter. All the companies were found to be offering assistance and requesting information and providing general information/explanation/advice. Customers were also directed to Twitter’s DM channel or email/telephone numbers when personal details were required for addressing complaints. PS2 and PS3 demonstrated a greater preference for directing customers to other channels in their conversations. PS2 directed customers to DM in 44.3% of their customer service conversations and PS3 directed customers to email in 42.1% of their interactions. In general, the companies varied on their alternative channel preference for private communications (email was the preferred alternative channel for most companies). Customers rarely returned to the platform to acknowledge resolutions of problems directed to other channels, thus limiting the opportunity for companies to use these instances to publicly demonstrate the quality of their efforts. Across the seven companies, customer service agents used different approaches to make the communications seem less impersonal. Names and initials were used to end customer service tweets. Smiley and sad face symbols were used to demonstrate empathy.

Table 5: Characteristics of responses from the parcel shipping companies (preferred alternative channels highlighted in bold)

Characteristics		Parcel shipping companies						
		PS1	PS2	PS3	PS4	PS5	PS6	PS7
Offering assistance, requesting information, providing information /explanation/advice		X	X	X	X	X	X	X
Personalization and empathy (initials/names and/or smiley/sad faces, etc.)		X	X	X	X	X	X	X
Minimum response time (minutes)		0.07	0.10	0.07	0.37	0.05	0.33	0.05
Maximum response time (minutes)		5,760	5,700	5,753	5,582	5,747	5,564	5,730
Median response time (minutes)		1,872	228	246	187	52	63	146
% of conversations in which consumers are directed to other channels	DM	4.6%	44.3%	0.0%	8.0%	6.9%	1.1%	4.3%
	Email	6.3%	1.0%	42.1%	0.4%	8.1%	8.6%	7.5%
	Phone numbers	1.6%	2.5%	0.1%	0.2%	0.5%	0.1%	3.0%

As shown in Table 5, customer service response times varied from a few seconds to about four days for customer tweets received on Friday evenings or over the weekend. Figure 2 shows the relative frequency distribution of responses from companies. PS1 has the highest median response time of 1,872 minutes while PS5 has the lowest median response time of 52 minutes.

