Understanding Engagement Behavior in Online Brand Communities: How Social Identity relates to Frequency of Interaction and Tweet Sentiment

CANDY REEBROEK

Candy Reebroek (candyreebroek@hotmail.com) is Master student of Communication Science, University of Twente, Enschede, 7500 AE.

ABSTRACT

This study explains engagement behavior in online brand communities based on data of Twitter users who present different types of social identities. For this, we examined fifteen online brand communities that are popular on Twitter and originated from fashion, fast-food, gaming, cars, and sports sectors. In total, 27,143 twitter messages were analyzed from 22,333 unique Twitter users. We used the Twitter user's profile descriptions to classify their social identity with the help of computational methods such as Machine Learning and Natural Language Processing. To study the engagement behavior of the Twitter users, we calculated the tweets sentiment and the frequency of interaction between Twitter users and online brand communities. With this method, we investigated the relationship between different social identity groups and engagement behavior in different online brand communities. We found that tweet sentiment and frequency of interaction vary significantly between different social identity groups when mentioning different online brand communities. This result is important for online brand community managers to understand what kind of Twitter users interact with their online brand community and how these users engage with the community. Right now, they might only investigate demographics about the users but do not consider the user's self-presentation online. Furthermore, we made a theoretical contribution by including a larger dataset, by performing computational methods and by exploring multiple online brand communities from different sectors.

Keywords: online brand communities, social identities, engagement behavior

INTRODUCTION

For years, social network sites proved to be suitable platforms for online brand community managers to build and maintain long-term relationships with their consumers (Lee, Kim, and Kim 2011; Zaglia 2013). Through online brand communities, these managers are able to advertise, promote and communicate offerings to potential buyers and influence the perceptions and behavior of the consumers (Muniz and O'Guinn 2001). Conversely, consumers benefit from being engaged in online brand communities. For instance, they can satisfy their social needs by interacting with like-minded individuals. Also, consumers can preserve a positive (social) identity by creating associations through the content of online brand communities (Belk 1988, 2013; Muniz and O'Guinn 2001).

Although online brand communities are often studied, there is still a lack of research with the focus on the self-concept (Belk 2015; Mayshak et al. 2017). In this research area, social identity proved to be a key determinant of engagement behavior in online brand communities (Lee et al. 2011). To clarify, social identity describes how individuals identify their selves as members of online brand communities. Moreover, this identification process proved to strengthen the motivation to engage in activities of the community (Bagozzi and Dholakia 2006). We contribute to this concept by describing how different social identities of individuals engage in the online brand.

Another contribution to research is that we performed computational methods such as Machine Learning and Natural Language Processing. This is important because engagement behavior in the online brand community is often studied using qualitative methods (Brodie et al. 2011) while computational methods enable the researcher to study a larger and more geographically diverse community (Belk 2013). Also, former research often studied relatively small online brand communities. With computational methods, we are able to examine larger online brand communities across different product categories (Brodie et al. 2011).

Twitter is a rich data source when performing computational methods like Machine Learning and Natural Language Processing (Priante et al. 2016). Unlike other popular social network sites such as Facebook and Instagram, Twitter is a microblog which means that users are restricted to post small bits of content. These small bits of content are called tweets and have a maximum of 280-characters (Ibrahim, Wang, and Bourne 2017). In former research, tweets proved to be suitable for studying engagement behavior. For example, Ibrahim, Wang, and Bourne (2017) and Dessart, Veloutsou, and Morgan-Thomas (2015) used tweets for studying the emotional and behavioral dimensions of engagement. The emotional dimension of engagement was studied by performing a sentiment analysis on tweets (Angiuoli et al. 2011; Ibrahim et al. 2017). The behavioral dimension of engagement was indicated by calculating the frequency Twitter users interact with online brand communities (Hodas, Butner, and Corley 2016; Ibrahim et al. 2017; Jansen and Zhang 2009; Vargo 2016). In addition to researching tweets, social identity can be studied using the 160-character profile description of Twitter users (Priante et al. 2016). On these profile descriptions, Twitter users often present an essential expression of the social identity.

Because of the properties of Twitter, we used its data to explore how Twitter users with different presented social identities engage in online brand communities. For this, we formulated the following research question*: What is the relationship between the social identity of Twitter users and their engagement behavior in online brand communities on Twitter?* To answer the research question, we categorized Twitter users in different social identity groups. Next, we studied engagement behavior by analyzing the tweet sentiment and the frequency of interactions between Twitter users and online brand communities.

The following study is organized as follows: (1) reviewing literature about online brand communities, brand community engagement, social identity and self-presentation, (2) analyzing Tweets mentioning online brand communities via the Twitter API, (3) categorizing the profile description of Twitter users by using a social identity classifier and (4) measuring the engagement behavior of each Twitter user in terms of sentiment and frequency. Finally, (6) the results will be discussed for future research and marketing implications.

THEORETICAL FRAMEWORK

To answer the research question, the theoretical framework draws on prior studies on brand communities (Algesheimer, Dholakia, and Herrmann 2005; Muniz and O’Guinn 2001), social identity (Mousavi, Roper, and Keeling 2017; Tajfel and Turner 1986), classification of social identities on Twitter (Priante et al. 2016), engagement behavior in online brand communities (Brodie et al. 2011;

Dessart et al. 2015) and techniques to measure this engagement behavior (Ibrahim et al. 2017; Jansen and Zhang 2009).

Introduction to online brand communities on Twitter

First, we explain how online brand communities on Twitter are built and how Twitter users interact with these communities. In 2001, Muniz and O'Guinn (2001, 412) introduced the idea of online brand communities and described it as followed: "A specialized, non-geographically bound community, based on a structured set of social relationships among users of a brand". Lee et al. (2011) added to this theory that online brand communities are created by online brand community managers or by consumers. In both types of online brand communities, Twitter users interact, discuss, and share experiences with the community. By using the "@"-sign, the Twitter user can reply on tweets or mention usernames in their tweets (Marwick and Boyd 2011). This enables the users to start conversations with online brand community managers or like-minded users. Also, it is possible for users to repost tweets of other Twitter users. This is called retweeting. Through retweeting, Twitter users generate new content for their followers and can associate themselves with online brand communities.

Online brand community identification: Self-categorization and Social identity

The strength of the relationship between Twitter users and the online brand community depends on the identification with the brand community

(Algesheimer et al. 2005). To understand how Twitter users identify with online brand communities, studying the social identity is key (Mousavi et al. 2017). Social identity refers to the part of the self which is adopted from and defined by social groups (Tajfel and Turner 1986; Turner and Reynolds 2016). It refers to the knowledge of individuals that they are members of certain social groups or categories (Burgess 2002). These social groups vary from nations, organizations, and brand communities, to religion, and politics. These groups also differ from age to gender (Shamir, House, and Arthur 1993; Tajfel and Turner 1986). To feel part of these social groups, the individuals categorize themselves and others in groups. For this, individuals evaluate the similarities between the members in contrast to other social groups (Turner et al. 1992). These similarities are based on similar emotional involvement and a similar evaluation of the group. This categorization process is called self-categorization. Self-categorization explains that individuals ask themselves if they belong to the community. If this is the case, the individual agrees with the online brand community's norms, traditions, rituals, and objectives (Algesheimer et al. 2005). The study of Mousavi et al. (2017) shows that self-categorization is the most important step for group identification. If the individuals see themselves as members of the online brand community, the individuals determine if they feel a strong or weak connection with the online brand community. Afterward, they evaluate the association with the online brand community as positive or negative. This evaluation determines if the individuals want to be associated with the social group and if they want to use this group to preserve a positive social identity (Tajfel and Turner 1986).

Presenting online brand communities on Twitter

When individuals enhance their image by associating themselves with online brand communities, this refers to self-presentation. According to Goffman (1959) and Leary and Kowalski (1990), self-presentation refers to image management in order to control how others perceive the individual. In other words, which impression would the individual like to create. In 1988, Belk discovered that individuals create the desired impression by presenting possessions. With the rise of online media, Belk (2013) added the following ways to present possessions: dematerialization ("i.e.", virtual possessions as music on Spotify or objects in games), re-embodiment ("i.e.", physical appearance through avatars, photos and videos), sharing, distributed memory ("i.e.", narratives of the self through Facebook's timeline) and co-construction of self ("i.e.", aggregate possessions like shared messages composed by couples or groups). Following the study of Belk (2013), tweets composed in online brand communities are part of the co-construction of self. To clarify, the co-construction of the self refers to the constructing of the desired image with the help of others (Belk 2015). In online brand communities, this happens through interactions with other like-minded individuals. Together, individuals improve their self-presentation by posting and replying to text messages, photos, videos and geo-locations in the online brand community.

