Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter

Article *in* International Journal of Information Management · October 2017

DOI: 10.1016/j.ijinfomgt.2017.09.007

3 authors:

Hamid Shirdastian
Concordia University Montreal

Michel Laroche
John Molson School of Business, Concordia University

Marie-Odile Richard
State University of New York Polytechnic Institute

ARTICLE IN PRESS

International Journal of Information Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Information Management

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

Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter

Hamid Shirdastian[a], Michel Laroche[a,*], Marie-Odile Richard[b]

[a] Department of Marketing, John Molson School of Business, Concordia University, 1455 de Maisonneuve West, Montréal, Québec, H3G 1M8, Canada
[b] Department of Business Management, State University of New York Polytechnic Institute, Donovan Hall 1264, 100 Seymour Road, Utica, NY 13502, USA

ARTICLE INFO

Keywords:
Brand sentiment analysis
Brand authenticity
Social media
Big data analytics
Support Vector Machine (SVM)
Latent semantic analysis (LSA)

ABSTRACT

There is a strong interest among academics and practitioners in studying branding issues in the big data era. In this article, we examine the sentiments toward a brand, via brand authenticity, to identify the reasons for positive or negative sentiments on social media. Moreover, in order to increase precision, we investigate sentiment polarity on a five-point scale. From a database containing 2,282,912 English tweets with the keyword 'Starbucks', we use a set of 2204 coded tweets both for analyzing brand authenticity and sentiment polarity. First, we examine the tweets qualitatively to gain insights about brand authenticity sentiments. Then we analyze the data quantitatively to establish a framework in which we predict both the brand authenticity dimensions and their sentiment polarity. Through three qualitative studies, we discuss several tweets from the dataset that can be classified under the *quality commitment, heritage*, *uniqueness,* and *symbolism* categories. Using latent semantic analysis (LSA), we extract the common words in each category. We verify the robustness of previous findings with an in-lab experiment. Results from the support vector machine (SVM), as the quantitative research method, illustrate the effectiveness of the proposed procedure of brand authenticity sentiment analysis. It shows high accuracy for both the brand authenticity dimensions' predictions and their sentiment polarity. We then discuss the theoretical and managerial implications of the studies.

1. Introduction

Social networks (e.g., Facebook), microblogs (e.g., Twitter and Tumblr), blogs (e.g., Blogger and WordPress), social bookmarking (e.g., Delicious and StumbleUpon), and review sites (e.g., Epinions.com, Yelp, TripAdvisor) are considered to be very important in the big data era (Barbier & Liu, 2011; Gandomi & Haider, 2015). Each of these platforms has many users and fans; however, among these platforms, people all around the world are more engaged with social media, more specifically with Twitter and Facebook. Companies invest heavily in developing a social media community not only to strengthen customer–firm relationships, but also to increase the firms' revenues and profits (Kumar, Bezawada, Rishika, Janakiraman, & Kannan, 2016). In recent years, due to the vast potential of social media, and the companies' interests in utilizing it for branding purposes, social media studies have garnered much attention in the branding and information management literatures (Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013; Habibi, Laroche, & Richard, 2014; Hajli, Sims, Zadeh, & Richard, 2017; Hajli, Shanmugam, Papagiannidis, Zahay-Blatz, & Richard, 2017; Laroche, Habibi, & Richard, 2013; Mahrt & Scharkow, 2013; Naylor, Lamberton, & West, 2012). Today, brands could easily be affected if there is a mismatch between consumer expectations and product characteristics from a variety of sources including social media (Hofacker, Malthouse, & Sultan, 2016). This requires firms to constantly monitor brand health, compare it with that of their competitors, and periodically examine customer mindset measures to guide marketing decisions (Ailawadi, Lehmann, & Neslin, 2003).

This research stream contributed to the literature in terms of illustrating the possibility of analyzing branding issues on social media. However, the literature on brand studies on social media is still limited with regards to brand sentiment analysis. Indeed, most of the existing research studied brand sentiments with a three-point scale (positive, negative and neutral). This simple scale is not able to provide more precise information about the polarity of positive or negative attitudes towards a brand (Hu, Koh, & Reddy, 2014). For instance, it could not show differences between those who like/dislike a brand with those who love/hate a brand. Moreover, it does not provide any insight on the reasons for either a positive, negative, or a neutral sentiment towards a brand. Practically speaking, with these kinds of metrics, a brand manager could not determine which brand characteristics lead to better or

* Corresponding author.
E-mail addresses: hamid.shirdastian@concordia.ca (H. Shirdastian), michel.laroche@concordia.ca (M. Laroche), richarm3@sunyit.edu (M.-O. Richard).

http://dx.doi.org/10.1016/j.ijinfomgt.2017.09.007
Received 3 April 2017; Received in revised form 8 August 2017; Accepted 28 September 2017
0268-4012/

worse sentiments. Thus, the analysis should go beyond positive or negative classifications and provide clearer evidence and explanations (Gaspar, Pedro, Panagiotopoulos, & Seibt, 2016).

In this article, given the necessity of monitoring the perceived value of brand authenticity, to protect a popular brand against the heartbreak of *genericide* (Walsh, 2013), we examine sentiments towards brand authenticity on Twitter with a five-point scale. This would fill the current gaps in the literature, and contribute both in terms of better precision, and of providing firms with valuable insights about the way people interact with their brands. We investigate brand authenticity sentiments both qualitatively and quantitatively through four studies. Before conducting the four studies, we establish the descriptions of the associated items of brand authenticity dimensions. Then, we code tweets in reference to the items of brand authenticity construct to be used in study 1, 2, and 4. In study 1, we explore the coded tweets for each of the brand authenticity dimensions in order to not only check the validity of the coding process but also to gain qualitative insights from each tweet. In study 2, we use latent semantic analysis (LSA) to extract the common words in each of the brand authenticity categories. This would again help to check the validity of our tweet coding, and would lead to obtaining a group of common words for each of the brand authenticity dimensions. In study 3, we check the robustness of the coding process and the findings from study 1 and 2 with an in-lab sharing sentiment study. Once we are assured of the coding robustness of studies 1–3, in study 4 we use the coded tweets to train and validate two models through support vector machine (SVM) analyses. This helps us automatically classify/regress tweets regarding both their sentiment polarities and their brand authenticity dimensions. Finally, we use the established models to predict any new tweet.

The research findings on brand authenticity sentiment analysis could facilitate further inquiries into sentiment analysis in branding contexts and also in several related domains, such as e-Word-of-Mouth studies. This research also provides marketing practitioners with a reliable and valid decision support system to evaluate the level of sentiment towards a brand more specifically, and propose appropriate strategies to strengthen it.

2. Literature review

2.1. Brand sentiment analysis

There is a growing interest among marketing researchers in studying branding issues on social media. On one hand, more and more companies engage with social media, carefully broadcast sentiments to entertain consumers, and promote brands (Gopaldas, 2014). On the other hand, customers themselves publicly share how they feel and how they evaluate different specifications of brands on social media, forums, and websites. Research suggests that in forecasting consumer behavior, it is reasonable to assume a relatively stable link between online and offline attitudes (Huberty, 2015). In order to provide more personalized offers to customers, marketers should know the users' emotional states toward different aspects of the brand (Ortigosa, Martín, & Carro, 2014). So far, brand managers, both those who have formally engaged with social media and those who have not, are very interested in getting insights about the effectiveness of their branding campaigns through social media contexts.

Arguing that more and more consumers rely on online contents when they want to get information about brands, He, Wu, Yan, Akula, and Shen (2015) focused on brand sentiments and proposed a social media competitive analytics framework. Fuchs, Höpken, and Lexhagen (2014) studied knowledge generation in the tourism field based on customers searching, booking, and providing feedback in websites and social media. Similarly, Xiang, Schwartz, Gerdes, & Uysal (2015) explored the utility of a big database (www.expedia.com reviews) to better understand the relationship between the hotel guests experiences and their level of satisfaction. Smith, Fischer, and Yongjian (2012) illustrated that the brand sentiments of user-generated content is not predictably different across social media platforms (Facebook, Twitter, and YouTube). Lee and Bradlow (2011) justified an automated marketing research model to uncover the customer voice, using six years of online customer reviews for digital cameras. Extending the existing knowledge about brand sentiments (i.e. Ghiassi, Skinner, & Zimbra, 2013), we examine the sentiments towards a brand via brand authenticity to address the reasons behind positive or negative sentiments. Without this theoretical basis, one could not provide evidence to address questions about different sentiments. In other words, the brand authenticity helps brand managers address why some customers love or hate their brand. Next, we discuss the brand authenticity concept, its associated dimensions and measurement scale.

2.2. Brand authenticity

2.2.1. Brand authenticity concept

According to the Oxford Advanced Learner's *Dictionary*, the word 'authentic' comes from the Greek *"authentikos,"* meaning "principal" and "genuine." The dictionary also provided the following three definitions for "authentic": 'known to be real and genuine and not a copy', 'true and accurate', and 'made to be exactly the same as the original'.

In the literature, there is little congruency among the proposed definitions, leading to interpreting authenticity in different ways (Choi, Ko, Kim, & Mattila, 2015). Beverland and Farrelly (2010) invited researchers to view authenticity as "a socially constructed interpretation of the essence of what is observed rather than properties inherent in an object" (p. 839). Gathering these aspects together, Beverland (2005) suggested that "brand authenticity can be inherent in an object, come from a relation between an object and/or a historical period, an organizational form, or nature, or be given to an object by marketers and consumers. Authenticity can also be true and/or contrived" (p. 1006).

Although authors defined authenticity in different ways, the literature is unanimous regarding its significant effects and advantages in marketing and branding. Eggers, O'Dwyer, Kraus, Vallaster, and Güldenberg (2013) established the linkages among brand authenticity, brand trust, and SME growth from a CEO perspective. Assiouras, Liapati, Kouletsis, and Koniordos (2015) found that brand authenticity predicts brand attachment, while brand attachment influences consumer purchase intentions, willingness to pay more, and to promote the brand. Kadirov (2015) focused on the perceived authenticity gap between national brands and private labels to explore whether and how this factor influences the effect of marketing and manufacturing variables on willingness to pay. Johnson, Thomson, and Jeffrey (2015) believed that if consumers judge brands to be less authentic, the brand is considered to be of lower quality, less socially responsible, and they are less likely to join the corresponding brand community. Arguing that what customers want are memorable experiences rather than products, Gilmore and Pine (2007) suggested that the success of brands, such as Starbucks, no longer depends on its operational prowess or taste superiority; it relies solely on sustaining coffee drinkers' perceptions of the Starbucks experience as being authentic (p. 2).

2.2.2. Brand authenticity dimensions

According to Fournier and Avery (2011), brands have a strategy of openness on social media to establish their authenticity. This strategy led to an increased attention towards the factors which bring perceived authenticity to customers. Here, we conceptualize four brand authenticity dimensions, and then we use exploratory and confirmatory factor analysis (EFA and CFA) to establish them.

First, Gilmore and Pine (2007) believed that while customers previously perceive low quality products as *junks*, today they do not tolerate products with poor quality and call them *fakes* (p. 2). As such, Napoli et al. (2014) established brands' quality commitment as an important factor of brand authenticity. The authors suggest that producing to the most exacting standards and making the products by a master

Table 1
Brand authenticity items and dimensions.

Items	Dimension	Adapted from
Only the finest ingredients/materials are used in the manufacture of Starbucks's products.	Quality commitment	Napoli et al. (2014)
Quality is central to the Starbucks.		
Starbucks's product is made to the most exacting standards, where everything is aimed at improving quality.		
Starbucks's products are manufactured to the most stringent quality standards.		
Artisan skills and customized manufacturing processes used in the production of Starbucks' products.		
Starbucks' products are made by a master craftsman who pays attention to detail and is involved throughout the production process.		
Starbucks has strong connections to a historical time period, a culture and/or a specific region.	Heritage	Napoli et al. (2014)
Starbucks has a strong link to the past, which is still perpetuated and celebrated to this day.		
Starbucks reminds me of a golden age.		
Starbucks exudes a sense of tradition.		
Starbucks reinforces and builds on long-held traditions.		
I miss or remember a regular habit that I had in the past with Starbucks.		
The way Starbucks fulfills its brand promise is very different from competing brands.	Uniqueness	Schallehn et al. (2014)
The way Starbucks fulfills its brand promise is unique.		
Starbucks fulfills its brand promise in a distinct way.		
Starbucks is a brand that adds meaning to people's lives.	Symbolism	Morhart et al.(2015)
Starbucks is a brand that reflects important values that people care about.		
Starbucks is a brand that connects people with their real selves.		
Starbucks is a brand that connects people with what is really important.		
Starbucks is a brand that cares about protecting people's identity.		
Starbucks is a brand that fulfills my life in providing my favorite product.		

craftsman would be signals for customers to perceive *quality commitments*, and therefore authenticity for a brand. As such, we use *quality commitments* as the first dimension of brand authenticity.

Second, brands may have connections to particular places, times and also specific methods of production, designs, and styles, which reflect their concrete referents and cultural associations (Spiggle, Nguyen, & Caravella, 2012). Being called *heritage* in the literature, it could be achieved by using marketing-mix variables that invoke the history of a particular brand, including all its personal and cultural associations (Brown, Kozinets, & Sherry, 2003). Moreover, previous research suggests that through building links to cultural events and also drawing on historical and past events associations, brands could be perceived as authentic (Beverland, 2005). Therefore, in this research, *heritage* is the second corresponding factor for brand authenticity.