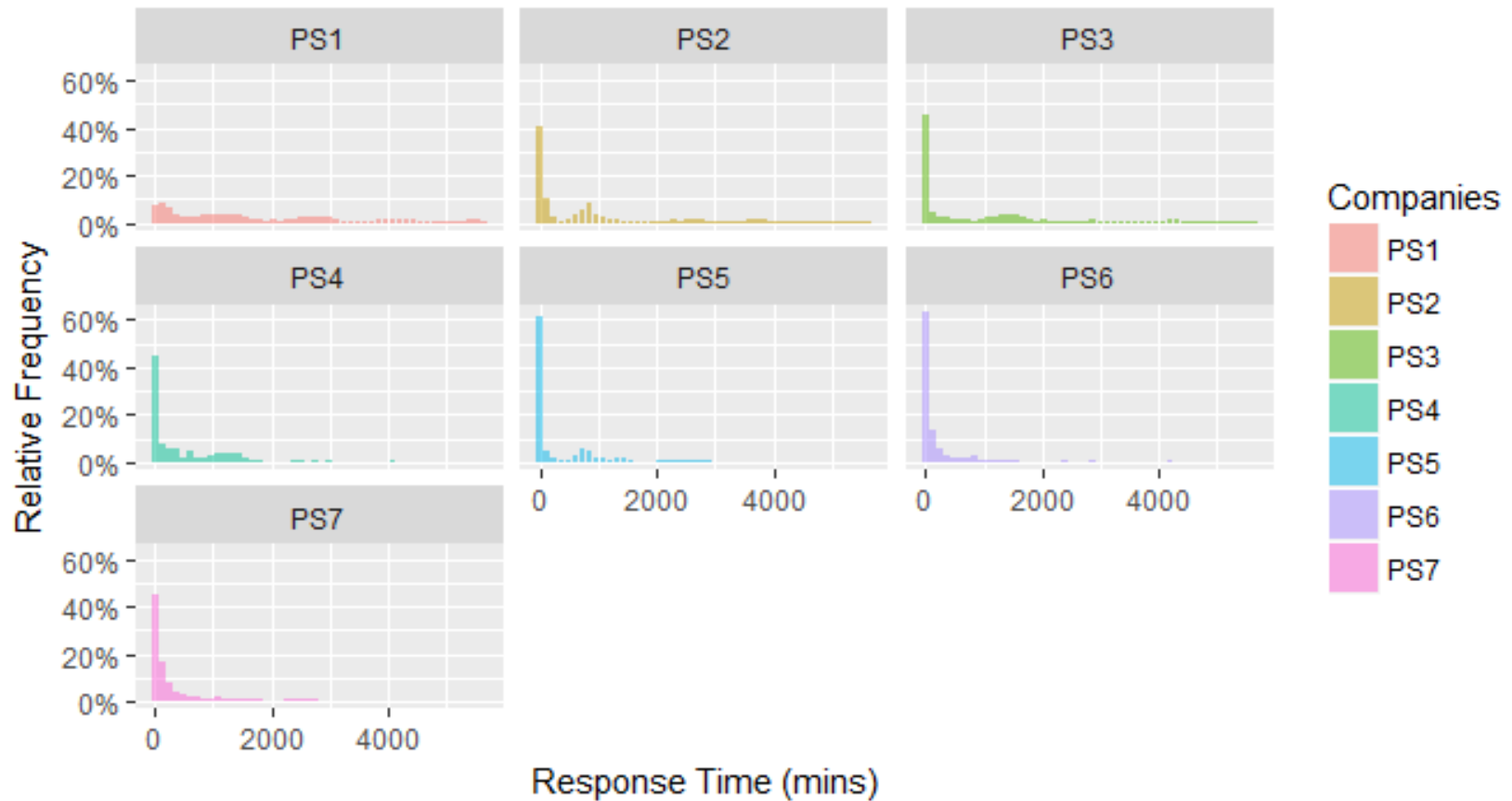


Figure 2: Relative frequency distributions of responses from companies

Discussions on the utilisation of the richness of the medium by users

The study examines the customer service interactions of parcel shipping companies from the perspective of media richness theory which characterises the richness of a medium of communication based on four key criteria: rapidity of mutual feedback, multiple cues, language variety and personal focus (Daft et al., 1987). Based on the cases examined, while customers used multiple cues or language devices to elicit quicker responses, they had little control over the rapidity of responses from companies. For example, PS1 generally had a considerably lower response time than other companies. While the medium is certainly capable of transmitting responses almost instantly, companies appear to be either taking longer to address customer service queries or directing customers quickly onto more traditional channels. Both the rapidity and the extent of mutual feedback are affected by the companies' approach to utilising the features of the platform.

Table 6 illustrates both a shortest-length conversation involving PS6 and a longer conversation involving PS1. The medium allows customers to engage with multiple supply chain partners simultaneously as seen in the example related to PS1. In such cases, rapidity of response can be clearly seen to vary among these partners. In the conversation initiated by a PS1 customer, the relevant retailer, Retailer1 (also addressed in the customer's tweets), was seen to respond before PS1.

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

Table 6: A shortest-length and a longer conversation

Username	Tweet
<i>A shortest-length conversation (two tweets)</i>	
CustomerK	<i>@PS6 disappointed 2nd delivery attempt that I took a morning off for. Tracker says delivery was attempted but doorbell never rang!</i>
PS6	<i>@CustomerK Sorry to hear that! We can address this if you email the tracking# to [email address]. <PS6CSOfficer></i>
<i>A longer conversation (more than two tweets)</i>	
CustomerL	<i>Love that @PS1 didn't deliver one of my parcels from my @RetailerR order & won't allow me to lodge a missing item claim...</i>
CustomerL	<i>@PS1 @RetailerR how is it that they can deliver ONE box & not both at the same time??</i>
RetailerR	<i>@CustomerL @PS1 will get onto it as soon as PS1 is open and either fix or replace your parcel</i>
PS1	<i>@CustomerL I'm really sorry you haven't receive the last order. What is the tracking number? I'll look into it for you <PS1CSOfficer>.</i>
RetailerR	<i>@CustomerL @PS1 Hi CustomerL, if u can DM your full name & order number, we can communicate to our dispatch team 4 the quickest outcome</i>
CustomerL	<i>@RetailerR @PS1 Hi, the second parcel has arrived! Not sure why it wasn't delivered with the 1st box, thanks</i>
PS1	<i>@CustomerL Hi! Glad it arrived! The 2 parcels would come in 2 separate batches <smiley face> <PS1CSOfficer></i>

The Twitter platform facilitates the use of multiple cues in communications as suggested by media richness theory. Customers were found to use a range of different cues when contacting the parcel shipping companies regarding their tracking and delivery related problems. The hashtag is a platform-specific feature that was used by customers to highlight negative experiences with services (e.g., #badservice, #fail). Customers also used the attachment feature to include snapshots of tracking information and damaged packaging. Sad and smiley faces were used to express emotion. The companies made limited use of cues and their messages generally took a minimalist approach to addressing the issues raised in the queries. PS1 made a greater use

of sad and smiley face symbols than other parcel shipping companies while PS4 customer service officers made very limited use of such symbols.

