Food Quality and Preference 75 (2019) 145–156

Contents lists available at ScienceDirect

Food Quality and Preference

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

# Consumers' associative networks of plant-based food product communications

Anne O. Peschel[a,*], Shahab Kazemi[b], Martina Liebichová[b], Stine Cecilie Mangaard Sarraf[a], Jessica Aschemann-Witzel[a]

[a] MAPP Centre, Department of Management, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark
[b] Department of Management, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark

ARTICLE INFO

*Keywords:*
Concept mapping
Flexitarians
Text mining
Network analysis
Potato protein

ABSTRACT

Food producers respond to the current consumer trend of clean label products and reducing meat consumption by increasingly offering plant-based food products and transparent, understandable ingredient lists. However, consumer interest can be driven by various motives and food producers face the challenge of identifying the most effective motive to address. We analyze concept maps of 90 consumers who received information that positioned plant-based food products as sustainable, healthy, or with a transparent ingredient focus. To assess the applicability of text mining with a view to reducing coder bias and the duration of qualitative data analysis, we compared the results of text mining versus a human coder approach.

Our results show that human coder analysis results in more detail, however the advantage of the text mining procedure is that it can run independently and analyze qualitative data more objectively. When a high degree of control and depth of analysis is necessary to satisfy the study objective, human coding might have its rewards. For the current study, both approaches draw a similar picture of the associative networks and are therefore equally suitable to satisfy the study objective. When plant-based diets are communicated solely based on the ingredient used for substituting animal-based ingredients, associative networks are less complex and associations are primarily concerned with taste. A health communication perspective results in more complex networks with a focus on other food product properties such as processing degree and nutrition. A sustainability communication also results in higher complexity, with fewer associations concerning the product properties itself, but rather with the environmental impact and the authenticity of the product. The in-depth understanding of consumers' associations evoked by communicating different perspectives of plant-based food products can be used by practitioners in tailoring their marketing activities to the characteristics of their product offerings.

## 1. Introduction

Consumers today are increasingly willing to reduce their meat consumption and adopt plant-based alternatives in their diet. This trend has been described as a flexitarian diet and is driven by a diversity of factors (e.g., Lazzarini, Visschers, & Siegrist, 2017). Another trend closely linked to the growing interest in plant-based eating is the so-called clean-label trend, where consumers demand or prefer ingredients that sound familiar and natural, without any negative associations or allergenic potential (Asioli et al., 2017; Roman, Sanchez-Siles, & Siegrist, 2017). An example is animal-derived gelatin, which is thus exchanged for plant-based proteins, and the exchange highlighted with a 'free from' claim on the front of the package. In response to this development, food producers are offering more and more plant-based products in the market and seek out new plant proteins that can be used as a substitution ingredient for animal-based protein sources (Banovic et al., 2018). One of these relatively new protein sources is potato protein, which offers a nutritional quality similar to egg or soy protein (Ju, Mu, & Sun, 2017; Waglay & Karboune, 2016). Importantly, potato proteins do not bear the risk of causing allergic reactions as soy and other beans (Vanga & Raghavan, 2017), and for many consumers, it is a well-known and locally produced crop (Wood, Carragher, & Davis, 2017). Thus, in respect to both the plant-based movement and the clean label trend, potato protein is a potentially attractive substitution protein for consumers.

Previous research has shown that consumers can have very different

* Corresponding author.
*E-mail addresses:* peschel@mgmt.au.dk (A.O. Peschel), shahab.kazemi@mgmt.au.dk (S. Kazemi), scms@mgmt.au.dk (S.C.M. Sarraf), jeaw@mgmt.au.dk (J. Aschemann-Witzel).

