Co-occurrence networks of Twitter content after manual or automatic processing. A case-study on “gluten-free”

Patricia Puerta¹, Laura Laguna¹, Leticia Vidal², Gastón Ares², Susana Fiszman¹, Amparo Tárrega^{1*}

¹Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC). Agustín Escardino, 7, Paterna, Valencia, Spain

²Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Uruguay

*Corresponding author's email: atarrega@iata.csic.es

Abstract

Gathering information from social networks such as Twitter has emerged to obtain spontaneous and direct opinions of users about a topic. This study focuses on using co-occurrence networks to analyse Twitter information. The objectives were to study the impact of text pre-treatment (codification based in qualitative analysis or just pre-cleaning) and to apply co-occurrence networks for analysing what is said on Twitter about specific topics like “gluten-free”. As such, 16,386 tweets in Spanish containing terms “sin-gluten” and “gluten-free” were collected. A subset of 3,000 tweets was used to make co-occurrence networks two ways: i) from the manually coded text and ii) from pre-cleaned text. Results indicate that the co-occurrence network from pre-cleaned text provides meaningful information showing structure and relevance for terms like the network from coded text. The whole set of tweets was used to explore Twitter information about gluten-free, showing users share information about products, occasions, social situations, and places but also to product characteristics, sensations, and diet or health issues related to the products. Five product categories, critical for the lack of gluten (bread, cake, cookie, beer, and pizza) occupied most tweets, and according to the related terms were intended to recommend how to get (buying or cooking) these gluten-free products and to exhibit what (how, when, and where) they prepare and eat. These aspects were different among products and separated co-occurrence networks allowed better identification.

Keywords: Gluten-free; co-occurrence networks; social media; consumers; Twitter

1. Introduction

1.1. Gluten-free consumers

In recent years, an increase in consumer demand has been observed for gluten-free products (Christoph, Larson, Hootman, Miller, & Neumark-Sztainer, 2018; Missbach et al., 2015; Molina-Rosell, 2013). Research on gluten-free products has focused on strategies dealing with the negative impact a lack of gluten has on the quality properties of these products. Manufacturing gluten-free cereal products is a challenging task for the food industry (Capriles, dos Santos, & Arêas, 2016; Houben, Höchstötter, & Becker, 2012). Besides, according to Naqash, Gani, Gani, & Masoodi (2017), most approaches include the addition of functional ingredients to the formulation (gluten-free flours, starches, hydrocolloids, proteins, fats, and fibres) or the adoption of alternative processing methods (high pressure, extrusion, and sourdough fermentation) to produce gluten-free products with good sensory quality, especially a texture comparable to those containing gluten (Marston, Khouryieh, & Aramouni, 2016; Matos & Rosell, 2012; O'Shea, Doran, Auty, Arendt, & Gallagher, 2013; Penjumras et al., 2019).

However, according to Do Nascimento, Fiates, & Teixeira (2017), consumers concerns for gluten-free products include sensory quality of products and the issues they experience trying to have a “normal life”, especially in a social context. Still, information on the relevance of extrinsic properties of products, context aspects, and individual attitudes and opinions of gluten-free consumers is scarce. This is for the difficulty in finding coeliac participants for consumer studies, accounting for an estimated 1 to 2% of population (Sapone et al., 2012). Therefore, we believe what is said on social media networks may be a way to obtain opinions of this target group of consumers, allowing us to understand their motivations and interests when consuming gluten-free products.

1.2. Using Twitter as a source for gathering consumers' opinion

Among social media platforms, Twitter is one of the most popular and dynamic microblogging services, with 500 million text-based messages, called “tweets”, generated by active users per day (Chae, 2015; Da Silva, Hruschka, & Hruschka Jr, 2014; Mention, 2018; Vidal et al., 2015). The informal and colloquial nature of tweets, together with the ease and instant access of the platform make its use widespread, giving rise to a huge volume of rapidly generated data (Fried, Surdeanu, Kobourov, Hingle, & Bell, 2015; Moe & Schweidel, 2017). Unlike other opinion gathering methods for consumers (surveys), social media users spontaneously post what they want when they want, avoiding forced biases to express their opinion.

Food represents one of the key themes discussed on Twitter (Platania & Spadoni, 2018) and consequently, tweets are potentially valuable data sources for gaining insight on food-related consumer studies. To date, the exploration of user-generated content on Twitter has been useful

to study food-related topics (food in general, influence of food choices, language of food, food chains, health food, different eating situations, and emotional responses to food and beverages) (Chen & Yang, 2014; Fried et al., 2015; He, Zha, & Li, 2013; Platania & Spadoni, 2018; Samoggia, Bertazzoli, & Ruggeri, 2019; Vidal, Ares, & Jaeger, 2016; Vidal et al., 2015). However, no study has addressed the exploration and interpretation of a topic like gluten-free.

Different approaches have been made to analyse tweets; automatic word counting is the simplest method of gathering information from users. Calculating the frequency or occurrence of mentions for an individual word, is simple and rapid for summarising the text according to the terms that are frequently mentioned. Nevertheless, the frequency of occurrence of individual words has several important limitations. It may not represent the meaning of the word isolated in the dataset and can lead to misleading conclusions because of the loss of the words' context (Hsieh & Shannon, 2005; Vidal et al., 2015; Zhao, Zhang, Qian, & Zhou, 2013). Therefore, previous qualitative analysis of tweet contents, with individual reading, was proposed to analyse tweets in the context of which the words are mentioned. Thus, the content was classified into themes and sub-themes related to the specific topic (Nguyen et al., 2019; Platania & Spadoni, 2018; Samoggia et al., 2019; Vidal et al., 2015). Although implementing manual content analysis can be tedious and time consuming for the large amounts of text to be read, it proved to be successful at gaining better interpretation of Twitter content (He et al., 2017; He et al., 2013; Vidal et al., 2015). As an automatic alternative, text analysis based on machine learning algorithms have been used to extract meaningful information from the textual data, recording themes already established or commonly studied (Constantinides & Holleschovsky, 2016; Sengupta & Ghosh, 2020; van Zoonen & van der Meer, 2016). However, for the correct performance of these models, machine learning algorithms usually require a large external source of coded dataset to analyse the text units (Vidal, Ares, & Jaeger, 2018). Thus, the development and adjustment of the algorithms for new topics is complex or require added information.

1.3. Co-occurrence networks technique

Co-occurrence networks have been proposed as an approach to facilitate the understanding and visualisation of the structure of different text items and their content. Co-occurrence networks graphically represent the relevance of terms and the relatedness among them, identifying and displaying patterns of co-occurrence within the text (Ruiz & Barnett, 2015; Su & Lee, 2010). Although broadly applied in studies of bibliometric analysis, to identify and visualise the existing connections among data (Skaf, Buonocore, Dumontet, Capone, & Franzese, 2020; van Eck & Waltman, 2018; Wen, Horlings, van der Zouwen, & van den Besselaar, 2017), co-occurrence networks can also be used for exploring connections of terms in different text documents.