Third, customers may believe a brand to be different from competing brands. *Uniqueness* refers to the extent to which customers feel the relative distinction between a brand and its competitors (Netemeyer et al., 2004). Lewis and Bridger (2001) suggested that consumers, by emphasizing brand authenticity, and even when their purchases are not the same, expect unity in terms of product uniqueness and originality. Schallehn, Burmann, and Riley (2014) found brand individuality, the unique way in which a brand fulfills its promise, to be a factor of brand authenticity. Accordingly, we expect that *uniqueness* also forms the brand authenticity construct.

Finally, *symbolism* reflects the symbolic quality of a brand that defines consumers as who they are or who they are not. Morhart, Malär, Guèvremont, Girardin, and Grohmann (2015) found that authentic brands reflect values that customers consider important and may thus help shape who they are. Napoli et al. (2014) expected that symbolism should also form part of the brand authenticity construct, and called for further research. Hence, we select *symbolism* as the fourth dimension of the brand authenticity construct.

In summary, motivated by Napoli et al. (2014)'s call for further studies about the factors of brand authenticity, we conducted factor analyses (both exploratory and confirmatory) to establish the brand authenticity factorial model, using adapted items from the existing literature namely: *quality commitment* (9 items), *heritage* (7 items), *uniqueness* (3 items), and *symbolism* (6 items). Before going through the four main brand authenticity sentiment studies on Twitter, we present the EFA and CFA study, and the findings.

2.2.3. Brand authenticity factorial model

Two hundred undergraduate students (51% males) participated in the study in exchange for extra course credit. Participants rated their level of agreement with 25 items on a seven-point scale about the Starbucks brand (1 = strongly disagree, 7 = strongly agree). Previous research shows different levels of perceived authenticity among Starbucks customers (Thompson, Rindfleisch, & Arsel, 2006). Following Morhart et al. (2015), we eliminated participants who indicated they were unfamiliar, not knowledgeable, or inexperienced with Starbucks (with a mean score of less than 2 on the five-point scale: "Please indicate your level of experience with Starbucks": strongly unfamiliar/strongly familiar, not knowledgeable at all/strongly knowledgeable, strongly inexperienced/strongly experienced, $\alpha = 0.89$), resulting in a final sample size of 188.

Using principal component EFA with oblimin rotation, a four-factor model with eigenvalues greater than one was obtained, *symbolism* (8.75), *quality commitment* (1.93), *uniqueness* (1.26), and *heritage* (1.17). The scree plot examination also confirmed the existence of four major factors. The four factors explained 62.43% of the total variance, 41.65, 9.21, 6.02, and 5.55 respectively, which is in the acceptable range (Hinkin, 1998). We removed three items with low factor loadings on their main factors (< 0.4) and one item with high cross-loading (> 0.2). Then, we conducted a series of CFAs to see which model best fit the data (Appendix A). Comparing the fit indices of different measurement models shows the presence of a four-factor correlated model ($\chi^2(183) = 372.91$, $p < 0.01$, $\chi^2/df = 2.04$, CFI = 0.945, NNFI = 0.938, GFI = 0.924, SRMR = 0.061, RMSEA = 0.072; Bollen, 1989). As a result, we used this model and its corresponding items for further steps in brand authenticity sentiment analysis. Table 1 shows and defines the emerged items of each brand authenticity dimension.

3. Research design

3.1. Twitter as the research platform

To reach our research goals, the brand authenticity sentiments were analyzed from contents generated on Twitter, which is one of the most popular social media platforms. In 2006, two years after Facebook, Twitter introduced itself as a microblogging social media platform and recently surpassed 500 million tweets per day on average (Internet live stats, 2015). On this platform, users are able to share publicly with their followers on a variety of devices up to 140-character texts (tweets). This limited number of characters leads users to express their updates (thoughts, news, emotions and so on) in smaller phrases, which could help content analyzing researchers deal with lesser amounts of unnecessary information (Milstein et al., 2008).

3.2. Starbucks coffee as the brand

Given that our research is about brand-related issues, and that we are going to illustrate the practical effectiveness of our proposed procedure, we had to select a brand for this study. Considering the *Business Week*'s Top 100 Brands and the American Customer Satisfaction Index, and taking into account statistics regarding the most popular brands on Twitter (Social bakers, 2015), we selected Starbucks Coffee.

The Starbucks Corporation, commonly known as Starbucks Coffee, opened its first coffeehouse in Seattle (US) in 1971. Today, it owns a vast coffeehouse chain in many countries (21,536 stores in 64 countries) and it has become the largest coffeehouse company in the world ahead of its UK rival, Costa Coffee (Wikipedia, 2015). Starbucks Coffee joined Twitter in 2006, and at the time of our research, was the second most popular brand on Twitter with 8.3 million followers. It received replies (comments), retweets (forwarding), favorites (likes), updates (tweets) and mentions (listed) much more than even the first popular brand, Samsung Mobile (Twitter counter, 2015). Previous research found Starbucks Twitter account as a place for a combination of customer testimony, complaining, feedback, and Q & A (Jansen, Zhang, Sobel, & Chowdury, 2009). This means that people communicate with Starbucks Coffee more than with other brands. In other words, users use social media to express feelings and attitudes towards this brand (Lee, Han, & Suh, 2014), which is exactly the main focus of our research.

3.3. Data collection

There are two methods for gathering the required data from Twitter. The first, which is free of charge, is provided by Twitter (Twitter Application Programming Interface (API) version 1.0 and recently version 1.1). Through Twitter's streaming API, one can obtain real-time access to tweets, replies, and mentions created by public accounts in a sampled and filtered form (Bifet & Frank, 2010). Driscoll and Walker (2014) suggested using the 'fire hose' of other data providers (e.g. Gnip PowerTrack) as the second method of tweet collection, and illustrated some advantages compared with tweets provided APIs, mostly in terms of tweets volume.

Although some authors (e.g. Boyd & Crawford, 2012) expressed some uncertainty about how tweets are provided through Twitter API, and introduced *rich big data* and *poor big data* terms as a symbol of the differences between those who have access to more and fewer data in the new era, Twitter API is commonly used by scholars and no one reported unexpected results due to using this method (Shi, Rui, & Whinston, 2013). Thus, we collected stream tweets with the 'Starbucks' keyword through Twitter API version 1.1 fire hose; using the Python programming language. We collected the data for two months; starting on July 20; 2015 at 10:00:00 p.m. (GMT) to Sep 20; 2015 at 9:59:59 p.m. (GMT). We obtained 2,988,560 tweets; which contain the keyword *Starbucks*; approximately 50,000 per day. The data was messy but absolutely rich; including information regarding its created time; sender's name; ID; location; language; number of followers; number of friends; number of favorites; number of statuses if the tweet is in reply to others; who sent the tweet if it was a retweet; when the sender joined Twitter and so on (see Appendix B for a sample of received raw data for two tweets).

First, we created a query to filter the tweets based on its language to remove non-English ones. That led to the total number of 2,282,912 English tweets for further analyses. As we were interested in examining user sentiments, we ignored tweets by Starbucks accounts such as @ Starbucks, @Starbucks − OA, @Starbucks Arizona, @Starbucks Australia, @Starbucks Help, @Starbucks Partners, @Starbucks UK, @ Starbucks view, @Starbucks Belleville, @Starbucks Rosental. Removing these tweets from the database gave us 2,242,231 tweets. From the final dataset, we randomly chose 3000 tweets, representative of the whole period,[1] for our qualitative and quantitative studies. Next, we provide details regarding the proposed procedure for brand authenticity sentiment analysis.

4. Brand authenticity sentiment analysis: coding algorithm

While some authors used a prepared lexicon (e.g., Jansen et al., 2009), we did not exploit these available sources because they are not able to properly address our research objectives, due to its novelty in terms of studying brand authenticity sentiments rather than just brand sentiments in general, and also in terms of the use of the five-point scale rather than a three-point one. Thus, two graduate students evaluated individually the tweets through the following algorithm, step by step and tweet by tweet.

1. First, they judged sentiments towards Starbucks on a five-point scale: 2 = strongly positive, 1 = positive, 0 = neutral, −1 = negative, −2 = strongly negative. This evaluation helps us have an overall judgment about the brand sentiment before investigating authenticity sentiment dimensions.
2. Second, following detailed guidelines on how items on each dimension of brand authenticity (*quality commitment*, *heritage*, *uniqueness*, and *symbolism*) should be captured from the tweets, they categorized each tweet under the proper dimension among the four dimensions of brand authenticity. The detailed guidelines were written based on the items and descriptions of each brand authenticity dimension, previously introduced in Table 1. There were three more options for tweets which could not be classified under the four dimensions. For tweets about brand authenticity but not classifiable under one of the above-mentioned dimensions, we created a fifth category, *none of them*. The sixth category was for tweets, which did not provide sufficient information to judge the users' brand authenticity sentiments. Also, for those tweets which were irrelevant to brand authenticity, we created another category.
3. Third, based on the n-gram analysis technique (Sidorov, Velasquez, Stamatatos, Gelbukh, & Chanona-Hernández, 2014), evaluators detected the most meaningful uni-gram (i.e. *#amazing*), bi-gram (i.e. *#Starbucks_addiction*), and tri-gram (*#do_not_like*) and assessed its brand authenticity sentiments towards Starbucks Coffee on a five-point scale (2 = strongly positive, 1 = positive, 0 = neutral, −1 = negative, −2 = strongly negative).

One of the authors trained the coders to work with the platform (Appendix C) and continuously monitored their tasks. In the platform, evaluators read the tweet, decided on its overall sentiments about the brand, chose its class among the seven categories, evaluated its brand authenticity sentiments, and selected its most meaningful unigram, bigram, and trigram from the database just by typing its initials. Evaluations which were not unanimous between the evaluators were discarded, resulting in 2204 tweets incorporated into the final dataset, and showing a high level of inter-rater agreement (0.73). To verify the coding process, one of the authors checked the final coded data, and found the evaluations completely in accordance with the provided guidelines. Our analysis of brand authenticity sentiment consists of two phases, the qualitative (with three studies) and the quantitative (with one study). Fig. 1 presents a picture of the research design and the four studies. In the next section, we provide details for each of the four studies and present the results and discussion.

[1] We verified the 3000 tweets to span the whole period of time and to be relatively the same number as the total number of received tweets per day.

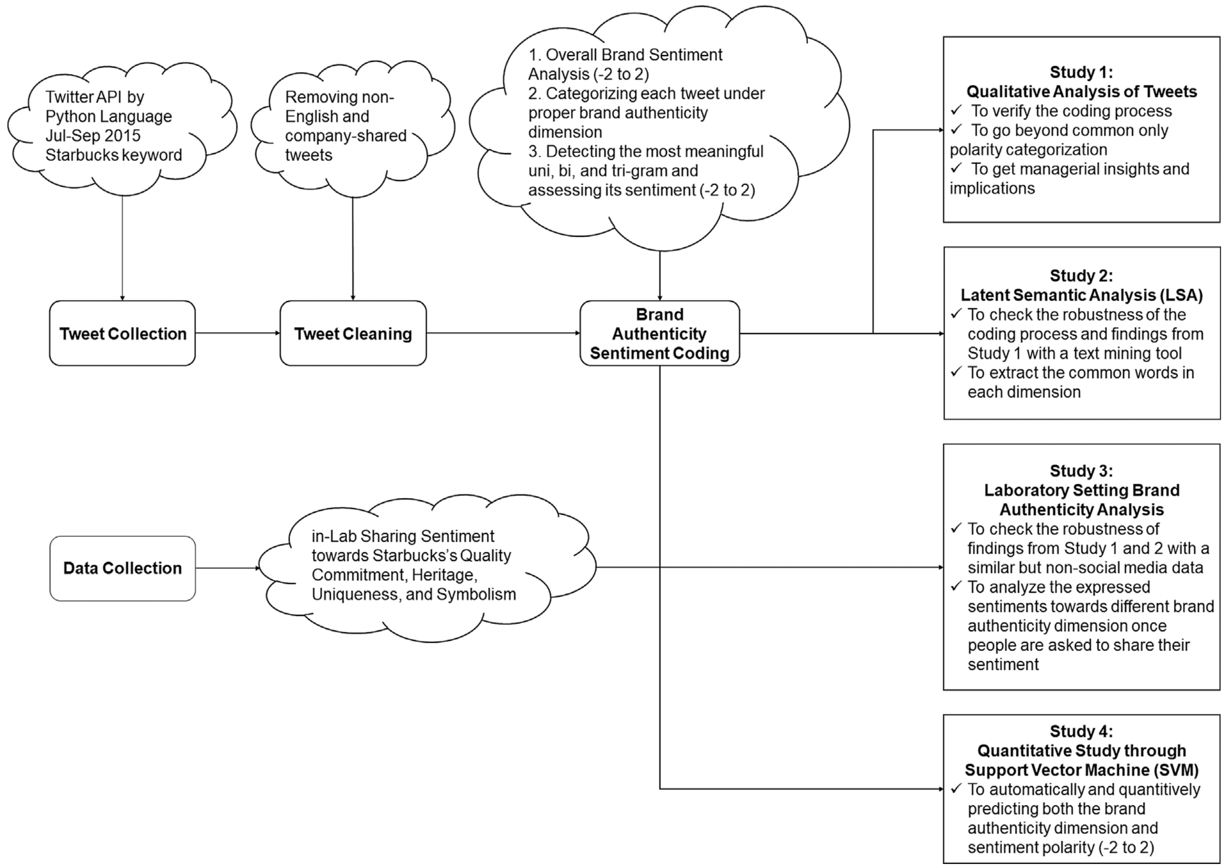

Fig. 1. Overall research design and the four studies.