Language variety, the third factor in determining media richness, refers to the range of meaning that can be conveyed with natural language. While companies tended to keep the language used in their tweets simple and focussed on the issues presented, customers often embedded emotion in their use of language. Capitalisation and exclamation marks were used to express emotion. Wording with underlying sarcasm (e.g., *“If someone at @PS6 could tell me how you misplace a package four times in a week, that’d be just dandy [image]”*) and abusive words were also used to refer to add emphasis when reporting negative experiences. These could be used in conjunction with one or more cues (e.g., a snapshot of the parcel tracking page in the case of the above tweet).

The fourth media richness aspect of personal focus is about the usage of personal feelings and emotions to convey a message more fully. In the customers’ use of the Twitter platform the infusion of emotion in messages was achieved through the use of multiple cues and devices available in natural language. While customer service officers made limited use of emotion in their messages, they were found to add sympathetic wording to otherwise standard messages and sign tweets with their names or initials. Table 5 summarises the utilisation of the platform by parcel shipping companies and their customers from a media richness perspective.

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

Table 7: A media richness perspective on parcel shipping related customer service interactions on the Twitter platform

Richness criteria in media richness theory perspective	Customers	Parcel Shipping Companies
Feedback rapidity	Little control over speed of exchange but only occasional complaints about delays	Rapidity of response varied across companies
Multiple cues	Combinations of the following cues: <ul style="list-style-type: none"> - words - hashtags - smiley and sad faces - images 	Messages usually limited to words and punctuation Use of smiley and sad face symbols when receiving positive feedback and complaints respectively
Language variety	Meanings conveyed through natural language used the following devices: <ul style="list-style-type: none"> - exclamation marks - capitalisation of text - text with underlying sarcasm - abusive words Numbers used were related to tracking or expected shipping period	Messages generally kept simple with a direct focus on the issue
Personal focus	Personal emotions were infused in the messages using multiple cues and devices available in natural language	Demonstrated focus on the customer by including sympathetic wording in otherwise simple and formulaic messages Customer service officers often signed their tweets to add an element of personalisation

Implications

Implications for research

The media richness perspective explains the attractiveness of the Twitter platform (and other social media platforms) to consumers. Since companies have largely followed consumers on to social media platforms out of necessity, they tend to be much more cautious in utilising richness of the medium. Media richness theory (Daft et al., 1987; Daft and Lengel, 1986) could be used in conjunction with service-dominant logic (Vargo and Lusch, 2004) in supply chains (Lusch, 2007) to better understand this perceived necessity. With the evolution in consumer expectations, the provision of service has had to broaden to encompass a wider range of media. However, companies have adopted different approaches to media usage. For instance, the choice between email and Twitter's DM feature for customer service communications may depend on the relative richness of the two media from the perspectives of these companies as well as policies requiring email trails for specific types of customer service issues.

The customers' use of hashtags and Twitter handles of relevant retailers and other companies in the parcel shipping chain is likely to be a product of a heightened sense of social presence (Klaus, 2013; Short et al., 1976) in conjunction with the awareness of the richness of the medium. Cultural influences on this sense of social presence and associated behaviour on Twitter may also be explicated in future research.

Implications for practice

The management of rapidly growing volumes of unstructured big data from social media and other platforms poses a significant challenge for companies used to working with traditional structured supply chain data sources. To gain actionable intelligence, companies should also be able to examine large volumes of unstructured data created from their own operations as well as similar data from other companies (competitors and partners). The extraction, storage and analysis of this data require new techniques as well as the combination of multiple new and existing techniques as demonstrated in this paper. This implies the need for improved infrastructure and the development of in-house skills or enlisting the assistance of consultants to

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

develop these capabilities. The strategies of similar companies around interactions that create such data and any reputational consequences also need careful consideration.