Classifying presented social identity on Twitter

An essential and direct presentation of the social identity can be classified in the profile description of Twitter users (Priante et al. 2016). To create the social identity classifier, Priante et al. (2016) used the social identity theory (Tajfel and Turner 1986) to define different social identity groups. In total, five social identity groups have been defined: relational, occupational, stigmatized, political and ethnic/ religious. In short, relational identity refers to self-definition based on relationships with other individuals or social roles (Sluss and Ashforth 2007). Occupational identity is the self-definition based on careers, avocations, interests, and hobbies (Phelan and Kinsella 2009; Skorikov and Vondracek 2011). Next, political identity defines the self-definition based on political preferences, parties, and groups, but also memberships of social movements or participating in collective action (Xiaomei and Shimin 2014). An ethnic or religious identity refers to self-definition based on being a member of ethnic or religious groups (Ysseldyk, Matheson, and Anisman 2010). Finally, stigmatized identity describes the self-definition based on being a member of a stigmatized group (Crocker and Major 1989). Or in other words, individuals with a stigmatized identity consider themselves different from what the society describes as normal, compared to the social and cultural norms.

Engagement behavior in the online brand communities

Studying engagement behavior is key when investigating how the different presented social identities of Twitter users relate to different behavior in online brand communities (Martínez-López et al. 2017). To define engagement behavior in online brand communities, Brodie et al. (2011, p. 3) created the following definition: *“ Engagement in a virtual brand community involves specific interactive experiences between consumers and the brand, and/or other* *members of the community"* In addition to this, *"Engagement is a multidimensional concept comprising cognitive, emotional, and/ or behavioral dimensions".* With Twitter data, it is possible to test the emotional and behavioral dimensions. To clarify, the emotional dimension of engagement refers to the level of emotions which individuals experience when they engage with online brand communities. To test the emotional dimension of engagement, recent studies analyzed sentiments (Angiuoli et al. 2011; Hung 2017; Ibrahim et al. 2017; Mayshak et al. 2017). By performing a sentiment analysis with the help of Natural Language Processing, the sentiment in tweets mentioning online brand communities can be evaluated as positive, negative or neutral (Vargo 2016). Next, the behavioral dimension of engagement will be explored. This dimension of engagement refers to the behavioral manifestation towards a brand, beyond purchase, and can be evaluated by calculating the frequency of interaction (Dessart et al. 2015). On Twitter, types of behavioral manifestations are posting or retweeting content. By combining the emotional and behavioral dimensions of engagement, engagement behavior can be investigated by measuring the tweet sentiment and calculating the frequency of interactions between the Twitter users and the online brand communities (Ibrahim et al. 2017).

Research model

This study is interested in understanding engagement behavior in online brand communities of Twitter users from different social identity groups. To construct the research model, the most important founding from the theoretical framework will be summarized.

Following the social theory and self-categorization theory, Algesheimer et al. (2005) found that the strength of the relationship between the individuals and the online brand community is influenced by the identification with the online brand community. This identification is based on the community its norms, traditions, rituals, and objectives. In the study of Algesheimer et al. (2005), qualitative data from offline car communities were studied. However, by using computational methods, we can study a larger amount of online brand communities from different sectors. With the use of Twitter data, we have the possibility to explore if Twitter users with similar social identities mention similar online brand communities. We expect that Twitter users who present similar social identities agree with similar norms, traditions, rituals, and objectives and therefore mention similar online brand communities (Algesheimer et al. 2005).

Proposition 1: There is a relationship between how many times Twitter users of a social identity group mentions an online brand community and the type of online brand community they mention.

The social identity of individuals determines if they want to be associated with the online brand community (Mousavi et al. 2017). On Twitter, users associate themselves with social groups by mentioning them in tweets. To contribute to the study of Mousavi et al. (2017), we use computational methods to study the sentiment in tweets instead of qualitative measurements. By performing a sentiment analysis, using natural language processing, we can automatically analyze the sentiment of a huge number of tweets. Afterward, this data is used to investigate the relationship between the presented social identity of Twitter users and the sentiment in tweets mentioning online brand communities.

Proposition 2a: There is a relationship between the social identity presented by Twitter users and the difference in sentiment when engaging in online brand communities.

Proposition 2b: Sentiment scores vary among tweets which mention different online brand communities.

Proposition 2c: There is an interaction effect in Tweet sentiment between the type of social identity and the type of online brand communities.

Apart from differences in tweet sentiment, we also expect to find a relationship between the presented social identity of Twitter users and the frequency of interaction with online brand communities. We derived this from the study Algesheimer et al. (2005). They found out that the more individuals engage with online brand communities, the more the brand fits their social identity. As stated above, a contribution to the study of Algesheimer et al. (2005) can be made by performing computational methods to analyze a larger dataset including a larger number of online brand communities from different sectors. With Twitter data, it is possible to quickly analyze a large number of interactions between Twitter users within an online brand community.

Proposition 3a: There is a relationship between the social identity presented by Twitter users and the frequency they interact with online brand communities.

Proposition 3b: The frequency of mentioning an online brand community on Twitter varies among different online brand communities.

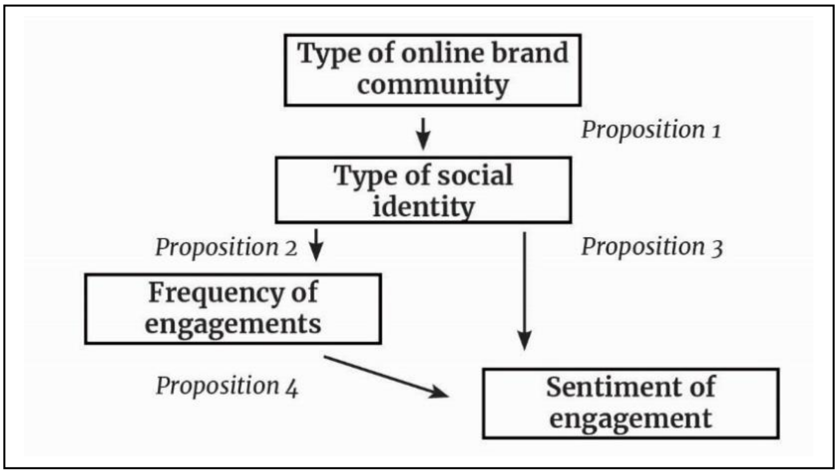
Proposition 3c: There is an interaction effect in the frequency of interaction between type of social identity and the type of online brand communities

Since we studied the tweet sentiment and frequency of interactions with online brand communities, we also aim to investigate the relationship between sentiment and frequency. Recent studies discovered that a high number of interactions with an online brand community positively affects the tweet sentiment (Dessart et al. 2015; Ibrahim et al. 2017). We contribute to these studies, by using a larger dataset with more different types of online brand communities which also operate in different sectors.

Proposition 4: Twitter users who frequently mention online brand communities in their tweets, tweet more positive about these communities.

FIGURE 1:

RESEARCH DESIGN



METHOD

Research design

The main goal of this study is to investigate the relationship between engagement behavior in online brand communities and the social identity presented by Twitter users. For this, a descriptive research will be conducted to describe the relationship. We are using quantitative data from Twitter which is retrieved from the Twitter API. These tweets will be analyzed with computational methods such as Machine Learning, and Natural Language Processing. To test the first proposition of this study, we count the number of times Twitter users of a social identity category mention an online brand community. By using this method, we aim to discover if the user's social identity relates to the online brand community they mention. The second proposition will be answered by calculating the sentiment of each tweet. We compare these scores among the different types of presented social identities and different types of online brand communities. To confirm the third proposition, we measure the frequency that a Twitter user mentioned an online brand community. These results are also compared to the different types of presented social identities and the different types of online brand communities. Finally, we answer the fourth proposition by categorizing the frequency a Twitter user mentioned an online brand community into the following groups: one time, light, medium and heavy. Among these groups, we compare the average tweet sentiment.

Procedure

Data collection

For the data collection, we used the Twitter API to collect tweets and profile descriptions of the Twitter users who posted the tweets. In the Twitter API, we applied searchTwitter to search for tweets which contain the "@"-sign for mentioning one of the online brand communities. Also, we limited our search to only English tweets. For each request to the Twitter API, we were able to receive 1,500 tweets. To prevent duplicates, we made a request once or twice a day. From each tweet, we extracted the username and used it as a tag to request profile descriptions from the Twitter API. In total, 27,143 tweets and 22,333 unique profile descriptions were collected over a period of four months (from January 2018 to April 2018).

Cleaning of the data

All retrieved tweets which did not fit the research model were deleted from the dataset. These tweets were duplicates or not written in English. Also, we deleted tweets which were posted by Twitter users which profile descriptions was not written in English or not fitted the conditions of the social identity classifier. The data was cleaned using R and Python. We deleted in each tweet and profile description the stop words, Unicode characters, retweet headers, URLs, unnecessary spaces and replaced capitalization.

