Using Twitter to explore consumers' sentiments and their social representations towards new food trends

Emilio Pindado¹ and Ramo Barrena²

¹ Department of Agricultural Economics, Statistic and Business Management, ETSIAAB, Technical University of Madrid, Madrid, Spain, emilio.pindado@upm.es

² Department of Business Administration, IS-FOOD Research Institute, Public University of Navarre, Pamplona, Spain, ramo.barrena@unavarra.es

Document Version:
Accepted author manuscript

Published in:
British Food Journal

DOI:
https://doi.org/10.1108/BFJ-03-2020-0192

Citation for published version:

Publisher's copyright statement:
This author accepted manuscript is deposited under a Creative Commons Attribution Non-commercial 4.0 International (CC BY-NC) licence. This means that anyone may distribute, adapt, and build upon the work for non-commercial purposes, subject to full attribution. If you wish to use this manuscript for commercial purposes, please contact permissions@emerald.com.
Using Twitter to explore consumers’ sentiments and their social representations towards new food trends

Abstract

Purpose: This paper investigates the use of Twitter for studying the social representations of different regions across the world towards new food trends.

Design/methodology/approach: A density based clustering algorithm was applied to 7,014 tweets to identify regions of consumers sharing content about food trends. The attitude of their social representations was addressed with sentiment analysis and grid maps were used to explore sub-regional differences.

Findings: Twitter users have a weak-positive attitude towards food trends and significant differences were found across regions identified which suggests that factors at regional level such as cultural context determine users’ attitude towards food innovations. The sub-regional analysis showed differences at the local level, which reinforces the evidence that context matters in consumers’ attitude expressed in social media.

Research limitations/implications: Social media content is sensitive to spatiotemporal events. Therefore, research should take into account content, location and contextual information to understand consumers’ perceptions. The methodology proposed here serves to identify consumers’ regions and to characterise their attitude towards specific topics. It considers not only administrative but also cognitive boundaries in order to analyse subsequent contextual influences on consumers’ social representations.

Practical implications: The approach presented allows marketers to identify regions of interest and localize consumers’ attitudes towards their products using social media data, providing real time information to contrast with their strategies in different areas and adapt them to consumers’ feelings.

Originality/value: This study presents a research methodology to analyse food consumers’ understanding and perceptions using not only content but also geographical information of social media data, which provides a means to extract more information than the content analysis applied in the literature.

Keywords: Food trends; Consumer behaviour; Big data; Twitter analytics; Opinion mining; Density-based clustering
1. Introduction

Over the past decade the variety and number of innovative food trends like new products, new packages, new forms of consumption and commerce have increased considerably due to changes in consumer habits and globalization. However, failure rates are still high (Aqueveque, 2016). Some of these food trends generate insecurity and suspicion, whereas others are seen as already familiar, which directly affects consumers’ behaviour and acceptance of innovations (Siegrist et al., 2013). Therefore, a vast amount of research has been conducted to understand, conceptualize and measure food consumers’ perception and adoption of innovations (Nazzaro et al., 2019).

The acceptance of these innovative food trends may be due to factors of the trend itself, consumers’ personal characteristics, or the influence of the social understanding of the food innovations (Bartels and Reinders, 2010). Indeed, the literature suggests that a common understanding of food innovations is a strong predictor of innovation adoption decisions by food consumers (Bäckström et al., 2004; Huotilainen et al., 2006; Onwezen and Bartels, 2013). In response to changing environments, consumers socially create and share common knowledge that constitutes common sense about unfamiliar topics, which allows them to deal with the novelty (Huotilainen et al., 2006). This type of knowledge including a practical vision of a common trend and developed through socio-cognitive processes is what the literature has referred to as social representations (Howarth, 2006; Jodelet, 2008). Social representations express an attitude (positive or negative) towards an object or topic and are shaped by the social interactions and the cultural context of social groups and communities (Howarth, 2006; Moscovici, 2001).

A series of studies have focused on how different social representations of food innovations vary across different communities and cultures (Barrena et al., 2015; Bartels and Reinders, 2010; Onwezen and Bartels, 2013). However, the specificity of studies on food consumers’ social representations of innovations –based mainly on limited samples – has prevented the obtaining of generalizable results, thus offering multiple research opportunities (Casini et al., 2015; Mäkiniemi et al., 2014).

These studies have been traditionally based on surveys, case studies and focus groups requiring a large amount of time and resources. The growth of social media has changed the way in which consumers communicate and, therefore, the way in which researchers and organizations access consumers’ information (Wamba et al., 2015). As a result of
this, a growing literature on consumer behaviour is exploring the use of social networks
to measure food consumers’ attitudes, and company managers recognize the use of big
data analytics as a powerful tool to extract information about their products, customers
and markets (Mostafa, 2018, 2019; Ruggeri and Samoggia, 2018; Samoggia et al., 2020).
Twitter is one of the most popular social platforms, and the attitudes and sentiments of
individuals can be easily found in their tweets (Chamlertwat et al., 2012). Additionally,
this platform allows the geolocation of Twitter users, which can provide information of
great interest for the study of communities and their social phenomena (Herdağdelen et
al., 2013; Widener and Li, 2014).