https://doi.org/10.1016/j.foodqual.2019.02.015
Received 18 December 2018; Received in revised form 25 February 2019; Accepted 25 February 2019
Available online 26 February 2019
0950-3293/ 

reasons for reducing their meat consumption, which can be related to own health or concerned with topics that affect others such as sustainability and animal welfare (Verain, Sijtsema, & Antonides, 2016). Even though consumers perceive a certain relation between both (Aschemann-Witzel, 2015), health-related motives are in essence self-centered reasons, while sustainability-related motives are primarily centered on others (Thøgersen, 2011). Therefore, the product evaluation process might arrive at different results, depending on how the new product is presented to the consumer. In spite of the relatively high degree of innovation in offering plant-based alternatives using new protein sources, little is known about how to communicate these products effectively to consumers. It is well known that many new product innovations 'flop' in the market in the first year (Stewart-Knox & Mitchell, 2003). In addition, consumers dedicate only limited attention to each product in the supermarket aisle or when perceiving advertising content (e.g., Gidlöf, Anikin, Lingonblad, & Wallin, 2017; Peschel, Orquin, & Mueller Loose, 2019). Therefore, the communication must be targeted and to the point, focusing on the most important message or benefit depending on the product in question and the consumer group in focus. Either it might be worthwhile to reinforce the motives of choosing plant-based food products due to health, or it might be important to underline that the product is more sustainable. Another approach might be to focus on the substitution ingredient that transformed the product to qualify as "free from" a negatively associated animal-based ingredient. This more functional perspective improves the ingredient list by increasing transparency and highlighting the familiar and "harmless" new ingredient (Aschemann-Witzel, Varela, & Peschel, 2019).

To understand the effectiveness of the above-mentioned communication strategies, we draw on associative techniques rooted in spreading activation theory. Spreading activation theory explains how information is organized and retrieved from long-term memory (Anderson, 1983; Collins & Loftus, 1975). Underlying is the assumption that all beliefs, for example about ingredients, products, health and sustainability issues, are stored in associative networks. Depending on the type of communication chosen, different sets of associations might be evoked by the consumer, which can then influence the product evaluation process (e.g., Cowley & Mitchell, 2003) with implications for product choice later on. Studying how precisely consumers perceive communication and process new information is, however, intricate and time consuming. This holds in particular when wishing to apply methods that attempt not merely to look at the result, such as perceived quality dimensions, but at the cognitive processes involved, via exploring the associations that preceded the quality assessment. One method of approximating and visualizing consumers' associative networks is concept mapping, as it allows exploring consumers' associative processes (Novak & Cañas, 2008). Few studies so far have applied concept mapping in a food marketing context with slight variations in the analysis of the data (Askelson et al., 2015; Donaldson, Reimers, Brophy, & Nicholson, 2019; Grebitus & Bruhn, 2008, 2011; Hasimu, Marchesini, & Canavari, 2017; Seitz & Roosen, 2015a, 2015b). However, none of these studies applied text mining to the analysis of the concept maps. This type of textual data on consumer associations is usually analyzed by human coders, which apart from being time consuming in part depends on the analyst's subjective evaluation. With the technological advancement of text mining techniques, however, it is possible with a computerized analysis of consumer associations to achieve similar results to those by human coder approaches (e.g., Christensen et al., 2018).

Therefore, the aim of our study is twofold. We want to answer the research questions below with focus on RQ 1:

RQ 1: How do consumers' associative networks differ when communicating a plant-based diet as more sustainable, healthier or with a focus on the substitution ingredient?

RQ 2: What are the differences between text mining and a human coder approach when analyzing associative networks?

We explore the above-mentioned research questions using plant-based food products containing potato protein as a substitute for animal-based ingredients. The study uses a sample of Danish consumers looking to reduce their meat consumption via a concept mapping approach, followed by a short, multi-method design focus group. Concept maps are analyzed by both traditional human coding and a text mining procedure, and findings triangulated by analyzing respondents' statements in the focus groups. Our research thus contributes to understanding how consumers form associative networks during product perception and how these associative networks can best be studied. Findings allow deriving recommendations that can support food producers' effective and targeted communication choice based on a better understanding of consumers' associative networks.