Selecting online brand communities

Following the study of Jansen and Zhang (2009), we investigated online brand communities in which individuals can identify. For that reason, we selected leading online brand communities (see table 1) which produce or sell products or services for daily use. We assumed these online brand communities are most relatable for Twitter users and therefore highly mentioned on Twitter. Besides, we assumed that these online brand communities would have a high score on prestige (high status) and distinctiveness (brand uniqueness). According to Mousavi et al. (2017), when online brand communities score high on these attributes, this positively influence the identification with the online brand community. In total, five sectors (fashion, game consoles, sports clubs, fast food services and cars) were selected. For all the sectors, three brands were selected; one major online brand community, the competitor of the major online brand

TABLE 1

ONLINE BRAND COMMUNITIES ORDERED BY SECTOR

Sector	Leading twitter brand	Competitor	Leading brand different segment	Target group
Fashion	H&M @hm 8.71m followers	Topshop @Topshop 1.33m followers	Chanel @CHANEL 3.4m followers	Fashionable and trendy individuals. Group 1/2: upper middle class Group 3: upper class
Game consoles	PlayStation @PlayStation 15.1m	Xbox @Xbox 12.4m	Nintendo @NintendoAmerica 8.43m	Gamers. Group 1/2: older audience Group 3: younger audience
Sport clubs	Real Madrid CF @realmadrid 27.6m followers	FC Barcelona @FCBarcelona 25.7m followers	Los Angeles Lakers @Lakers 7.09m followers	Supporters of sports clubs. Group 1/2: soccer Group 1/2: basketball
Fast food services	McDonald's @McDonald's 3.52m followers	Burger King @Burger King 1.58m followers	Starbucks @Starbucks 11.9m followers	Individuals who eat fast-food. Group 1/2: Lower and middle working class (who eat fries and burgers) Group 3: middle and upper class (who drink coffee)
Cars	Tesla @Tesla 2.04m followers	BMW @BMW 1.54m followers	Toyota @Toyota 689k	Individuals who like cars. Group 1/2: upper class (cars with a business executives' attitude) Group 3: middle upper class (cars with a family size and a sporty attitude)

community and one online brand community of the same sector but with a different target group. In this way, we were able to compare major online brand communities with similar products or services or similar target groups. This makes the online brand communities more comparable (Jansen and Zhang 2009). To improve the comparability, we only compared online brand communities from the same sectors. To study if the results differ among other online brand communities, we calculated the mean in each sector and compared this to other sectors.

After the data collection and data cleaning, the dataset consisted of 27,143 tweets which mentioned an online brand community. The distribution of the online brand communities in this dataset was as followed: 8,7% of the tweets mentioned H&M (N=2,372), 6,7% of the tweets mentioned Topshop (N=1,823), 5,2% of the tweets mentioned Chanel (N=1,417), 8,6% of the tweets mentioned PlayStation (N=2,336), 9,8% of the tweets mentioned Xbox (N=2,671), 8,2% of the tweets mentioned Nintendo (N=2,214), 3,4% of the tweets mentioned Real Madrid CF (N=936), 2,9% of the tweets mentioned FC

FIGURE 2

DISTRIBUTION OF ONLINE BRAND COMMUNITIES PER SECTOR

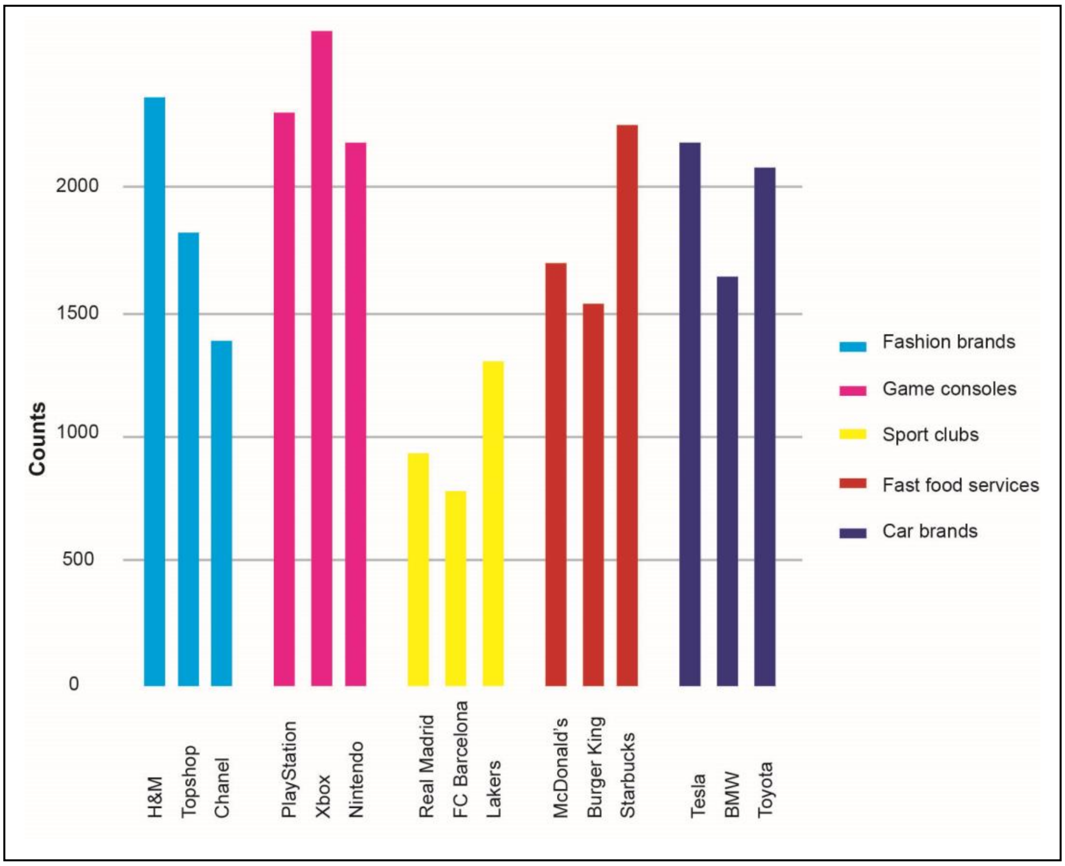


TABLE 2

CODEBOOK 5-CATEGORY SOCIAL IDENTITY CLASSIFIER

Social identity	Description	Example
Relational	Self-definitions based on relationships with other individuals and social roles.	*"Mother of baby boy", "Married to @Peter", "Lady Gaga fan", "Ajax is the best".*
Occupational	Self-definitions based on careers, avocations, interests, and hobbies.	*"Data specialist", "love to game and watch YouTube", "Like to eat nacho's and bike through the mountains".*
Political	Self-definition based on political preferences, parties, and groups, but also memberships of social movements or participating in collective action.	*"Feminist Activist", "Democrat", "Council candidate", "Mobro in #movember", "#BlackLivesMatter".*
Ethnic/religious	Self-definition based on being a member of ethnic or religious groups.	*"Loves Jesus", "#Christian", "Atheism", "Native New Yorker", "I am Dutch and German? "*
Stigmatized	Self-definition based on being a member of a stigmatized group.	*"Call me an affectionate idiot", "Okay to call me a dork", "Idiot savant", "#Workaholic woman with ADHD", "I am a nerd".*

Barcelona (N=775), 4,8% of the tweets mentioned Los Angeles Lakers (N=1,304), 7,8% of the tweets mentioned McDonald's (N=2121), 6% of the tweets mentioned Burger King (N=1622), 7,5% of the tweets mentioned Starbucks (N=2023), 6,3% of the tweets mentioned Tesla (N=1,704), 5,7% of the tweets mentioned BMW (N=1552) and 8,4% of the tweets mentioned Toyota (N=2,273).

Classification of social identities

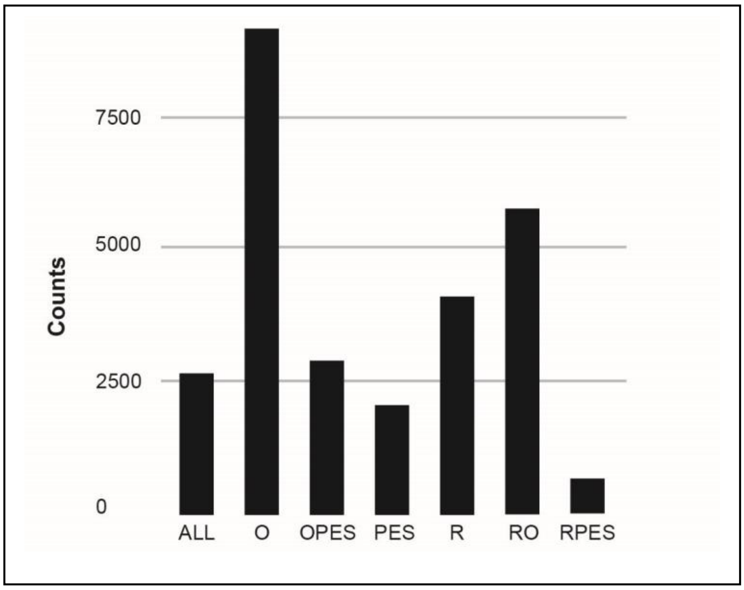
In this study, we used the 5-category social identity classifier. This classifier is developed by Priante et al., (2016) and explicitly categorizes Twitter users in terms of social identity. The social identities, which were presented in

Twitter user's profile descriptions, were categorized as followed: relational identity, occupational identity, political identity, stigmatized identity and ethnic/ religious identity (see table 2). According to Deaux, Reid, Mizrahi, and Ethier (1995), politicized, stigmatized, and religious or ethnic groups behave differently than relational and occupational groups. This is because these groups are action-oriented instead of social statuses like relational and occupational groups. Because of the overlapping behavior of Twitter users with a political, stigmatized and ethnic/ religious identity, Priante et al. (2016) merged those groups into one category called PES. By merging those groups, Priante et al. (2016) managed to improve the predictive performances of the classifiers. The same approach is used in this study.