The existing tools available to researchers and practitioners focus on the extraction of tweets rather than conversations from Twitter. Given the interactive nature of social media, this only allows for partial insight into the engagement of companies and customers over Twitter. The data extraction tool developed for this study overcomes this problem and facilitates the mining and analysis of tweets from both parcel shipping companies and customers. The data analysis has involved the use of both a commercially available package and scripting within an open source environment. This demonstrates the potential for the use and development of multiple tools and techniques in practice-oriented big data analytics research that may be transferred from a research environment to a business context.

Conclusions

The emergence of communication services, such as Twitter, used by consumers to discuss products and services in public has pushed companies towards establishing a presence on these platforms. In doing so companies are able to promote customer goodwill, mitigate (or redirect) customer complaints and provide a means to assess the company's performance from the consumer perspective. One consequence of a social media presence is the greater visibility that end consumers have into communication/communication gaps between different companies in a supply chain. If not handled effectively, this could potentially have reputational consequences. When discussions about specific service issues are directed off the public channel, few customers return to acknowledge positive outcomes. Companies could potentially track such cases more effectively and acknowledge the outcomes themselves in simple confirmation messages to relevant customers.

References

- Abney, A.K., Pelletier, M.J., Ford, T.-R.S. and Horkey, A.B. (2017), “#IHateYourBrand: adaptive service recovery strategies on Twitter”, *Journal of Services Marketing*, Vol. 31 No. 3, pp. 281–294.
- Aral, S., Dellarocas, C. and Godes, D. (2013), “Social Media and Business Transformation: A Framework for Research”, *Information Systems Research*, Vol. 24 No. 1, pp. 3–13.
- Arya, V., Sharma, P., Singh, A. and De Silva, P.T.M. (2017), “An exploratory study on supply chain analytics applied to spare parts supply chain”, *Benchmarking*, Vol. 24 No. 6, pp. 1571–1580.
- Ashktorab, Z., Brown, C., Nandi, M. and Culotta, A. (2014), “Tweedr: Mining twitter to inform disaster response”, *11th International Conference on Information Systems for Crisis Response and Management*, No. May, pp. 352–356.
- Barratt, M. and Choi, T. (2007), “Mandated RFID and Institutional Responses: Cases of Decentralized Business Units.”, *Production & Operations Management*, Vol. 16 No. 5, pp. 569–585.
- Bhattacharjya, J. and Ellison, A.B. (2015), “Building Business Resilience with Social Media in B2B Environments: The Emergence of Responsive Customer Relationship Management Processes on Twitter”, in Camarinha-Matos, L.M., Bénaben, F. and Picard, W. (Eds.), *16th IFIP Working Conference on Virtual Enterprises PRO-VE 2015, Albi, France, October 5-7, 2015: Proceedings*, Springer, Heidelberg, pp. 170–177.
- Bhattacharjya, J., Ellison, A.B. and Tripathi, S. (2016), “An exploration of logistics-related customer service provision on Twitter: The case of e-retailers”, *International Journal of Physical Distribution & Logistics Management*, Vol. 46 No. 6/7, pp. 659–80.
- Boyd, D. and Crawford, K. (2012), “Critical Questions for Big Data”, *Information, Communication & Society*, Vol. 15 No. 5, pp. 662–679.
- Chae, B. (Kevin). (2015), “Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research”, *International Journal of Production Economics*, Elsevier, Vol. 165, pp. 247–259.
- Culnan, M., McHugh, P.J. and Zubillaga, J.I. (2010), “How large U.S. companies can use