Co-occurrence networks can be obtained by specific software as VOSviewer and Gephi or by using the Python programming language. In the VOSviewer software used in this study, the construction of a map comprises three steps: i) A similarity matrix (association strength as a measure of similarity) is obtained from a co-occurrence matrix (van Eck & Waltman, 2007; van Eck, Waltman, van den Berg, & Kaymak, 2006). The similarity between two terms is calculated as the ratio: the number of co-occurrences of two terms *i* and *j* divided by the product of the total number of co-occurrences of *i* and the total number of co-occurrences of *j*. ii) The visualisation of similarities (VOS) mapping technique constructs a two-dimensional map in which the items are located in such a way that the distance between any pair of items reflects their similarity. The base for doing so is minimising a weighted sum of the squared Euclidean distances between all pairs of items. The higher the similarity between two items the higher the weight of their squared distance in the sum. iii) The obtained map is translated, rotated, and reflected to obtain consistent results (always the same map) regardless of the different solutions that can be reached in the optimisation process.

In the obtained network, the size of the label representing a term is proportional to its frequency of appearance in the text (occurrence). The thickness of the line connecting two terms indicates how often they co-occur within the same text unit. The distance between two terms offers an approximate indication of the relatedness of the terms (Cunillera & Guilera, 2018; Marinho, Hirst, & Amancio, 2017; Sharma, Bairwa, Gowthamghosh, Gupta, & Mangal, 2018; van Eck, Waltman, Noyons, & Buter, 2010). A dataset of 70 text documents describing flowers has been created to illustrate the explanation (Figure 1). The table in the figure includes the occurrences and co-occurrences of the seven terms. Each term is represented in the network as a circle of size proportional to the number of occurrences; for example, the labels and circles of terms *red colour* and *pink colour* are the largest and the smallest because they are the most and least mentioned terms, respectively. Distribution of terms on the map responds to the relationships between items. *Spring* was a general term, co-mentioned with many terms (*red*, *rose*, *poppy*, and *jasmine*) and thus, appears located in the centre. The links with the four terms have the same thickness because the number of co-occurrences is the same (five). Terms that do not show co-occurrences among them are separated in the extremes, and close to the terms with higher co-occurrence. In the top-left appears the term *poppy* related to *red*, while in the bottom right, the term *jasmine* relates to *fragrance*. *Rose* shows a strong link with *red* and *fragrance*, and appears in the bottom-left. The term *pink* links to *rose* but does not show co-occurrence with any other term, appearing separated on the bottom-left extreme of the network.

In this study, we propose using co-occurrence networks as a tool for analysing terms in tweets to give more structured information than word counting. Using of raw text directly from tweets would make the analysis almost automatic, however, a previous qualitative analysis of tweets

and the corresponding coding of text could be necessary to provide the relevance of ideas expressed in many different terms to avoid misunderstanding of the text.

Therefore, the first aim was to study how pre-processing of tweet text (coding through qualitative analysis or just pre-cleaning) influences co-occurrence networks to determine if the process can be automated without losing relevant information. The second aim was to analyse tweets about “gluten-free” to gather information about the aspects that are relevant for this specific group of consumers, in general and in relation to specific products.

2. Materials and methods

2.1. Retrieval of tweets

A total of 16,386 tweets containing “sin-gluten” or “gluten-free” terms posted by users writing in Spanish, between September 2017 and January 2018, were retrieved with the rtweet package (Kearney, 2018) from R software (R Core Team, 2017) via the Twitter’s Application Programming Interface (API). Re-tweets and repeated tweets were removed. Each retrieved tweet included an ID number, username of the person posting and the date and time when the tweet was published, among other information. Gender and age of Twitter users was not provided by rtweet package, and geographical location data was shown when specified by users in their personal profiles. Finally, the tm package (Feinerer & Hornik, 2017) was used for automatically lowering character case, removing punctuation, numbers, and stop words in the tweets text.

2.2. Qualitative analysis of tweets and manual coding

Because of time restrictions, a subset of the first 3,000 tweets was extracted. This subset was used for comparing networks obtained from coded and un-coded tweets and to determine the need of previous qualitative analysis when obtaining co-occurrence networks.

By a first independent reading, two researchers identified, by consensus, the most relevant and frequent type of terms (sub-themes). Considering both the type of terms and the meaning in the sentence, these sub-themes were grouped into main themes. For example, the tweet “Gluten-free party with school friends” was coded into the sub-themes: “social event” and “family-friends”, both belonging to the “Social context” main theme.

2.3. Co-occurrence networks

VOSviewer software (Version 1.6.7, Centre for Science and Technology Studies, Leiden University, The Netherlands) was used to obtain networks based on the frequency of mentioned terms (occurrence) and on their co-occurrence within the same tweet (i.e., when both terms occur together in one tweet). Based on the interconnection of terms within each tweet, this tool constructed, analysed, and visualised keyword networks by using the visualisation of similarities (VOS) technology (van Eck & Waltman, 2010; van Eck et al., 2010). The VOSviewer software created a two-dimensions map where distance between terms can be interpreted as an indication of their relatedness.

For evaluating the impact of pre-treatment of tweets, two different co-occurrence networks were obtained from the subset of 3,000 tweets. A first network was obtained from the coded text. Here, the database uploaded comprised the sub-themes identified after manually coding the tweets. A second network was obtained from directly the same 3,000 tweets text after just a pre-cleaning process, which included steps: i) Bi-grams (i.e. words that co-occur together and need to be treated as a single entity (e.g. “lactose” and “free” are converted into “lactose_free”) were identified using an N-gram analyser tool (<http://guidetodatamining.com/ngramAnalyzer/>), to be later replaced by the corresponding single word unit. ii) NVivo software (Version 12, QSR International Pty Ltd., Victoria, Australia) was used for identifying related words sharing a common stem (e.g. different forms of the same verb), to be later replaced by a single word selected as representative. Replacements of both bi-grams and stem words were performed by including a thesaurus list in the VOSviewer software when creating the co-occurrence networks. Alternatively, these steps can be performed automatically using the tm and ngram packages in R software.

Finally, to analyse Twitter content on gluten-free, the co-occurrence networks for the 16,386 tweets and those mentioning specific products (*bread, cookie, cake, beer, and pizza*) were obtained using the pre-cleaned text of the tweets.

3. Results

3.1. Themes and sub-themes in tweets on gluten-free

Table 1 shows the content of the subset of 3,000 tweets summarised into nine main themes: products, places, culinary preparations, product-related characteristics, ingredients, occasions, social context, diet / health, and sensory characteristics / sensations. Tweets relating to the themes product, places, and culinary preparation were the most frequent (> 30%). Although with lower frequency (< 10%), other themes related to diet or health issues and sensory characteristics / sensations were also found. The sub-theme *positive_sensation* referred to expressions of positive ideas and opinions such as *good, like, well, excellent, and enjoy*. The

sub-theme *negative_sensation* referred to expressions of negative ideas and opinions such as *bad*, *worse*, and *awful*.

