Journal of Advertising

ISSN: 0091-3367 (Print) 1557-7805 (Online) Journal homepage: http://www.tandfonline.com/loi/ujoa20

An Investigation of Brand-Related User-Generated Content on Twitter

Xia Liu, Alvin C. Burns & Yingjian Hou

Journal of Advertising, 0(0), 1–12

ISSN: 0091-3367 print / 1557-7805 online
DOI: 10.1080/00913367.2017.1297273

Routledge
Taylor & Francis Group

An Investigation of Brand-Related User-Generated Content on Twitter

Xia Liu
Rowan University, Glassboro, New Jersey, USA

Alvin C. Burns
Louisiana State University, Baton Rouge, Louisiana, USA

Yingjian Hou
MMY Digital, Cherry Hill, New Jersey, USA

The big data of user-generated content (UGC) on social media are laden with potential value for brand managers. However, there are many obstacles to using big data to answer brand-management questions. This article presents a framework that automatically derives latent brand topics and classifies brand sentiments. It applies text mining with latent Dirichlet allocation (LDA) and sentiment analysis on 1.7 million unique tweets for 20 brands across five industries: fast food, department store, footwear, electronics, and telecommunications. The framework is used to explore four brand-related questions on Twitter. There are three main findings. First, product, service, and promotions are the dominant topics of interest when consumers interact with brands on Twitter. Second, consumer sentiments toward brands vary within and across industries. Third, separate company-specific analyses of positive and negative tweets generate a more accurate understanding of Twitter users' major brand topics and sentiments. Our findings provide brand managers with actionable insights in targeted advertising, social customer relationship management (CRM), and brand management.

Social media marketing has grown to rival traditional promotion techniques and has become a viable component of integrated marketing communication (Keller 2016; Lipsman et al. 2012). At the same time, consumers are actively engaged with brands on social networks such as Twitter and Facebook and create a huge amount of data about their experiences with brands and products. The big data of user-generated content (UGC) have significant potential business value in targeted adverting (Zhang and Katona 2012), brand communication (De Vries, Gensler, and Leeflang 2012), and customer engagement (Calder, Malthouse, and Maslowska 2016).

However, there are many obstacles to using big data to answer brand-management questions. We list two of the most challenging ones here. First, most big data from social media are textual or graphical in nature. Current research methods in advertising and marketing are not adequate to extract meanings from large-scale unstructured or semiunstructured textual data (Bendle and Wang 2016). Second, researchers need to figure out how to identify brand insights from big data quickly and correctly. For instance, Twitter has 310 million active users per month, and more than 500 million new tweets are posted each day (Statista 2016). Many tweets pertain to products and brands. While it is easy for a marketing manager to decipher the topic and valence of a single tweet, it would take months to read through tweets numbering in the millions.

These challenges demonstrate the need for a framework that can easily transform textual big data into brand insights. Hence, we propose an automatic framework (see Figure 1) that effectively integrates latent Dirichlet allocation (LDA) and sentiment analysis. We explain in detail how we apply this framework on 1.7 million unique tweets for 20 brands across five industries. More specifically, we demonstrate how to answer four brand-related research questions.

Address correspondence to Xia Liu, William G. Rohrer College of Business, Rowan University, 201 Mullica Hill Rd, Glassboro, NJ 08028. E-mail: liul@rowan.edu

Xia Liu (PhD, Louisiana State University) is an assistant professor of marketing, William G. Rohrer College of Business, Rowan University.

Alvin C. Burns (PhD, Indiana University) is a professor of marketing, E.J. Ourso College of Business, Louisiana State University.

Yingjian Hou (MISM, Brigham Young University) is a data scientist at MMY Digital.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujoa.

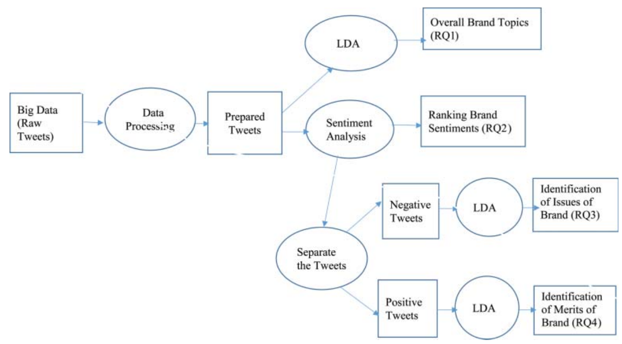

FIG. 1. The framework. LDA = Latent Dirichlet Allocation; RQ = research question.

RQ1: What brand-related topics do consumers discuss on Twitter?

RQ2: What are the rankings of brand sentiments within and across industries?

RQ3: How can we identify specific product and service issues that consumers complain about?

RQ4: How can we identify the merits of products and services that consumers feel good about?

The answers to these four questions reveal several important findings that provide brand managers with actionable insights in targeted advertising, social customer relationship management (CRM) and brand management.

LITERATURE REVIEW

UGC from social media is an important source of big data. There are different types of UGC from a variety of venues, such as tweets or Facebook pages, pictures (Pinterest), blogs, microblogs, and product reviews (Amazon, Yelp). Previous empirical findings show that UGC has significant effects on brand images, purchase intentions, and sales (De Vries, Gensler, and Leeflang 2012; Jin and Phua 2014; Naylor, Lamberton, and West 2012). Most extant business research on UGC uses summary numerical values (Netzer et al. 2012), such as rating scores or the number of views or retweets. Recently, a few marketing and advertising scholars have examined the textual content of UGC, often with text mining or sentiment analysis (see Table 1). Lee and Bradlow (2011) combine marketing methods and text mining to extract attributes of products and determine positioning of brands from product reviews. Okazaki et al. (2014) provide some guidelines on how to classify brand tweets and use a small number of tweets to show their classification methods. Tirunillai and Tellis (2014) extend the LDA model to mine quality dimensions from about 350,000 customer product reviews. Ma, Sun, and Kekre (2015) empirically analyze the relationship between consumer tweets and firm intervention and find that consumers' primary motivation for complaining on Twitter is to ask firms to resolve issues. However, few scholars have used text