5. The four studies, results and discussions

5.1. Study 1: qualitative analysis of tweets

In study 1, we explore the dataset to ensure the reliability and validity of the classification and coding process. Once we noticed a great passion towards Pumpkin spice latte (PSL) among users, we decided to focus the study 1 on the first week of fall 2015 (September 1 to September 8), the days around the time Starbucks returned that popular product to the menu. It was in line with previous research, which found strong sentiment towards PSL but didn't investigate its reasons (Ghiassi, Zimbra, & Lee, 2016). In fact, similar to the previous qualitative studies on Twitter (Gaspar et al., 2016), we narrowed down the window to have the chance to review all tweets one by one. This qualitative study helped us go beyond common sentiment analysis by providing many more insights about the brand authenticity sentiments. Although we chose days around the official return of PSL to the Starbucks menu, we did not constrain ourselves to PSL related tweets. In other words, we analyzed all tweets in that time frame, totaling 328 tweets, even those which belonged to other brand aspects or products. Here, we present and discuss some tweets by categorizing them in their appropriate brand authenticity dimensions.

5.1.1. Quality commitment

Building on the emerged items of the quality commitment dimension (Table 1), tweets which deal with product ingredients, improving quality, quality standards, artisan skills, customized manufacturing processes, and involving a master craftsman (barista) are categorized under the quality commitment dimension. Customers usually share on social media their positive or negative evaluations of Starbucks's quality commitment in-store through their smart phones or tablets. They also talk about the ingredients and how they taste. At the time of data collection, Starbucks was promoting the PSL by emphasizing its coming back "with real pumpkin." This led to questioning PSL quality in previous years, and might be the reason for negative comments regarding the new PSL taste. However, PSL still had its own fans, sharing their positive sentiment about their first ever PSL or their first PSL of the year. Considering both the positive and negative tweets, one could argue about the importance of quality consistency over time. This is consistent with the finding obtained by Kim and Sullivan (1998) suggesting that if quality varies and does not match customer expectations, a consumer will be more likely to switch to another product or brand.

There were also several tweets complaining about service and/or product quality. In most of the complaint cases, Starbucks Help (@starbuckshelp) came back to the customers in a timely manner, and asked for more details regarding the complaint to follow up on it. This observation is also reported about Amazon Help, suggesting the importance of having a proper communication with customers through online platforms (Ibrahim, Wang, & Bourne, 2017). However, a proficient store manager or barista could handle the case even better. This may not only impede and prevent the decrease in perceived quality commitment, but may also bring positive tweets. Table 2 provides several tweets regarding customers' perceived quality commitments.

5.1.2. Heritage

Based on the descriptions about the heritage dimension (Table 1), tweets which show Starbucks's connection to a historical period in time, culture and/or a specific region, or tweets which remind the customers of a golden age are categorized under the heritage dimension. Tweets in this category mostly came from shares about Seattle, Washington, where Starbucks was founded in 1971, or about the Seahawks, a professional American football franchise based in Seattle. In line with another heritage item, we found that customers also tweet about missing a regular habit they had in the past in Starbucks. They also share

Table 2
Examples of quality commitment tweets.

Description	Tweet
Instantly sharing sentiments	*"Tried my first ever caramel macchiato from Starbucks. Yummy considering I don't usually like coffee."*
	"Congratulations Starbucks. This is the weirdest shet I've ever drank ever. #tasteslikepoison… https://t.co/3FrlKNoh1T*"*
Talking about the ingredients	*"White chocolate, coconut and lime cookie from Starbucks is the best thing in the entire world…"*
Promoting PSL by "with real pumpkin"	*"The Pumpkin Spice Latte is back at Starbucks! Now made with actual pumpkin in it! Wait, what have I been drinking all these years?"*
Negative sentiments towards PSL	*"This year's #PSL recipe doesn't taste as good as last years! Sorry, @Starbucks! At least this saves me a lot of money"*
	"Pumpkin scones @Starbucks are different this season. No taste. Zero. #ButIAteWholeThing"
	"I don't get why Starbucks had to start using real pumpkin this year …it was way better last year"
Positive sentiments towards PSL	*"@Starbucks just tried the new #PSL. Love the new recipe!* http://t.co/psSArukYuJ*"*
	"Had the #PSL today for the first time ohh my god it's amazing now I know why everyone likes it haha @Starbucks #tobeapartner"
	"Just had my first @Starbucks PSL of the year. Decaf and skimmed milk mind due to remaining a good girl. Still tastes amazing"
Complains about Starbucks services	*"TERRIBLE SERVICE…don't ask for personal drink order…they don't read them… #disappointed (@ Starbucks)* https://t.co/C93Rd34kMu*"*
	"I'm not one to complain, but I just found a finger/toe nail in my @Starbucks this morning. So gross."
	"So this happened at @Starbucks this morning. I mobile ordered, it wasn't ready and then I get this. #disappointed http://t.co/yyEGutNj82*"*
Taking care of complains	*"This lady at Starbucks was so boss at taking care of a customer complaint."*

Table 3
Examples of heritage tweets.

Description	Tweet
Starbucks connection to Seattle and its football team	*"Starbucks first store at Pike Place #Starbucks #Seattle #coffee @Starbucks #Washington #sightseeing…* https://t.co/yrxuVIWUdI*"*
	"Seattle Seahawks Starbucks. https://t.co/CCLF7CVO83"
Missing or remembering a past connection with the Starbucks	*"I've been back in LA for 36 hours and have already stumbled upon a live reading in the back of a Starbucks. I think I maybe missed this?"*
	"Working from Starbucks, sipping a London fog, cloudy day… makes me feel like I'm in college again which is kinda tight"

contacting with Starbucks highlights from their past memories. These findings are consistent with previous research in the field of tourism that suggests heritage creation from "surviving memories, artifacts, and sites of the past to serve contemporary demand (Chhabra, Healy, & Sills, 2003)". Table 3 shows a few examples of these kinds of tweets.

While we expected more tweets about brand heritage, the coding process resulted in 66 tweets. This may be because users are more engaged with present events for a brand, like Starbucks, which they are in touch daily. It may also be because Starbucks itself has not promoted its heritage features for millennials. We validate these explanations with the in-lab setting data in study 3.

5.1.3. Uniqueness

Following the guidelines about the uniqueness dimension (Table 1), we classified the tweets concerning the ways Starbucks is fulfilling its brand promises in comparison with competing brands under the uniqueness dimension. Practically, we examined if customers perceive Starbucks services and products in a very different, unique or distinct way from competitors. Customers usually mention the competing brand, use the *than* conjunction to compare, or express their attitudes and reasons for the comparison. In many cases, when users reveal their product desires, they highlight their sentiments towards Starbucks uniqueness by using the phrase *from Starbucks.* For example, a customer orders any frozen drink as long as it comes from Starbucks and not from anywhere else. This specification is also observed in other tweets about requiring a specific drink and cake from Starbucks, i.e. a Frappuccino with chocolate cake. These findings correspond with empirical research which supports the association of brand uniqueness and being loyal to a brand (Valta, 2013).

The other stream of tweets in this category comes from Starbucks's pricing strategy. In fact, Starbucks differentiates itself from competing brands by its ability to safely offer specialty drinks at higher prices (Miller, 2009). Research also shows that perceived brand uniqueness is associated with willingness to pay premium prices (Netemeyer et al., 2004). However, price-sensitive customers might not appreciate this pricing strategy. That is why we found a bunch of tweets comparing Starbucks prices with that of competing brands, or stating their intentions to go elsewhere as they could not afford Starbucks prices. The last aspect of Starbucks brand uniqueness is directly related to the way this brand exploits the power of social media. To influence consumers, positive e-WOM is considered a powerful marketing medium for companies (Jansen et al., 2009). That is exactly what Starbucks uses from its popularity on social media. The way that Starbucks customers share their positive sentiments on social media, not only promotes the brand uniqueness itself, but also attracts other users to purchase Starbucks products. Table 4 contains several tweets about different sentiments about Starbucks's perceived uniqueness.

5.1.4. Symbolism

In line with our conceptualization of symbolism as one of the brand authenticity dimensions, we explored our dataset to catch tweets that are associated with symbolism items (Table 1). These tweets suggest that Starbucks is/is not adding meaning to people's lives, reflecting important values that people care about, connecting people with their real selves, connecting people with what is really important, caring about protecting people's identity, and fulfilling one's life in providing his/her favorite products. In essence, customers purchase Starbucks beverages and foods not only for consuming them, but for adding meaning to their lives. We observed several tweets that appreciate a Starbucks product or the Starbucks itself because of wishing for a better life. This notion is supported by previous research suggesting that customers are actively looking for authenticity to make their lives meaningful (Beverland & Farrelly, 2010).

Based on identity theory (Stryker & Burke, 2000), one could expect that customers care about the way they are named and called by a salesperson. Previous research reports how students interpret misspelling or mispronouncing by cashiers or workers at an on-campus Starbucks store (Kanemoto & Dai, 2015). In accordance with that research, we found several customers wishing to be known by baristas or complaining about being named wrongly.

As the last stream of tweets in this dimension, we noticed several calls to boycott Starbucks because they believe that Starbucks does not

Table 4
Examples of uniqueness tweets.

Description	Tweet
Comparing Starbucks with competing brands	*"@TheRealPSL @Starbucks I don't need one at @DunkinDonuts I enjoyed my nice Iced pumpkin latte & greeted w/[=with] awesome smiles #mademyday"* *"@DutchBrothers is FAR superior to Starbucks anyway! Better coffee and much friendlier, unpretentious customer service!* https://t.co/5N7eXPal1O" *"RT @JuicyZac: Tim Hortons over Starbucks anytime anywhere any day"* *"Nothing better than Starbucks and a shower … & music … Hella feeling myself right now"* *"RT @PreventionMag: This homemade pumpkin spice latte is healthier, tastier, and even quicker than Starbucks:* http://t.co/FQjZ20WkX6"
Only from Starbucks	*"A frozen drink from Starbucks sounds great right now."* *"If someone can bring me a vanilla bean frappe with 2 chocolate cake pops from Starbucks, that'd be great"*
Starbucks price	*"When you're too broke for Starbucks so you go to McDonalds for ice coffee"* *"Might have to take out a loan to be able to afford Starbucks"*
Starbucks popularity on social media	*"Starbucks don't need ads cuz [=because] teenage girls Instagram accounts are their ads…"* *"Social media users go nuts for return of Starbucks pumpkin spice latte #Durham* http://t.co/S0m0bUE3Fn" *"Genius marketing by @Starbucks cuz [=because] I'm gonna go & get a pumpkin spice latte 1st thing in the morning* http://t.co/jyex1lPmID"

Table 5
Examples of symbolism tweets.

Description	Tweet
Starbucks & a better life	*"RT @KaileyDaggitt: Life is so much better when Starbucks has pumpkin spice."* *"When Starbucks is life. @jiffpom* http://t.co/69mfuemTbf" *"#PSL is back at Starbucks…. Fall weather just needs to come, my life will be complete"*
Starbucks & calling customers' name	*"Life goal is the Starbucks worker to know my name and order"* *"Starbucks has really outdone themselves with misspelling my name…#starbucksfail #notmyname #fail…* https://t.co/SNIbVd5sJ3"
Not taking care of customers' values & calling for Starbucks's boycott	*"Boycott Starbucks… Let your wallet speak your values and faith. Time for Starbucks to learn a business & lesson* http://t.co/Ez8mC1Pd4G" *"THIS IS ASININE IT'S TIME 2 BOYCOTT. Starbucks CEO: If You Support Traditional Marriage We Don't Want Your Business* http://t.co/0yDC5ACuqo" *"@D_Rob317 @donna_jacobsen @FreeAmerican100 @Starbucks I'm asking conservative Christians to boycott Starbucks. As the CEO said"* *"RT @estera8763: AFTER 16 YEARS OF WHITE CHOCOLATE MOCHA AND WARM BUTTERED CROISSANT, I AM DONE. #BOYCOTT* https://t.co/Wttzivxqm1"

reflects values they care about. For example, in response to Starbucks CEO's statement against shareholders who support traditional marriage, we found many tweets containing invitations to boycott Starbucks. The tweets show how consumers react to a brand once they perceive their values are not being respected. Since it is well documented that boycotts devalue customer perceptions of a brand (Klein, Smith, & John, 2004), managers should attempt to avoid boycott campaigns. In Table 5, a sample of tweets about different items of symbolism is available.