Our classifier was trained using the codebook and training set (a set of examples which fit the codebook) of Priante et al. (2016). After the classifier

FIGURE 3

DISTRIBUTION OF SOCIAL IDENTITIES CATEGORIES



NOTE. − ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity, OPES = occupational PES identity.

was trained, the classifier was conducted on the retrieved profile descriptions. The classifier labels each profile description as a binary variable (True/False, 1/0) for each social identity category. It is possible that the profile descriptions fit into multiple categories. To distinguish the mixed profile descriptions, this study labeled each combination as followed: NO (not fitted conditions of classifier), R (relational identity), O (occupational identity), PES (political, ethnic/ religious and stigmatized identity), RO (relational-occupational identity), RPES (relational-PES identity), OPES (occupational-PES identity), ALL (all social identities presented). Next, a TF-IDF weighting (Salton and Buckley 1988) was used to measure the importance of a word in the profile descriptions. By combining this result with the Chi-Square feature, the term which correlated the most with a certain social identity was selected. Logistic Regression is the classification algorithm which we used for this study. In comparison to the algorithms Bernoulli Naïve, Random Forest, and Polynomial kernel, Logistic Regression is the most precise and complete classifier (Priante et al. 2016). Overall, Logistic Regression scores a good to excellent result for classifying relational (F=0.735) and occupational (F= 0.803) identities. Only, an acceptable score appeared when classifying the PES identity (F=0.604).

In total, 22,333 profile descriptions contained a presentation of the social identity which fitted the conditions of the classifier. The distribution of these social identities were as followed: 15,1% of the Twitter users presented a relational identity (N=4,097), 33,7% of the Twitter users presented an occupational identity (N=9,145), 7,5% of the Twitter users presented a PES identity (N=2,039), 10,5% of the Twitter users presented an occupational-PES identity

(N=2,845), 21,2% of the Twitter users presented a relational-occupational identity (N=5,751), 2,4% of the Twitter users presented a relational-PES identity (N=644) and 9,7% of the Twitter users fitted all social identity groups (N=2,622).

Sentiment analysis

To understand the engagement behavior of Twitter users, we studied the sentiment in tweets and frequency of interaction. The sentiment in tweets was analyzed to explore the emotional dimension of engagement behavior. To measure the sentiments in tweets, a sentiment analysis was conducted using the AFINN sentiment lexicon. We selected the AFINN sentiment lexicon after reviewing literature about the following sentiment lexicons: General Inquier (11,789 words), Micro-WNOp (1960 words), Opinion Lexicon (6786 words), SenticNet (15,143 words), SentiSence (4404 words), SentiWordNet (155,287 words) , SO-CAL (6306 words), Subjectivity Lexicon (8221 words), WordNetAffect (4552 words) and Hedonometer (10,000 words) (Cho et al. 2014). The main reason for selecting the AFINN sentiment lexicon is because the lexicon is created from Twitter data and includes internet slang, like WTF, LOL, and ROFL (Nielsen 2011). Another reason for using this lexicon is its high accuracy score (75%) in comparison to the other sentiment lexicons (Cho et al. 2014; Sharma, Srinivas, and Balabantaray 2015).

The lexicon is built by Nielsen (2011) and contains 2,477 unique words. All these words contain a score ranging from negative five (very negative) to five (very positive). To measure the tweet sentiment, we calculated for each tweet the mean sentiment score.

Frequency of interaction

Next, we measured the frequency of interaction to explore the behavioral dimension of engagement behavior. In this study, we used two different methods to investigate the frequency of interactions. The first method was used to test proposition 3 by calculating the average frequency score and compare this score between the different online brand communities, different sectors, and different social identity groups. The second method was used to test proposition 4 by categorizing the frequency scores into the following groups: one time (one tweet), light (two - five tweets), medium (six - twenty tweets), heavy (21 or more tweets) (Tumasjan et al. 2010). Then, we compared the tweet sentiment among the frequency groups.

Analysis

The chi-square test was performed to study proposition 1. In this proposition, we proposed the relationship between the number of times an online brand community is mentioned in a social identity group and the type of online brand community they mention. With the chi-square test, we test if the two categorical variables are associated with each other (Teetor 2011). Next, we proposed in proposition 2 and 3 that the tweet sentiment, but also the frequency of interaction, differs among different social identities groups and different types of online brand communities. The One-Way ANOVA-test was applied to test if the means of these groups are significantly different (Casella, Fienberg, and Olkin 2006). To investigate which groups are significantly different, the Bonferroni correction was performed. After this step, we studied if there is an interaction effect between the online brand communities and the social identity groups. For this, the Two-Way ANOVA test was performed to test the difference in means between two independent groups. At last, we performed the One-Way ANOVA test and Bonferroni correction to study proposition 4. In this proposition, we proposed the relationship between the frequency of interaction and the tweet sentiment.

RESULTS

In short, the dataset of this study includes information about the Twitter user's social identity and engagement behavior in online brand communities. We explain engagement behavior by analyzing data about the frequency Twitter users mention online brand communities in their tweets and the average sentiment of these tweets. In this dataset, 13,020 tweets were positive, 9,157 tweets were neutral, and 4.966 tweets were negative. Concluding, Twitter users mentioned the online brand communities more positive than negative. Moreover, Twitter users mentioned online brand community 19,672 one time, 1,009 medium, 6,002 light and 460 heavy. In view of this, most Twitter users mentioned an online brand community one time. On average, the Twitter users lightly interact with online brand communities (M=2.76, SD=9.18) and mentioned

TABLE 3

SUMMARY OF SIGNIFICANCE SCORES FOR SOCIAL IDENTITY AND ONLINE BRAND COMMUNITY

Sectors	*df*	X^2	*p*
Fashion	12	108.86	0.001
Games	12	52.74	0.001
Sports	12	33.50	0.001
Fast food	12	44.89	0.001
Cars	12	100.96	0.001
sectors combined	24	700.26	0.001

NOTE. – Mean effects between different social identity groups for each sector. Df = degrees of freedom, X^2 = *chi-square*, *p* = *p-value.*

these communities slightly positive (M=.86, SD=2.07). The online brand community which was mentioned the most by Twitter users was Xbox (N=2671). Furthermore, most Twitter users presented an occupational identity (N=5751).

How social identity groups vary in mentioning online brand community

The relationship between the number of times an online brand community was mentioned by a social identity group and the type of online brand community that was mentioned is studied by performing the chi-square test (see table 3). We found that the number an online brand community was mentioned by a social identity group is significant different between all online brand communities. This result supports proposition one. In table 4 it appeared that most online brand communities were mentioned by Twitter users who presented an occupational identity. On the contrary, all online brand communities were the least mentioned by Twitter users who presented a relational-PES identity.

TABLE 4

NUMBER OF TIMES SOCIAL IDENTITY GROUPS MENTION ONLINE BRAND COMMUNITIES

Sector	OBCs	ALL	O	O PES	PES	R	RO	R PES
Fashion	H&M	186	841	237	171	403	475	59
	Topshop	74	847	230	116	203	329	24
	Chanel	74	536	140	91	237	306	33
		334	2224	607	378	843	1110	116
Games	PlayStation	234	815	272	127	336	522	30
	Xbox	339	861	299	142	384	607	39
	Nintendo	172	781	247	152	352	458	52
		745	2457	818	421	1072	1587	121
Sports	Real Madrid	70	267	57	70	236	206	30
	FC Barcelona	68	160	43	80	206	186	32
	Lakers	103	346	102	136	267	298	52
		241	773	202	286	709	690	114
Fast food	McDonald's	253	620	247	228	299	401	73
	Burger King	152	481	154	153	252	390	40
	Starbucks	242	640	212	182	330	351	66
		647	1741	613	563	881	1142	179
Cars	Tesla	210	695	188	115	121	358	17
	BMW	142	566	158	131	192	323	40
	Toyota	303	689	259	145	279	541	57
		655	1950	605	391	592	1222	114

NOTE. − ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity, OPES = occupational PES identity.

How the sentiment in tweets differs among online brand communities and social identity groups

The One-Way ANOVA-test was applied as exploratory tool to understand the relationship between tweet sentiment and the social identities of Twitter users (see table 5). There are three sectors which supported this proposition,

TABLE 5

ANALYSIS OF VARIANCES (ANOVA) TEST AND BONFERONNI TEST FOR SENTIMENT SCORES IN SOCIAL IDENTITY GROUPS

Sector	One-Way ANOVA			Bonferonni	
	df	*f*	*p*	Social identity	*p*
Fashion	6	3.57	<.001	O-ALL	.047
				O-PES	.007
Games	6	1.37	.222	-	
Sports	6	1.61	.141	-	
Fast food	6	2.88	.008	RO-PES	.011
Cars	6	9.33	<.001	OPES-ALL	.004
				OPES-O	.004
				OPES-R	.003
				OPES-RO	<.001

NOTE. – Results from One-Way ANOVA Test and Bonferroni test to study the significance scores of sentiment scores between social identity groups. For the Bonferroni test, only significant scores have been denoted in table. *Df* = degrees of freedom, *f* = f-score, *p* = p-score, ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity, OPES = occupational PES identity.

namely the fashion, game consoles and cars sector. In these sectors, there was a significant difference in tweet sentiment between the social identity groups. To investigate which groups were significantly different, the Bonferroni correction was performed. As shown in table 5 and table 6, Twitter users with an occupational identity tweet significantly less positive about online brand communities in the fashion sector than Twitter users who presented all the social identities and Twitter users who presented a PES identity. Conversely, Twitter users who presented a relational-occupational identity tweeted significantly less positive in the fast-food service than users who presented a PES identity. In the car sector, Twitter users who presented an occupational-identity tweeted significantly more positive than other users.