TABLE 1
Text Mining and Sentiment Analysis in Marketing Literature

Method	Validation
Lee and Bradlow (2011) analyze Epinions reviews for digital cameras and automatically extract attributes of products and relative positions of brands.	Buying guides, lab survey, and correspondence analysis
Netzer et al. (2012) use messages from Edmunds and a drug forum and extract market structure with text mining and network analysis from messages.	Survey and sales data
Moon, Park, and Seog Kim (2014) use information for 121 movies and 9,380 IMDb reviews and combine clusters of text reviews with rating scores to predict box-office sales.	Increases in the explanatory power of sales models for box office
Schweidel and Moe (2014) analyze 7,565 posts with sentiment analysis.	Linguistic inquiry, word count, and sampled and coded 200 reviews
Tirunillai and Tellis (2014) analyze 350,000 reviews with LDA and extract product dimensions and valence.	Human raters (face validity); *Consumer Reports* (external validity)
Homburg, Ehm, and Artz (2015) analyze 115,000 online messages with sentiment analysis.	Field experiment (internal validity)
Ma, Sun, and Kekre (2015) use a panel data of tweets from 714 customers and classify tweets as negative, neutral, or positive with text mining.	No validation for text mining results
This article presents a framework that automatically derives brand topics and classifies brand sentiments. It explores four brand-related questions on Twitter by applying both LDA and sentiment analysis to 1.7 million tweets.	Benchmarking against ACSI; expertise from industry experts

mining to specifically examine large-scale, brand-related UGC from Twitter.

For this article, we chose to study UGC from Twitter because it is well established as an effective consumer communication vehicle (Lin and Peña 2011). Firms can improve customer relationships by listening to and handling specific customer complaints on Twitter (Ma, Sun, and Kekre 2015). Tweets and retweets are considered to be the online equivalent of word-of-mouth (WOM) communications, sometimes referred to as electronic word of mouth (eWOM; Gruen, Osmonbekov, and Czaplewski 2006), which is an important artifact of Web 2.0 technologies (Li 2011; Liu, Burns, and Hou 2013). Berger and Iyengar (2013) find that communication channels impact the kinds of discussions consumers have about brands and products. In our search of the literature on brand-related UGC from Twitter, at least three streams of research emerged. The first examines why consumers post tweets. By analyzing the content and sentiments in tweets, Jansen et al. (2009) find that consumers use Twitter as a tool for eWOM communication. Using tweets from the top 100 global brands, Araujo, Neijens, and Vliegenthart (2015) find that the retweeting of brand content is largely motivated by informational rather than emotional aspects of the original tweets.

Second, several articles explain how to derive business insights from tweets, but there is scant use of large-scale UGC data. Most current advertising and marketing studies on Twitter UGC have utilized surveys, experiments, or relatively small numbers of tweets. For example, Jin and Phua (2014) conduct two experiments to explore the impact of eWOM on consumer behavior on Twitter. Okazaki et al. (2015) provide steps on how to use opinion mining to analyze the content of 4,000 tweets. Third, several studies explore the impact of Twitter UGC on business outcomes, such as brand reputation, customer satisfaction, and firm performance. Bollen, Mao, and Zeng (2011) use the moods reflected in tweets to predict the performance of stocks. Their approach reaches 87.6% accuracy in predicting the daily ups and downs of Dow Jones Industrial Average closing values.

To sum up, our literature review reveals that although there is some current research about UGC on Twitter in marketing and advertising, few articles have examined the contents of brand-related tweets by using big data. Almost all these articles use small or moderate numbers of tweets; and when analyzing those tweets, human coders were often used. While there is consensus that UGC contains valuable brand-related information, it is extremely challenging for marketers to tame unstructured textual big data. Few have used both text mining and sentiment analysis to illustrate the benefits of social big data.

METHODOLOGY

In this article, we provide a valuable framework to use both LDA and sentiment analysis to study the big data of brand-related UGC from Twitter. To substantiate the applicability of our framework, we will demonstrate that it works on very large data sets of text. We draw our raw data from Twitter, where consumers often post their opinions on brands, products, and services. This social network is very popular and influential, allowing for the collection of large amounts of data. We demonstrate that UGC from Twitter is an important source of big data with potentially invaluable customer insights.

Brand Selection and Data Collection

To generalize our research findings, we tried to make our data set as representative of brand-related tweets in general as possible by choosing five industries and utilizing four brands (companies) in each industry to investigate industry-specific similarities and differences. The five industries were chosen based on the Global Industry Classification Standard (GICS). Our data set includes tweets for fast-food restaurants, department stores, footwear companies, consumer electronics products, and telecommunication carriers. For each industry, we selected one representative brand that met two criteria: first, it had to be owned by an S&P 500 company; second, it had to be closely related to consumers' daily lives. Then we used Hoover's to select the top three competitors for each brand.

We collected around 10 million tweets by using a combination of Twitter's API and customized Web crawling algorithms written in Java. To make the task of data collection manageable, we took advantage of Twitter's unique interactive features. On Twitter, one can place the symbol @ in front of a user name to deliver a tweet directly to the user. So we used "@company" to collect only those tweets which are directly targeted at brand companies' Twitter accounts. Our sample includes tweets for 20 brands over the period of six months from January 1 to June 30, 2015, and additional Comcast tweets for the period between July 1 and December 31, 2015. There were roughly 10 million tweets targeted at the chosen brands during this period, including a large number of retweets. However, we wanted to make sure that the content of the original tweet was used only once for analysis. After several steps of data processing (see the following section for details), about 1.7 million tweets were kept for data analysis. Table 2 presents a summary of tweets for the 20 brands.

The Framework

To tackle big data problems, we propose a framework (see Figure 1) that automatically extracts brand topics and classifies brand sentiments. We emphasize the criticality of data preparation in the first component. The other two key components of the framework are LDA (Blei, Ng, and Jordan 2003) and sentiment analysis (Liu 2012; Pang and Lee 2008). These two machine-learning methods have been used successfully in analyzing big data and are readily available in several

TABLE 2
Summary Tweet Information for Industries and Brands

Industry	Brands and Tweets			
Fast-food restaurant	McDonald's (318,003)	Burger King (122,075)	Wendy's (84,219)	KFC (70,533)
Department store	JCPenney (43,887)	Macy's (184,715)	Sears (33,005)	Kohl's (53,469)
Footwear	Nike (151,437)	New Balance (27,205)	Adidas (57,987)	Puma (23,427)
Electronics	LG (32,230)	Panasonic (4,286)	Samsung (14,857)	Sony (131,264)
Telecommunications	Comcast (261,914)	TWC (57,771)	Dish (35,603)	Cox (20,993)

open-source software packages. As a useful analogy, one might envision this framework as a pipeline with millions of tweets flowing through the framework like water. As the tweets "flow through" this pipeline, "unstructured" big data are organized into useful structures, then changed into summary data that are amenable to traditional marketing analytic tools and finally transformed into brand insights.