Thus, this work investigates the potential of Twitter for studying the social
representations of innovative food trends among different geolocated communities. We
attempt to shed light on the cross-cultural differences and spatial dynamics of food
consumers’ behaviour across the world. For this purpose, we used 7,014 messages
broadly containing the words “new foods” to capture multiple food related innovations,
defined here as innovative food trends. We developed a three-step approach for the
identification and characterisation of food trends communities based on the approach
proposed by Gao et al. (2017) to identify cognitive regions using social media users’
location and content analysis. First, we analysed the geographical distribution of tweets
on this topic based on the location provided by users in their profile. Second, we applied
a density based clustering algorithm on geolocated tweets to identify regions of
consumers sharing content about innovative food trends. Finally, following Widener and
Li (2014), we carried out a sentiment analysis of tweets published within these areas to
address the positive or negative attitude of their social representations towards food
novelties (Lipizzi et al., 2016).

2. Related Work

Over recent years, microblogging platforms like Twitter have become a fruitful source of
data for the study of human behavior applying data mining techniques (Hilbert, 2016). It
has been successfully used to address a variety of research topics such as consumers’
behavior regarding consumption restraints using manual content analysis (Paschen et al.,
2020), or people’s emotions on specific issues like urban green spaces using descriptive
and content analytics (Roberts, 2017). However, users of these platforms not only provide
the content of the message published but also secondary information such as localization,
which can be easily harvested (Stefanidis et al., 2013). This location is not only
geographic data per se; it also conveys contextual information about people’s perceptions and preferences which allows for the analysis of the human phenomena and its spatiotemporal distribution (Stefanidis et al., 2013).

Within this spatial perspective, there is an increasing amount of research identifying geolocated online communities (Gao et al., 2017; Gruzd et al., 2016). These communities can be defined as groups of users that are more densely connected to each other than to the rest of the users of the platform (Bakillah et al., 2015). These groups, furthermore, can be characterised as communities of interest where individuals share information and engage in social interactions based on their common interests (Chiu et al., 2006). Hence, these communities have a structural dimension (i.e. bounded location) and a socio-psychological one (i.e. cognitive) in the form of a sense of shared values and understandings regarding a specific topic (Porter, 2004).

The structural identification of these communities within social platforms can be seen as a partitioning or clustering problem (Croitoru et al., 2015) while applying sentiment analysis provides the means to determine the attitudes of these communities (socio-psychological dimension) through the content that they have shared (Deitrick and Hu, 2013; Pang and Lee, 2008). The combination of both approaches, therefore, leads to a more in-depth understanding of the social phenomenon studied (Deitrick and Hu, 2013). Consequently, there are an increasing number of studies based on Twitter adopting this joint approach of identification of geolocated communities applying density-based spatial clustering techniques and sentiment analysis of content published within them to address its public opinion dynamics. Hridoy et al. (2015), for example, applied DBSCAN cluster algorithm together with sentiment analysis to analyse public opinion across the USA regarding the iPhone 6. Likewise, Stojanovski et al. (2018) applied DBSCAN cluster algorithms and sentiment analysis on Twitter messages regarding the 2014 FIFA World Cup to address the emotional attitude of fans during this sport event and social hotspots within New York. Furthermore, recent studies have applied this clustering methodology to different social media platforms directly revealing consumers’ sentiments trough rating apps such as Dianping –a social platform for catering services--, all of which shows the accuracy of DBSCAN algorithm to identify consumer clusters (Fan et al., 2019).

However, much of the social media research regarding food consumers’ behaviour has focused on content analysis while research on the spatial dimension is still scarce. For example, Vidal et al. (2015) have addressed Twitter users’ opinion regarding the different...
meals of the day i.e. breakfast, lunch and dinner, applying content analysis and highlighting the importance of context in each situation (e.g. midnight and Sunday) in the content published. Likewise, Vidal et al. (2016) have analysed the non-verbal emotional response to these meals through emoticon and emoji analysis. Widener and Li (2014), on their part, explored the spatial distribution of tweets related to healthy food across the USA and carried out a sentiment analysis which revealed that the density of positive tweets is lower in areas with no access to this type of food. Other studies have studied specific food types such as halal food. Mostafa (2018), for instance, applied sentiment analysis and mapped geolocated tweets which revealed a positive attitude toward this type of food as well as the location of Muslim communities across the world. The same author used sentiment analysis and topic clustering techniques to segment halal consumers in four categories according to their concerns in terms of product authenticity, animal welfare, identity, and religiosity (Mostafa, 2019). Ruggeri and Samoggia (2018) analysed Twitter content generated by palm oil producers, chocolate manufacturers, and food retailers to address sustainability issues. Likewise, recent research has analysed Twitter users’ concerns regarding health and coffee by applying content and sentiment analysis to show how tweets express a low positive attitude towards the health benefits of coffee (Samoggia et al., 2020).