TABLE 6

THE AVERAGE SENTIMENT SCORES OF EACH SOCIAL IDENTITY GROUP PER SECTOR

Sector	Social Identity	*M*	*SD*
fashion	ALL	.57	2.22
	O	.94	1.99
	OPES	.78	2.03
	PES	.53	2.15
	R	.79	2.03
	RO	.73	2.20
	RPES	.83	2.04
		.81	2.07
games	ALL	1.05	1.95
	O	1.06	1.92
	OPES	1.04	1.90
	PES	.87	2.01
	R	1.02	2.00
	RO	1.15	1.94
	RPES	1.16	1.78
		1.06	1.94
sports	ALL	1.00	2.19
	O	.76	2.15
	OPES	1.04	2.11
	PES	.75	2.15
	R	.83	2.18
	RO	.90	2.11
	RPES	1.27	2.06
		.86	2.15
fast food	ALL	.46	2.16
	O	.41	2.14
	OPES	.34	2.7
	PES	.15	2.24
	R	.48	2.23
	RO	.54	2.15
	RPES	.12	.21
		.41	2.19
cars	ALL	1.20	2.01
	O	1.13	1.92
	OPES	.79	2.12
	PES	.96	2.04
	R	1.21	1.97
	RO	1.26	1.93
	RPES	.90	1.99
		1.12	1.98

NOTE. − Mean and standard deviation scores for each social identity group per sector. *M = Mean, SD = Standard Deviation,* ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity, OPES = occupational PES identity.

TABLE 7

ANALYSIS OF VARIANCES (ANOVA) TEST AND BONFERONNI TEST FOR SENTIMENT SCORES IN ONLINE BRAND COMMUNITIES

Sector	One-Way ANOVA			Bonferonni	
	df	*f*	*p*	Social identity	*p*
Fashion	2	35.23	<.001	HM-CH	< .001
				TS-HM	< .001
Games	2	89.36	<.001	PS-NI	< .001
				XB-NI	< .001
				XB-PS	< .001
Sports	2	26.24	<.001	FC-LA	< .001
				FC-RM	< .001
				LA-RM	.020
Fast food	2	49.35	<.001	SB-BK	< .001
				SB-MC	< .001
Cars	2	25.91	<.001	TY-BMW	< .001
				TY-TL	< .001
				TL-BMW	.011
All sectors com-pared	4	110.1	<.001	CAR-FASHION	<.001
				CAR-FOOD	<.001
				CAR-SPORTS	<.001
				FASHION-GAMES	<.001
				FASHION-FOOD	<.001
				FOOD-GAMES	<.001
				FOOD-SPORTS	<.001
				GAMES-SPORTS	<.001

NOTE. – Results from One-Way ANOVA Test and Bonferroni test to study the significance scores of sentiment scores between the different online brand communities per sector. Also, compared the overall scores between the sectors. For the Bonferroni test, only significant scores have been denoted in table. *Df* = degrees of freedom, *f* = f-score, *p* = p-score, *HM* = H&M, TS = Topshop, CH = Chanel, PS = PlayStation, XB = Xbox, NI = Nintendo, RM = Real Madrid, FC = FC Barcelona, LA = Lakers, MC = McDonald's, BK = Burger King, SB = Starbucks, TL = Tesla, BMW = BMW and TY = Toyota.

Next, we performed the One-Way ANOVA test to study the relationship between tweet sentiment and online brand communities. It appeared in table 7 that the difference in tweet sentiment is significant between online brand communities in all sectors. This result supports proposition 2b. In table 8 is shown

TABLE 8:

AVERAGE SENTIMENT SCORES OF ONLINE BRAND COMMUNITIES PER SECTOR

Sector	Online brand communities	*M*	*SD*
Fashion	H&M	.54	2.13
	Topshop	.98	1.97
	Chanel	1.04	2.05
		.81	2.07
Games	PlayStation	.65	1.95
	Xbox	1.37	1.89
	Nintendo	1.12	1.91
		1.06	1.94
Sports	Real Madrid	.85	2.22
	FC Barcelona	1.31	2.00
	Lakers	.61	2.13
		.86	2.15
Fast food	McDonald's	.23	2.23
	Burger King	.17	2.28
	Starbucks	.79	2.00
		.41	2.19
Cars	Tesla	.88	2.13
	BMW	1.08	1.94
	Toyota	1.33	1.86
		1.12	1.98

NOTE. – Mean and standard deviation scores for each online brand community per sector. *M = Mean, SD = Standard Deviation.*

that all online brand communities in the game console, sports club, and car sector significantly differ from each other. In these sectors, Twitter users posted on average the most positive about Starbucks, Xbox and FC Barcelona. Chanel scored on average the highest tweet sentiment in the fashion sector. This score

TABLE 9

TWO-WAY ANOVA TO TEST FOR INTERACTION EFFECTS BETWEEN THE SOCIAL IDENTITY GROUPS AND OBCS WITHIN EACH SECTOR

Sector	Two-Way ANOVA		
	df	*f*	*p*
Fashion	12	8.90	<.001
Games	12	1.31	.206
Sports	12	1.71	.059
Fast food	12	1.77	.046
Cars	12	3.56	<.001
All sectors compared	24	2.24	<.001

NOTE. – Results from Two-Way ANOVA Test to study the interaction between online brand communities and social identity groups based on sentiment scores. Df = degrees of freedom, f = f-score and p = p-score.

was significantly different from H&M. Next, Starbucks was significantly different from Burger King and McDonald's and scored on average the highest sentiment for the fast food sector. From all the online brand communities, Xbox scored the highest tweet sentiment and Burger King scored the lowest sentiment.

To test proposition 2c, we performed the Two-Way ANOVA test. By using this test, we showed an interaction effect in tweet sentiment between the online brand communities and social identity groups in the following sectors: fashion, fast food services, and cars (see table 9). In the fashion sector, Twitter users presenting an occupational identity tweeted the most positive (see table 10). Twitter users who presented a relational-occupational identity tweeted the most positive about online brand communities in the fast food sector and in the car sector. On the contrary, users who presented a kind of PES identity (PES in combination with relational or occupational identity) tweeted the least positive in all sectors.

TABLE 10

HIGHEST AND LOWEST SENTIMENT SCORES IN ONLINE BRAND COMMUNITIES BASED ON THE AVERAGE OF EACH SOCIAL IDENTITY GROUP

Sector	OBCs	Highest frequency scores			Lowest frequency scores		
		Social identity	*M*	*SD*	*Social identity*	*M*	*SD*
Fashion		O	.94	1.99	PES	.53	2.15
	H&M	RO	.89	2.21	PES	-.11	2.24
	Topshop	O	1.10	1.91	R	.60	2.15
	Chanel	R	1.31	1.94	RO	.17	2.27
Games		RPES	1.16	1.78	PES	.87	2.01
	PlayStation	RO	.86	1.99	RPES	.13	1.85
	Xbox	RPES	1.72	1.75	O	1.31	1.90
	Nintendo	RPES	1.33	1.54	PES	.89	2.01
Sports		RPES	1.27	2.06	PES	.75	2.15
	Real Madrid	ALL	1.23	2.11	PES	.36	2.11
	FC Barcelona	RPES	2.03	1.64	RO	1.09	2.06
	Lakers	RPES	1.15	2.02	R	.36	2.12
Fast food		RO	.54	2.15	RPES	.12	2.21
	McDonald's	RO	.45	2.20	PES	.06	2.25
	Burger King	RO	.51	2.21	PES	-.36	2.20
	Starbucks	R	1.03	1.85	RPES	.26	2.17
Cars		RO	1.26	1.93	OPES	.79	2.12
	Tesla	RO	.99	2.08	RPES	-.76	1.71
	BMW	RPES	1.75	1.66	OPES	.77	2.06
	Toyota	RO	1.57	1.75	OPES	.76	2.11

NOTE. − We select the highest and lowest mean of sentiment scores of social identities for each online brand community. *M = Mean, SD = Standard Deviation,* ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity and OPES = occupational PES identity.