Data Processing

The first component of the framework (see Figure 1) is data processing. In current big data research, the data preparation step is often underemphasized. In our approach, we stress the critical role data preparation plays in getting brand insights from text mining and sentiment analysis. Properly preparing and annotating unstructured texts can substantially increase the accuracy of the results of text mining and sentiment analysis. It is important to note that processing of social media textual data is usually time-consuming.

We adopted a combination of methods to remove tweets created by Twitterbots. Twitterbots are automated programs that regularly send out tweets, often in very large volumes to many accounts (Haustein et al. 2016). Malicious bots exploit Twitter's open platform and generate "false UGC," or spam, which poses a significant problem: It distorts the true voices of consumers. Researchers have suggested at least two ways to detect Twitterbot-created tweets: by analyzing the content of the tweets and by examining social graph features of accounts (Chu et al. 2012; Wang 2010). Wang (2010) points out that an account is likely to be a spam account if it sends out duplicate tweets in large volumes at regular intervals or its tweets contain only URL links or "unsolicited replies and mentions." An account's social features, such as the number of followers, the number of friends, and the follower-to-friend ratio, can also be applied in spam detection (Wang 2010).

Second, we filtered out hashtags and URLs. The analysis of metadata tags and links is beyond the scope of this article. Third, we annotated the tweets with tokenization and parts of speech (Manning and Schütze 1999). Annotation is a common natural language processing technique that creates metadata for unstructured texts and improves machine learning results. Fourth, we removed stop words, such as *the, to, is*, and *in*. These are very common in tweets, but their presence adds little value to machine-learning models such as LDA and sentiment analysis. Thus, by removing spam tweets, URLs, nonessential grammatical elements, and stop words, we sought to purify the data so that LDA and sentiment analysis could better reflect the true voices of the consumers.

Latent Dirichlet Allocation

Social big data are highly multidimensional. For instance, a single tweet can mention one or many of the numerous characteristics of a brand or product, ranging from brand reputation to specific product promotions. As such, a textual data matrix is extremely sparse. High dimensionality and sparse data cause traditional multivariate techniques, such as exploratory factor analysis (EFA), to produce unreliable results (Blei, Ng, and Jordan 2003; Tirunillai and Tellis 2014). Furthermore, the high dimensionality and massive volume of information make it impossible for human beings to analyze themes and summarize them within a reasonable amount of time. Because of these issues, we utilize LDA in our framework (see Figure 1) to discover latent brand topics from big data.

LDA, one of the most common probabilistic topic models, is a useful text mining technique for analyzing big data in text format (Tirunillai and Tellis 2014). Blei (2012) provides a nontechnical introduction to LDA and probabilistic topic models. For those interested in more technical details on LDA and clustering, Blei, Ng, and Jordan (2003) provide an advanced introduction to LDA and Xie and Xing (2013) shed some light on the connection between LDA and clustering. According to LDA's data-generation process, every word in a document is generated as follows: in the first step, a topic is randomly selected from the multinomial distribution of topics; in step two, a word is randomly picked from the selected topic, which consists of a multinomial distribution of words (Hong and Davison 2010). Both topic and word distribution have a Dirichlet prior, which explains the "D" in LDA (Hong and Davison 2010). "Latent" implies that topics are hidden in the data sets and can only be inferred from the observable data.

The LDA model requires the analyst to specify how many topics are to be discovered. (Grün and Hornik 2011). When LDA is used for prediction tasks, a maximum likelihood criterion is often used to determine the number of topics (Wallach et al. 2009). This article uses LDA for exploring brand topics, so in our approach we followed the recommendations by Chang et al. (2009), which can be described as informed trial and error. We ran LDA on our data set with different topic numbers. When the number was set at 15, the results appeared to be most concise and insightful. Then, industry experts with extensive domain knowledge were consulted as to the appropriate number of topics. As another technical point, to handle the text mining of more than 1.7 million tweets, we used a very fast LDA implementation called MALLET, written in the programming language Java (McCallum 2002). MALLET implements a fast Gibbs sampling for the approximation of the posterior distribution, so it scales up very well with big data.

The output for LDA in MALLET is a list of identified topics, and each topic has an associated list of top words. Each topic has a coefficient interpreted as the percentage of its presence in the document, and the word list that comprises the topic includes coefficients indicating the probability of each word belonging to the topic. We use Burger King as an example to illustrate the topics, top words, and labeling process.

The first column in Table 3 is not part of the output of LDA. These topic numbers (T0, T1, and so on) are assigned to facilitate discussion. The second column in Table 3 is the proportion of the specific topic in the set of Burger King tweets under analysis. Using T2, for example, the LDA output shows that among all the tweets about Burger King, 2.9% are about topic T2. Figure 2 shows the distribution of all 15 topics. The vertical axis shows the proportion (probability) of the topics.

Column 3 in Table 3 contains the top 10 words with the probability of association with their respective topic in parentheses following each word. As can be seen in Table 3, these words are listed in decreasing order by their probabilities. While they are provided in column 4 of Table 3, the descriptive labels for the topics are not produced by LDA. It is a task for the analyst to devise the descriptive labels based on his or her domain knowledge, assisted by domain experts if necessary. Some labels are obvious, but others are considerably less so. Thus, the labeling process is subjective and involves some judgment. If an analysis team is involved, interjudge agreement procedures might be employed for the more obscure topic labels. After the researchers assign labels to those topics, it might be wise to employ the assistance of domain experts to verify the validity of problematic labels.

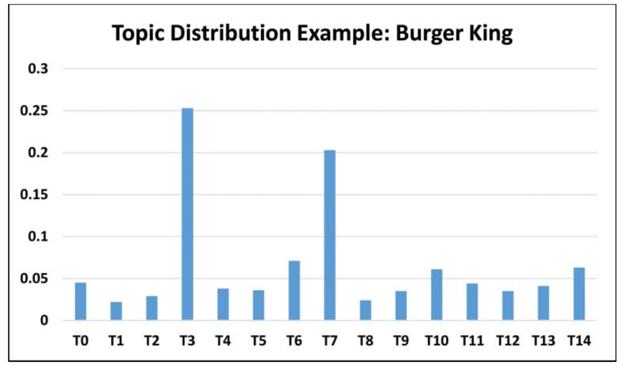

FIG. 2. Burger King: Visualizing topic distribution.