Within each theme, four to fifteen sub-themes were identified. Product was the category with the highest number of sub-themes (15): *cookie*, *bread*, *pizza-dough-patty*, *beer*, *cereal*, *pasta*, *cake-pastry*, *meat*, *fish-seafood*, *fruit-vegetable*, *dairy*, *drink*, *mix*, *snack*, and *others*. Within the themes places and ingredients, a high number of sub-themes were also identified (12 each). For places, the sub-themes were *city*, *country*, *restaurant*, *chain*, *bakery*, *supermarkets*; *Carrefour*, *Lidl*; and *Mercadona*, *bar*, and *home*. Culinary aspects included *recipe*, *meat meal*, *fish_meal*, *vegetable-rice-legume_meal*, *pasta-meal*, *creams-soup-sauce*, *drinks*, and *eggs*. Tweets referring to ingredients contained *fruits-vegetable*, *cocoa*, *flour*, *cereal* as *corn*, *rice* or *wheat*, *nut*, *dairy products*, *sweetener*, and *meat*. The most relevant moments were eating occasions along the day, and the name of the days. Social themes referred to family, friends, and events. For diet and health, the sub-themes were *coeliac* and other disease aspects.

3.2. Comparison of co-occurrence networks from coded and pre-cleaned text

Figures 2 and 3 show the co-occurrence networks for coded and just pre-cleaned text from the subset of 3,000 tweets. In this plot, the size of circles represents the frequency of mentioned terms (or sub-themes, in the case of coded text). Co-occurrence is shown by the line connecting two terms, which is thicker when more times are co-mentioned together within the same tweet. Furthermore, the position of a term in the network is determined by the co-occurrence regarding all the other terms. Printed versions of the network make it difficult to see the connections, still the VOSviewer software can show every single term and the connections they have with others by just clicking on the word. For example, Figure 4 shows the highlighted connections of the term *flour* with its co-mentioned words (in the same network of Figure 3).

Co-occurrence networks from coded text showed, in agreement with Table 1, the most frequent sub-themes (bigger circles) were related to products (*cake-pastry*, *bread*, *cookie* and *pizza-dough-patty*), culinary preparations (*recipe*), places (*city*, *bakery* and *restaurant*), and product associated characteristics (*brand*). Finally, although the theme sensory / sensation was one of the least mentioned, *positive* was one of the most frequently mentioned sub-themes.

Relationships among sub-themes and their position show products at the ‘nodes’ of the network structure. The products *cake-pastry*, *bread*, and *cookie* appeared separated from the others and related to *recipe* and *flour*, which was connected to different cereals and other flour sources like *wheat*, *spelt*, *corn*, *chickpea*, *carob*, *rice*, *brown_rice* and *almond*. Few ingredients (*cocoa* / *chocolate* and *fruit-vegetables*) also appeared connected with these three products. *Pizza-dough-patty* and *beer* appeared related to occasions (*weekend* and *snack_time*), places (*restaurant*, *city*,

food chain and *supermarket*), social (*family-fiend*), and with *brand* and *coeliac*. Besides, *coeliac* was related to *restaurant*, *city*, *food chain*, and *diet*. *Pasta* related to places (*restaurant* and *city*), social aspects (*family-friend*), the word *coeliac* and to other groups of sub-themes like ingredients (*fruit-vegetable* and *fish-seafood*) and occasions (*dinner* and *weekday*). Sub-themes *positive sensations*, *brand*, and *bakery* appeared in the central area of the network indicating that they were equally related to most other sub-themes.

The co-occurrence network obtained from the pre-cleaned text of the same 3,000 tweets (Figure 3) showed many terms represented by circles of intermediate size. The frequency of mention (also indicated in Table 2) was high for products (*bread*, *cake*, *cookie*, *beer*, *pizza* and *pasta*), ingredients (*cocoa / chocolate*, *flour*, *cream*, *rice* and *sugar*), product-related features (*vegan*, *lactose_free*, *new*, *sugar_free* and *organic*), positive sensations (*tasty*, *good*, *delicious* and *enjoy*) and other terms related to different aspects like *recipe*, *coeliac*, *meal*, *eat*, *restaurant*, *dessert*, *diet* and *cook*.

Figure 3 shows that products like *cookie*, *bread*, and *cake* were frequently co-mentioned with *recipe* and *vegan*. *Bread* and *cake* were also co-mentioned with *flour*, which related to different cereals and flour sources. Many small-size terms comprising fruits, nuts, and condiments related to *cake*. *Pasta* also related to *recipe*, *vegan*, *flour*, and specific ingredients such as *cheese*, *sauce*, *ham*, and *fungi*. *Coeliac* was related to *restaurant*, *product*, *diet*, *meal / menu* and *suitable*.

Here, the terms *tasty*, *taste*, *delicious*, *enjoy*, *lactose-free* and *eat* were in the centre of the network (equally related to most other terms).

3.3. Co-occurrence networks from tweets on “gluten-free” in general and from tweets on specific products.

The co-occurrence network (Figure 5) was obtained for the totality of tweets (16,386) after just the pre-cleaning step. In this network, more mentioned terms and connections were the same as the subset of pre-cleaned text of 3,000 tweets (Figure 3). However, two main effects were observed when increasing the number of tweets in the analysis. First, the differences in frequency (label size) among terms were more evident in the network with totality of tweet. Second, some terms that initially showed low frequency (some specific ingredients and brand and company names) became relevant, showing more detailed information.

Individual co-occurrence networks were obtained for the products *bread*, *cake*, *cookie*, *beer*, and *pizza*. Frequencies of mention for terms related to these products are shown in Table 4. These products were selected given the relevance of the presence of gluten in their composition and for their high frequencies of mention in the totality of tweets (Table 3).

In the network for *bread* (Figure 6), the terms *recipe* and *flour* were highly mentioned, and also others related to many relevant different types of flours (*chickpea*, *teff*, *quinoa*, *linseed*, *rice*, *buckwheat*, *seed*, *yucca*, *wheat*, *corn*, *spelt*, *rye* and *whole_grain*). Other terms like *breakfast*, bread for *burger* and *home_made* (relating to *delicious* and *oven*) were also relevant.

For *cake* (Figure 7) and *cookie* (Figure 8), networks were similar. *Recipe* and *cocoa / chocolate* were the most relevant terms and many types of ingredients, mostly fruits and nuts (*carrot*, *apple*, *strawberry*, *cherry*, *orange*, *walnut*, and *blueberry*) became relevant. Terms relating to context such as celebrations (*Christmas* and *Halloween*) and *sugar_free* were associated to both products. *Birthday* was importantly mentioned in the *cake* network, while *snack_time* was mentioned in the *cookie* network.