Thus far, food researchers have addressed the spatial and socio-psychological dimensions separately for the most part, which shows that this platform is a rich data source when it comes to understanding food consumer behaviour. In this regard, there have been recent calls for further investigation on the potential of social media platforms to discover food consumers’ openness to innovations (Bartels and Onwezen, 2014; Carr et al., 2015). Understanding what consumers think or believe about specific situations plays a critical role in the acceptance of innovations (Huotilainen et al., 2006) particularly in terms of the configuration of attitudes, emotions, and knowledge that are built and shared by social groups defined as social representations (Howarth, 2006; Jodelet, 2008). This socio-psychological process is directly related to the combined analysis of community identification and sentiment analysis developed in Twitter studies.

Therefore, in light of the above, the aim of this study is the identification of food trends geolocated communities, defined here as dense groups of people broadcasting information and opinions about innovative food trends (i.e. any food-related concept implying novelty), and the characterization of their attitudes towards this topic as a pillar
of its social representation (Moscovici, 1961). Below, we describe the methodology employed to address these issues.

3. Data and Methods

3.1. Data collection

The data for this paper were acquired using R software and the “Twitter application programming interface” (API) through the "twitteR" package (Gentry, 2015). The data collection methodology consisted in retrieving the tweets containing the English words “new foods” over January 11-31 2016. We thus captured messages related to the broad topic of innovative food trends in a simple way (see Figure 1 for examples), trends understood here as new directions in which a topic is developing or changing (Celi and Rudkin, 2016). The Twitter site allows for the retrieval of tweets published up to one week before the search in samples of 1,500 tweets. Therefore, several searches were conducted during this period to cover the maximum number of tweets possible.

An initial sample of 18,911 tweets was obtained. In order to clean the dataset for further analyses, the next step was data pre-processing, which is necessary as tweets frequently contain noise e.g. links, non-Latin characters, numbers, and users. Likewise, due to the gathering procedure and the possibility of redundancies, duplicate tweets were removed.

The clean data contained information regarding the username of the author of each tweet, which can be used to get the location provided by users in their own profile using the twitteR package. Once we obtained the location, we were able to geolocate each tweet author using the "Google geocoding web-service" through the "dismo" package (Hijmans et al., 2017). After data geolocation, we obtained a localized sample of 7,014 messages about innovative food trends.

3.2 Spatial analysis of georeferenced tweets and community identification

To analyse the spatial distribution of tweets about food trends across the world, we performed a kernel density analysis using the tweet’s user location. Kernel density is a

---

1 Guerrero et al. (2009) identified “novelty-change” as one of the main dimensions that consumers relate to innovative food products. This, together with the fact that microblogging services like Twitter only allow for the sharing of short pieces of information usually written in an informal language, led us to select the terms new and foods to capture the tweets referring to innovative food trends (Thelwall et al., 2011).

2 Data and R code used for analyses in this study are available from the authors upon request.
non-parametric method used to estimate the intensity of points by calculating a smooth surface based on a bivariate normal probability distribution (Bailey and Gatrell, 1995). This approach has been extensively used to explore spatial patterns in consumer behaviour and Twitter’s sentiment analysis (Roig-Tierno et al., 2013; Widener and Li, 2014).

We applied spatial clustering techniques on the geolocated tweets in order to analyse the spatial patterns of the users more in depth and to discover the underlying distribution of geolocated communities. Density-based cluster methods are the most frequent approach for assessing non-random geospatial patterns (Kriegel et al., 2011). These aggregating mechanisms have the capacity to identify high density regions separated from regions with low densities in spatial datasets. Among these methods, the Density Based Spatial Clustering of Application with Noise (DBSCAN) algorithm, proposed by Ester et al. (1996), is one of the most used due to its capacity to find spatial clusters with arbitrary shapes in spatial data with noise (Bakillah et al., 2015; Ben Khalifa et al., 2017). Many authors have used this approach to detect spatial clusters among Twitter users based on the advantages offered by DBSCAN such as the identification of noise points and clusters of different sizes and shapes without the pre-assumption on the number of clusters (Ben Khalifa et al., 2017; Hridoy et al., 2015; Stojanovski et al., 2018).

DBSCAN is based on the idea that the density of each neighbourhood (i.e. cluster) has to exceed a given threshold defined by the cluster core point, the radius of the cluster ($\epsilon$) and the minimum number of points inside the cluster ($MinPts$), which determines density. Thus, DBSCAN algorithms are sensitive to the $\epsilon$ and $MinPts$ thresholds.