The evaluation criteria for LDA topics depend on the modeling task. It is useful to evaluate the topic models with maximum likelihood measures for prediction tasks (Wallach et al. 2009). However, when the main goal of LDA is to explore the latent topics in available texts, Chang et al. (2009) suggest that the interpretation and evaluation of LDA results depend on the requirements of real-world applications. Thus we use several ways to evaluate LDA results in this article, including comparing our results with trade reports and consulting with industry experts.

Sentiment Analysis

Sentiment analysis, a subfield in natural language processing, is a means to automatically classify texts by valence (Liu 2012; Pang and Lee 2008) and find the authors' views on specific entities (Feldman 2013). Sentiment analysis of tweets has attracted much attention from researchers. For instance, the mood on Twitter has been used to predict the ups and downs of the stock market (Bollen, Mao, and Zeng 2011). The core

TABLE 3
Burger King Topic Proportion, Top Words, and Labels

Number	Proportion	Top 10 Words	Label
T0	.045	order (.080); drive (.040); wrong (.032); home (.031); time (.023); minutes (.020); wait (.016); delivery (.015); cream (.015); line (.015)	Customer service
T1	.022	menu (.101); kids (.086); meals (.048); meal (.045); soda (.035); drinks (.028); keyboard (.024); kid (.020); soft (.012); item (.012)	Products
T2	.029	McDonald's (.145); Wendy's (.066); Taco Bell (.054); shake (.045); red (.040); velvet (.040); Oreo (.039); KFC (.036); Subway (.030); Pizza Hut (.022)	Competitors
T3	.253	fries (.298); chicken (.223); back (.138); nuggets (.031); happy (.021); day (.019); life (.013); forever (.011); strips (.010); fry (.009)	Promotions

technology of sentiment analysis is classification (Pang and Lee 2008). In the context of categorizing tweet valence, we assume that the message in each tweet explicitly expresses the writer's opinion on aspects of the company, such as brands, products, and services. This assumption is largely supported because Twitter generates a good deal of opinion-based UGC. In this study, the classification task is to label the message as neutral, negative, or positive.

Our framework (see Figure 1) gauges consumer sentiment toward a brand and ranks brand sentiments across and within industries. Following the recommendation by Eguchi and Lavrenko (2006), we tallied the numbers of negative, positive, and neutral comments and determined the overall sentiment toward a brand by calculating the proportion of negative tweets relative to the total number of tweets. We used a sentiment analysis toolkit provided by the Stanford CoreNLP software package (Manning et al. 2014) that is considered state-of-the-art open-source software for this purpose. This toolkit uses a deep learning algorithm and achieves a high classification rate.

Data Analysis

Having described the basics of LDA and sentiment analysis, we will demonstrate how we applied our framework to a 20-company big data set of brand-related tweets to answer each of the four research questions.

Result for Research Question 1

What brand-related topics do consumers discuss on Twitter? We ran LDA and generated 15 topics for each of the 20 brands in the data set, resulting in 300 separate topics. One of the critical tasks after running LDA is to give labels to these topics based on research requirements and domain knowledge. Because our purpose here is to compare brands within and across industries, we opted to use a generic list of topic labels: product, service, promotion, competitors, news/trend, shows/games, price, and location. We then categorized the 300 topics using these labels and averaged the incidences of each topic label within each industry. As can be seen in Table 4, about 50% of the topics are product based; the range is a low of 24% for department stores and a high of 71% for electronics companies. Approximately one-third of the topics relate to either service or promotion. Again, the range is considerable, with department stores and telecom companies close to 30% while footwear and electronics are substantially lower in service topic mentions. Inspection of Table 4 reveals some similarities and differences across industries. For example, consumers are more interested in topics regarding products and promotions than in service topics for athletic shoe brands. Promotion topic mentions are highest for department stores and significantly lower for telecom and electronic companies. Across the topics, standard deviations vary substantially. Other topics such as shows/games, price, and location are mentioned with relatively low incidence and are not included in Table 4.

Discussion: Research Question 1

Understanding topics in brand-related UGC is an efficient way for marketers to learn about their customers' interests and preferences. Overall, product, service, and promotion are the three topics that consumers are most interested in. Consumers tend to tweet about various aspects of products, such as quality, features, and innovations. A typical tweet about products is "@BurgerKing Your french fries are nice and fresh!" The topics are identified at the general level in Table 4, but at the brand level the topics should be industry specific. For instance, fast-food tweets categorized as "product" pertained to menu items like burgers, French fries, meals (e.g. breakfast), kids' meals, and so on. For department stores, the product topic involves specific merchandise items, such as clothes, appliances, and jewelry. Smart TVs, tablets, mobile phones, and smartwatches are the most talked about products in tweets for electronics brands.

In addition, consumers are interested in promotions, such as discounts, gift cards, and coupons. When they learn about good deals or find products with reasonable prices, they are excited about them. Users also tend to compare quality and innovation of products among competitor brands in their tweets. For instance, the following tweet clearly favors McDonald's: "@BurgerKing give up u will never be McDonald's." Furthermore, consumers are interested in current news or trends related to the brands, especially for footwear products, which have celebrities such as Kobe Bryant

TABLE 4
User-Generated Content Topics: Industry Similarities and Differences

Topics/Industry	Fast Food	Department Store	Footwear	Telecommunications	Electronics	Average
Product	47.9%	23.5%	55.9%	45.6%	70.6%	48.7% (17.2%)
Service	20.4%	29.6%	1.8%	28.5%	8.0%	17.6% (12.4%)
Promotion	15.7%	24.6%	15.3%	5.0%	5.8%	13.3% (8.1%)
Competitors	4.8%	5.4%	7.7%	5.7%	9.0%	6.5% (1.8%)
News/trends	6.7%	9.2%	9.2%	2.6%	5.1%	6.6% (2.8%)

Note. Standard deviation in parentheses.

and Kanye West as their spokespersons. Celebrity news can generate conversations among consumers.

Result for Research Question 2

What is the relative ranking of brand sentiments within and across industries? We performed sentiment analysis on all the tweets in the data set. As a result, each tweet was classified as negative, neutral, or positive. Then, for each brand, the total numbers for each sentiment category were tabulated, and the percentages of the negative and positive tweets were calculated for each brand and averaged by industry (see Table 5).