For *beer* (Figure 9), the relevant terms were *craft* (relating to *burger* and *taste*), *alcohol_free* (relating to *new*), and many different brands and types of beer (*Estrella Galicia*, *Cruzcampo*, *Daura*, *Ambar*, *Mahou*, *IPA*, and *Pilsen*). Terms like *bar*, *restaurant*, *Saturday*, *home*, *day*, *have* and *now*, appeared related to *enjoy*, *delicious*, *good*, *taste* and *happy*. Regarding places, *bars* and *restaurants* were mentioned (places to drink), also *supermarkets* and *shops* (places to buy beer) were named.

For *pizza* (Figure 10), *dough* was the most mentioned term, while others related to *recipe*, *home_made*, and different ingredients were shown in this network. Restaurants, food chains (*Tele Pizza*, *Pizza Hut* and *Domino's Pizza*) and aspects related to buying and supermarkets (*promo*, *new* and *novelty*) were also relevant.

All products showed connections to positive ideas and concepts, like *good*, *enjoy*, *delicious*, *tasty*, and *happy*. *Vegan* and *lactose-free* were terms that appeared frequently for all products except for *beer*.

4. Discussion

4.1. Treatment of the information obtained from Twitter

In this study, the analysis of tweet information was conducted using co-occurrence networks. The tweets were pre-processed in two ways, either manually coding the content of tweet or using direct raw tweet text (after just a pre-cleaned step). The networks from the manual coded and pre-cleaned text from the subset of 3,000 tweets revealed similar main ideas about the topic “gluten-free” but they were differently represented. When coding the information after reading, concepts and ideas were grouped into the same term in the network from coded text, increasing the frequency of mention and, as it was intended, highlighting the differences in the relevance of sub-themes.

In many cases, the terms and sub-themes with a high frequency of mention were the same in both networks. However, few sub-themes highly cited in the coded text network were not found in the pre-cleaned text network because the frequency was split into different terms. This is observable for the names of weekdays, cities, and countries and for positive sensations. Therefore, manual reading and coding text into themes, organises the information more and structures separate topics together; however, can lead to a loss of detailed or meaningful information for interpretation.

Notably, the network obtained from pre-cleaned text of the totality of 16,386 tweets showed a structure and size distribution of terms that resembled those observed for the coded text network. Seemingly increasing the sample size (tweets) for analysis, highlighted the differences in the relevance of terms, like what was achieved by manually coding the text of the tweets. In this study, with a large volume of tweets, only a cleaning of text is needed to obtain co-occurrence networks with relevant and meaningful terms, representing what people tweet about for a given topic. Furthermore, using pre-cleaned text avoids limitations of coding, subjectivity bias, and mistakes. Here, both *vegan* and *lactose_free* terms showed a high frequency of mention in the pre-cleaned text network and were not found in the coded text. This mistake in coding (*vegan* and *lactose_free* were not considered when reading the tweets and identifying the relevant themes) made that this information was not considered and therefore lost. In addition, interpreting the information provided by coded text, without knowing the coding criteria, can be complex. Finally, saving time is also an advantage as manual coding of 3,000 tweets took around 120 hours of work of two researchers and the text pre-cleaning process, in turn, takes only 2-4 hours of one researcher, regardless of the volume of tweets used for the analysis.

To date, co-occurrence networks had been scarcely applied in the analysis of Twitter than other methods (Eriksson-Backa, Holmberg, & Ek, 2016; He et al., 2017; Kang et al., 2017). The most direct and simplest method to analyse tweets' content is the study of the frequencies of mentioned words and using wordclouds to visualise its relevance (Carr et al., 2015; Fried et al., 2015; Le & Worch, 2014; Vidal et al., 2018). However, as demonstrated by Vidal et al. (2015) qualitative analysis of text is necessary to interpret each word in its context, avoiding misinterpretation and providing full meaning of ideas. Thus, co-occurrence networks that show the frequency of mentions of a word and the connections with the co-mentioned words, can play the role of qualitative analysis and place words into their context automatically. Text mining techniques based on machine learning algorithms are also used to analyse Twitter text, using themes already established or commonly studied (Mishra & Singh, 2016; Shah et al., 2016). In contrast, co-occurrence networks are less complex and do not require the adjustment of algorithms for new topics.

4.2. Information found on Twitter about gluten-free

Twitter exploration using co-occurrence networks has been useful to gather knowledge about what is relevant, in general, for a specific target of consumers, in this case gluten-free consumers. However, from an industry perspective, it is sometimes interesting to get more detailed information on what consumers say about a topic regarding a specific product category, brand, or attribute. Here, obtaining the individual networks from tweets, including “gluten-free” and the specific product, allowed better visualisation of the more relevant topics for each of the products.

When tweeting about gluten-free, people mainly talk about five product categories, which are critical for the lack of gluten (*bread*, *cake*, *cookie*, *beer*, and *pizza*), share information about how to prepare food (recipes and ingredients), where to get food (restaurants, supermarkets, food chains), describe the context of consumption (social events, places, and occasions), or talk about diet and health aspects related to coeliac disease.

Bakery goods and pizza are essential products for the people tweeting about gluten-free. The large amount of information about recipes, ingredients, and culinary preparations suggest that people commonly prepare these categories of products at home. The importance of recipe sharing in online communities has also been highlighted by Blackburn, Yilmaz, & Boyd (2018), who observed that the process of cooking, sharing instructions for recipes, and the ingredients used are key topics of discussion on the social networking site Reddit.

The mention of occasions, together with people; places; and positive sensations indicated that describing the eating / drinking situation was also a frequent topic in tweets. People associated the consumption of *pizza* or *beer* to a given time (weekend) or place (restaurant / bar / home), while the consumption of *bread*, *cake*, and *cookie* was linked to breakfast, and *cake* was especially related to celebrations. The importance of context of consumption has also been reported by previous works in other product categories and situations. Beyond the scope of social media, Spinelli et al. (2017) found that coffee consumers focus more on the pleasantness of the related social situation in which the coffee is consumed rather than on the experience of enjoying the product by itself. When exploring food topics on Twitter, Vidal et al. (2015) observed, agreeing with our findings, that people usually mention the food they consume, places, or the people they share the meal with when tweeting about different eating situations like breakfast, lunch, dinner, and snack. Here, the eating situation and social context was a relevant part of the information obtained, but for people tweeting about gluten-free, it was of great importance how and where to find the products for this specific population with restrictions in their diet.

The high frequencies of mention found for positive sensations in this study reveals that people talk positively, avoiding negative aspects or complaints about gluten-free. These observations

are in line with other studies showing a positive attitude when mentioning or remembering specific eating situations (Vidal et al., 2015) or “memorable” meals (Piqueras-Fiszman & Jaeger, 2015), especially those involving family, friends, good food, and positive emotions.