How the frequency of an interaction differs among online brand communities and social identity groups

TABLE 11

ANALYSIS OF VARIANCES (ANOVA) TEST AND BONFERONNI TEST TO STUDY FREQUENCY SCORES BETWEEN SOCIAL IDENTITY GROUPS

Sector	One-Way ANOVA			Bonferonni	
	df	*f*	*p*	Social identity	*p*
Fashion	6	61.43	<.001	RO-ALL	< .001
				RO-O	< .001
				RO-OPES	< .001
				RO-PES	< .001
				RO-R	< .001
				RO-RPES	< .001
Games	6	16.59	<.001	ALL-O	< .001
				ALL-OPES	< .001
				ALL-PES	< .001
				ALL-R	< .001
				ALL-RO	< .001
				ALL-RPES	< .001
				O-PES	< .001
				O-R	< .001
				O-RO	.020
Sports	6	3.54	.002	RO-ALL	.003
Fast food	6	65.47	<.001	RO-ALL	< .001
				RO-O	< .001
				RO-PES	< .001
				RO-R	< .001
				RO-RPES	< .001
				RO-OPES	.002
Cars	6	9.33	<.001	OPES-ALL	.002
				OPES-O	< .001
				OPES-PES	.011
				OPES-R	<.001
				OPES-RPES	.003
				RO-R	.029
				RO-PES	<.001

each social identities group per sector. *M = Mean, SD = Standard Deviation,* ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity and OPES = occupational PES identity.

As shown in table 11, the One-Way ANOVA test showed a significant difference in the frequency of interaction between the social identity groups in all sectors. This result is in support of proposition 3a. It appeared in table 12, that in the sectors fashion, sports club and fast food services, Twitter users who

TABLE 12

AVERAGE SCORES BETWEEN SOCIAL IDENTITY GROUPS BASED ON FREQUENCY SCORE

Sector	Social identity	*M*	*SD*
Fashion	ALL	1.76	2.27
	O	2.96	8.20
	OPES	2.84	4.88
	PES	1.55	1.80
	R	1.81	2.52
	RO	11.57	29.50
	RPES	1.96	1.96
		4.29	14.7
Games	ALL	3.82	8.24
	O	2.68	6.82
	OPES	2.32	3.33
	PES	1.44	.89
	R	1.84	2.27
	RO	2.12	3.32
	RPES	1.21	.53
		2.41	5.26
	ALL	1.81	2.17
Sports	O	1.87	2.56
	OPES	1.97	2.01
	PES	1.56	1.71
	R	1.82	1.88
	RO	2.00	2.58
	RPES	1.57	1.25
	ALL	1.76	2.27
		.86	2.15
Fast food	ALL	1.21	.60
	O	1.34	1.12
	OPES	1.37	1.09
	PES	1.41	1.13
	R	1.46	1.89
	RO	8.70	24.98
	RPES	1.99	2.63
		2.83	11.55
Cars	ALL	2.06	3.44
	O	1.93	3.53
	OPES	2.84	5.23
	PES	1.60	1.24
	R	1.89	2.59
	RO	2.46	3.64
	RPES	1.47	.84
		2.13	3.56

NOTE. – Mean and standard deviation of frequency scores for each social identities group per sector. *M = Mean, SD = Standard Deviation,* ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity and OPES = occupational PES identity.

TABLE 13:

ANALYSIS OF VARIANCES (ANOVA) TEST AND BONFERONNI TEST TO STUDY SENTIMENT SCORES BETWEEN ONLINE BRAND COMMUNITIES

Sector	One-Way ANOVA			Bonferonni	
	df	*f*	*p*	Social identity	*p*
Fashion	2	12.02	<.001	HM-CH	< .001
				HM-TS	< .001
				CH-TS	< .001
Games	2	67.06	<.001	PS-NI	< .001
				PS-XB	< .001
				NI-XB	< .001
Sports	2	28.91	<.001	FC-LA	<.001
				FC-RM	<.001
				LA-RM	<.001
Fast food	2	123.62	<.001	BK-MC	< .001
				BK-SB	< .001
Cars	2	18.66	<.001	TL-BMW	<.001
				TL-TY	< .001
All Sectors Compared	4	56.60		CAR-FASHION	<.001
				CAR-FOOD	<.001
				FASHION-FOOD	<.001
				FASHION-GAMES	<.001
				FASHION-SPORTS	<.001
				SPORTS-FOOD	<.001
				SPORTS-GAMES	.026

NOTE. – Results from One-Way ANOVA Test and Bonferroni test to study per sector the significance between online brand communities based on frequency scores. Afterward, we compared the overall scores between the sectors. For the Bonferroni test, only significant scores have been denoted in table. *Df* = degrees of freedom, *f* = f-score, *p* = p-score*, HM* = H&M, TS = Topshop, CH = Chanel, PS = PlayStation, XB = Xbox, NI = Nintendo, RM = Real Madrid, FC = FC Barcelona, LA = Lakers, MC = McDonald's, BK = Burger King, SB = Starbucks, TL = Tesla, BMW = BMW and TY = Toyota.

presented a relational-occupational identity mentioned online brand communities the significantly more frequent (see table 12). Twitter users who presented all the social identities mentioned online brand communities from the game console sector significant more frequent. In the car sector, Twitter users who presented an occupational-PES identity mentioned online brand communities the most frequent.

TABLE 14

THE DIFFERENCE BETWEEN OBCS AND THE FREQUENCY THOSE OBCS ARE MENTIONED IN TWEETS

Sector	OBCs	Mean	SD
Fashion	H&M	1.67	2.62
	Topshop	3.93	9.49
	Chanel	9.13	26.34
		4.29	14.70
Games	PlayStation	3.32	7.40
	Xbox	2.33	4.93
	Nintendo	1.54	1.30
		2.41	5.26
Sports	Real Madrid	1.75	2.31
	FC Barcelona	1.38	1.01
	Lakers	2.12	2.51
		1.81	2.17
Fast food	McDonald's	1.39	2.96
	Burger King	6.58	21.01
	Starbucks	1.34	1.38
		2.83	11.55
Cars	Tesla	1.70	2.31
	BMW	2.41	4.23
	Toyota	2.25	3.79
		2.13	3.56

NOTE. – Mean and standard deviation scores based on frequency score for each online brand community per sector. *M = Mean, SD = Standard Deviation.*

Another One-Way ANOVA test was performed to test the relationship between the frequency of mentioning online brand communities and the type of online brand community. The result of this test showed a significant difference in the frequency of interaction among the different online brand communities for all sectors (see table 13). In view of this (see table 14), we discovered that the fashion sector scored on average the highest frequency score, due to the high frequency score of Chanel. Conversely, the sports club sector scored on average the lowest frequency score. From the Bonferroni correction, we conclude that all online brand communities in the fashion, game console, and sports club significantly differed from each other. In these sectors, the online brand communities Chanel (fashion sector), The Lakers (sports clubs sector) and PlayStation (game console sector) scored the highest frequency score in comparison to other communities of their sectors.

In table 15 the results from the Two-Way ANOVA are shown. We found that the frequency Twitter users mention an online brand community interacts with the social identity group of the users and the online brand communities they mentioned. In table 16 appeared that in the fashion, sports club and fast food services sector, Twitter users who presented a relational occupation identity tweeted the most frequent about online brand communities. Within the game console sector, Twitter users who presented all the social identities scored the highest frequency score. In the car sector, Twitter users who presented an occupational PES identity showed the highest frequency score. Surprisingly, in all sectors, Twitter users who presented a kind of PES identity (PES identity in

TABLE 15

RESULTS TWO-WAY ANOVA TEST INTERACTIONS BETWEEN SOCIAL IDENTITY GROUPS AND ONLINE BRAND COMMUNITIES BASED ON FREQUENCY SCORES

Sector	Two-Way ANOVA		
	df	*f*	*p*
Fashion	12	11.07	<.001
Games	12	8.05	<.001
Sports	12	7.40	<.001
Fast food	12	7.65	.<.001
Cars	12	14.22	<.001
All sectors compared	24	4.77	<.001

NOTE. − Results from Two-Way ANOVA Test to study the interaction between online brand communities and social identity groups based on frequency scores. Df = degrees of freedom, f = f-score and p = p-score.

TABLE 16

HIGHEST AND LOWEST FREQUENCY SCORES IN ONLINE BRAND COMMUNITIES BASED ON THE AVERAGE OF EACH SOCIAL IDENTITY GROUP

Sector	OBCs	Highest frequency scores			Lowest frequency scores		
		Social identity[1]	*Mean*	*SD*	*Social identity*	*Mean*	*SD*
Fashion		RO	11.57	29.50	PES	1.55	1.80
	HM	RPES	1.41	.72	PES	1.19	.39
	TS	OPES	1.71	.96	RPES	1.29	.46
	CH	RO	2.22	1.36	PES	1.26	.44
Games		ALL	3.82	8.24	RPES	1.21	.53
	PS	ALL	1.73	1.10	RPES	1.10	.31
	XB	ALL	1.44	.65	RPES	1.10	.31
	NI	OPES	1.37	.56	RPES	1.21	.41
Sports		RO	2.00	2.58	ALL	1.39	.96
	RM	O	1.36	.65	ALL	1.13	.34
	FC	RPES	1.28	.46	ALL	1.13	.34
	LA	OPES	1.64	.74	O	1.27	.52
Fast food		RO	8.70	24.98	ALL	1.21	.60
	MC	RPES	1.27	.61	RO	1.13	.38
	BK	RO	1.93	1.22	RPES	1.18	.38
	SB	RPES	1.35	.75	OPES	1.08	.28
Cars		OPES	2.84	5.23	RPES	1.47	.84
	TL	RO	1.35	.63	ALL	1.20	.48
	BMW	O	1.49	.76	PES	1.28	.45
	TY	OPES	1.60	.99	RPES	1.19	.40

NOTE. − We select the highest and lowest mean of frequency scores of social identities for each online brand community. *M = Mean, SD = Standard Deviation,* ALL = All the social identities presented, R = relational identity, O = occupational identity, PES = PES identity, RO = relational occupational identity, RPES = relational PES identity and OPES = occupational PES identity.

combination with relational or occupational identity) mentioned online brand communities the least frequent in their tweets.