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

- Twitter and other social media to gain business value”, *MIS Quarterly Executive*, Vol. 9 No. 4, pp. 243–259.
- Daft, R.L. and Lengel, R.H. (1986), “Organizational Information Requirements, Media Richness, and Structural Design”, *Management Science*, 32, 5, Vol. 32 No. 5, pp. 554–571.
- Daft, R.L., Lengel, R.H. and Trevino, L.K. (1987), “Message equivocality, media selection, and manager performance: Implications for information systems”, *MIS Quarterly*, Vol. 11 No. 3, pp. 355–366.
- Drake, B. (2013), “Twitter served as a lifeline of information during Hurricane Sandy”, *Pew Research Center*, available at: <http://www.pewresearch.org/fact-tank/2013/10/28/twitter-served-as-a-lifeline-of-information-during-hurricane-sandy/> (accessed 28 September 2017).
- Dutta, D. and Bose, I. (2015), “Managing a Big Data project: The case of Ramco Cements Limited”, *International Journal of Production Economics*, Vol. 165, pp. 293–306.
- Eisenhardt, K.M. (1989), “Building Theories from Case Study Research.”, *Academy of Management Review*, Vol. 14 No. 4, pp. 532–550.
- Fan, W. and Gordon, M.D. (2014), “The Power of Social Media Analytics”, *Communication of the ACM*, Vol. 57 No. 6, pp. 75–81.
- Fan, W., Wallace, L., Rich, S. and Zhang, Z. (2006), “Tapping the Power of Text Mining”, *Communications of the ACM*, Vol. 49 No. 9, pp. 77–82.
- Fan, Y. and Niu, R.H. (2016), “To tweet or not to tweet? Exploring the effectiveness of service recovery strategies using social media”, *International Journal of Operations & Production Management*, Vol. 36 No. 9, pp. 1014–1036.
- Freelon, D. and Karpf, D. (2014), “Of Big Birds and bayonets: Hybrid Twitter interactivity in the 2012 presidential debates”, *Information, Communication & Society*, Vol. 18 No. 4, pp. 396–406.
- Freitag, A.R. and Picherit-Duthler, G. (2008), “A hierarchical model for employee benefits communication based on media richness theory”, *Public Relations Journal*, Vol. 2 No. 1, pp. 1–17.
- Gandomi, A. and Haider, M. (2015), “Beyond the hype: Big data concepts, methods, and analytics”, *International Journal of Information Management*, Vol. 35 No. 2, pp. 137–144.
- Glaser, B.G. and Strauss, A.L. (1967), *The Discovery of Grounded Theory: Strategies for*

- Qualitative Research*, AldineTransaction, New Brunswick.
- Goes, P.B. (2014), “Editor’s Comments: Big Data and IS Research.”, *MIS Quarterly*, Vol. 38 No. 3, pp. iii–viii.
- Gunton, L. and Davis, K. (2012), “Beyond broadcasting: Customer service, community and information experience in the Twittersphere”, *Reference Services Review*, Vol. 40 No. 2, pp. 224–227.
- Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P. and Monostori, L. (2015), “Advanced predictive-analysis-based decision support for collaborative logistics networks”, *Supply Chain Management*, Vol. 20 No. 4, pp. 369–388.
- Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009), “Twitter power: Tweets as electronic word of mouth”, *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2169–2188.
- Kaplan, A.M. and Haenlein, M. (2010), “Users of the World, Unite! The Challenges and Opportunities of Social Media”, *Business Horizons*, Vol. 53, pp. 59–68.
- Kiron, D., Palmer, D., Phillips, A.N. and Kruschwitz, N. (2012), “Social Business: What Are Companies Really Doing?: 2012 Social Business Global Executive Study and Research Project”, *Sloan Management Review*, Vol. 53 No. 4, pp. 1–32.
- Klaus, P. (2013), “The case of Amazon.com: towards a conceptual framework of online customer service experience (OCSE) using the emerging consensus technique (ECT)”, *Journal of Services Marketing*, Vol. 27 No. 6, pp. 443–457.
- Lamba, K. and Singh, S.P. (2017), “Big data in operations and supply chain management: current trends and future perspectives”, *Production Planning & Control*, Vol. 28 No. 11–12, pp. 877–890.
- Lee, C.-H. and Chien, T.-F. (2013), “Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking”, *Journal of Information Science*, Vol. 39, pp. 523–543.
- Lengel, R.H. and Daft, R.L. (1988), “The Selection of Communication Media as an Executive Skill”, *Academy of Management Executive*, Vol. 2 No. 3, pp. 225–232.
- Liau, B.Y. and Tan, P.P. (2014), “Gaining customer knowledge in low cost airlines through text