The presence of terms like *bio*, *organic*, *natural*, *artisan*, *vegan*, *sugar-free* and *eco* in central areas of the network reveals that people tweeting about gluten-free are especially interested in these aspects. Regardless of the reasons behind this, people associate gluten-free diet to a particular style of food, evoking a sense of natural foods and lifestyle.

Regarding the essential importance of avoiding gluten in disorders related to this protein, concepts associated to diet, health, and food intolerances revealed the special concern of people about these aspects. Previous studies showed the importance, among consumers with special requirements in their diet, of consulting the available information online regarding food allergies and intolerances (Hamshaw, Barnett, & Lucas, 2018). Therefore, sharing information about food chains for eating or ordering *pizza* was especially relevant. Given that eating out presents a particular challenge for people suffering food allergies and / or intolerances (Begen et al., 2016), they pay special attention to the places where they can eat, limiting the potential risk to their health and where they can find a tasty and compatible gluten-free product.

Other social networks have been used for exploring consumer behaviour using different approaches depending on the type of information that users share. For example, the visual analysis of images on Instagram, related to food products posted by users, showed more detailed information about the context surrounding consumption, and also the brand, type, and flavour of the product (Gómez-Corona, Ares, Spinelli, Veflen, & Stathopoulou, 2019). Other Instagram studies have focused on the analysis of the hashtags posted associated to a particular topic. Pilař, Balcarová, Rojík, Tichá, & Poláková (2018) explored the farmers’ market context, which showed customer clusters according to their attitudes or motivations, finding segments oriented by product, emotions, and social aspects. The same authors also explored Instagram interactions and hashtags to study consumer perceptions and feelings regarding organic food production; also they identified communities of users motivated by areas like healthy living, vegetarian diets, clean eating, and active healthy living (Pilař, Stanislavská, et al., 2018). Arellano-Covarrubias, Gómez-Corona, Varela, & Escalona-Buendía (2019) studied the relationship between flavours in beer pairings across different cultures and countries, using Synthesio® for retrieving data from Twitter, Facebook, forums, and blogs. This pay-platform allows access to geographical information (as well as gender and age when specified by users in their public profile) of the messages posted, which made possible to perform the study by country. In this study, it would be interesting to analyse potential differences across countries, gender, and age groups. However, the rtweet package did not show this information from the users.

5. Conclusions

Co-occurrence networks allow the understanding of the information on Twitter showing the relevance of terms and how they are structured through co-occurrence connections.

This study shows that co-occurrence networks can be used, almost directly, from pre-cleaned data without losing relevant information. Furthermore, the study highlighted the importance of the number of tweets when making relevant and dependable information.

This approach almost automatic based on co-occurrence networks from pre-cleaned text can be a tool for analysing, in an easy and fast manner, the massive information continuously available and updated from online platforms like Twitter. This would enable monitoring of what people think, feel, and talk about regarding product categories, services, and topics and how they evolve.

This strategy allows exploration of topics concerning a certain group of consumers. Here, people talking about gluten-free in Twitter mentioned five categories of products (*bread, cake, cookie, beer, and pizza*) plus the context of consumption and purchasing. Tweets were posted to share eating or preparing food situations and also to recommend how to get gluten-free products: with bakery goods, recipes and ingredients for preparing them; while for beer and pizza, recommendations were related to brands, supermarkets and restaurants.

Acknowledgements

Authors are grateful to the Spanish Ministry of the Economy and Competitiveness for financial support (project AGL-2016-75403-R) and for the Juan de la Cierva contract for author Laura Laguna (IJCI-2016-27427). Furthermore, to Generalitat Valenciana (Project Prometeo 2017/189). Authors are also grateful to Dr. Waltman for his valuable advice related to using VOSviewer technique.

References

- Arellano-Covarrubias, A., Gómez-Corona, C., Varela, P., & Escalona-Buendía, H. B. (2019). Connecting flavors in social media: A cross cultural study with beer pairing. *Food Research International*, 115(July 2018), 303–310. <https://doi.org/10.1016/j.foodres.2018.12.004>
- Begen, F. M., Barnett, J., Payne, R., Roy, D., Gowland, M. H., & Lucas, J. S. (2016). Consumer preferences for written and oral information about allergens when eating out. *PLoS ONE*, 11(5), 1–12. <https://doi.org/10.1371/journal.pone.0156073>
- Blackburn, K. G., Yilmaz, G., & Boyd, R. L. (2018). Food for thought: Exploring how people think and talk about food online. *Appetite*, 123, 390–401. <https://doi.org/10.1016/j.appet.2018.01.022>
- Capriles, V. D., dos Santos, F. G., & Arêas, J. A. G. (2016). Gluten-free breadmaking: Improving nutritional and bioactive compounds. *Journal of Cereal Science*, 67, 83–91. <https://doi.org/10.1016/j.jcs.2015.08.005>
- Carr, J., Decreton, L., Qin, W., Rojas, B., Rossochacki, T., & Yang, Y. wen. (2015). Social media in product development. *Food Quality and Preference*, 40(PB), 354–364. <https://doi.org/10.1016/j.foodqual.2014.04.001>
- Chae, B. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247–259. <https://doi.org/10.1016/j.ijpe.2014.12.037>
- Chen, X., & Yang, X. (2014). Does food environment influence food choices? A geographical analysis through “tweets.” *Applied Geography*, 51, 82–89. <https://doi.org/10.1016/j.apgeog.2014.04.003>
- Christoph, M. J., Larson, N., Hootman, K. C., Miller, J. M., & Neumark-Sztainer, D. (2018). Who Values Gluten-Free? Dietary Intake, Behaviors, and Sociodemographic Characteristics of Young Adults Who Value Gluten-Free Food. *Journal of the Academy of Nutrition and Dietetics*, 118(8), 1389–1398. <https://doi.org/10.1016/j.jand.2018.04.007>
- Constantinides, E., & Holleschovsky, N. I. (2016). Impact of online product reviews on purchasing decisions. *WEBIST 2016 - Proceedings of the 12th International Conference on Web Information Systems and Technologies*, 1(Webist), 271–278.
- Cunillera, T., & Guilera, G. (2018). Twenty years of statistical learning: from language, back to machine learning. *Scientometrics*, 117(1), 1–8. <https://doi.org/10.1007/s11192-018-2856-x>
- Da Silva, N. F., Hruschka, E. R. E. R., & Hruschka, E. R. E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66, 170–179.