5.1.5. None of them

In addition to the four dimensions of brand authenticity – *quality commitment, heritage, uniqueness,* and *symbolism* – we added an alternative option for our coders to choose from. When they found that a tweet is related to brand authenticity but the four dimensions are not able to capture it, they chose the *none of them* category. They labeled 69 tweets in the *none of them* category which means 3.1% of all tweets did not belong to any of the four dimensions. It is similar to results of from an average variance extracted (AVE), which is not 100% (Fornell & Larcker, 1981).

Tweets in this category are covering a wide range of issues, from difficulties in finding stores, stores being crowded or having long lines, to appreciating Starbucks because of product unrelated issues. Here are some examples of these tweets in Table 6.

Table 6
Examples of none of them tweets.

Description	Tweet
Unassociated tweets with the four dimensions	*"So, @Starbucks why can't I login to the website to manage my card?!"* *"When Starbucks is so crowdedyou don't even want to go in"* *"There are like no Starbucks anywhere"* *"I'd like to thank Starbucks for calling 911 and giving me all the stuff to give to the doctor on scene! You guys rock!"* *"Loving the new look of our favorite Starbucks! @Starbucksph"*

5.1.6. Irrelevant to brand authenticity

Since the research object was to analyze the tweets' brand authenticity sentiments, we grouped the tweets which were not about brand authenticity in the *irrelevant to brand authenticity* category. This helped us have a lexicon for tweets which could be ignored in sentiment analysis. The focus of these tweets was not about Starbucks while they talk about it. There were also some tweets in this category where Starbucks is introduced as a prize or a Starbucks gift card is offered. In Table 7, a few examples of irrelevant to brand authenticity tweets are shown.

5.1.7. Not sufficient information to judge brand authenticity

The last category consists of tweets where coders could not judge the brand authenticity sentiments due to a lack of sufficient information or cues. This might come from having few words or it might be because of our evaluation outside the context of the tweets. For instance, in the tweets mentioned in Table 8, while we know that the users are sharing some sort of feelings towards Starbucks, we do not have cues to help us categorize them under a proper brand authenticity dimension.

Table 7
Examples of Irrelevant to brand authenticity tweets.

Description	Tweet
Not about Starbucks brand	*"When your best friend knows your exact order at Starbucks and brings it to you after a long night out. Thanks @_TheChosenJuan_ #TrueMVP"* *"Cam fell asleep in a chair in Starbucks and twitched so hard he woke himself up lmao"*
Starbucks as a prize, Starbucks coupon and gift cards	*"2 Winners for Starbucks! Also join online community to earn Amazon! At @CrunchyBchMama #giveaway!* http://t.co/SkYg72TTBm" *"L #Gifts #Cards New 2015 Starbucks Coffee Taiwan Gift Card FRAPPUCCINO… #Coupon #BuyItNow…"*

Table 8
Examples of not sufficient information to judge brand authenticity tweets.

Description	Tweet
Not sufficient information	*"@KIRSTlN I love you more than Ariana and Nathan loves Starbucks"* *"RT @Wonder_Buns: @LGP4july @Just_a_Texan @helpsosme same here! I tell everyone I know that @Starbucks is a piece of crap company"* *"RT @hansclm49: That s easily fixed. Bye Starbucks!* https://t.co/xA582BnXEt" *"At Starbucks* https://t.co/y3Js1QvJb7*"*

5.2. Study 2: latent semantic analysis

To support the coding process and also the study 1 analysis, we approached the whole data with a text mining tool. In order to achieve these aims, finding the common words in each category would be beneficial. In fact, if the results from the coding process and the study 1 are legitimate, the emerged words in each category should be meaningful and relevant. Therefore, in order to establish the key common themes of the tweets, we used latent semantic analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) to determine associated words in each of the seven categories. This helped us support the coding process and also the findings from the qualitative study using another perspective.

Evangelopoulos (2011) compared Latent Dirichlet Allocation (LDA) and LSA methods, and found LSA results to be more accurate. LSA, instead of counting the frequency of words, co-occurrences, or simple correlations in usage, creates a new semantic space where deeper relations among words/documents are inferred (Ahmad & Laroche, 2017). Once we conducted our LSA, we set the singular value decomposition (SVD) to seven. Here, we report and discuss the extracted words in each category.

For the *quality commitment* dimension, we found that the established common words are related to perceived quality issues, such as: great, make, crave, taste, really, good, love, pumpkin, coffee, and like. These words are representative of the tweets, which we discussed earlier in the section regarding the quality issue. Similarly, the words such as: day, life, basic, boycott, fall, white_girl, first, back, pumpkin_fall, and name remind us of the positive or negative attitudes towards Starbucks in terms of *symbolism*; adding meaning to its customer lives, or reflecting important values that customers care about. Likewise, the common words in the *uniqueness* dimension, present its singularity by for example coffee comparison and admitting addiction. Those words are: better, addict, just, coffee, girl. For the *heritage* dimension, as expected, we found Seattle, Seahawks, miss, and remember as the frequent words. Finally, results of LSA were also consistent with those of qualitative analysis for the tweets which were *Irrelevant to brand authenticity*. Supporting the discussed tweets earlier in this section, we found gift, coupon, and eBay as the common words for this dimension. Table 9 shows the most common words in each dimension. Due to the wide range of tweets in the *not sufficient information to judge brand authenticity* and *none of them* categories, we did not get common words for these two categories.

Table 9
Most common words in each dimension.

Quality commitment	Symbolism	Uniqueness	Heritage	Irrelevant to brand authenticity
great	day	better	Seattle	gift
make	life	addict	Seahawks	coupon
crave	basic	just	miss	eBay
taste	boycott	coffee	remember	
really	fall	girl		
good	white_girl			
love	first			
pumpkin	back			
coffee	pumpkin_fall			
like	name			

5.3. Study 3: laboratory setting brand authenticity analysis

As the last step in the qualitative phase, we validated our findings from study 1 and 2 with data from a laboratory setting. 180 undergraduate students (52% female) participated in study 3 in exchange for having a chance of winning one of six Amazon's $25 gift cards. At the beginning of each of the four sessions, one of the authors asked the participants to think about their own experience with Starbucks, and answer four questions about the brand. Then, they saw following scenario on the screen:

> *"Suppose you want to share your sentiments and attitudes (either very positive, positive, negative, very negative or neutral) towards Starbucks in social media (i.e. Twitter, Facebook, Instagram, or so on). What are the words and/or phrases, which you might use about each of the following situations? (There is no right or wrong answer! Also, do not worry about possible spelling errors!)*

1. *You want to share something about Starbucks's product quality:*
2. *You want to share something about Starbucks's products, values, behaviors, and so on that adds meaning to your life:*
3. *You want to share something about Starbucks's ordinary/unique characteristics:*
4. *You want to share something about Starbucks's connection to a historical period in time, culture and/or specific region:"*

We asked demographic questions along with questions regarding their level of social media usage, and the social media platforms in which they have an account, which they use at least once a day, and at least once a week. Finally they were thanked and debriefed. Here, we present and discuss the findings from this study.

Our sample had 3.76 social media accounts (including Facebook, Instagram, Twitter, LinkedIn, Google+, Pinterest, Tumblr, and Snapchat among others) on average (median = 4, S.D. 1.66). Regarding the time spent on social media, 61% use social media more than one hour (0–30 mins: 10%, 31–60 mins: 29%, 61–90 mins: 22%, 91- 120 mins: 13%, 121- 150 mins: 10%, and more than 150 mins: 16%). In the next sections, we quote some of the shared sentiments by our participants respecting Starbucks's perceived *quality commitment*, *heritage*, *uniqueness*, and *symbolism*.

5.3.1. Starbucks's product quality

We received 150 valid responses about the situation, in which respondents want to share something about Starbucks's product quality on social media. In line with our findings from the tweets, there are different sentiments towards products' quality. For example:

- *"Their coffee doesn't taste good and I think Starbucks is overrated."*
- *"I'm a coffee lover and I love the taste of their coffees."*
- *"Very disappointing: at first, I thought that Starbucks was a good brand, but then I heard that they ally with Monsanto, so the quality of the products is more than disappointing, it's poisoning us. It really changed the way I like Starbucks."*
- *"The Starbucks product quality is amazing! I am a person that likes strong coffee and Starbucks is able to fulfill my consumer needs!"*

Although we gathered this data in early spring, far from the fall – the season of PSL- one of our participants expressed her interest in the PSL. She wrote:

- *"Nothing better than a Pumpkin Spice Latte!"*

As another source of perceived quality commitment, several participants appreciate its consistency across different branches. For example:

- *"Starbucks's product quality is really good, with a consistent product across branches. They source their ingredients from a variety of geographic regions, and are open about it, which I appreciate."*
- *"Consistent, however not the best. You always know what you're getting."*

However, someone reports a contradictory observation among Middle Eastern branches and North American ones:

- "*Not consistentacross countries. Way better quality in the Middle East than in North America."*

Taking these examples into consideration and comparing them to the tweets (discussed in section 5.1.1), and the extracted words (shown in Table 3), we see consistency between findings from studies 1 and 2 and the current study. This shows the validity of our findings about quality commitment across different samples and different research methods.

5.3.2. Starbucks's connection to a historical period in time, culture and/or specific region

111 valid responses received under the question regarding sharing about Starbucks's connection to a historical period in time, culture and/or specific region. Among the answers, nearly one third were declines of knowledge about Starbucks's heritage. We consider the following quotes as the reasons of lower amount of heritage related tweets in the database.

- "*I never heard about Starbuckshistorical time or culture.*"
- "*Starbucks has become atrendy hang out placefor theyounger generation. Don't know much about the past history.*"
- *"To me, Starbucks is a global brand that doesn't really reflect a specific history."*
- *"I don't see any link to this. It really doesn't feel like a 'traditional' brand to me. Very modern and in today's time. Not any historical period."*

However, the rest of the participants mentioned its connection with the American culture, the 1970s time period, and its first store in Seattle's historic Pike Place Market. For example:

- *"Starbucks does have a connection to the 1970s era when coffee became highly stipulated into the North American lifestyle and has been ever present since then."*
- *"It's a good cultural reference to Italian 'bars' in Italy, where they serve coffee and espressos. They took that concept and brought the cafe experience to North America."*
- *"Starbucks's linked with North American's culture, people don't have time to sit and drink a coffee in the morning. In North America (vs. European take the time in the morning), so Starbuck's did think about a new and practical way to facilitate worker's life."*
- *"They still have their first Starbucks in the Pike Place Market, Seattle, and it's very lovely."*

Responding to this question, as expected, participants also mentioned remembering or missing an experience with Starbucks. For example, one of the participants wrote about her first experience with Starbucks, and another shared her favorite childhood sweet roll:

- *"Always keeps its traditions but still caters to new cultures and traditions that form and develop every day. First time I entered a Starbucks was unlike any other experience. It was a very relaxed environment. My order was made personal and I felt as part of the community. They say you never forget first impressions and this experience is no exception. Every time I pass by a Starbucks, I remember the first time I got something from there."*
- *"Starbucks is the only brand that makes Cinnamon rolls, which reminds me of my childhood in Sweden and when my mom would make them later in France. As I am not a coffee drinker, my main connection to Starbucks is to this sole product. The quality has remained constant, not as good as home baked, but good enough for me to buy them frequently."*

In summary, in this study we checked the robustness of our findings about heritage from studies 1 & 2. We also reported more explanations for the relatively low heritage related tweets.

5.3.3. Starbucks's ordinary/unique characteristics

We got 130 valid responses on sharing something about Starbucks's ordinary/unique characteristics. A few participants mentioning not being aware of unique characteristics for Starbucks, e.g.:

- *"Starbucks is not unique in my opinion."*
- *"Not much uniqueness, standard business model."*
- *"Starbucks is not that unique of a brand, and the price of their food is pretty expensive."*

However, most of the responses contained different kinds of sentiments with regard to Starbucks uniqueness. As expected, they compared Starbucks characteristics with other coffeehouses. For example:

- *"The ambiance is what differentiates them from the Second cup and others."*
- *"Starbucks was my favorite store previously. Their coffee was better than other brands and the slightly more expansive coffee was worth it for the atmosphere but it has greatly changed lately."*
- *"I prefer Tim Hortons to be honest and barely go to Starbucks."*

Moreover, some participants mentioned product's customization, its friendly atmosphere, providing seasonal drinks, changing cups' color for the holiday season, and also its marketing strategies as features for its uniqueness, e.g.:

- *"The drinks are customizable and are flexible to your needs, which not many coffee places have."*
- *"Good environment, like a living room space where you can relax, talk with people conduct business or do work."*
- *"I like how they change the color of their cups for Christmas."*

- *"Very very good marketing is the best characteristic they have."*

Participants also specified the higher price of Starbucks products as another source of its perceived uniqueness. For instance:

- *"I love Starbucks but I find it very expensive, therefore it is rare that I can afford it."*
- *"So expensive, God damn!"*
- *"Expensive addiction, expensive habit."*

In sum, study 3 helped us replicate findings from studies 1 and 2 with other data, which provides support for the coding scheme.