To describe the DBSCAN method, we need the following concepts based on a dataset D:
- The neighbourhood within a radius $\epsilon$ of a given point $p$ ($p \in D$) is the subset called $\epsilon$-neighbourhood (denoted by $N(p)$) defined as: $N_\epsilon(p) = \{ q \in D | \text{dist}(p,q) \leq \epsilon \}$
- A point $p$ ($p \in D$) is denoted as a core point if the $\epsilon$-neighbourhood of $p$ contains at least $MinPts$ points.
- A point $p$ ($p \in D$) is denoted as a noise point if the $\epsilon$-neighbourhood of $p$ contains less than $MinPts$ points.
- A point $q$ ($q \in D$) is denoted as a directly density reachable point from the point $p$ ($p \in D$) if $p$ is a core point and $q$ is in the $\epsilon$-neighbourhood of $p$. 


- A point $p$ ($p \in D$) is denoted as a density reachable point from the point $q$ ($q \in D$) if $p$ is in the $\epsilon$-neighbourhood of $q$ and $q$ is not a core point but they are reachable through a chain of directly density reachable points.

- Two points $p$ ($p \in D$) and $q$ ($q \in D$) are denoted as density connected points, with respect to $\epsilon$ and $MinPts$, if there exists a point $o$ ($o \in D$) such that $p$ and $q$ are density-reachable from $o$ with respect to $\epsilon$ and $MinPts$.

DBSCAN algorithm starts with an arbitrary point $p$ ($p \in D$), finds all density-reachable points from $p$ and if $q$ ($p \in D$) is a core point a cluster will be created. The algorithm iteratively adds points that do not correspond to any cluster and are directly density reachable from the new cluster’s core points. When the new cluster cannot be expanded, a cluster is completed. Then, DBSCAN arbitrarily selects a remaining unvisited point and the clustering procedure continues until all points are visited and new clusters cannot be created. Those points excluded from the clusters identified are marked as noise. As previously stated, the algorithm requires the definition of two parameters ($\epsilon$ and $MinPts$), which requires a sorted k-dist graph analysing the nearest distances between points to find the suitable value of $\epsilon$ (Ester et al., 1996). We used the “dbscan” package (Hahsler et al., 2019) to implement the DBSCAN algorithm and to create the k-dist graph.

3.3 Sentiment analysis

To explore the general orientation (i.e. attitude that expresses a position) of the social representations of innovative food trends across the communities, a sentiment index was developed using the tweets generated by users within the identified clusters (Lipiţzi et al., 2016; Widener and Li, 2014). Sentiment analysis can be defined as the field of study that deals with people’s opinions about products, organizations, events, or topics expressed in written texts (Liu, 2015). As Pang and Lee (2008) indicate, sentiment analysis methods are based on two basic steps: opinion extraction and sentiment classification. Opinion extraction is the task of obtaining subjective texts, while sentiment classification is the task of classifying opinion words into sentiment categories. These tasks can be carried out to different levels of detail such as word, sentence, document and feature (Kumar and Sebastian, 2012).

A broad approach to sentiment classification is to use a pre-existing lexicon with information about which words and sentences are positive and which are negative (Wilson et al., 2009). These approaches are usually called dictionary-based methods; the semantic orientation score (i.e. the degree of positive or negative sentiment scaled from
+1 to -1 respectively) is calculated by point-wise mutual information measures (Widener and Li, 2014). However, it is important to note that the sentiment polarity of a given sentence may be different from the prior polarity of the words that compose that sentence (Wilson et al., 2009).

Hence, as Twitter allows users to share pieces of information limited to 140 characters containing several phrases, we adopted a sentence-level sentiment approach (Yu et al., 2013). Specifically, we followed a dictionary-based approach based on Syuzhet dictionary (Jockers, 2017), using the "sentimentr" package (Rinker, 2017), which calculates the average score sentiment taking into account contextual valence shifters of the sentences contained in each tweet. This approach has been used in previous studies to analyse sentiments in CEOs’ oral communications as well as Twitter data (Choudhury et al., 2019; Meier et al., 2019).

4. Results and discussion

Before examining the geolocated communities of Twitter users concerned with innovative food trends and their sentiments, it is important to offer an overview of the spatial distribution of data collected across the world. Thus, Figure 2 shows the data points on a world map. As expected, there is a greater concentration of tweets among English speaking countries due to the nature of the dataset and, furthermore, we can appreciate several areas with a high intensity of points. Although the points are plotted in a partly transparent way, the patterns observed are in accordance with population density areas, which suggests “overplotting” problems (Poorthuis and Zook, 2015). In other words, “points are layered on top of one another to the point” which makes it difficult to discern meaningful spatial patterns” (Shelton, 2017). Furthermore, this reflects the fact that urban residents are more likely to use Twitter than rural users (Smith and Brenner, 2012).