<https://doi.org/10.1016/j.dss.2014.07.003>

- do Nascimento, A. B., Fiates, G. M. R., & Teixeira, E. (2017). We want to be normal! Perceptions of a group of Brazilian consumers with coeliac disease on gluten-free bread buns. *International Journal of Gastronomy and Food Science*, 7(January), 27–31. <https://doi.org/10.1016/j.ijgfs.2017.01.001>
- Eriksson-Backa, K., Holmberg, K., & Ek, S. (2016). Communicating diabetes and diets on Twitter - a semantic content analysis. *International Journal of Networking and Virtual Organisations*, 16(1), 8. <https://doi.org/10.1504/IJNVO.2016.075133>
- Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2015). Analyzing the language of food on social media. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, (Section II), 778–783. <https://doi.org/10.1109/BigData.2014.7004305>
- Hamshaw, R. J. T., Barnett, J., & Lucas, J. S. (2018). Tweeting and Eating: The Effect of Links and Likes on Food-Hypersensitive Consumers' Perceptions of Tweets. *Frontiers in Public Health*, 6(April), 1–12. <https://doi.org/10.3389/fpubh.2018.00118>
- He, W., Tian, X., Tao, R., Zhang, W., Yan, G., & Akula, V. (2017). Application of social media analytics: A case of analyzing online hotel reviews. *Online Information Review*, 41(7), 921–935. <https://doi.org/10.1108/OIR-07-2016-0201>
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33, 464–472. <https://doi.org/10.1016/j.ijinfomgt.2013.01.001>
- Houben, A., Höchstötter, A., & Becker, T. (2012). Possibilities to increase the quality in gluten-free bread production: An overview. *European Food Research and Technology*, 235(2), 195–208. <https://doi.org/10.1007/s00217-012-1720-0>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, 35(29), 3621–3638. <https://doi.org/10.1016/j.vaccine.2017.05.052>
- Le, S., & Worch, T. (2014). *Analyzing Sensory Data with R - CRC Press Book*. <https://doi.org/10.1201/b17502>
- Marinho, V. Q., Hirst, G., & Amancio, D. R. (2017). Authorship Attribution via Network Motifs Identification. *Proceedings - 2016 5th Brazilian Conference on Intelligent Systems*,

BRACIS 2016, 355–360. <https://doi.org/10.1109/BRACIS.2016.071>

Marston, K., Khouryieh, H., & Aramouni, F. (2016). Effect of heat treatment of sorghum flour on the functional properties of gluten-free bread and cake. *LWT - Food Science and Technology*, 65(2016), 637–644. <https://doi.org/10.1016/j.lwt.2015.08.063>

Matos, M. E., & Rosell, C. M. (2012). Relationship between instrumental parameters and sensory characteristics in gluten-free breads. *European Food Research and Technology*, 235(1), 107–117. <https://doi.org/10.1007/s00217-012-1736-5>

Mention. (2018). *The Twitter Engagement Report*. Retrieved September 2019, from <https://mention.com/en/reports/twitter/>

Mishra, N., & Singh, A. (2016). Use of twitter data for waste minimisation in beef supply chain. *Annals of Operations Research*, 1–23. <https://doi.org/10.1007/s10479-016-2303-4>

Missbach, B., Schwingshackl, L., Billmann, A., Mystek, A., Hickelsberger, M., Bauer, G., & König, J. (2015). Gluten-free food database: The nutritional quality and cost of packaged gluten-free foods. *PeerJ*, 2015(10). <https://doi.org/10.7717/peerj.1337>

Moe, W. W., & Schweidel, D. A. (2017). Opportunities for Innovation in Social Media Analytics. *Journal of Product Innovation Management*, 34(5), 697–702. <https://doi.org/10.1111/jpim.12405>

Molina-Rosell, C. (2013). Alimentos sin gluten derivados de cereales. In L. Rodrigo & A. Peña (Eds.), *Enfermedad celíaca y sensibilidad al gluten no celíaca* (pp. 447–461). Barcelona, España: OmniaScience. <https://doi.org/10.3926/oms.27>

Naqash, F., Gani, A., Gani, A., & Masoodi, F. A. (2017). Gluten-free baking: Combating the challenges - A review. *Trends in Food Science and Technology*, 66, 98–107. <https://doi.org/10.1016/j.tifs.2017.06.004>

Nguyen, T. T., Criss, S., Allen, A. M., Glymour, M. M., Phan, L., Trevino, R., ... Nguyen, Q. C. (2019). Pride, love, and twitter rants: Combining machine learning and qualitative techniques to understand what our tweets reveal about race in the us. *International Journal of Environmental Research and Public Health*, 16(10), 1–19. <https://doi.org/10.3390/ijerph16101766>

O'Shea, N., Doran, L., Auty, M., Arendt, E., & Gallagher, E. (2013). The rheology, microstructure and sensory characteristics of a gluten-free bread formulation enhanced with orange pomace. *Food & Function*, 4(12), 1856. <https://doi.org/10.1039/c3fo60294j>

Penjumras, P., Janmeesup, C., Umnat, S., Chokeprasert, P., Wattananapakasem, I., & Phaiphon, A. (2019). Development of gluten-free cream puff with the addition of carboxymethylcellulose and carrageenan. *Food Research*, 3(2), 138–144.

548 [https://doi.org/10.26656/fr.2017.3\(2\).133](https://doi.org/10.26656/fr.2017.3(2).133)

549 Pilař, L., Balcarová, T., Rojik, S., Tichá, I., & Poláková, J. (2018). Customer experience with
 550 farmers' markets: What hashtags can reveal. *International Food and Agribusiness*
 551 *Management Review*, 21(6), 755–770. <https://doi.org/10.22434/IFAMR2017.0039>

552 Pilař, L., Stanislavská, L. K., Rojik, S., Kvasnička, R., Poláková, J., & Gresham, G. (2018).
 553 Customer experience with organic food: Global view. *Emirates Journal of Food and*
 554 *Agriculture*, 30(11), 918–926. <https://doi.org/10.9755/ejfa.2018.v30.i11.1856>

555 Piqueras-Fiszman, B., & Jaeger, S. R. (2015). Emotions associated to mealtimes: Memorable
 556 meals and typical evening meals. *Food Research International*, 76(P2), 243–252.
 557 <https://doi.org/10.1016/j.foodres.2014.12.004>

558 Platania, M., & Spadoni, R. (2018). How people share information about food: Insights from
 559 tweets regarding two Italian Regions. *International Journal on Food System Dynamics*,
 560 9(2), 149–165. <https://doi.org/10.18461/ijfsd.v9i2.924>

561 R Core Team. (2016). R: A language and environment for statistical computing. Vienna,
 562 Austria: R Foundation for Statistical Computing, 860–864.