### *1.1. Spreading activation theory and associative networks*

In order to understand the impact of different communication approaches, it is beneficial to understand which associations or schemata these communication approaches trigger. Schemata refer to humans' cognitive structures of how memory is organized (e.g., Axelrod, 1973). Both semantic memory of subjective beliefs as well as episodic memory involving experiential knowledge (Tulving, 1972) are of relevance in a consumer shopping situation as they influence the product evaluation process (e.g., Cowley & Mitchell, 2003; Erasmus, Bishoff, & Rousseau, 2002). According to spreading activation theory, which information is available to influence the product evaluation process depends on whether it was activated (Anderson, 1983; Collins & Loftus, 1975). According to the theory, schemata can be understood as networks consisting of associative concepts, which are linked based on their role in memory. When one concept is retrieved from memory, activation spreads to all linked concepts. On surpassing a certain threshold of activation, those activated concepts are retrieved from memory for information processing (Anderson, 1983; Collins & Loftus, 1975). The more complex a network (with many interlinkages between concepts), the faster the information flow (Cowley & Mitchell, 2003). Being able to understand which type of communication activates which type of information in consumers' memory can therefore be seen as a competitive advantage for marketing communication efforts.

### *1.2. Exploring the applicability of text mining in analyzing associative networks*

The technological developments in analyzing big data sets through data-mining techniques has paved the way for analyzing text data through text mining also in an academic setting (e.g., Evans & Aceves, 2016; Weiss, Indurkhya, Zhang, & Damerau, 2010). Text mining can be applied as unsupervised learning technique to transform large amounts of more or less structured text data into smaller units of text that can be interpreted directly or used for further statistical analysis. Unsupervised techniques based on machine learning algorithms allow independent classification of large amounts of text into smaller units and are applied when no a priori hypothesis about data classification exist. Common procedures include tokenization (i.e. identification of words that belong together, e.g. "social responsibility"), lemmatization (i.e. reducing words to their roots: "health", "healthiness", "healthier" are reduced to "health") and the removal of superfluous stop words (e.g. "and", "in", etc.). The advantage of text mining techniques lies in the increased amount of text data that can be processed more efficiently and often cheaper than by employing human coders (for a more detailed methodological account, see e.g., Christensen, Nørskov, Frederiksen, & Scholderer, 2017; Evans & Aceves, 2016; Weiss et al., 2010). In the area of food research, more and more studies are published using more or less computer-supported analysis of structured and unstructured text, such as posts in online communities (e.g., Christensen, Liland et al., 2017), Twitter posts (e.g., Chae, 2015; Vidal, Ares, Machín, & Jaeger, 2015) or open-ended survey questions (e.g., Spinelli et al., 2017). Their approaches showcase that text data can be analyzed in large amounts, faster and to a certain extent more objectively than by involving human coders in the procedure. This is confirmed by a validation study in the brewing industry aimed at identifying ideas for innovations from an online user platform. Their results showed that text mining can achieve similar patterns to human coder analysis (Christensen et al., 2018). To the best of our knowledge, no previous research has applied text mining to the analysis of consumers' associative networks, and consequently we aim to explore its applicability in this domain. We therefore advance the development of an efficient research methodology by providing a better understanding of the differences between machine-based text mining and traditional human coder analysis.

## 2. Materials and methods

### 2.1. Recruitment and sample

The main interest of this study is to understand the associations and network structures of consumers looking to reduce their meat consumption. The recruitment procedure accounted for the higher share of educated and younger consumers in this segment. We screened for consumers between 18 and 50 years of age, who were familiar with the Danish market (had lived in Denmark for at least one year) and "agreed" or "somewhat agreed" to the statements: "I have considered or am considering eating less meat" or "I have bought at some point/sometimes buy vegetarian products". About 50% of the sample was recruited through the university's laboratory participant pool, while the other share was recruited via social media posts or leaflets at local sports clubs, schools and day-care institutions to enroll also non-students in the study. Based on the screening criteria, a balanced sample (in terms of age, gender and children in the household) of 90 participants was invited to the university's research laboratory for a study about "consumer perceptions of plant-based food products"; the participants were divided into three study groups. Due to missing data, three concept maps had to be excluded from the data analysis. Group 1 had 27 participants, group 2 29 and Group 3 31 participants. The sample consisted of 63% females, with a mean age on 28.2, 53% students, and 47% of Danish nationality. Task instructions were given in either English or Danish, depending on the participants' preference. It should be noted here that participants in this study also participated in a projective mapping task. The results of this task are described in Aschemann-Witzel et al. (2019).