As can be seen in Table 5, 16.9% of the tweets were categorized as positive and 47.8% were categorized as negative. The remaining 35.3% were neutral. The 20 brands demonstrate substantial variance across different categories of sentiments. Comcast has the highest negative sentiment (66.7%), while New Balance has the lowest (34.3%). In addition, the percentages of positive, negative, and neutral sentiments are different across the five industries. For instance, the athletic shoes industry has the highest average percentage of positive sentiments (20.2%) while the telecommunications industry has the highest average percentage of negative sentiments (61.1%).

Moreover, consumers expressed different sentiments toward competitor brands. Among Burger King, KFC, McDonald's, and Wendy's within the fast-food industry, consumers hold the strongest positive sentiment toward Wendy's and the strongest negative sentiments toward KFC, though the percentages are similar for all fast-food brands. For the electronics brands of Sony, Samsung, LG, and Panasonic, Panasonic has the highest positive brand sentiments (20.6%), while Sony has the highest negative sentiment (52.4%). Inspection of the industry averages finds that the telecommunications industry receives the greatest percentage of negative tweets (61.1%); the others tend to be 15% to 20% lower in negative tweets.

Discussion: Research Question 2

Notably, the percentage of negative tweets is much higher than that of positive tweets among all brands. To validate that the sentiment analysis percentages are truly reflecting the negative and positive feelings toward the brands on Twitter, we consulted the American Customer Satisfaction Index (ACSI) and obtained the 2015 ACSI ratings for the companies examined in this study. The ACSI is a well-respected proprietary

TABLE 5
Brand Sentiments Across Industries

Industry	Company	Negative	Neutral	Positive	Negative (%)	Positive (%)
Fast-food restaurant	Burger King	38,123	36,633	12,071	43.9	13.9
	KFC	27,105	20,093	7,422	49.6	13.6
	McDonald's	108,686	81,698	35,693	48.1	15.8
	Wendy's	26,493	23,975	10,190	43.7	16.8
	Industry average				**46.3**	**15.0**
Department store	JCPenney	10,589	8,558	3,937	45.9	17.1
	Kohl's	9,103	9,971	6,187	36.0	24.5
	Macy's	21,980	20,155	9,919	42.2	19.1
	Sears	8,537	6,007	2,972	48.7	17.0
	Industry average				**43.2**	**19.4**
Footwear	Adidas	21,870	16,450	7,853	47.4	17.0
	New Balance	7,218	8,859	4,968	34.3	23.6
	Nike	55,980	40,595	22,383	47.1	18.8
	PUMA	6,953	7,534	3,928	37.8	21.3
	Industry average				**41.7**	**20.2**
Telecommunications	Comcast	70,280	24,959	10,050	66.7	9.5
	Cox	10,111	4,531	2,096	60.4	12.5
	Dish	16,015	11,056	2,327	54.5	7.9
	TWC	27,337	11,549	4,757	62.6	10.9
	Industry average				**61.1**	**10.2**
Electronics	LGUS	6,534	4,027	2,663	49.4	20.1
	Panasonic	950	1,079	528	37.2	20.6
	Samsung	5,024	3,423	2,045	47.9	19.5
	Sony	55,329	31,469	18,774	52.4	17.8
	Industry average				**46.7**	**19.5**

weighted average of answers to survey questions spanning perceived value, perceived quality, and customer expectations for brands. The correlation for the positive tweets' percentages and brands' ACSI ratings is .77, while the correlation for negative tweets' percentages and ACSI ratings is −.78. Brands with higher overall satisfaction exhibit proportionately more positive tweets, whereas brands with lower overall satisfaction are associated with proportionately more negative tweets. These correlations corroborate our findings.

In addition, the finding that the percentage of negative sentiments is much higher is consistent with previous research findings. For instance, Richins (1983) shows that dissatisfied customers are more likely to engage in word of mouth than happy customers. Anderson (1998) tests the relationship between customer satisfaction and word of mouth and finds that dissatisfied customers engage in greater word of mouth than satisfied customers. Smith, Fischer, and Yongjian (2012) finds that there are quite a lot of negative sentiments toward some brands on Twitter. However, no previous research has found out how much more likely unhappy customers are to spread word of mouth than happy customers. Our findings provide empirical support that unhappy customers are about three times more likely to engage in negative eWOM than happy customers are to engage in positive eWOM. Obviously, negative word of mouth is detrimental to a brand's image and reputation, so marketers are advised to promptly respond to negative word of mouth to reduce the negative impact on their brands.

Result for Research Question 3

How can we identify the specific issues of products and services that consumers complain about? The focus of research question 3 was on the individual brand. Based on findings from research questions 1 and 2, we selected Comcast as a brand in considerable need of analyzing customer complaints, because it had such a high percentage of negative tweets. To answer research question 3, we ran LDA only on the Comcast tweets that were labeled as negative. The most prevalent negative topic is billing issues (32.5%), and the next is Internet and cable outages issues (12.8%), followed closely by disappointment with Xfinity (12.5%), and then by worst customer service (11.8%). Combining "billing issues," "worst customer service," and the other related topics, more than half of the topics are related to complaints about services.

To this point, we had used the year 2015 as the unit of analysis, but it is possible to use sentiment analysis and LDA with a much shorter time interval and as a monitoring tool. We ran sentiment analysis on Comcast-related tweets from 2015 on a monthly basis (see Figure 3). The left-hand panel of Figure 3 reveals that negative sentiments toward Comcast are relatively constant: There is minimal variation across the 12 months of 2015. To show the temporal dynamics of topics, we can use the analysis result for one topic (e.g., "worst customer service") month by month. When LDA analysis is applied to all Comcast-targeted tweets collected on a monthly basis, "worst customer service" shows up in every month, and it tends to account for between 20% and 30% of all tweets. The right-hand panel of Figure 3 shows the dynamics of the topic "worst customer service."