563 Ruiz, J. B., & Barnett, G. A. (2015). Exploring the presentation of HPV information online: A
 564 semantic network analysis of websites. *Vaccine*, 33(29), 3354–3359.
 565 <https://doi.org/10.1016/j.vaccine.2015.05.017>

566 Samoggia, A., Bertazzoli, A., & Ruggeri, A. (2019). Food retailing marketing management:
 567 social media communication for healthy food. *International Journal of Retail and*
 568 *Distribution Management*. <https://doi.org/10.1108/IJRDM-08-2018-0178>

569 Sapone, A., Bai, J. C., Ciacci, C., Dolinsek, J., Green, P. H. R., Hadjivassiliou, M., ... Fasano,
 570 A. (2012). Spectrum of gluten-related disorders: Consensus on new nomenclature and
 571 classification. *BMC Medicine*, 10. <https://doi.org/10.1186/1741-7015-10-13>

572 Sengupta, A., & Ghosh, A. (2020). Mining social network data for predictive personality
 573 modelling by employing machine learning techniques. *Lecture Notes in Electrical*
 574 *Engineering*, 575, 113–127. https://doi.org/10.1007/978-981-13-8687-9_11

575 Shah, S., Kumar, K., & Saravanaguru, R. K. (2016). Sentimental analysis of twitter data using
 576 classifier algorithms. *International Journal of Electrical and Computer Engineering*, 6(1),
 577 357–366. <https://doi.org/10.11591/ijece.v6i1.8982>

578 Sharma, N., Bairwa, M., Gowthamghosh, B., Gupta, S. D., & Mangal, D. K. (2018). A
 579 bibliometric analysis of the published road traffic injuries research in India, post-1990, 1–
 580 11. <https://doi.org/10.1186/s12961-018-0298-9>

- Skaf, L., Buonocore, E., Dumontet, S., Capone, R., & Franzese, P. P. (2020). Applying network analysis to explore the global scientific literature on food security. *Ecological Informatics*, 56(January), 101062. <https://doi.org/10.1016/j.ecoinf.2020.101062>
- Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *Scientometrics*, 85(1), 65–79. <https://doi.org/10.1007/s11192-010-0259-8>
- van Eck, N. J., & Waltman, L. (2007). VOS: A new method for visualizing similarities between objects. *Advances in Data Analysis*, (April 2006), 299–306. <https://doi.org/10.1007/978-3-540-70981-7>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- van Eck, N. J., & Waltman, L. (2018). VOSviewer Manual version 1.6.9. *Universiteit Leiden*, (August), 1–51. <https://doi.org/10.3402/jac.v8.30072>
- van Eck, N. J., Waltman, L., Noyons, E. C. M., & Buter, R. K. (2010). Automatic term identification for bibliometric mapping. *Scientometrics*, 82(3), 581–596. <https://doi.org/10.1007/s11192-010-0173-0>
- van Eck, N. J., Waltman, L., van den Berg, J., & Kaymak, U. (2006). Vizualizing the computational intelligence field. *IEEE Computational Intelligence Magazine*, 1(4), 6–10. <https://doi.org/10.1109/MCI.2006.329702>
- van Zoonen, W., & van Der Meer, T. G. L. A. (2016). Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior*, 63, 132–141. <https://doi.org/10.1016/j.chb.2016.05.028>
- Vidal, L., Ares, G., & Jaeger, S. R. (2018). Application of Social Media for Consumer Research. In G. Ares & P. Varela (Eds.), *Methods in Consumer Research, Volume 1: New Approaches to Classic Methods* (pp. 125–155). Woodhead Publishing.
- Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference*, 49, 119–128. <https://doi.org/10.1016/j.foodqual.2015.12.002>
- Vidal, L., Ares, G., Machín, L., & Jaeger, S. R. (2015). Using Twitter data for food-related consumer research: A case study on “what people say when tweeting about different eating situations.” *Food Quality and Preference*, 45, 58–69. <https://doi.org/10.1016/j.foodqual.2015.05.006>
- Wen, B., Horlings, E., van der Zouwen, M., & van den Besselaar, P. (2017). Mapping Science

Through Bibliometric Triangulation: An Experimental Approach Applied to Water
Research. *Journal of The Association for Information Science and Technology*, 68(3),
724–738.

Zhao, B., Zhang, Z., Qian, W., & Zhou, A. (2013). Identification of collective viewpoints on
microblogs. *Data and Knowledge Engineering*, 87, 374–393.
<https://doi.org/10.1016/j.datak.2013.05.003>

Table 1

Themes and sub-themes identified in Twitter messages on gluten-free by qualitative analysis of a subset of 3,000 tweets. For each sub-theme, frequency of mention is indicated in brackets.

Theme	Sub-themes	Frequency (%)
Products	<i>cake-pastry</i> (14.4), <i>bread</i> (7.2), <i>fruit-vegetable</i> (4.7), <i>cookie</i> (4.5), <i>pizza-dough-patty</i> (4.1), <i>cereal</i> (3.8), <i>snack</i> (2.5), <i>dairy</i> (1.7), <i>beer</i> (1.5), <i>meat_product</i> (1.4), <i>pasta</i> (1.4), <i>mix</i> (0.6), <i>fish-seafood</i> (0.5), <i>drink</i> (0.2)	48.7
Places	<i>city</i> (9.1), <i>bakery</i> (4.4), <i>restaurant</i> (3.7), <i>chain</i> (2.3), <i>country</i> (1.0), <i>supermarket</i> (1.0), <i>Mercadona</i> (0.7), <i>bar</i> (0.4), <i>home</i> (0.4), <i>Lidl</i> (0.3), <i>Carrefour</i> (0.1)	32.1
Culinary preparations	<i>recipe</i> (11.6), <i>vegetable-rice-legume_meal</i> (2.6), <i>meat_meal</i> (1.9), <i>cream-soup-sauce</i> (1.9), <i>pasta_meal</i> (1.4), <i>fish_meal</i> (0.7), <i>eggs</i> (0.1)	31.8
Product-related characteristics	<i>brand</i> (5.7), <i>eco</i> (1.9), <i>bio-organic</i> (0.8), <i>craft</i> (0.7), <i>natural-home-made</i> (0.6), <i>price</i> (0.4)	19.9
Ingredients	<i>fruit-vegetable</i> (4.7), <i>cereal</i> (3.8), <i>flour</i> (2.3), <i>dairy</i> (1.7), <i>cocoa/chocolate</i> (1.4), <i>nut</i> (1.1), <i>sweetener</i> (0.7), <i>corn</i> (0.3), <i>rice</i> (0.2), <i>wheat</i> (0.2)	16.9
Occasions	<i>weekend</i> (3.4), <i>breakfast</i> (2.9), <i>weekday</i> (2.2), <i>snack_time</i> (1.3), <i>lunch</i> (1.1), <i>month-season</i> (1.0), <i>dinner</i> (0.6), <i>day_moment</i> (0.3)	13.0
Social context	<i>family-friends</i> (3.1), <i>social_event</i> (1.3), <i>professional_event</i> (2.7), <i>platform</i> (0.8), <i>professional_area</i> (0.4), <i>social_policy</i> (0.3), <i>association</i> (0.2)	12.8
Diet / Health	<i>coeliac</i> (6.6), <i>diet</i> (1.5), <i>drug-treatment</i> (0.3), <i>coeliac_disease</i> (0.2), <i>associated_diseases</i> (0.1)	9.8
Sensory characteristics / Sensations	<i>positive_sensation</i> (6.2), <i>complaint</i> (1.5), <i>negative_sensation</i> (0.8), <i>craving-hunger</i> (0.5)	9.1

Table 2

Frequency of terms mentioned on Twitter when talking about gluten-free obtained from the subset of 3,000 tweets.