The relationship between the frequency of interaction and tweet sentiment

Concluding from table 17, a significant difference in tweet sentiment is found between different frequency groups (one time, light, medium, heavy). For that

TABLE 17

AVERAGE SENTIMENT SCORE AMONG DIFFERENT FREQUENCY GROUPS

Frequency	*M*	*SD*
One time	.86	2.09
Light	.92	2.03
Medium	.71	1.94
Heavy	.50	2.02

NOTE. – Mean and standard deviation of sentiment scores for each frequency group. M= Mean and Df = degrees of freedom

reason, we support proposition 4. In addition, in table 18 is shown, that Twitter users who mention online brand communities heavily tweet significantly less positive about online brand communities than users who mention these communities lightly or one time. Next, Twitter users who mention online brand com-

TABLE 18

ANALYSIS OF VARIANCES (ANOVA) TEST AND BONFERONNI TEST TO STUDY SENTIMENT SCORES BETWEEN FREQUENCY GROUPS

	One-Way ANOVA			Bonferonni	
	df	*f*	*p*	Social identity	*p*
Frequency groups	3	8.09	<.001	Heavy-light	< .001
				Heavy-one time	< .001
				Light-medium	< .001

NOTE. – Results from One-Way ANOVA Test and Bonferroni test to study the significance between frequency groups based on sentiment scores. For the Bonferroni test, only significant scores have been denoted in the table. *Df* = degrees of freedom, *f* = f-score and *p* = p-score.

munities' medium tweet significantly less positive about the online brand communities than users who mention the online brand communities lightly.

Discussion and implications

Several insights were found when studying the relationship between social identities presented by Twitter users and their engagement behavior in online brand communities. In this study, fifteen leading online brand communities on Twitter from five different sectors were explored. By using computational methods, we were able to indicate for each Twitter users which social identity they presented, the average tweet sentiment and the frequency of mentioning online brand communities in their tweets. By combining this data, we found that the tweet sentiment and the frequency of interaction varied not only between the different social identity types but also varied between the different types of online brand communities. To explain this, we further discuss the results per proposition.

First, we counted the occurrence of Twitter users mentioning online brand communities for each social identity group. Next, we compared this result among the different online brand communities. With this data, we found that the number of times Twitter users of a social identity group mentioned an online brand community significantly differed between the different types of online brand communities. Overall, Twitter users from the occupational identity group turned out to mention online brand communities far more than other social identity groups. Conversely, the relational-PES identity group turned out to mention online brand communities far less than other social identity groups. This result can be explained by the study of Priante et al. (2016). They suggest that most Twitter users present themselves in terms of an occupation or a hobby and mostly not in terms of political preferences, ethnicity, religion or a membership to a stigmatized group or social movement. Our study contributes to the study of Priante et al. (2016) by not only using a larger dataset but also by improving their 5-category social identity classifier. In addition, we performed this classifier specifically on Twitter users who interact with online brand communities. For this reason, our findings are useful for future marketing implications. For example, brand community managers should keep in mind that Twitter users mostly present an occupational identity and therefore should target their marketing implications on this social identity group. In addition to our findings, future studies should combine quantitative and qualitative data to discover if Twitter users with similar social identities have similar evaluations of the online brand community and if they have similar emotional involvement with this community. In this way, future studies could explore brand community identification in terms of social identity (Algesheimer et al. 2005; Kozinets 1997; Mousavi et al. 2017). Another contribution of this study can be made by studying multiple social network sites. So far, only Twitter data was used to classify social identities and study the relationship between different social identity groups and different types of online brand communities (Priante et al. 2016).

Next, we studied the engagement behavior of Twitter users in online brand communities based on the emotional and behavioral dimensions of engagement. We discovered that the presented social identity of Twitter users interacts with the type of online brand community when investigating the tweet sentiments. For instance, fashion brands were mentioned most positive by Twitter users with an occupational identity, game consoles and sports clubs were most positive mentioned by Twitter users with a relational-PES identity and fast food services and cars were mentioned most positive by Twitter users with a relational-occupational identity. Next, we found that the presented social identity of Twitter users also interacts with the type of online brand community when testing the frequency of interaction. We found that Twitter users who presented a relational-occupational identity mentioned fashion brands, sports clubs, and fast food services more often than other users, Twitter users who presented all the social identities mentioned games consoles more often and Twitter users who presented an occupational-PES identity mentioned car brands more often. On the contrary, in all sectors, Twitter users who presented a PES identity (also in combination with relational and occupational identity) not only mentioned the online brand communities the least positive but also mentioned the online brand communities least frequent. This finding contributes to the studies of Ren, Kraut, and Kiesler (2007) and Goffman et al. (1989). Unlike these studies, we found with computational methods that Twitter users who present a PES identity engage differently with the online brand community than Twitter users who present relational or occupational identities. We suggest Twitter users who presented a PES identity engage more negative because these users are more action-oriented than Twitter users who present social statuses (Broek, Tijs van den; Need, Ariana; Ehrenhard, Michel; Priante, Anna and Hiemstra 2015; Dessart et al. 2015; Goffman 1959; Ren et al. 2007). Also, Twitter users who presented a PES identity identify with stigmatized groups. This might explain why these users do not behave corresponding to the social and cultural norms and might want to evoke negative behavior. To test this, future research should combine qualitative and quantitative methods. This method should also be used to further investigate if individuals with similar social identities identify similar with online brand communities. To clarify, Algesheimer et al. (2005) found out that when individuals mention online brand communities more frequent, they identify more with the online brand community. This could explain why our findings show a relationship between the type of social identity and frequency score.

Moreover, our findings contribute to future marketing implications. By classifying the presented social identity, brand community managers enhance their strategies for personalized marketing. This is important because brand community managers should invest in long-term relationships with individuals by responding to the needs, wants or desires of the consumer (Vargo 2016). By investigating the presented social identity of an individual and how this relates to engagement behavior in the online brand communities, the online brand community manager acquires knowledge about which type of Twitter user identifies positive or negative with the online brand community. This knowledge can be used to enhance the personalizing of current marketing tools. Furthermore, our method is helpful for online brand community managers to monitor their competitors and to learn from these competitors (Jansen and Zhang 2009). For example, other online brand communities could learn from Chanel, Xbox, and Toyota in relation to the handling of consumer relationship. All three online brand communities show a positive and frequent relationship with the Twitter users. Especially, the online brand communities in the fast food sector should improve their relationship with the Twitter users, since Twitter users mention these communities less positive in comparison to other sectors.

Also, we researched how sentiment in tweets relates to the frequency a Twitter user mentions an online brand community. In contrast to the study of Ibrahim et al (2017), this study discovered that the more frequent Twitter users mention an online brand community, the more negative the sentiment in tweets is. This could be explained due to how complaints are handled in organizations. Ibrahim et al. (2017) suggest that Twitter users show a higher intensity of engagement when they are satisfied with the way organizations handle complaints to enhance bonding with the organization. In this case, negative tweets are considered as complaints or dissatisfaction. Further research should examine if tweets become more positive after a conversation in the online brand community. For this, a graph could be added to examine how negative tweets are handled from the beginning to the end of the conversation (Ibrahim et al. 2017).

Limitations and future studies

In this study, we came across some limitations. The first limitation is that we only analyzed online brand communities which are created by online brand community managers. Future studies should explore how different social identity groups engage in online brand communities which are created by Twitter users. This is important since Twitter users have different motives to engage in these communities (i.e. no profit exploitation, corporation image enhancement, discounts) (Lee et al. 2011). Also, we only investigated tweets which are written in English and tweets which mention online brand communities of Western Cultures. Future studies can extend their knowledge by including a more geographical approach. Next, the AFINN sentiment lexicon only consisted of 2,477 words. Due to this, some words in the tweets might not receive a score. This could be solved using other lexicons for analyzing the sentiment of tweets. Another limitation is that this study used a descriptive statistical approach. For that reason, we were not able to test if variables correlated with each other. Future studies should improve this.

CONCLUSION

By investigating how Twitter users behave in online brand communities in terms of emotional (the sentiment of the tweets) and behavioral dimensions (frequency of tweeting) (Brodie et al. 2011), we aimed to understand the engagement patterns of Twitter users in online brand communities. In most online brand communities, the communities were most often mentioned by Twitter users who presented an occupational identity. These online brand communities were the least mentioned by Twitter users who presented a combination of relational and PES identity. When studying the sentiment in tweets, we found out that the sentiment in tweets mentioning an online brand community relates to the presented social identity of Twitter users. The same result was found for the frequency that a Twitter user mentions the online brand community. When we combine both results, we conclude that Twitter users who present some kind of relational or occupational identity mention online brand communities more frequent and positive than Twitter users who present some kind of PES identity.