Discussion: Research Question 3

The analysis of the negative tweets as identified by sentiment analysis reveals the primary reasons for dissatisfaction with Comcast. Apparently, many customers have posted comments and complaints about billing issues in Comcast customer service. Approximately one-third of negative posts are

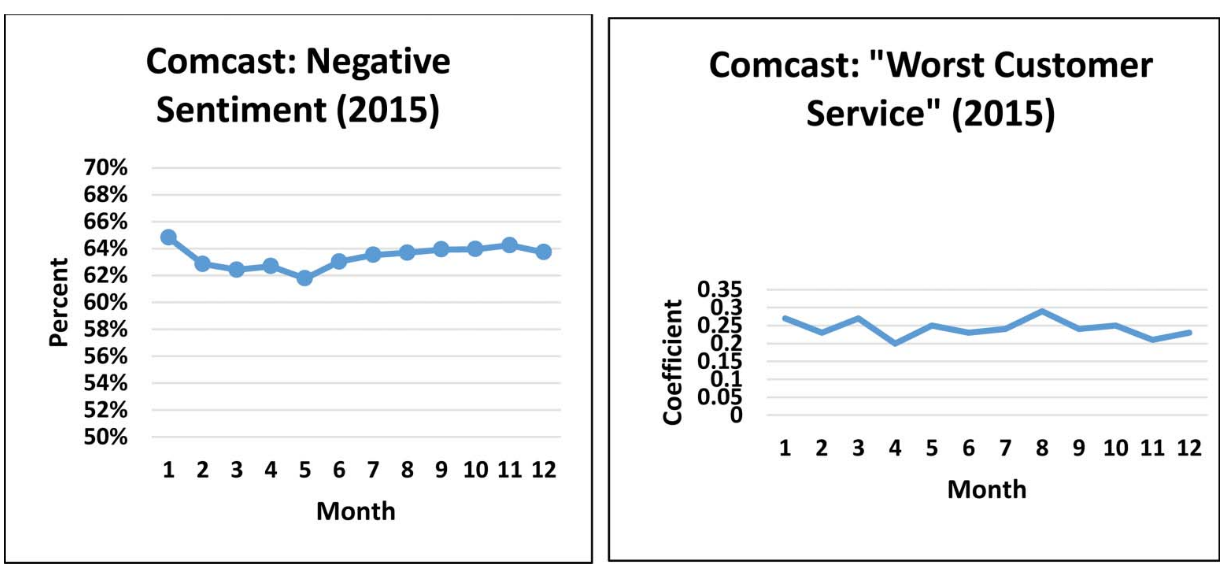

FIG. 3. Comcast in 2015, left: dynamics of negative sentiment; right: dynamics of brand topic.

concerned with the company's charges and prices, which are seen as unreasonably high and prone to appear incorrect. Moreover, customers are frustrated with both Internet and cable issues mostly in the form of outages.

To evaluate the LDA topics for Comcast, we compared our findings against the 2015 ACSI report that pertains to the telecommunication industry. The LDA topics correspond well with the analysis from the report. Second, we consulted three telecommunications industry experts, who confirmed the validity of the findings from the LDA topics. Third, we compared the LDA findings against more than 2,000 formal complaints against Comcast, which were filed with the Federal Communications Commission (FCC) in April, May, and June of 2015. We found that the LDA topics covered all the major categories of complaints. Thus, our approach's findings are consistent with the findings and opinions of independent monitors of the telecommunications industry.

Figure 3 demonstrates how a company can observe its customers' perceptions of its performance. The graph is relatively flat, meaning that Comcast experienced practically no change in its poor customer service reputation during the entire year of 2015. If Comcast had experienced a dramatic deterioration of its customer service, the graph would show a spike or pronounced trend upward, meaning more expressed negativity concerning its customer service. On the other hand, if Comcast implements strategies to bolster or otherwise address its poor customer service, assuming it is successful in this endeavor, the graph should exhibit a corresponding decrease in the proportion of posts about poor customer service. To reiterate the managerial usefulness of our approach: When combined with sentiment analysis to identify negative and positive tweet groups, a company can use trend analysis to diagnose shifts and implement strategies accordingly. Or, after strategies are implemented, longitudinal analysis can be applied to assess their impacts on managerially important topics.

Result for Research Question 4

How can we identify the merits of products and services that consumers feel good about? To answer research question 4, we ran LDA analysis on positive Comcast tweets for the year 2015 to reveal the company's strengths. Using the same approach as used in the negative tweets section, we discovered Comcast's Internet service to be its most significant positive aspect. For the top two positive topics (25.2% and 16.8%), *Internet* is a dominant word. Customer service is the third most prevalent topic (7.2%). Given the low percentage of positive tweets among all Comcast tweets, the numbers for positivity are significantly smaller.

Discussion: Research Question 4

For those Comcast customers who post positive tweets, two aspects of Comcast's Internet service are commonly applauded. Specifically, Comcast's high-speed Internet is often paired with price, charge, or bill references, so it seems that for a portion of Comcast's customer base, there is satisfaction with its Internet speed. Also, Comcast's Internet service is often paired with service-quality references such as "help" and "fix," suggesting that some customers are quite happy with the responsiveness of Comcast technicians when Internet outages and issues arise for them. Last, not all Comcast customers experience poor customer service, and roughly 1 out of 10 positive tweets praise its customer service with words such as *great* and *amazing*. As noted earlier, negative sentiments far exceed positive ones, and from a managerial standpoint the discovery of a group of customers who express their satisfaction with Comcast's customer service on Twitter would seem to be a fertile area for best practices investigation. Investigating positive tweet topics is essentially brand- or company-strengths analysis: It identifies key strengths, provides a measure of their relative sizes, and, when tracked over time, potentially reveals insightful dynamics.

DISCUSSION

In this study, we present a framework that automatically derives brand topics and classifies brand sentiments. This framework effectively employs both LDA and sentiment analysis, which are two highly acclaimed machine-learning methods that are specifically designed to tackle the challenge of big data in textual formats. Using 1.7 million unique tweets, this article illustrates how to follow the framework to transform big data into brand insights and answers four brand-related research questions. It provides brand managers with several important findings regarding the topics and sentiments in brand-related UGC on Twitter.

We provide detailed steps for uncovering and interpreting latent topics gleaned by LDA from brand-related UGC. We also show how to extract dimensions of product and customer service and to investigate specific brand strengths and issues. Furthermore, the temporal development of brand topics is presented. However, although LDA is indeed a useful text mining technique for analyzing big data of electronic documents, its learning curve is steep. To help advertising researchers and professionals who want to benefit from LDA, we demonstrate in great detail how to label and evaluate the derived topics. In addition, we recommend using a software tool called MALLET, which extracts topics quickly from a huge amount of texts.

This article demonstrates how to apply sentiment analysis to quickly and automatically categorize tweets as either positive, negative, or neutral. This classification technique is very useful in helping brand managers understand consumers' overall attitudes toward their companies as expressed on social media. This article provides the steps for labeling each tweet with its valence and ranking sentiments toward brands across and within industries. As a technical note for practical big data analysis, we recommend using a sentiment analysis software tool designed and implemented by the Stanford CoreNLP group. This tool helps us classify our tweets with high accuracy.