Term	Frequency (%)	Term	Frequency (%)	Term	Frequency (%)
bread	7.2	dessert	2.4	buy	1.4
vegan	6.7	flour	2.4	eco	1.3
recipe	6.6	menu	2.4	healthy	1.3
coeliac	5.8	diet	2.4	craft	1.3
cake	4.9	dairy_free	2.3	easy	1.3
cocoa/chocolate	4.3	pizza	2.3	home_made	1.3
lactose_free	4.0	food	2.1	bakery	1.3
product	3.7	cook	2.0	home	1.2
cookie	3.6	beer	2.0	better	1.2
tasty	3.5	christmas	2.0	sweet	1.2
meal	3.4	option	1.8	vegetarian	1.1
eat	3.3	organic	1.8	rice	1.1
good	3.0	day	1.7	pancake	1.1
taste	3.0	enjoy	1.7	love	1.0
new	2.7	pasta	1.7	prepare	1.0
restaurant	2.7	cream	1.4	vegetable	1.0
delicious	2.6	breakfast	1.4	want	1.0
sugar_free	2.5				

Only frequencies higher than 1% are listed.

Table 3

Frequency of terms mentioned on Twitter when talking about gluten-free obtained from the totality of 16,386 tweets.

Term	Frequency (%)	Term	Frequency (%)	Term	Frequency (%)
bread	6.7	day	2.4	craft	1.5
coeliac	6.0	beer	2.4	enjoy	1.4
vegan	5.7	delicious	2.3	better	1.4
recipe	5.3	menu	2.3	dairy_free	1.4
eat	3.9	pizza	2.2	option	1.3
product	3.8	Christmas	2.2	breakfast	1.2
cake	3.7	flour	2.1	want	1.2
lactose_free	3.7	organic	2.1	home	1.2
cocoa/chocolate	3.2	sugar_free	1.9	sweet	1.1
tasty	3.2	cook	1.9	dessert	1.1
meal	3.1	restaurant	1.8	now	1.1
diet	3.1	food	1.6	cream	1.1
new	2.9	healthy	1.6	easy	1.1
cookie	2.8	eco	1.6	morning	1.0
taste	2.7	pasta	1.5	happy	1.0
good	2.5	buy	1.5		

Only frequencies higher than 1% are listed.

657 **Table 4**

658 Frequency of terms mentioned on Twitter when talking about gluten-free obtained from the
 659 totality of 16,386 tweets), related to different products.

Bread		Cake		Cookie		Beer		Pizza	
Term	F (%)	Term	F (%)	Term	F (%)	Term	F (%)	Term	F (%)
recipe	6.2	cocoa/chocolate	14.5	recipe	12.5	craft	10.8	dough	8.3
eat	5.5	recipe	13.0	cocoa/chocolate	11.9	cruzcampo	8.2	dominospizza	8.1
coeliac	5.1	cupcake	10.0	vegan	9.1	alcoholfree	8.0	telepizza	6.8
flour	4.9	lactose_free	7.4	Christmas	6.9	coeliac	7.0	vegan	6.5
tasty	4.7	vegan	6.4	tasty	6.7	taste	7.0	taste	4.3
day	3.8	tasty	5.1	delicious	6.5	new	4.6	eat	4.0
burger	3.7	dessert	4.8	lactose_free	5.8	burger	3.9	recipe	4.0
vegan	3.6	delicious	4.4	taste	5.2	good	3.4	flour	3.8
dairy_free	2.8	apple	4.2	new	4.1	estrellagalicia	3.1	lactose_free	3.8
good	2.8	flour	4.2	buy	3.5	better	2.8	new	3.5
home_made	2.8	sugar_free	3.9	sugar_free	3.5	bar	2.6	Tuesday	3.5
delicious	2.8	birthday	3.8	eat	3.2	daura	2.6	bread	3.3
lactose_free	2.8	cream	3.8	good	3.2	day	2.6	want	3.3
taste	2.6	coeliac	3.5	sweet	3.2	enjoy	2.6	coeliac	3.0
sweet	2.3	sweet	3.0	flour	2.8	novelty	2.6	good	3.0
buy	2.2	dairy_free	2.9	coconut	2.6	now	2.6	vegetarian	2.8
dough	2.1	taste	2.9	cream	2.6	bread	2.3	dinner	2.5
easy	2.1	prepare	2.7	healthy	2.6	mahou	2.3	home_made	2.5
wheat	2.1	carrot	2.6	prepare	2.4	drink	2.1	menu	2.5
milk	2.0	cheesecake	2.3	sugar	2.4	menu	2.1	lactose	2.3
		coconut	2.3	want	2.4	tasty	2.1	promo	2.3
		christmas	2.1	apple	2.2			tasty	2.3
		celebrate	2.0	coeliac	2.2			craft	2.0
		easy	2.0	rice	2.2			enjoy	2.0
		eat	2.0	snack_time	2.2			Friday	2.0
		Halloween	2.0						

660 Only frequencies higher than 2% are listed.

Figure Captions

Figure 1. Example of a co-occurrence network from a dataset created from 70 documents describing flowers and containing seven terms whose occurrence and co-occurrences are indicated in the table.

Figure 2. Co-occurrence network of Twitter information on “gluten-free” obtained from manually coded text of the subset of 3,000 tweets.

Figure 3. Co-occurrence network of Twitter information on “gluten-free” obtained from pre-cleaned text of the subset of 3,000 tweets.

Figure 4. Connections for the term “*flour*” in the co-occurrence network (pre-cleaned text, Figure 3) highlighted by VOSviewer software when clicking on the term.

Figure 5. Co-occurrence network of Twitter information on “gluten-free” obtained from pre-cleaned text of the totality of tweets (16,386).

Figure 6. Co-occurrence network of Twitter information on “gluten-free” and “*bread*”.

Figure 7. Co-occurrence network of Twitter information on “gluten-free” and “*cake*”.

Figure 8. Co-occurrence network of Twitter information on “gluten-free” and “*cookie*”.

Figure 9. Co-occurrence network of Twitter information on “gluten-free” and “*beer*”.

Figure 10. Co-occurrence network of Twitter information on “gluten-free” and “*pizza*”.

Figure 1. Example of a co-occurrence network from a dataset created from 70 documents describing flowers and containing seven terms whose occurrence and co-occurrences are indicated in the table.

Figure 2. Co-occurrence network of Twitter information on “gluten-free” obtained from manually coded text of the subset of 3,000 tweets.

Figure 5. Co-occurrence network of Twitter information on “gluten-free” obtained from pre-cleaned text of the totality of tweets (16,386).

Figure 9. Co-occurrence network of Twitter information on “gluten-free” and “beer”.

Figure 10. Co-occurrence network of Twitter information on “gluten-free” and “*pizza*”.