5.3.4. Starbucks adding meaning to consumers' lives

We received 131 responses to the question about Starbucks's products, values, behaviors, and so on that add meaning to one's life. Few people stated that Starbucks is a coffeehouse, offering coffee and tea, so they do not think that it would add anything to their lives. As an example:

- *"I don't believe Starbucks products add meaning to life as it is only coffee and teas etc."*

Moreover, we found that being a brand's consumer is not sufficient to feel a brand's symbolism. But, if someone considers a brand is adding meaning to his life, he would also being attached to that brand. For example:

- *"I am a Starbucks consumer but I don't consider myself the attached customer that only has Starbucks products. So it doesn't really add meaning to my life."*

In this study our participants shared cases where they perceive Starbucks is adding meaning to their life. For instance:

- *"Starbucks cookies are life."*
- *"Makes my day brighter."*

They also had several observations supporting their feelings. For example:

- *"Potential links to political factors put a damper on its alleged values and good behaviour."*
- *"Some of the shares of their products go to charities."*
- *"I'd like to see more giving back to communities."*
- *"The only reason I go to Starbucks is for one specific product. I try to boycott since I believe they have sacrificed freshness and quality for profits."*
- *"It provides yummy vegan drinks!"*

Similar to what we previously found about the importance of being correctly named and called, we got several comments on this subject. For example:

- *"I love the "name feature, makes people feel like individuals and not consumers."*
- *"When you order a drink at Starbucks, the baristas write your name on the cup which makes the customer feel special and nice."*
- *"The cashier often messes up the names of their clienteles… Although I spell out my name instead of simply saying it, they end up misspelling the name. But the cashiers and baristas are still very nice and welcoming."*

Finally, we found several quotes indicating consuming Starbucks in a conspicuous manner, or calling the girls going there as the white girls, e.g.:

- *"Starbucks in known to be more luxurious than the other companies (it's more expansive). Also, when a girl now buys from Starbucks people may call her a basic white girl."*
- *"Starbucks coffees are expensive, bad quality and are only a way for people to associate themselves with the brand and show off!"*
- *"Never been to Starbucks, but I guess they do something right with the product since every white girl go."*

Taking findings from study 3 into account, we provide support for findings from studies 1 and 2 from a new perspective.

5.4. Study 4: quantitative study through support vector machine (SVM)

Studies 1, 2 and 3, using different samples and/or methods, show the robustness of the coding scheme for the brand authenticity sentiment analysis. While the findings have several theoretical and managerial implications, discussed further in the next section, these studies are still qualitative in nature and might be cumbersome for big data analysis. In other words, while we think that without conducting previous studies, one could not verify the robustness of the coding process, for a big data study there is need for a framework to perform the analysis automatically. As such, the object of study 4 is to automatically predict the brand sentiments both in terms of associated authenticity dimensions–among the seven options – and their polarity–on a scale of −2 to +2. To this end, in the quantitative study, and as the last step in the brand authenticity sentiment analysis, we used a machine learning algorithm,[2] introduced in the next section, to analyze the efficiency and accuracy of created features in the coding process.

5.4.1. SVM method

We used support vector machine (SVM) analysis (Cortes & Vapnik, 1995). SVM is recognized as a fundamental algorithm for classification and regression problems (Chen & Zhang, 2014). Findings from different fields, specifically on sentiment analysis (Marafino, Davies, Bardach, Dean, & Dudley, 2014; Mostafa, 2013; Sidorov et al., 2014) suggest its usefulness and acceptable results accuracy. Once SVM is trained by the training data set, it can map a new data point into a space, where it belongs, and classify it into a specific category (Chan, Li, & Zhu, 2015). In other words, by using the established model, any new tweet, even those not coded by the judges, can be predicted accordingly (Cortes & Vapnik, 1995).

Practically, SVM finds hyperplanes that separate the classes with the largest margins. In the simplest case, suppose the training data set is $\{x_i, y_i\}$, $i = 1, 2, 3, \ldots, l$, $y_i \in \{-1, +1\}$, $x_i \in R^d$, and we have the following hyperplane, which divides the positive from the negative examples:

[2] There are two types of machine learning techniques, namely supervised and unsupervised. Supervised machine learning refers to receiving labeled data to achieve the ultimate goal of predicting unknown data points. However, in unsupervised machine learning methods, the machine itself learns from the patterns and associations among the data points, and then builds a model accordingly. From another perspective, supervised and unsupervised machine learning models differs in addressing research questions of having a specific purpose or target or not (Provost and Fawcett, 2013). In study 4 we use a supervised machine learning for our specific purpose, which is sentiment analysis. As said before, the supervised machine learning algorithm produces a model based on a training data set. Then, the machine classifies/regresses the testing data set according to what it has learned from the training sample (Provost and Fawcett, 2013). Regarding the training sample size in a supervised machine learning, methodologically speaking, it is not necessary to label all data (here the total final datasets containing 2.24 million tweets) or a large percentage of the whole dataset. Research shows that increasing the training size from 6% to 20% (233% growth), only elevates the overall prediction accuracy from 74.2 to 75.6 (less than 2% growth). However, increasing the input variables from 3 to 7, raises significantly the overall prediction accuracy from 65.5 to 74.2 (Huang, Davis, & Townshend, 2002). Practically, coding more data for the purpose of training the machine would be expensive, both in terms of labeling and the time needs the machine to learn and build the model. Therefore, once a supervised machine learning method is used in a big data analytics, the percentage of training size would be less important than the number of independent variables.

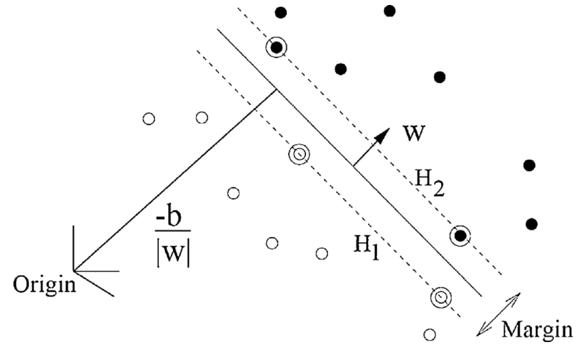

Fig. 2. An example of support vector machine classification. The support vectors are circled (Burges, 1998).

$$\vec{w}.\vec{x} + b = 0, \tag{1}$$

where the $\vec{x}$ are those which lie on the hyperplane, $\vec{w}$ is normal to the hyperplane, $\frac{|b|}{\left\|\vec{w}\right\|}$ is the perpendicular distance from the hyperplane to the origin, and $\|\vec{w}\|$ is the Euclidean norm of $\vec{w}$ (Burges, 1998). As depicted in Fig. 2, H_1 represents all vectors that fit in the $\vec{x_i}.\vec{w} + b = +1$ and H_2 shows all vectors in the $\vec{x_i}.\vec{w} + b = -1$ equation. All data points should fall outside of these hyperplanes, and not inside the margin. In other words, the data points should satisfy the following inequality:

$$(2) y_i(\vec{x_i}.\vec{w} + b) - 1 \geq 0 \quad \forall\, i$$

The SVM job is to maximize the margin. The distance of positive vectors (the black ones in Fig. 2), from the origin is $\frac{|1-b|}{\|\vec{w}\|}$. Similarly, the distance of negative vectors (the white ones), from the origin is $\frac{|-1-b|}{\|\vec{w}\|}$. So, the distance between the positive and negative pair vectors, which is the margin, is $\frac{2}{\|\vec{w}\|}$. So by minimizing $\|\vec{w}\|$, or $\frac{1}{2}\|\vec{w}\|^2$ for mathematical convenience, we would maximize the margin (Burges, 1998).

In order to minimize the margin subject to the inequality (2) constraint, using the Lagrange multipliers is a solution. To form the Lagrangian, the constraint should be multiplied by positive Lagrange multipliers (α_i) and subtracted from the objective function. This gives:

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_{1} \alpha \left[y_i \left(\vec{x_i}.\vec{w} + b \right) - 1 \right] \tag{3}$$

Then in order to find the extremum, partial derivatives in respect to $\vec{w}$ and b should be set to zero.

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum_{1} \alpha_i y_i \vec{x_i} = 0 \Rightarrow \vec{w} = \sum_{1} \alpha_i y_i \vec{x_i} \tag{4}$$

$$\frac{\partial L}{\partial b} = -\sum_{1} \alpha_i y_i = 0 \Rightarrow \sum_{1} \alpha_i y_i = 0 \tag{5}$$

Substituting equations of (4) and (5) in the Lagrangian equation gives:

$$L = \frac{1}{2} - \left(\sum_{1} \alpha_i y_i \vec{x_i} \right) . \left(\sum_{1} \alpha_j y_j \vec{x_j} \right) - \left(\sum_{1} \alpha_i y_i \vec{x_i} \right) . \left(\sum_{1} \alpha_j y_j \vec{x_j} \right) - b \sum_{1} \alpha_i y_i + \sum_{1} \alpha_i$$

$$L = \sum_{i} \alpha_i - \frac{1}{2} \sum_{i} \sum_{j} \alpha_i \alpha_j y_i y_j (\vec{x_i}.\vec{x_j}) \tag{6}$$

As shown in equation (6), maximizing the margin is only dependent on the dot product of the vectors, with respect to α_i and its positivity, and also subject to constraints (5) and with the solution given by (4) for a linear separable support vector training (Cortes & Vapnik, 1995). It is also the case for nonlinear separable support vectors (Burges, 1998).

Since most of the real datasets are not 100% linearly separable, Cortes and Vapnik (1995) introduced positive slack variables ξ_i for the constraints (equation 2), and C constant, $0 < \alpha_i < C$, as the positive multiplier for the $\sum_i \xi_i$. Practically, when C is low, it makes a smooth decision surface, while when it is high, it classifies examples more correctly by having freedom to select more samples as support vectors, but it makes the model less smooth (Pedregosa et al., 2011). Moreover, Boser, Guyon, and Vapnik (1992) showed that for nonlinear separable datasets it is a good solution to map the dataset to some other (possibly infinite dimensional) Euclidean space. As established before in section 4.3, the SVM model is only dependent on the dot products of the support vectors. By mapping into a new space, the SVM model is still dependent on the dot products of the support vectors in the new space. So, using a Kernel function (Aizerman, Braverman, & Rozoner, 1964) such that $K(x_i,x_j) = \Phi(x_i).\ \Phi(x_j)$, it is only needed to use K in the training process (Burges, 1998).

5.4.2. Dependent and independent variables in SVM

As mentioned, the object in this phase was to automatically categorize the tweets in accordance with the sentiment polarity, which has five possibilities, and the brand authenticity which has seven categories. Therefore, the dependent variables in SVM analysis were the class designation of each tweet among the possible five categories for brand sentiment polarity, and the seven options for brand authenticity dimensions.

For the independent variables, we built on the results from the coding algorithm (section 4.1). Based on the frequency of containing the labeled words–unigram, bigram, and trigram – we obtained a number containing 35 digits for each tweet. Each five numbers respectively represent the frequency of words in a dimension from strongly negative to strongly positive. In fact, we have five sets of frequencies; respectively for *heritage*, *none of them*, *symbolism*, *uniqueness*, *irrelevant to brand authenticity*, *quality commitment*, and for *not sufficient information to judge brand authenticity*. As an example, for a tweet like "*#PSL is back at Starbucks…. Fall weather just needs to come, my life will be complete*", we have the following code: [(0, 0, 0, 0, 0); (0, 0, 0, 0, 0); (0, 1, 0, 4, 5); (0, 0, 0, 3, 1); (0, 0, 0, 0, 0); (0, 0, 0, 2, 2); (0, 0, 0, 0, 0)]. It means that the tweet does not contain any of the words associated with *heritage* (0, 0, 0, 0, 0), *none of them* (0, 0, 0, 0, 0), *irrelevant to brand authenticity* (0, 0, 0, 0, 0)*, and not sufficient information to judge brand authenticity* (0, 0, 0, 0, 0). However, its n-grams are matched with some coded n-grams, which are associated with sentiment strength of *symbolism* (0, 1, 0, 4, 5), *uniqueness* (0, 0, 0, 3, 1), and *quality commitment* (0, 0, 0, 2, 2).

5.4.3. Running the SVM and findings

We used the Statistica 13 software to conduct SVM analysis. In this study, the one-versus-all type, with the winner-takes-all strategy, is applied (Zhou & Tuck, 2007). It means that for each classification, the classifier with the highest output value is selected as the corresponding class label (Yun, Sim, & Kim, 2000). To train and validate the SVM model, we applied v-fold cross-validation technique (v-value = 10). In this method of model validation, the whole data set is randomly subdivided into v subsets of equal size. Each model runs v times in the way that v-1 subsets train the model and the other one validates the model. In fact, each data point trains the model v-1 times, and validates the model once (Arlot, & Celisse, 2010). Previous research illustrates that the v-fold cross-validation method not only validates the model, but also prevents overfitting problems (Hsu, Chang, & Lin, 2003). Through grid search and v-fold cross-validation, which are considered the best way to optimize C and gamma (Hsu & Lin, 2002), we came into the optimized values for C and gamma in Kernel type of Radial Basis Function (RBF) of 5.0 and 0.445 respectively.

The results show a high level of cross-validation accuracy. The

Table 10
Classification summary.