Therefore, to avoid the problem of overplotting and visually identify high aggregation areas, we generated a heat map (see Figure 2) using kernel density estimation (Poorthuis and Zook, 2015). In Figure 2 we show the three main regions, or global hotspots, corresponding to a higher density of users tweeting about innovative food trends. Specifically, we can identify several areas within the United States (namely, California, Texas Triangle/Gulf Coast, Florida, Great Lakes/Piedmont Atlantic/Northeast, and Cascadia), the United Kingdom area, and finally the area of Malaysia. Overall, these patterns seem to correspond to areas of higher density of English speakers and urban
agglomerations (Mostafa, 2018). However, it is important to notice the low densities shown in the Australian region and other areas with high population of English speakers. This reinforces the idea that users make different uses of social media to share their experiences, in this case about food, depending on the sociocultural values of their environment (Hodeghatta and Sahney, 2016).

Nevertheless, heat maps assume that the underlying spatial processes are continuous, which may hinder the identification of social phenomena like geolocated communities (Poorthuis and Zook, 2015). Thus, in order to address how individuals discussing innovative food trends are spatially distributed in a meaningful way, we applied the DBSCAN algorithm to group the geolocated tweets into spatial communities based on users’ proximity (Bakillah et al., 2015).

Before running the DBSCAN algorithm it is necessary to select its parameters $\epsilon$ and $MinPts$. Here, we employ the heuristic method presented in Birant and Kut (2007), which suggests that, $n$ being the size of the dataset, the parameter $MinPts \approx \ln(n)$ and $\epsilon$ must be estimated depending on the value of $MinPts$. This method requires the calculation of the distances to the $k$-nearest neighbours for each point (i.e. each located tweet), $k$ being equal to $MinPts$. The threshold point is determined as the first “valley” of the sorted $k$-dist graph, the optimum $\epsilon$ being the distance between this point and its $k$th nearest neighbour. Figure 3 shows the k-dist graph ($k=9$) for the dataset, which indicates a threshold point of $\epsilon = 3.5$ degrees (i.e. a cluster’s radius of approximately 385 km)$^3$ and $MinPts = 9$. We empirically tested the selection of the $\epsilon$ parameter based on the minimum noise generated (Villatoro et al., 2013).

We applied a DBSCAN algorithm based on the above parameters, obtaining 32 clusters for the global dataset of geolocated tweets. Table 1 summarizes the characteristics of the clusters identified, as well as the sentiment analysis statistics, and Figure 4 shows their

---

$^3$ 1 degree is approximately 110 km (Harari and Ferrara, 2018).
spatial distribution. A detailed map of the clusters’ distribution can be found in the Appendix.

--Insert Figure 4 Here--

Results of Table 1 show three very crowded clusters distributed in North America. The first and biggest (cluster 2) is located in the east of North America accounting for 55.52% of the total located messages about food trends, followed by a second region in the southwest (cluster 6) accounting for 12.25% of the tweets, and finally a third region on the north-west coast (cluster 7) accounting for 4.05% of the located messages. Additionally, we can identify the other main community located in Europe (cluster 5), covering the United Kingdom, Western Europe, the east of Spain and north of Italy, which represents 11.25% of the geolocated tweets. Finally, based on the density-based cluster analysis carried out, the community geolocated in Malaysia (cluster 3) stands out with 3.81% of the messages. We can also appreciate how the rest of clusters identified across the world are distributed heterogeneously with few observations per cluster (27 of the 32 clusters identified account for 9.65% of all of the geolocated tweets).

The pattern and size of communities identified are in line with the nature of the Twitter data, users tend to share messages in the predominant language in their area, and the density of population and technological development (i.e. greater penetration of Twitter) directly shape the size and localization of the communities (Graham et al., 2014). However, the generated clusters differ substantially from the initial heat maps distribution, which confirms a need for spatial clustering techniques to identify spatial patterns of human behaviour affected by different overlapped processes (Han et al., 2001; Pei et al., 2009). Therefore, the spatial distribution of these clusters suggests that, like other human behaviours, the identification of food trends communities depends not only on geographic and administrative boundaries but on cultural aspects, human interactions and the scale of analysis (Gao et al., 2017).

--Insert Table 1 Here--

As Veltri and Atanasova (2015) stressed, the emotional component of Twitter messages provides valuable information to understand the social representations (i.e. social configuration of attitudes, beliefs, knowledge, and emotions) developed by distinct social
groups towards a specific phenomenon. Therefore, and in order to gain insight of the
social understanding of innovative food trends, we developed a dictionary-based
sentiment analysis that calculates the average polarity score of the tweets shared within
the identified communities. This methodology enabled us to analyse the positive or
negative attitude of communities’ social representations towards food innovations across
the world.

The sentiment scores (mean and standard deviation) of the communities identified are
provided in Table 1. The overall sample has a mean sentiment score of 0.320 and standard
deviation of 0.251, which indicates that the users of this social platform tend to have a
weak-positive social attitude towards food trends. This is consistent with the fact that
Twitter users are “information seekers”, which is positively related to the propensity to
adopt innovations, as well as to the large number of tweets that are simply informative
due to the strong presence of the media in Twitter (Rogers, 1995; Samoggia et al., 2020).