Brand-related UGC on Twitter includes diverse topics, such as product, service, promotions, competitors, news, and location. About half of the topics are product based and one-third of the topics relate to either service or promotion. Product, service, and promotions are the predominant consumer concerns when consumers interact with brands on Twitter. By taking advantage of the detailed information in each topic of interest, marketers and advertisers can gain better understanding of customer interests and knowledge and then develop and implement effective marketing and advertising strategies.

The results of sentiment analysis show that brand-related tweets commonly express consumer emotions. Among the 1.7 million tweets, two-thirds are categorized as either negative (47.8%) or positive (16.9%). Furthermore, the different types of sentiments among the 20 brands we investigated vary greatly. The percentage of negative sentiments is much higher than that of positive sentiments; unhappy customers are almost three times more likely to tweet about their grievances than happy customers are to compliment the strengths and merits of the brands on Twitter. More important, the results from LDA analysis of brand topics also confirm that there are consistent complaint topics for specific brands, especially for those in the telecom industry. The large proportion of negative (sometimes even virulent) tweets points to the urgent need to figure out how to properly respond to negativity in social media. By identifying the specific issues that consumers discuss, marketers can acquire the necessary knowledge to improve product and service qualities and better manage social CRM (Malthouse et al. 2013).

Finally, our findings are relevant to targeted advertising on social media. Online targeted advertising assigns more relevant ads based on the content of Web pages and browsing behavior (Zhang and Katona 2012). It has been found to increase purchase intention (Goldfarb and Tucker 2011). The impact of attitudes toward advertising is particularly important to social media Web sites because of the potentially damaging effects of advertising obtrusiveness on those platforms. As Twitter has rolled out its own advertising platform, advertisers can request Twitter to promote their accounts and tweets or present their tweets as trends. The effectiveness of those advertising forms depends heavily on an accurate understanding of Twitter users' major topics of interest and overall sentiments. For instance, a deeper analysis of those tweets that specifically mention brand competitors sometimes indicate consumers' strong desires to switch to competitor products.

LIMITATIONS AND FUTURE DIRECTIONS

Because of the newness of the phenomenon and the dearth of prior research on it, our study is situated as early understanding of brand-related UGC on Twitter. Although we endeavored to maximize the quality of our work with appropriate procedures and validation techniques, there are at least three limitations to point out. First, LDA is a bag-of-words method and does not take into consideration correlations between documents. It is worth noting that there are several important extensions to LDA, such as the author-topic model (Rosen-Zvi et al. 2004) and nested hierarchical Dirichlet process (Paisley et al. 2015). Second, both LDA and sentiment analysis are in need of more effective and accurate means of validation for their findings. As a caution, no matter how "big" it is, big data may give very misleading information (Lazer et al. 2014; Lin, Lucas, and Shmueli 2013). Grimmer and Stewart (2013) warn researchers about the threats to validity in automatic text analysis and point out that quantitative text analysis should be used with convincing validation methods. Third, our data set consists of 20 brands of very large companies. It remains to be seen whether our findings are generalizable to less popular brands, and a much broader representation of industries is needed.

Future researchers can extend our work in at least three ways. First, they can experiment with our framework on other social big data to ascertain brand insights. If our research findings are confirmed with other sources of big data text messages, such as online blogs and Facebook posts, brand management implications will be more generalizable. Second, we need to explore whether big data can help enhance our theoretical understanding of the differential impacts of social media contents. For instance, why does some UGC spread to much larger audiences and become more influential (Berger and Milkman 2012), while other UGC fails to gain such traction? Third, scholars can attempt to help companies better interpret and track social big data to develop innovative products and improve customer service delivery.

REFERENCES

Anderson, Eugene W. (1998), "Customer Satisfaction and Word of Mouth," *Journal of Service Research*, 1 (1), 5–17.

Araujo, Theo, Peter Neijens, and Rens Vliegenthart (2015), "What Motivates Consumers to Re-Tweet Brand Content?," *Journal of Advertising Research*, 55 (3), 284–95.

Bendle, Neil T., and Xin Shane Wang (2016), "Uncovering the Message from the Mess of Big Data," *Business Horizons*, 59 (1), 115–24.

Berger, Jonah, and Raghuram Iyengar (2013), "Communication Channels and Word of Mouth: How the Medium Shapes the Message," *Journal of Consumer Research*, 40 (3), 567–79.

———, and Katherine L Milkman (2012), "What Makes Online Content Viral?," *Journal of Marketing Research*, 49 (2), 192–205.

Blei, David M. (2012), "Probabilistic Topic Models," *Communications of the ACM*, 55 (4), 77–84.

———, Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011), "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2 (1), 1–8.

Calder, Bobby J., Edward C. Malthouse, and Ewa Maslowska (2016), "Brand Marketing, Big Data, and Social Innovation as Future Research Directions for Engagement," *Journal of Marketing Management*, 32 (5–6), 579–85.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L. Boyd-Graber, and David M. Blei (2009), "Reading Tea Leaves: How Humans Interpret Topic Models," in *Advances in Neural Information Processing Systems*, Yoshua Bengio, Dale Schuurmans, John D. Lafferty, Christopher K. I. Williams, and Aron Culotta, eds., La Jolla, CA: NIPS Foundation, pp. 1–9.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia (2012), "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?," *IEEE Transactions on Dependable and Secure Computing*, 9 (6), 811–24.

De Vries, Lisette, Sonja Gensler, and Peter S.H. Leeflang (2012), "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing," *Journal of Interactive Marketing*, 26 (2), 83–91.

Eguchi, Koji, and Victor Lavrenko (2006), "Sentiment Retrieval Using Generative Models," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing: Association for Computational Linguistics*, Dan Jurafsky and Eric Gaussier, eds., Stroudsburg, PA: Association for Computational Linguistics, pp. 345–54.

Feldman, Ronen (2013), "Techniques and Applications for Sentiment Analysis," *Communications of the ACM*, 56 (4), 82–89.

Goldfarb, Avi, and Catherine Tucker (2011), "Online Display Advertising: Targeting and Obtrusiveness," *Marketing Science*, 30 (3), 389–404.

Grimmer, Justin, and Brandon M. Stewart (2013), "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 21, 267–97.