	Total	Correct (%)	Incorrect (%)
Heritage	66	92.42	7.58
Irrelevant to brand authenticity	146	91.78	8.22
None of them	69	94.20	5.80
Not sufficient information to judge brand authenticity	90	73.33	26.67
Quality commitment	798	96.87	3.13
Symbolism	579	96.20	3.80
Uniqueness	455	89.67	10.33

average accuracy for training and validating the model through ten-fold cross-validation is 93.2%. To derive more confidence on the performance estimate, we ran the cross-validation SVM model multiple rounds, and did not get significant differences. Table 10 summarizes the correct and incorrect classification for each category. The highest accuracy, 96.9%, comes from the *quality commitment* class. However, the lowest one, 73.3%, comes from *not sufficient information to judge brand authenticity*. The low accuracy in this class might be due the wide range of tweets in this category, from a single word to a long sentence with lots of similar n-grams with other categories.

After obtaining predictions of the associated brand authenticity dimensions, we ran another SVM algorithm to analyze the overall sentiments towards the brand. This time we used SVM regression in order to predict the sentiments. Again, we used grid search and cross-validation method to train and test the SVM model for multiple times. Using ten-fold cross validation, we obtained C = 4, epsilon = 0.10 and gamma = 0.384 for Kernel type of Radial Basis Function.

Our results establish a high level of correlation between the observed and predicted values for our dependent variable, brand sentiment. The 90.8% correlation coefficient shows that the model can predict the polarity of brand sentiment very accurately. Table 11 provides details about the regression findings, including means, and standard deviations of observed versus predicted data. (Appendix D has more details about the SVM results.)

5.4.4. Out-of-sample validation

Using each of the established classification and regression SVM model, we can now predict brand authenticity category and brand sentiment polarity for any tweet, including those which are in the database but not coded by human judges. The process for predicting the brand authenticity category and brand sentiment polarity for unlabeled tweets is as follows (Fig. 3). First, using n-gram technique and based on the frequency of containing the labeled unigram, bigram, and trigram words, each new tweet gets the 35-digit code. Second, once the 35-digit code is assigned to each tweet, it will be used as the input to the models. Then, the established SVM models predict both the brand authenticity dimension and its sentiment polarity for each unlabeled tweet as the output. Since the utilized SVM models are validated with ten-fold cross validation, the new predictions are also considered accurate as well.

Table 11
SVM Regression results.

	Overall
Observed mean	0.48
Predictions mean	0.46
Observed S.D.	1.01
Predictions S.D.	0.89
Mean squared error	0.18
Error mean	0.02
Error S.D.	0.42
Abs. error mean	0.33
S.D. ratio	0.42
Correlation	0.91

To use the validated SVM models in big data analytics, we randomly chose 3000 tweets from those the judges did not code. Using custom prediction feature of the Statistica 13, the associated brand authenticity dimension and the level of brand sentiment were predicted. For a tweet like *"The pumpkin bread from Starbucks is the only good thing"*, the n-grams (n = 3) extracted and from the extracted list, those which are associated with each of the polarity and authenticity dimension were automatically counted. The emerged 35-digit code is: [(0, 0, 0, 0, 0); (0, 0, 0, 0, 0); (0, 0, 0, 1, 1); (0, 0, 2, 4, 1); (0, 0, 0, 0, 0); (0, 1, 2, 3, 0); (0, 0, 0, 0, 0)]. Once we entered the code in the SVM classification model, the SVM categorized the tweet under *uniqueness* dimension. The presence of *from Starbucks* phrase and the *only* word support this prediction. Similarly, the SVM regression model predicts its sentiment as 0.842, which looks reasonable. This example shows the way any tweet (data point) can be predicted (mapped) by utilizing established SVM models.

6. Discussion and implications

6.1. Theoretical implications

This article demonstrates a new way to analyze brand sentiment. The results from the four studies contribute to the literature of analyzing brand authenticity in social media. While previous research supports the necessity of monitoring online medium to avoid inauthenticity through a *doppelgänger* brand image (Thompson et al., 2006), in this research we add to this notion by examining how brand authenticity is perceived by customers and is shared on social media in terms of *quality commitment*, *heritage*, *uniqueness*, and *symbolism*. Moreover, we extend the current knowledge about brand sentiment analysis by addressing the reasons for getting positive or confronting negative sentiments. This insight contributes to different marketing theories, including customer engagement. It is proposed by Harmeling, Moffett, Arnold, and Carlson (2017) that both *customer network asset*, "the number, diversity, and structure of a customer's interpersonal ties within his or her social network", and *customer knowledge store*, "a customer's accumulation of knowledge about the product, brand, firm, and other customers" are essential elements to the idea of customer engagement theory. In this regard, our findings help know the level of sentiment and its rationality, which could be translated into whether firms can fully take advantage of customer-owned resources in customer engagement. Furthermore, we improve the sentiment measurement accuracy by labeling and predicting the sentiments on a five-point-scale. Moreover, we contribute to the existing literature of decision support systems (DSS), which looks for quantitative intuitions from qualitative data of user-generated contents (Fersini, Messina & Pozzi, 2014). Methodologically speaking, we obtain high accuracies in training and testing the dataset, suggesting the usefulness of the proposed algorithm for predicting both brand authenticity dimensions and its sentiment polarity.

6.2. Managerial implications

As companies' interest in monitoring and getting insights from brand sentiment continues to increase, the research findings provide important managerial implications. First, our findings show that while PSL has its own fans – waiting for fall throughout the year – once they found the taste is different from their expectations, they begin to share negative sentiments about the PSL and the Starbucks brand. Therefore, changes to the specifications of popular products should be made with great caution. Second, companies should find ways to engage customers in sharing positive e-WOM on social media. Our results suggest that it could be promoted once customers feel improvements in different aspects of the perceived brand authenticity. For companies like Starbucks, with many followers on social media, highlighting services and products to reinforce ties with customers would be an appropriate

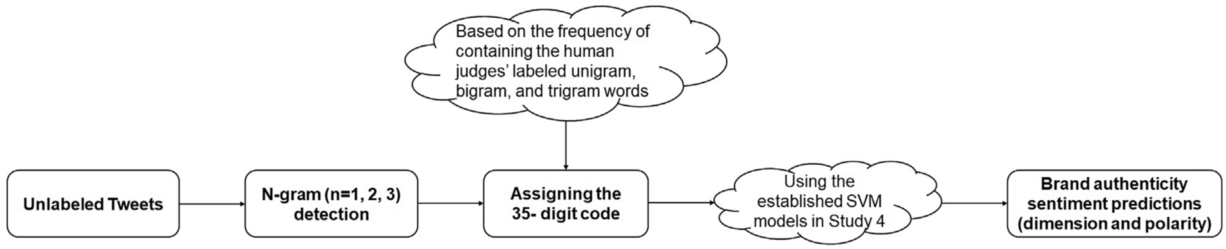

Fig. 3. Predicting the brand authenticity dimension and brand sentiment polarity for unlabeled tweets.

approach (Chu & Kim, 2011). Third, while some issues might look ordinary, i.e. misspelling customer names by Starbucks baristas, firms should be aware of their potential viral effects on social media and the possible negative impacts on customers perceived authenticity. Fourth, managers should be aware of their statements and behaviors specifically about issues that are strongly related to social values and beliefs. In this study, we found evidence that compromising these issues not only decreases perceived brand authenticity, but it also might lead to calls for boycott campaigns on social media. Finally, firms should not only continuously monitor sentiments toward their perceived brand authenticity, they should also look for ways to improve it in comparison to their competitors. To this end, our framework provides a useful way for understanding the tweet's related brand authenticity dimensions and sentiment polarity. This would help brand managers effectively observe what is happening on social media about themselves or their rivals, gain insights from the data in order to take proper actions to increase customer perceptions, and finally, formulate new insights for designing short-term and long-term strategies.

7. Conclusion

Right now, hundreds of thousands of new data are generated and added to the accumulated data. This immense amount of real-time and retrospective data has helped change previous paradigms, and led to a new one: big data analytics. Big data became widespread as recently as 2011 and attracted many researchers and organizations from different disciplines not only due to its potentially broad range of applications, but also because of its high involvement in the day-to-day life of people all over the world, including researchers and organizational stakeholders.

Not surprisingly, like other disciplines, marketing has been engaged in big data analytics to contribute to the knowledge of analyzing big data from its own perspective, and also to make itself able to propose more precise practical implications (Feldman, 2013). Nevertheless, the marketing intelligence literature is still far behind the big data mainstream and needs research in order to gain more insights for marketing practitioners. More theoretical and empirical research could help business managers stay one step ahead of the competition by obtaining and analyzing big data which could be converted into both short-term and long-term strategic planning in the new era (Wright & Calof, 2006).

In this research, we propose a new algorithm to analyze brand sentiments on social media. To the best of our knowledge, this research is the first attempt to bridge the brand sentiment and brand literature constructs. We empirically shed light on how brands could investigate sentiments towards them in terms of perceived brand authenticity. Results from the four studies demonstrate the usefulness of proposed procedure. Through study 1, we discussed several tweets from our dataset which our coders classified under the *quality commitment, heritage, uniqueness,* or *symbolism* categories. We found a few tweets which were about brand authenticity but not associated with our four categories, and placed them in the *none of them* class. We also provided examples of tweets that were categorized under *irrelevant to brand authenticity*, and *not sufficient information to judge brand authenticity*. In study 2 and by using LSA, we extracted the common words in each category. The extracted words supported the coding process and also the study 1's discussion. However, due to the diversity of tweets in *none of them* and *not sufficient information to judge brand authenticity* categories, we did not receive significant common words for these two categories. In study 3, we checked the robustness of our findings from studies 1 and 2 with data from a laboratory setting. Asking students to imagine intention of sharing sentiment towards Starbucks, we got similar feelings observed in studies 1 and 2, verifying findings from studies 1 and 2 with another data source. Finally, in study 4, the results of SVM analyses show high accuracy for the prediction of the brand authenticity dimensions and their sentiment strength. With the prepared lexicon and ten-fold cross validation, the accuracy was above 0.90 for both analyses. Moreover, we used the established SVM models to predict new and unlabeled tweets from the dataset. Results show the applicability of analyzing big data sets with the proposed methodology.

In summary, this research contributed both theoretically and managerially to the brand authenticity and brand sentiment literatures. The proposed procedure, and the research findings on brand authenticity sentiment analysis could facilitate further inquiries into sentiment analysis for all other brand constructs and in several related domains, such as e-Word-of-Mouth studies. Practically speaking, this research could provide marketing practitioners with a reliable and valid decision support system to evaluate the level of sentiment towards a brand more specifically and more accurately, which could lead to proposing appropriate strategies to strengthen their brand authenticity.

8. Limitations and future research

This study has some limitations that offer opportunities for further research. First, we studied brand authenticity sentiments with cross-sectional data, while we acknowledge that customer sentiments could even change over a short period of time. Thus, future research could examine how the sentiments towards brand authenticity vary over time. Second, we did not take into account the brand's follow-up interventions about the shared tweet. We think it could be interesting to see if this kind of intervention has positive effects on perceived brand authenticity. Moreover, we expect that this potential benefit would vary depending on the time lapse between the sharing moment and the brand's follow-up. Third, Twitter Company and researchers estimate that between 8.5% and 15% of all active users (up to 48 million accounts) might be bots, according how they differently act and contact twitter servers in comparison with true human accounts. Research suggests aggregating 100 features of each tweet to detect Twitter bots in a fuzzy scale (Chu, Gianvecchio, Wang, & Jajodia, 2012). Accordingly, we tried to rule out the spam tweets (which are mostly tweeted by bots) with classification under *irrelevant to brand authenticity* dimension; however, we acknowledge that some tweets in other categories might also be created by bots. While it might not affect the proposed process of brand authenticity sentiment analysis, future research could either clean the dataset from potential bots or compare sentiments towards a brand between tweets created by bots and by humans. While the earlier approach would lead to a cleaner dataset, the later one would be more realistic and fruitful for firms to understand how supportive bots and confrontative bots/human interact on social media. Finally, as we collected the data in real-time, we did not have the data about the number of likes and retweets a post received. Future studies could use retrospective data to access this sort of data to see if there are sentiment differences between popular tweets and others.

Acknowledgement

The authors gratefully acknowledge the financial support of the Fonds Québécois sur la societé et la culture.

Appendix A

We conducted a series of CFAs to see which model fit the data best. Table A1 presents and compares the fit indices results for different factorial models.

Table A1
Fit indices results for different factorial models.

Model Number	Model name	χ^2	df	P	χ^2/df	CFI	NNFI	RMSEA	SRMR
0	Null-model	1741	210	< 0.01	8.29	NA	NA	NA	NA
1	One-factor (Q-H-U-S)	673.73	189	< 0.01	3.56	0.735	0.706	0.117	0.085
2a	Two-factor uncorrelated (Q-U and H-S)	194.24	54	< 0.01	3.60	0.855	0.823	0.118	0.220
2b	Two-factor correlated (Q-U and H-S)	132.99	53	< 0.01	2.51	0.917	0.897	0.090	0.064
3	Three-factor correlated (Q, U, H-S)	184.80	87	< 0.01	2.12	0.920	0.903	0.078	0.062
4a	Four-factor uncorrelated (Q, U, H, S)	646.00	189	< 0.01	3.42	0.750	0.723	0.114	0.286
4b	Four-factor correlated (Q, U, H, S)	372.91	183	< 0.01	2.04	0.945	0.938	0.072	0.061

Q = Quality commitment, U = Uniqueness, H = Heritage, and S = Symbolism. NA = not applicable.
Model 1: All items forced to load on one factor. Model 2a and 2b: Due to some similarities between Q and U, and also H and S, we forced the items of these four dimensions to load on two, the first one made from Q and U, and the second one from H and S. Model 3: Due to some similarities between H and S, we forced the items of these two under one factor and items of Q and U on two other factors. Model 4a and 4b: Items of each dimension loaded on its own factor.