Next, we found that the more frequent Twitter users mention an online brand community the more negative the sentiment scores are on average. In conclusion of all results, the social identity groups interact with the online brand community when comparing the groups in terms of frequency and sentiment. For that reason, we suggest that there is a relationship between the presented social identity and the engagement behavior of Twitter users who mentioned online brand communities.

REFERENCES

Algesheimer, René, Utpal M Dholakia, and Andreas Herrmann (2005), "The Social Influence of Brand Community : Evidence from European Car Clubs," *Journal of Marketing*, 69(July), 19–34.

Angiuoli, Samuel V., James R. White, Malcolm Matalka, Owen White, and W. Florian Fricke (2011), "Resources and Costs for Microbial Sequence Analysis Evaluated Using Virtual Machines and Cloud Computing", *PLoS ONE*, 6(10).

Bagozzi, Richard P. and Utpal M. Dholakia (2006), "Antecedents and Purchase Consequences of Customer Participation in Small Group Brand Communities," *International Journal of Research in Marketing*, 23(1), 45–61.

Belk, Russell W. (2015), "Co-Construction of the Digital Self," *Advances in Consumer Research*, 43, 111–16.

——— (2013), "Extended Self in a Digital World," *Journal of Consumer Research*, 40(3), 477–500.

——— (1988), "Possessions and the Extended Self," *The Journal of Consumer Research*, 15(2), 139–68.

Brodie, Roderick J., Ana Ilic, Biljana Juric, and Linda Hollebeek (2011), "Consumer Engagement in a Virtual Brand Community: An Exploratory Analysis," *Journal of Business Research*, 66(1), 105–14.

Broek, Tijs van den; Need, Ariana; Ehrenhard, Michel; Priante, Anna and Hiemstra, Djoerd (2015), "The Influence of Prosocial Norms and Online Network Structure on Prosocial Behavior," *Social Media, Activism, and*

Organisations, 2566, 1–15.

Burgess, J. Peter (2002), “What’s So European About the European Union?,” *European Journal of Social Theory*, 5(4), 467–81.

Casella, George, Stephen Fienberg, and Ingram Olkin (2006), "*An Introduction to Statistical Learning"*, 102 Design.

Cho, Heeryon, Songkuk Kim, Jongseo Lee, and Jong Seok Lee (2014), “Data-Driven Integration of Multiple Sentiment Dictionaries for Lexicon-Based Sentiment Classification of Product Reviews,” *Knowledge-Based Systems*, 71, 61–71.

Crocker, Jennifer and Brenda Major (1989), “Social Stigma and Self-Esteem: The Self-Protective Properties of Stigma,” *Psychological Review*, 96(4), 608–30.

Deaux, Kay, Anne Reid, Kim Mizrahi, and Kathleen A. Ethier (1995), “Parameters of Social Identity,” *Journal of Personality and Social Psychology*, 68(2), 280–91.

Dessart, Laurence, Cleopatra Veloutsou, and Anna Morgan-Thomas (2015), “Consumer Engagement in Online Brand Communities: A Social Media Perspective,” *Journal of Product & Brand Management*, 24(1), 28–42.

Goffman, Erving (1959), *The Presentation of Self in Everyday Life*, Anchor Books, New York.

Hodas, Nathan O., Ryan Butner, and Court Corley (2016), “How a User???S Personality Influences Content Engagement in Social Media,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10046 LNCS, 481–93.

Hung, Chihli (2017), “Word of Mouth Quality Classification Based on Contextual Sentiment Lexicons,” *Information Processing & Management*, 53(4), 751–63.

Ibrahim, Noor Farizah, Xiaojun Wang, and Humphrey Bourne (2017), “Exploring the Effect of User Engagement in Online Brand Communities: Evidence from Twitter,” *Computers in Human Behavior*, 72, 321–38.

Jansen, Bernard. J. and Mini Zhang (2009), “Twitter Power: Tweets as Elyectronic Word of Mouth,” *Journal of the Association for Information Science and Technology*, 60(11), 2169–88.

Kozinets, Robert V (1997), “‘I Want To Believe’: A Netnography of The X-Philes’ Subculture of Consumption,” *Advances in Consumer Research*, 24, 470–475.

Leary, M R and Robin M Kowalski (1990), “Impression Management: A Literature Review and Two-Component Model,” *Psychological Bulletin*, 107(I), 34–47.

Lee, Doohwang, Hyuk Soo Kim, and Jung Kyu Kim (2011), “The Impact of

Online Brand Community Type on Consumer's Community Engagement Behaviors: Consumer-Created vs. Marketer-Created Online Brand Community in Online Social-Networking Web Sites," *Cyberpsychology, Behavior, and Social Networking*, 14(1–2), 59–63.

Martínez-López, F. J., Anaya-Sánchez, R., Molinillo, S., Aguilar-Illescas, R. & Esteban-Millat, I. (2017), "Consumer Engagement in an Online Brand Community," *Electronic Commerce Research and Applications*, 23, 24–37.

Marwick, Alice E. and Danah Boyd (2011), "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience," *New Media and Society*, 13(1), 114–33.

Mayshak, Richelle, Stefanie J Sharman, Lucy Zinkiewicz, and Alexa Hayley (2017), "Computers in Human Behavior The influence of Empathy and Self-Presentation on Engagement with Social Networking Website Posts," *Computers in Human Behavior*, 71, 362–77.

Mousavi, Sahar, Stuart Roper, and Kathleen A. Keeling (2017), "Interpreting Social Identity in Online Brand Communities: Considering Posters and Lurkers," *Psychology and Marketing*, 34(4), 376–93.

Muniz, Albert M. Jr. and Thomas C. O'Guinn (2001), "Brand Community," *Journal of Consumer Research, Inc.*, 27(4), 412–32.

Nielsen, Finn Årup (2011), "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs," *CEUR Workshop Proceedings*, 718(March 2011), 93–98.

Phelan, Shanon and Elizabeth Anne Kinsella (2009), "Occupational Identity: Engaging Socio-Cultural Perspectives," *Journal of Occupational Science*, 16(2), 85–91.

Priante, Anna, Djoerd Hiemstra, Tijs Van Den Broek, Aaqib Saeed, Michel Ehrenhard, and Ariana Need (2016), "#WhoAmI in 160 Characters? Classifying Social Identities Based on Twitter Profile Descriptions," in *EMNLP 2016, Workshop on Natural Language Processing and Computational Social Science*, 55–65.

Ren, Yuqing, Robert Kraut, and Sara Kiesler (2007), "Applying Common Identity and Bond Theory to Design of Online Communities," *Organization Studies*, 28(3), 377–408.

Salton, Gerard and Christopher Buckley (1988), "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, 24(5), 513–23.

Shamir, Boas, Robert J. House, and Michael B Arthur (1993), "The Motivational Effects of Charismatic Leadership: A Self-Concept Based Theory," *Organization Science*, 4(4), 577–94.

Sharma, Shashank, P. Y.K.L. Srinivas, and Rakesh Chandra Balabantaray (2015), "Text Normalization of Code Mix and Sentiment Analysis," *2015 International Conference on Advances in Computing, Communications* *and Informatics, ICACCI 2015*, 1468–73.

Skorikov, Vladimir B., and Fred W. Vondracek (2011), “Occupational Identity,” in *Handbook of Identity Theory and Research*, New York, NY: Springer New York, 693–714.

Sluss, David M and Blake E. Ashforth (2007), “Relational Identity and Identification: Defing Ourselves Through Work Relationships,” *Academy of Management Review*, 32(1), 9–32.

Tajfel, Henri and J.C. Turner (1986), “The Social Identity Theory of Intergroup Behavior,” in *Psychology of Intergroup Relations*, Nelson-Hall, 7–24.

Teetor, Paul (2011), "*R Graphics Cookbook,"* O'Reilly Media.

Tumasjan, Andranik, To Sprenger, Pg Sandner, and Im Welpe (2010), “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment,” *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–85.

Turner, John C. and Katherine J. Reynolds (2016), “Self-Categorization Theory,” in *Handbook of Theories of Social Psychology*, London, EC1Y 1SP United Kingdom: SAGE Publications Ltd, 399–417.

Turner, John C, Penelope J Oakes, S Alexander Haslam, and Craig Mcgarty (1992), “Personal and Social Identity: Self and Social Context,” Princeton: Department of Psychology, Princeton University.

Vargo, Chris J. (2016), “Toward a Tweet Typology: Contributory Consumer Engagement with Brand Messages by Content Type,” *Journal of Interactive Advertising*, 0(0), 1–12.

Xiaomei, Zhan and Wang Shimin (2014), “Political Identity: A Perspective from Cultural Identity,” *Social Sciences in China*, 35(2), 155–73.

Ysseldyk, Renate, Kimberly Matheson, and Hymie Anisman (2010), “Religiosity as Identity: Toward an Understanding of Religion from a Social Identity Perspective,” *Personality and Social Psychology Review*, 14(1), 60–71.

Zaglia, Melanie E (2013), “Brand Communities Embedded in Social Networks.,” *Journal of business research*, 66(2–2), 216–23.