Additionally, we verified whether there is significant variance across the identified
communities in terms of the sentiment score (Bliese, 2000). For this purpose, we carried
out a likelihood test (LRT) with a significant effect being found with a LRT of 813.16
(p<0.01). This indicates that users within the same spatial community are usually more
similar to each other than other users in a different community, which reinforces the
validity of the proposed clustering approach for identifying located thematic communities
with shared values and social understanding (Porter, 2004).

The analysis of the differences across these communities revealed interesting results. For
example, the highest-scoring polarity was found for Malaysia (cluster 3) with a mean of
0.740, which reveals a strong positive attitude towards new food trends. Additionally, the
standard deviation of 0.165 shows low variation for this attitude across the tweets
analysed within this region. On the other hand, the communities identified within North
America (clusters 2, 6 and 7) present a low positive attitude towards new food trends, all
with similar scores (mean of each cluster about 0.30) and greater variation than Malaysia
(standard deviation of each cluster around 0.22). The European region (cluster 5) presents
similar scores (mean of 0.315 and standard deviation of 0.263) as those obtained for North
America. The small number of observations per cluster of the rest of regions identified
condition the interpretation of its sentiment analysis due to the fact that it may have been
affected by peculiarities of tweets harvested; for example, Hawaii (cluster 30) has a mean
of 0.156 and standard deviation of 0.23, which implies high variability across the sentiments expressed.

These findings revealed a complex picture of the potential factors (e.g. territorial identity, multiculturalism and food policies) that determine the positivity of the social representations with regard to innovative food trends (Bäckström et al., 2004; Goodman, 2016). Nevertheless, the propensity of specific communities like Malaysia to have a positive attitude towards food trends show the important effect that the cultural environment, its diversity, and the information flows between members may exert on the building of positive social representations towards innovation in food consumption (Bartels and Reinders, 2010; Onwezen and Bartels, 2013; Pieniak et al., 2009). Further research is needed in order to determine the extent to which these socio-cultural factors influence social knowledge about food innovations.

Finally, it is important to note that Twitter users interact differently under specific conditions, which may imply different behaviours at the local scale (Lansley and Longley, 2016). Therefore, we aggregated the geolocated tweets’ sentiments and represented them in a grid map in order to analyse the spatial patterns within the communities. The results for the two largest communities in terms of surface area are presented in Figure 5 and Figure 6. We can observe that the concentrations of users with positive sentiments (i.e. hotspots) are heterogeneously distributed across the east of North America, whereas in Europe they are concentrated in the south-east. These patterns again reinforce the idea that social representations are complex phenomena and that the clustering techniques applied here at regional scale may mask the existence of sub-regional differences, which requires detailed analysis by future research at the sub-regional level to address the factors that may affect food consumers’ innovative behaviours at the local level (Bassanino et al., 2011; Jodelet, 2008).

5. Conclusion

The objective of this study was to propose a new methodology to analyse food consumers’ social representations regarding new trends and their spatial distribution using Twitter data. The research has demonstrated through a worldwide geospatial analysis that content and spatial data harvested from Twitter can provide a valuable source of information for the analysis of social representations about specific topics across different cultures. The
findings reveal that the approach based on density clustering and sentiment analysis is a valid research methodology suitable to identify communities with significantly different socio-psychological processes. Overall, our results show that Twitter users have a weak-positive social attitude towards food trends, which is consistent with the informative function of this social platform for food consumers noted by Samoggia et al. (2020). However, we found significant differences between regions such as a stronger positive attitude towards food innovations of users located in Malaysia. This finding is in line with prior studies highlighting the important role that contextual factors at the regional level such as social diversity and high intensity of flows of information may play in shaping consumers’ positive attitudes towards food innovations and extends this to the social media domain (Bartels and Reinders, 2010; Pieniak et al., 2009). Additionally, our study reveals sub-regional differences in food consumers’ social representations analysed through a grid map of the geolocated tweets’ sentiments, which shows the complexity of the social phenomena addressed and the influence of not only the regional but also of the local context on social media users’ attitudes.