Appendix B

Fig. A1 shows a sample of received raw data for the following two tweets:
"I've always loved & will remain loving Starbucks!"
"@Starbucks the worker was very rude too yikes I'm not going back to that one store".

```
*********************************************************************************************************************************
{"created_at":"Tue Jul 21 21:46:25 +0000 2015" id:623610002524274689 id_str:"623610002524274689" text:"I've always loved &amp; will remain loving Starbucks! \ud83d\ude18
https:\/\/t.co\/viguETV1nz1" source:"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\/a\u003e" truncated:false in_reply_to_status_id:null
in_reply_to_status_id_str:null in_reply_to_user_id:null in_reply_to_user_id_str:null in_reply_to_screen_name:null user:{"id":860191964 id_str:"860191964" name:"Anthony Goehring "
screen_name:"AnthonyGoehring" location:"" url:null description:null protected:false verified:false followers_count:136 friends_count:250 listed_count:0 favourites_count:184
statuses_count:1371 created_at:"Wed Oct 03 19:05:36 +0000 2012" utc_offset:null time_zone:null geo_enabled:false lang:"en" contributors_enabled:false is_translator:false
profile_background_color:"C0DEED" profile_background_image_url:"http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png"
profile_background_image_url_https:"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png" profile_background_tile:false profile_link_color:"0084B4"
profile_sidebar_border_color:"C0DEED" profile_sidebar_fill_color:"DDEEF6" profile_text_color:"333333" profile_use_background_image:true
profile_image_url:"http:\/\/pbs.twimg.com\/profile_images\/622562126650023937\/q513jFLQ_normal.jpg"
profile_image_url_https:"https:\/\/pbs.twimg.com\/profile_images\/622562126650023937\/q513jFLQ_normal.jpg"
profile_banner_url:"https:\/\/pbs.twimg.com\/profile_banners\/860191964\/1437265922" default_profile:true default_profile_image:false following:null follow_request_sent:null
notifications:null} geo:null coordinates:null place:null contributors:null quoted_status_id:622127994199515136 quoted_status_id_str:"622127994199515136" quoted_status:{"created_at":"Fri Jul
17 19:37:27 +0000 2015" id:622127994199515136 id_str:"622127994199515136" text:"Top 6 Mind-Blowing Starbucks Facts You Never Knew http:\/\/t.co\/VYmOvfhUSJ
http:\/\/t.co\/PAvhcPM1S3" source:"\u003ca href=\"https:\/\/about.twitter.com\/products\/tweetdeck\" rel=\"nofollow\"\u003eTweetDeck\u003c\/a\u003e" truncated:false in_reply_to_status_id:null
in_reply_to_status_id_str:null in_reply_to_user_id:null in_reply_to_user_id_str:null in_reply_to_screen_name:null user:{"id":3070181257 id_str:"3070181257" name:"Alex"
screen_name:"Alex122131" location:"" url:null description:"Fun Guy with Big Dreams." protected:false verified:false followers_count:2763 friends_count:0 listed_count:1 favourites_count:0
statuses_count:4 created_at:"Mon Mar 09 18:16:16 +0000 2015" utc_offset:19800 time_zone:"Chennai" geo_enabled:false lang:"en" contributors_enabled:false is_translator:false
profile_background_color:"C0DEED" profile_background_image_url:"http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png"
profile_background_image_url_https:"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png" profile_background_tile:false profile_link_color:"0084B4"
profile_sidebar_border_color:"C0DEED" profile_sidebar_fill_color:"DDEEF6" profile_text_color:"333333" profile_use_background_image:true
profile_image_url:"http:\/\/pbs.twimg.com\/profile_images\/581425659240558592\/OcoDssCq_normal.jpg"
profile_image_url_https:"https:\/\/pbs.twimg.com\/profile_images\/581425659240558592\/OcoDssCq_normal.jpg" default_profile:true default_profile_image:false following:null
follow_request_sent:null notifications:null} geo:null coordinates:null place:null contributors:null retweet_count:0 favorite_count:0 entities:{"hashtags":[] trends:[]
urls:[{"url":"http:\/\/t.co\/VYmOvfhUSJ" expanded_url:"http:\/\/bit.ly\/1TpBeVO" display_url:"bit.ly\/1TpBeVO" indices:[50 72]}] user_mentions:[] symbols:[]
media:[{"id":622127990718099457 id_str:"622127990718099457" indices:[73 95] media_url:"http:\/\/pbs.twimg.com\/media\/CKI-HIOUcAEfMPA.jpg"
media_url_https:"https:\/\/pbs.twimg.com\/media\/CKI-HIOUcAEfMPA.jpg" url:"http:\/\/t.co\/PAvhcPM1S3" display_url:"pic.twitter.com\/PAvhcPM1S3"
expanded_url:"http:\/\/twitter.com\/Alex122131\/status\/622127994199515136\/photo\/1" type:"photo" sizes:{"medium":{"w":600 h:333 resize:"fit"} small:{"w":340 h:189 resize:"fit"}
large:{"w":1009 h:562 resize:"fit"} thumb:{"w":150 h:150 resize:"crop"}}}]} extended_entities:{"media":[{"id":622127990718099457 id_str:"622127990718099457" indices:[73 95]
media_url:"http:\/\/pbs.twimg.com\/media\/CKI-HIOUcAEfMPA.jpg" media_url_https:"https:\/\/pbs.twimg.com\/media\/CKI-HIOUcAEfMPA.jpg" url:"http:\/\/t.co\/PAvhcPM1S3"
display_url:"pic.twitter.com\/PAvhcPM1S3" expanded_url:"http:\/\/twitter.com\/Alex122131\/status\/622127994199515136\/photo\/1" type:"photo" sizes:{"medium":{"w":600 h:333 resize:"fit"}
small:{"w":340 h:189 resize:"fit"} large:{"w":1009 h:562 resize:"fit"} thumb:{"w":150 h:150 resize:"crop"}}}]} favorited:false retweeted:false possibly_sensitive:false filter_level:"low"
lang:"en"} retweet_count:0 favorite_count:0 entities:{"hashtags":[] trends:[] urls:[{"url":"https:\/\/t.co\/viguETV1nz1"
expanded_url:"https:\/\/twitter.com\/alex122131\/status\/622127994199515136" display_url:"twitter.com\/alex122131\/sta\u2026" indices:[56 79]}] user_mentions:[] symbols:[]} favorited:false
retweeted:false possibly_sensitive:false filter_level:"low" lang:"en" timestamp_ms:"1437515185372"}
*********************************************************************************************************************************
{"created_at":"Tue Jul 21 21:48:02 +0000 2015" id:623610407870029824 id_str:"623610407870029824" text:"@Starbucks the worker was very rude too yikes im not going back to that one store"
source:"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\/a\u003e" truncated:false in_reply_to_status_id:623609658159140864
in_reply_to_status_id_str:"623609658159140864" in_reply_to_user_id:1535183485 in_reply_to_user_id_str:"1535183485" in_reply_to_screen_name:"shiningdeming" user:{"id":1535183485
id_str:"1535183485" name:"cindy" screen_name:"shiningdeming" location:"dtx" url:null description:"i like my eyelashes as dark as my soul and as long as the list of people i hate" protected:false
verified:false followers_count:521 friends_count:771 listed_count:2 favourites_count:4853 statuses_count:12511 created_at:"Thu Jun 20 22:22:58 +0000 2013" utc_offset:-18000
time_zone:"Central Time (US & Canada)" geo_enabled:false lang:"en" contributors_enabled:false is_translator:false profile_background_color:"080108"
profile_background_image_url:"http:\/\/pbs.twimg.com\/profile_background_images\/497853840369016832\/Mp03vhCE.jpeg"
profile_background_image_url_https:"https:\/\/pbs.twimg.com\/profile_background_images\/497853840369016832\/Mp03vhCE.jpeg" profile_background_tile:true profile_link_color:"7AD0D6"
profile_sidebar_border_color:"000000" profile_sidebar_fill_color:"DDEEF6" profile_text_color:"333333" profile_use_background_image:true
profile_image_url:"http:\/\/pbs.twimg.com\/profile_images\/622625867278086144\/2mjdkJUA_normal.jpg"
profile_image_url_https:"https:\/\/pbs.twimg.com\/profile_images\/622625867278086144\/2mjdkJUA_normal.jpg"
profile_banner_url:"https:\/\/pbs.twimg.com\/profile_banners\/1535183485\/1436332553" default_profile:false default_profile_image:false following:null follow_request_sent:null
notifications:null} geo:null coordinates:null place:null contributors:null retweet_count:0 favorite_count:0 entities:{"hashtags":[] trends:[] urls:[] user_mentions:[{"screen_name":"Starbucks"
name:"Starbucks Coffee" id:30973 id_str:"30973" indices:[0 10]}] symbols:[]} favorited:false retweeted:false possibly_sensitive:false filter_level:"low" lang:"en"
timestamp_ms:"1437515282014"}
*********************************************************************************************************************************
```

Fig. A1. Raw data for two tweets.

Appendix C

Fig. A2 shows the user-friendly PHP platform in which the coding process took place.

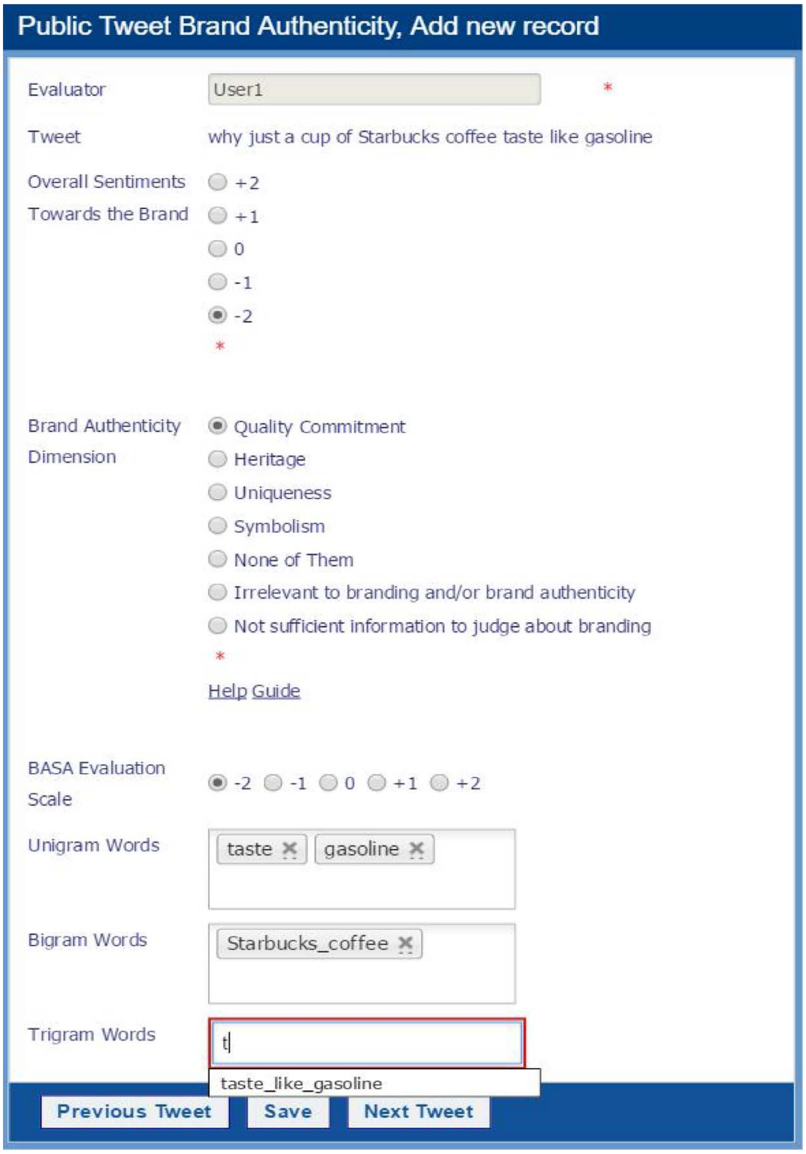

Fig. A2. The platform for coding process of brand authenticity sentiment analysis.

Appendix D

The confusion matrix in Table A2 shows the number of confused classifications for each dimension. For example, from the total of 455 tweets in the *uniqueness* category, one is incorrectly classified under *irrelevant to brand authenticity, none of them, not sufficient information to judge about branding*, 39 under *quality commitment* and five under *symbolism*. The elements on the main diagonal represent the tweets which were correctly classified.

Table A2
Confusion matrix: observed (rows) and predicted (columns).