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

- mining”, *Industrial Management & Data Systems*, Vol. 114 No. 9, pp. 1344–1359.
- Linville, D.L., McGee, S.E. and Hicks, L.K. (2012), “Colleges’ and universities’ use of Twitter: a content analysis”, *Public Relations Review*, Vol. 38 No. 4, pp. 636–638.
- Lusch, R.F. (2007), “Reframing Supply Chain Management : A Service-Dominant Logic Perspective”, *Journal of Supply Chain Management*, Vol. 47 No. 1, pp. 14–18.
- 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*, Vol. 37 No. 1, pp. 37–55.
- McAfee, A. and Brynjolfsson, E. (2012), “Big Data: The Management Revolution”, *Harvard Business Review*, Vol. 90 No. 10, pp. 61–68.
- Miller, C. (2009), “Dell says it has earned \$3 million from Twitter”, *The New York Times*, available at: <https://bits.blogs.nytimes.com/2009/06/12/dell-has-earned-3-million-from-twitter/?mcubz=1> (accessed 27 September 2017).
- Misopoulos, F., Mitic, M., Kapoulas, A. and Karapiperis, C. (2014), “Uncovering customer service experiences with Twitter: the case of airline industry”, *Management Decision*, Vol. 52, pp. 705–723.
- Papadopoulos, T., Gunasekaran, A., Dubey, R., Alay, N., Childe, S.J. and Fosso-Wamba, S. (2017), “The role of Big Data in explaining disaster resilience in supply chains for sustainability”, *Journal of Cleaner Production*, Vol. 142, pp. 1108–1118.
- Park, N., Chung, J.E. and Lee, S. (2012), “Explaining the Use of Text-Based Communication Media: An Examination of Three Theories of Media Use”, Vol. 15 No. 7.
- Richey, R.G., Morgan, T.R., Lindsey-Hall, K. and Adams, F.G. (2016), “A global exploration of Big Data in the supply chain”, *International Journal of Physical Distribution & Logistics Management*, Emerald Group Publishing Limited, Bradford, Vol. 46 No. 8, pp. 710–739.
- Rieder, B. (2013), “Studying Facebook via Data Extraction: The Netvizz Application”, *Proceedings of WebSci '13, the 5th Annual ACM Web Science Conference*, pp. 346–355.
- Sagiroglu, S. and Sinanc, D. (2013), “Big data: A review”, *International Conference on Collaboration Technologies and Systems*, San Diego, CA, pp. 42–47.
- SAS. (2012), *Getting Started with SAS*, SAS Institute Inc, Cary, NC.
- Schneider, C. (2016), “The biggest data challenges that you might not even know you have”,

- IBM, available at: <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/> (accessed 9 September 2017).
- Short, J., Williams, E. and Christie, B. (1976), *The Social Psychology of Telecommunications*, Wiley, New York, NY.
- Statista. (2017), “Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2017 (in millions)”, *Statista*, available at: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> (accessed 27 September 2017).
- Stone, G. and Lightbody, M. (2012), “The Nature and Significance of Listening Skills in Accounting Practice”, *Accounting Education*, Vol. 21 No. June 2014, pp. 363–384.
- Sun, P.C. and Cheng, H.S. (2007), “The design of instructional multimedia in e-Learning: a Media Richness Theory-based approach”, *Computers & Education*, Vol. 49 No. 3, pp. 662–676.
- Swani, K., Brown, B.P. and Milne, G.R. (2014), “Should tweets differ for B2B and B2C? An analysis of Fortune 500 companies’ Twitter communications”, *Industrial Marketing Management*, Elsevier Inc., Vol. 43 No. 5, pp. 873–881.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2011), “Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape”, *Social Science Computer Review*, Vol. 29 No. 4, pp. 402–418.
- Vargo, S.L. and Lusch, R.F. (2004), “Evolving to a New Dominant Logic”, *Journal of Marketing*, Vol. 68 No. January, pp. 1–17.
- Vis, F. (2012), “Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots”, *Digital Journalism*, Vol. 1 No. 1, pp. 27–47.
- Waller, M.A. and Fawcett, S.E. (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*, Vol. 34 No. 4, pp. 249–252.
- Wamba, S.F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015), “How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study”, *International Journal of Production Economics*, Vol. 165, pp. 234–246.

Creation of unstructured big data from customer service: The case of parcel shipping companies on Twitter

Bhattacharjya, Ellison, Pang, and Gezdur

- Wang, G., Gunasekaran, A., Ngai, E.W.T. and Papadopoulos, T. (2016), “Big data analytics in logistics and supply chain management: Certain investigations for research and applications”, *International Journal of Production Economics*, Amsterdam, Vol. 176, p. 98.
- Waters, R.D. and Jamal, J.Y. (2011), “Tweet, tweet, tweet: a content analysis of nonprofit organizations’ Twitter updates”, *Public Relations Review*, Vol. 37 No. 3, pp. 321–324.
- Waters, R.D. and Williams, J.M. (2011), “Squawking, tweeting, cooing, and hooting: analyzing the communication patterns of government agencies on Twitter”, *Journal of Public Affairs*, Vol. 11 No. 4, pp. 353–363.