5.2. Implications
Our study is a contribution to the extant research on food consumers’ behaviour using social media information as it reveals that the spatial information associated with social platforms could provide valuable information to improve our knowledge of human phenomena (Mostafa, 2018; Samoggia et al., 2020). In this regard, the research methodology used in this study, which combines community identification among users of social media using spatial clustering techniques and the analysis of the sentiments shared within these communities, provides useful insights in relation to cross-cultural perceptions of food consumers and regional dynamics that cannot be achieved in traditional studies. Moreover, this methodology goes beyond simple topic analyses and leads to capturing the attitudes of these social groups on specific objects. The application of social representations theory provides a well-established conceptual framework to address shared cultural values and meanings among Twitter users. Thus, the methodology presented here provides a valid and useful conceptual tool to analyse consumers’ understanding and perceptions regarding specific topics such as food trends as well as their spatial dynamics and contextual influences, considering cognitive rather than administrative boundaries.
Additionally, from a practical point of view, this approach allows marketers to identify, describe and monitor food consumers regions of interest (i.e. areas with positive attitude) to launch their new products based on social media data. Furthermore, this spatial analysis of online comments can provide valuable information to design tailored marketing strategies for their products taking into account not only the perceptions of its target communities, but also their distribution and spatial heterogeneity, revealed here to be an important factor to consider (Zarco et al., 2019). Data from microblogging platforms are easily accessible and represent consumers’ perceptions regarding specific products or brands in real time, thus offering a good source of information to benchmark against firms’ offers, prices, profits, sales and contextual information (Mostafa, 2018). The methodology proposed here is useful for marketers to identify regions and their attitude towards their products, as well as differences at the local scale, which can be contrasted with firms’ strategy and marketing actions within these zones. Consequently, marketers can quickly adapt advertising campaigns taking into account these localized opinions and attitudes (e.g. acting upon specific areas or retail stores), as well as design electronic word of mouth marketing (eWOM) strategies to intervene in and attempt to change consumers’ perceptions (Borah et al., 2020). Among the eWOM strategies, viral advertising has been recognized as one of the most effective for users of social media, therefore, marketers should create content aimed to be shared between Twitter users within communities or areas of interest in order to redirect their negative sentiments or take advantage of the positive attitudes (Kulkarni et al., 2020). This form of advertising benefits from established consumers’ networks and the inherent trust between them due to consumers’ distrust of traditional intrusive forms of advertising. However, it requires an adequate identification of potential seeds (i.e. consumers that initially share the content) as well as an appropriate incentive strategy to share the content crafted (Huh et al., 2020). A deep analysis of the content published and the spread of information between users within the given areas of interest should provide the basis to identify potential seed users and the message to share (Himelboim and Golan, 2019).

5.2. Limitations and future research

Despite these contributions, we have to acknowledge the limitations of this research. As Vidal et al. (2015) pointed out Twitter data are not a panacea but can provide a useful source of information for consumer researchers as long as their limitations are recognized. Among these are the facts that Twitter users are not representative of the general
population due to the use of this platform being elective and concentrated in urban agglomerations (Vidal et al., 2015). Likewise, the number of located tweets is small and the methods to infer users’ location (i.e. location mining methods), are valid but still imperfect (Garcia et al., 2018). Some tweets were geocoded to the centre of the country, a state, or a county, due to the fact that users only report broad addresses in their profile (e.g. Australia). Furthermore, the content of shared tweets is influenced by personal perceptions which may be sensitive to spatiotemporal events (Widener and Li, 2014). Regarding the methods used to analyse this information, sentiment analysis techniques may fail to capture ambiguous meanings and DBSCAN algorithms are sensitive to parameters’ values selection (Steiger et al., 2016).

Nevertheless, the approach presented provides a basis for future research using geospatial and content information from social media platforms on food consumers’ behaviour. In light of the multiple drivers that may impact food choices, which create a complex picture, further investigation should include multilevel strategies to address the heterogeneity and contextual influences revealed here. Furthermore, the approach proposed could serve to identify different food consumers’ communities posting on multiple issues like food crises or health issues on social media by analysing their behaviour and external influences.

References


Table 1. Characterization of clusters identified using DBSCAN and sentiment analysis.