	H	I	NOT	NSI	QC	S	U
Heritage (H)	**61**	0	0	0	3	1	1
Irrelevant to brand authenticity (I)	0	**134**	0	0	0	12	0
None of them (NOT)	0	1	**65**	0	1	0	2
Not sufficient information to judge about branding (NSI)	0	1	0	**66**	4	19	0
Quality commitment (QC)	0	0	0	1	**773**	3	21
Symbolism (S)	0	1	0	0	7	**557**	14
Uniqueness (U)	0	1	1	1	39	5	**408**

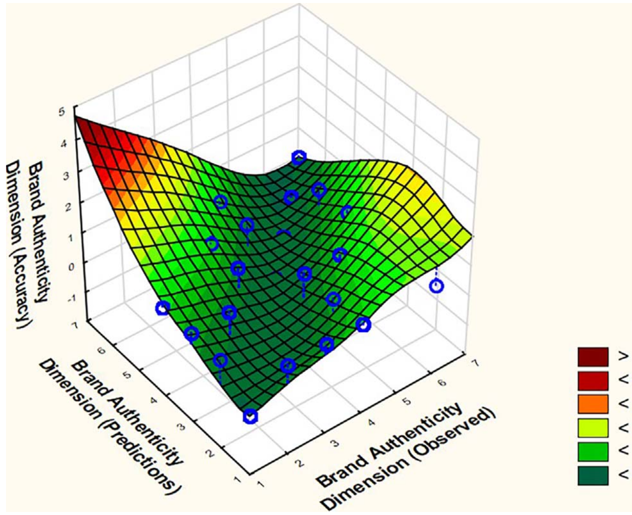

Fig. A3. Observed brand authenticity dimension (x) vs. predicted brand authenticity dimension (y) vs. its accuracy (z).

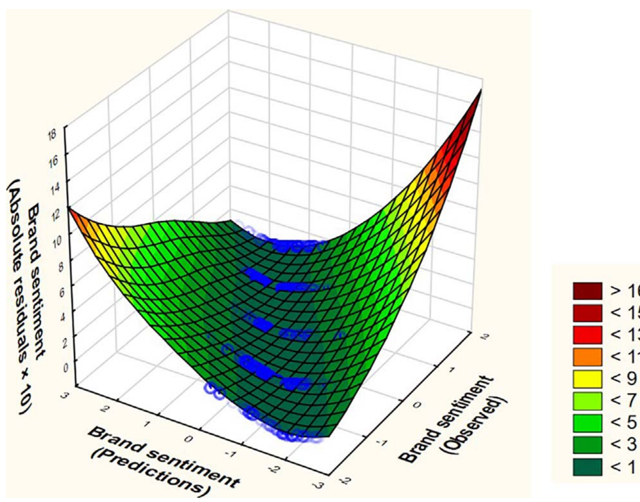

Fig. A4. Observed brand sentiment (x) vs. predicted brand sentiment dimension (y) vs. 10 times of Absolute residual (z).

To visualize the model's predictive accuracy, we graphed the observed brand authenticity vs. predicted vs. its accuracy with a surface plot. As shown in Fig. A3, a large amount of data is placed around z = 0, which translates to the model's highly predictive accuracy. Moreover, the surface plot in Fig. A4 demonstrates the observed brand sentiment vs. the predicted one while depicting its residual with ten times more scale. As shown, most tweets are correctly predicted within 0.5 error.

References

Ahmad, S. N., & Laroche, M. (2017). Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management, 37*(3), 202–213.

Ailawadi, K. L., Lehmann, D. R., & Neslin, S. A. (2003). Revenue premium as an outcome measure of brand equity. *Journal of Marketing, 67*(4), 1–17.

Aizerman, A., Braverman, E. M., & Rozoner, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control, 25*, 821–837.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79.

Assiouras, I., Liapati, G., Kouletsis, G., & Koniordos, M. (2015). The impact of brand authenticity on brand attachment in the food industry. *British Food Journal, 117*(2), 538–552.

Barbier, G., & Liu, H. (2011). *Data mining in social media. Social network data analytics.* Springer327–352.

Beverland, M. B., & Farrelly, F. J. (2010). The quest for authenticity in consumption: Consumers' purposive choice of authentic cues to shape experienced outcomes. *Journal of Consumer Research, 36*(5), 838–856.

Beverland, M. B. (2005). Crafting brand authenticity: The case of luxury wines. *Journal of Management Studies, 42*(5), 1003–1029.

Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. *International conference on discovery science* (pp. 1–15).

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303–316.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144–152).

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon Information. *Communication & Society, 15*(5), 662–679.

Brown, S., Kozinets, R. V., & Sherry, J. F., Jr (2003). Teaching old brands new tricks: Retro branding and the revival of brand meaning. *Journal of Marketing, 67*(3), 19–33.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121–167.

Chan, K. W., Li, S. Y., & Zhu, J. J. (2015). Fostering customer ideation in crowdsourcing community: The role of peer-to-peer and peer-to-firm interactions. *Journal of Interactive Marketing, 31*, 42–62.

Chen, C. P., & Zhang, C. (2014). Data-intensive applications, challenges techniques and technologies: A survey on big data. *Information Sciences, 275*, 314–347.

Chhabra, D., Healy, R., & Sills, E. (2003). Staged authenticity and heritage tourism. *Annals of Tourism Research, 30*(3), 702–719.

Choi, H., Ko, E., Kim, E. Y., & Mattila, P. (2015). The role of fashion brand authenticity in product management: A holistic marketing approach. *Journal of Product Innovation Management, 32*(2), 233–242.

Chu, S., & Kim, Y. (2011). Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. *International Journal of Advertising, 30*(1),

47–75.
Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 811–824.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.
Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391.
Driscoll, K., & Walker, S. (2014). Big data, big questions working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication, 8*, 20.
Eggers, F., O'Dwyer, M., Kraus, S., Vallaster, C., & Güldenberg, S. (2013). The impact of brand authenticity on brand trust and SME growth: A CEO perspective. *Journal of World Business, 48*(3), 340–348.
Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89.
Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian ensemble learning. *Decision Support Systems, 68*, 26–38.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research,* 39–50.
Fournier, S., & Avery, J. (2011). The uninvited brand. *Business Horizons, 54*(3), 193–207.
Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations–A case from Sweden. *Journal of Destination Marketing & Management, 3*(4), 198–209.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.
Gaspar, R., Pedro, C., Panagiotopoulos, P., & Seibt, B. (2016). Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior, 56*, 179–191.
Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of Interactive Marketing, 27*(4), 242–256.
Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266–6282.
Ghiassi, M., Zimbra, D., & Lee, S. (2016). Targeted twitter sentiment analysis for brands using supervised feature engineering and the dynamic architecture for artificial neural networks. *Journal of Management Information Systems, 33*(4), 1034–1058.
Gilmore, J. H., & Pine, B. J. (2007). *Authenticity: What consumers really want?* Harvard Business Press.
Gopaldas, A. (2014). Marketplace sentiments. *Journal of Consumer Research, 41*(4), 995–1014.
Habibi, M. R., Laroche, M., & Richard, M. (2014). Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities. *International Journal of Information Management, 34*(2), 123–132.
Hajli, N., Shanmugam, M., Papagiannidis, S., Zahay-Blatz, D., & Richard, M. O. (2017). Branding co-creation with members of online brand communities. *Journal of Business Research, 70*(January), 136–144.
Hajli, N., Sims, J., Zadeh, A. H., & Richard, M. O. (2017). A social commerce investigation of the role of trust in a social networking site on purchase intentions. *Journal of Business Research, 71*(February), 133–141.
Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science, 45*(3), 312–335.
He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management, 52*(7), 801–812.
Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121.
Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big data and consumer behavior: Imminent opportunities. *Journal of Consumer Marketing, 33*(2), 89–97.
Hsu, C. W., & Lin, C. J. (2002). A simple decomposition method for support vector machines. *Machine Learning, 46*(1-3), 291–314.
Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). *A practical guide to support vector classification*.
Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems, 57*, 42–53.
Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*(4), 725–749.
Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting, 31*(3), 992–1007.
Ibrahim, N. F., Wang, X., & Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from Twitter. *Computers in Human Behavior, 72*, 321–338.
Internet live stats (2015). *Twitter statistics.* Retrieved July 14, 2015, from http://www.internetlivestats.com/twitter-statistics/.
Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169–2188.
Johnson, A. R., Thomson, M., & Jeffrey, J. (2015). What does brand authenticity mean? Causes and consequences of consumer scrutiny toward a brand narrative. *Brand Meaning Management, 12*, 1–27.
Kadirov, D. (2015). Private labels ain't bona fide! Perceived authenticity and willingness to pay a price premium for national brands over private labels. *Journal of Marketing Management, 31*(17–18), 1773–1798.
Kanemoto, E., & Dai, Z. (2015). *Who cares about their names: Case study in Starbucks*.
Kim, B., & Sullivan, M. W. (1998). The effect of parent brand experience on line extension trial and repeat purchase. *Marketing Letters, 9*(2), 181–193.
Klein, J. G., Smith, N. C., & John, A. (2004). Why we boycott: Consumer motivations for boycott participation. *Journal of Marketing, 68*(3), 92–109.
Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing, 80*(1), 7–25.
Laroche, M., Habibi, M. R., & Richard, M. (2013). To be or not to be in social media: How brand loyalty is affected by social media? *International Journal of Information Management, 33*(1), 76–82.
Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research, 48*(5), 881–894.
Lee, H., Han, J., & Suh, Y. (2014). Gift or threat? An examination of voice of the customer: The case of MyStarbucksIdea.com. *Electronic Commerce Research and Applications, 13*(3), 205–219.
Mahrt, M., & Scharkow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media, 57*(1), 20–33.
Marafino, B. J., Davies, J. M., Bardach, N. S., Dean, M. L., & Dudley, R. A. (2014). N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. *Journal of the American Medical Informatics Association, 21*(5), 871–875.
Miller, C. (2009). *Will the hard-core starbucks customer pay more? the chain plans to find out. The New York Times*. Retrieved July 15, 2017, from http://www.nytimes.com/2009/08/21/business/21sbux.html.
Milstein, S., Lorica, B., Magoulas, R., Hochmuth, G., Chowdhury, A., & O'Reilly, T. (2008). *Twitter and the micro-messaging revolution: Communication, connections, and immediacy–140 characters at a time.* O'Reilly Media, Incorporated.
Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology, 25*(2), 200–218.
Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241–4251.
Napoli, J., Dickinson, S. J., Beverland, M. B., & Farrelly, F. (2014). Measuring consumer-based brand authenticity. *Journal of Business Research, 67*(6), 1090–1098.
Naylor, R. W., Lamberton, C. P., & West, P. M. (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), 105–120.
Netemeyer, R. G., Krishnan, B., Pullig, C., Wang, G., Yagci, M., Dean, D., ... Wirth, F. (2004). Developing and validating measures of facets of customer-based brand equity. *Journal of Business Research, 57*(2), 209–224.
Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior, 31*, 527–541.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*(October), 2825–2830.
Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.* O'Reilly Media, Inc. Retrieved February 20, 2017, from http://scholarworks.bgsu.edu/rbc/2015conference/panel7/3.
Schallehn, M., Burmann, C., & Riley, N. (2014). Brand authenticity: Model development and empirical testing. *Journal of Product & Brand Management, 23*(3), 192–199.
Shi, Z., Rui, H., & Whinston, A. B. (2013). Content sharing in a social broadcasting environment: Evidence from twitter. *MIS Quarterly, 38*(1), 123–142.
Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic N-grams as machine learning features for natural language processing. *Expert Systems with Applications, 41*(3), 853–860.
Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing, 26*(2), 102–113.
Social bakers (2015). *Starbucks statistics.* Retrieved July 14, 2015, from http://www.socialbakers.com/statistics/twitter/profiles/detail/30973-starbucks.
Spiggle, S., Nguyen, H. T., & Caravella, M. (2012). More than fit: Brand extension authenticity. *Journal of Marketing Research, 49*(6), 967–983.
Stryker, S., & Burke, P. J. (2000). The past, present, and future of an identity theory. *Social Psychology Quarterly, 63*(4), 284–297.
Thompson, C. J., Rindfleisch, A., & Arsel, Z. (2006). Emotional branding and the strategic value of the doppelgänger brand image. *Journal of Marketing, 70*(1), 50–64.
Twitter counter (2015). *Starbucks.* Retrieved July 14, 2015, from http://twittercounter.com/Starbucks.
Valta, K. S. (2013). Do relational norms matter in consumer-brand relationships? *Journal of Business Research, 66*(1), 98–104.
Walsh, M. G. (2013). Protecting your brand against the heartbreak of genericide. *Business Horizons, 56*(2), 159–166.
Wikipedia (2015). *Starbucks.* Retrieved July 14, 2015, from https://en.wikipedia.org/wiki/Starbucks.
Wright, S., & Calof, J. L. (2006). The quest for competitive, business and marketing intelligence: A country comparison of current practices. *European Journal of Marketing, 40*(5/6), 453–465.
Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120–130.
Yun, T., Sim, K., & Kim, H. (2000). Support vector machine-based inspection of solder joints using circular illumination. *Electronics Letters, 36*(11), 949–951.
Zhou, X., & Tuck, D. P. (2007). MSVM-RFE: Extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics (Oxford, England), 23*(9), 1106–1114.