Gruen, Thomas W., Talai Osmonbekov, and Andrew J. Czaplewski (2006), "eWOM: The Impact of Customer-to-Customer Online Know-How Exchange on Customer Value and Loyalty," *Journal of Business Research*, 59 (4), 449–56.

Grün, Bettina, and Kurt Hornik (2011), "Topicmodels: An *R* Package for Fitting Topic Models," *Journal of Statistical Software*, 40 (13), 1–30.

Haustein, Stefanie, Timothy D. Bowman, Kim Holmberg, Andrew Tsou, Cassidy R. Sugimoto, and Vincent Larivière (2016), "Tweets as Impact Indicators: Examining the Implications of Automated 'Bot' Accounts on Twitter," *Journal of the Association for Information Science and Technology*, 67 (1), 232–38.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, 52 (5), 629–41.

Hong, Liangjie, and Brian D. Davison (2010), "Empirical Study of Topic Modeling in Twitter," in *Proceedings of the First Workshop on Social Media Analytics*, Prem Melville, Jure Leskovec, and Foster Provost, eds., New York: Association for Computing Machinery, pp. 80–88.

Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury (2009), "Twitter Power: Tweets as Electronic Word of Mouth," *Journal of the American Society for Information Science and Technology*, 60 (11), 2169–88.

Jin, Seung-A. Annie, and Joe Phua (2014), "Following Celebrities' Tweets about Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers' Source Credibility Perception, Buying Intention, and Social Identification with Celebrities," *Journal of Advertising*, 43 (2), 181–95.

Keller, Kevin Lane (2016), "Unlocking the Power of Integrated Marketing Communications: How Integrated Is Your IMC Program?," *Journal of Advertising*, 45 (3), 286–301.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani (2014), "The Parable of Google Flu: Traps in Big Data Analysis," *Science*, 343 (March), 1203–1205.

Lee, Thomas Y., and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.

Li, Hairong (2011), "The Interactive Web," *Journal of Advertising Research*, 51, 13–34.

Lin, Jhih-Syuan, and Jorge Peña (2011), "Are You Following Me? A Content Analysis of TV Networks' Brand Communication on Twitter," *Journal of Interactive Advertising*, 12 (1), 17–29.

Lin, Mingfeng, Henry C. Lucas, Jr., and Galit Shmueli (2013), "Research Commentary—Too Big to Fail: Large Samples and the *P*-Value Problem," *Information Systems Research*, 24 (4), 906–17.

Lipsman, Andrew, Graham Mudd, Mike Rich, and Sean Bruich (2012), "The Power of 'Like': How Brands Reach (and Influence) Fans through Social-Media Marketing," *Journal of Advertising Research*, 52 (1), 40–52.

Liu, Bing (2012), "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, 5 (1), 1–167.

Liu, Xia, Alvin Burns, and Yingjian Hou (2013), "Comparing Online and In-Store Shopping Behavior towards Luxury Goods," *International Journal of Retail and Distribution Management*, 41 (11/12), 885–900.

Ma, Liye, Baohong Sun, and Sunder Kekre (2015), "The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, 34 (5), 627–45.

Malthouse, Edward C., Michael Haenlein, Bernd Skiera, Egbert Wege, and Michael Zhang (2013), "Managing Customer Relationships in the Social Media Era: Introducing the Social CRM House," *Journal of Interactive Marketing*, 27 (4), 270–80.

Manning, Christopher D., and Hinrich Schütze (1999), *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press.

———, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky (2014), "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Kristina Toutanova and Hua Wu, eds., Stroudsburg, PA: Association for Computational Linguistics, pp. 55–60.

McCallum, Andrew K. (2002), "MALLET: A Machine Learning for Language Toolkit," available at http://mallet.cs.umass.edu/.

Moon, Sangkil, Yoonseo Park, and Yong Seog Kim (2014), "The Impact of Text Product Reviews on Sales," *European Journal of Marketing*, 48 (11/12), 2176–97.

Naylor, Rebecca Walker, Cait Poynor Lamberton, and Patricia M. West (2012), "Beyond the 'Like' Button: The Impact of Mere Virtual Presence on Brand Evaluations and Purchase Intentions in Social Media Settings," *Journal of Marketing*, 76 (6), 105–20.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science*, 31 (3), 521–43.

Okazaki, Shintaro, Ana M. Diaz-Martfn, Mercedes Rozano, and Hector D. Menendez-Benito (2014), "How to Mine Brand Tweets," *International Journal of Market Research*, 56 (4), 467–88.

———, ———, ———, ———, and Jim Jansen (2015), "Using Twitter to Engage with Customers: A Data Mining Approach," *Internet Research*, 25 (3), 416–34.

Paisley, John, Chong Wang, David M. Blei, and Michael I. Jordan (2015), "Nested Hierarchical Dirichlet Processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 256–70.

Pang, Bo, and Lillian Lee (2008), "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, 2 (1–2), 1–135.

Richins, Marsha L. (1983), "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *Journal of Marketing*, 47 (1), 68–78.

Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth (2004), "The Author-Topic Model for Authors and Documents," in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, Christopher Meek and Joseph Halpern, eds., Arlington, VA: AUAI Press, 487–94.

Schweidel, David A., and Wendy W. Moe (2014), "Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.

Smith, Andrew N., Eileen Fischer, and Chen Yongjian (2012), "How Does Brand-Related User-Generated Content Differ across YouTube, Facebook, and Twitter?," *Journal of Interactive Marketing*, 26 (2), 102–13.

Statista (2016), “Number of Monthly Active Twitter Users Worldwide from 1st Quarter 2010 to 1st Quarter 2016 (in Millions),” available at https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

Tirunillai, Seshadri, and Gerard J. Tellis (2014), “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation,” *Journal of Marketing Research*, 51 (4), 463–79.

Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno (2009), “Evaluation Methods for Topic Models,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, New York: Association for Computing Machinery, 1105–12.

Wang, Alex Hai (2010), “Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach,” in *24th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy*, Sara Foresti and Sushil Jajodia, eds., Berlin: Springer-Verlag, 335–42.

Xie, Pengtao, and Eric P. Xing (2013), “Integrating Document Clustering and Topic Modeling,” in *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence*, Ann Nicholson and Padhraic Smyth, eds., Corvallis, OR: AUAI Press, pp. 694–703.

Zhang, Kaifu, and Zsolt Katona (2012), “Contextual Advertising,” *Marketing Science*, 31 (6), 980–94.