<table>
<thead>
<tr>
<th>Cluster id</th>
<th>Associated Region</th>
<th>Cluster Observations (geolocated tweets)</th>
<th>Sentiment Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total % of total tweets</td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>Noise</td>
<td>244</td>
<td>3.48%</td>
</tr>
<tr>
<td>1</td>
<td>Gujarat (India)</td>
<td>6</td>
<td>0.43%</td>
</tr>
<tr>
<td>2</td>
<td>East of North America</td>
<td>14</td>
<td>55.52%</td>
</tr>
<tr>
<td>3</td>
<td>Malaysia</td>
<td>1</td>
<td>3.81%</td>
</tr>
<tr>
<td>4</td>
<td>Sabah (Malaysia)</td>
<td>5</td>
<td>0.13%</td>
</tr>
<tr>
<td>5</td>
<td>Europe</td>
<td>9</td>
<td>11.25%</td>
</tr>
<tr>
<td>6</td>
<td>South-West of North America</td>
<td>3</td>
<td>12.25%</td>
</tr>
<tr>
<td>7</td>
<td>North-West of North America</td>
<td>0</td>
<td>4.05%</td>
</tr>
<tr>
<td>8</td>
<td>South Africa</td>
<td>0</td>
<td>0.23%</td>
</tr>
<tr>
<td>9</td>
<td>Kenya</td>
<td>0</td>
<td>0.24%</td>
</tr>
<tr>
<td>10</td>
<td>India</td>
<td>0</td>
<td>0.58%</td>
</tr>
<tr>
<td>11</td>
<td>Australia</td>
<td>0</td>
<td>0.34%</td>
</tr>
<tr>
<td>12</td>
<td>Chile</td>
<td>0</td>
<td>0.13%</td>
</tr>
<tr>
<td>13</td>
<td>Nigeria</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>14</td>
<td>Lagos (Nigeria)</td>
<td>0</td>
<td>0.93%</td>
</tr>
<tr>
<td>15</td>
<td>Canada</td>
<td>0</td>
<td>1.18%</td>
</tr>
<tr>
<td>16</td>
<td>Dubai- Qatar</td>
<td>4</td>
<td>0.23%</td>
</tr>
<tr>
<td>17</td>
<td>New Zealand</td>
<td>0</td>
<td>0.13%</td>
</tr>
<tr>
<td>18</td>
<td>Punjab (India)</td>
<td>1</td>
<td>0.46%</td>
</tr>
<tr>
<td>19</td>
<td>Sydney (Australia)</td>
<td>0</td>
<td>0.34%</td>
</tr>
<tr>
<td>20</td>
<td>Baltic States</td>
<td>6</td>
<td>0.23%</td>
</tr>
<tr>
<td>21</td>
<td>Ontario (Canada)</td>
<td>0</td>
<td>0.17%</td>
</tr>
<tr>
<td>22</td>
<td>Tokyo (Japan)</td>
<td>0</td>
<td>0.74%</td>
</tr>
<tr>
<td>23</td>
<td>Melbourne (Australia)</td>
<td>0</td>
<td>0.14%</td>
</tr>
<tr>
<td>24</td>
<td>Sweden</td>
<td>0</td>
<td>0.14%</td>
</tr>
<tr>
<td>25</td>
<td>Pretoria (South Africa)</td>
<td>2</td>
<td>0.21%</td>
</tr>
<tr>
<td>26</td>
<td>Karnataka (India)</td>
<td>0</td>
<td>0.27%</td>
</tr>
<tr>
<td>27</td>
<td>Philippines</td>
<td>0</td>
<td>0.44%</td>
</tr>
<tr>
<td>28</td>
<td>British Columbia (Canada)</td>
<td>0</td>
<td>0.33%</td>
</tr>
<tr>
<td>29</td>
<td>China</td>
<td>0</td>
<td>0.13%</td>
</tr>
<tr>
<td>30</td>
<td>Hawaii</td>
<td>0</td>
<td>0.19%</td>
</tr>
<tr>
<td>31</td>
<td>Brazil</td>
<td>0</td>
<td>0.14%</td>
</tr>
<tr>
<td>32</td>
<td>Bangladesh (India)</td>
<td>4</td>
<td>0.31%</td>
</tr>
</tbody>
</table>
Figure 1. Example of tweets harvested.

Eating lots of new foods is super fun, but I TRUST diners, they’re Stable
12:31 AM - 21 Jan 2016

LIVE on #Periscope: At fancy foods show finding new products you might want to try. pscp.tv/w/aW1mYzM1MzUz ... 3:01 PM - 18 Jan 2016

@deanfoods releases new calcium-enriched chocolate milk buff.ly/1T4N2NG 6:52 AM - 21 Jan 2016

Foods that make you go hmmmm? Definitely new pairings.

Sale-flavored Kit Kats, Earl grey tea gin, chocolate ramen and more weird food news. #FWA familyare/$5prrg 7:01 AM - 20 Jan 2016

I know this is weird cause I cook but I hate trying new foods lol 3:26 PM - 29 Jan 2016

If you try 5-10 new vegan foods every week, every week you’ll inevitably discover several great foods to add to your diet. 8:39 PM - 14 Jan 2016

I never try new foods bc my stomach always hurts and I did and what do u kno ….hurts. 9:07 PM - 17 Jan 2016
Figure 2. Geolocated tweets and kernel density surfaces.

Note: Location of tweets (red points) based on users’ profile information and kernel density surfaces where red tones correspond with higher densities and blue ones with lower densities.
Figure 3. Sorted 9-dist graph for points of the dataset.
Figure 4. Spatial distribution of clusters identified using the DBSCAN algorithm.
Figure 5. Cluster 2 “East of North America” spatial distribution of aggregated sentiments.
Figure 6. Cluster 5 “Europe” spatial distribution of aggregated sentiments.
APPENDIX

Figure 1A. Spatial distribution of clusters identified using DBSCAN algorithm in North America.

Figure 2A. Spatial distribution of clusters identified using DBSCAN algorithm in South America.
Figure 3A. Spatial distribution of clusters identified using DBSCAN algorithm in Europe/Asia.

Figure 4A. Spatial distribution of clusters identified using DBSCAN algorithm in Africa.
Figure 5A. Spatial distribution of clusters identified using DBSCAN algorithm in North Asia/Oceania.