### 2.2. Communication conditions

Participants performed the same study tasks, but received different accompanying communication about the benefit of plant-based food products, depending on to which group they belonged. Group 1 received information from a sustainability perspective, Group 2 received information from a health perspective and Group 3 received information with a focus on the substitution ingredient. From now on, we will refer to these groups as sustainability, health and substitution groups respectively. An overview of the information pieces can be found in Table 1.

### 2.3. Concept mapping approach

Concept mapping is a technique that is applied to elicit cognitive structures or associative networks. Compared to other associative techniques, such as free elicitation techniques or laddering (e.g., Gutman, 1982; Henderson, Iacobucci, & Calder, 1998; Reynolds & Gutman, 1988; Roininen, Arvola, & Lähteenmäki, 2006), concept mapping is a graphical task that enables participants to visualize the associative structures themselves. This reduces researcher bias in terms of interpreting potential links of associations. Information obtained through concept mapping can be used to represent and analyze

**Table 1**
Communicational conditions from a sustainability, health and a substitution perspective.

| |
|---|
| Sustainability – Group 1: |
| "Plant-based food products, which means products that do not contain any ingredients from animals, are more and more demanded in the market place, because they contribute to a more sustainable lifestyle. By eating less animal based products, we can contribute to reducing greenhouse gas emissions and thereby reduce our own negative impact on climate change." |
| Health – Group 2: |
| "Plant-based food products, which means products that do not contain any ingredients from animals, are more and more demanded in the market place, because they contribute to a healthier lifestyle. By eating less animal based products, we consume less saturated fatty acids, which has been found to prevent cardiovascular diseases and some forms of cancer." |
| Substitution – Group 3: |
| "Plant-based food products, which means products that do not contain any ingredients from animals, are more and more demanded in the market place. Potato proteins have been found to be a useful substitute for animal based ingredients in a range of products." |

information flow network structures as it is conceptually based on spreading activation theory (Anderson, 1983; Collins & Loftus, 1975). The method has its origin in education research (e.g., Novak & Cañas, 2008; Novak & Gowin, 1984) but has later been adopted in the field of marketing (Joiner, 1998). The application in the area of food marketing has recently become more popular (Askelson et al., 2015; Donaldson et al., 2019; Grebitus & Bruhn, 2008, 2011; Hasimu et al., 2017; Seitz & Roosen, 2015a, 2015b). Pioneering work by Grebitus and Bruhn (2008) elicited consumers' quality perceptions of pork. Seitz and Roosen (2015b) apply concept mapping to identify cultural differences in perception of traditional regional products. Similarly, Hasimu et al. (2017) seek to understand Chinese consumers' perceptions of organic products. Askelson et al. (2015) identify social marketing opportunities, while Donaldson et al. (2019) investigate drivers and barriers for sports sponsorship. All of these studies vary in their application of concept mapping. Grebitus and Bruhn (2008), for example, chose to supply participants with a set of possible associations and encouraged them to come up with their own associations. The method ensures a certain degree of conformity between concept maps, but at the same time might bias associations in a certain direction. Seitz and Roosen (2015a, b) or Hasimu et al. (2017) refrain from this option to ensure unbiased elicitation of associations. For this study, we chose to leave associations open to ensure that the most salient associations are elicited, even though results might be very diverse (e.g., Russo & Johnson, 1980). About half of the above-mentioned studies used network analysis to gain insight into the interrelationships between associations (Grebitus & Bruhn, 2008, Hasimu et al., 2017, Seitz & Roosen, 2015b). The remaining use methods such as clustering (Askelson et al. 2015, Donaldson et al., 2019) or count data regression (Grebitus & Bruhn, 2011, Seitz & Roosen, 2015a). We follow the network analysis approach as the one most closely accounting for the cognitive structures of associative networks.

In our study, the central concept was the communication piece as earlier presented in Table 1. Participants were instructed to write down everything that came to mind, link those associations and in the end value them as either positive or negative. An example of the instructions given verbally to the substitution group can be found in Fig. 1.

### 2.4. Focus groups

After the concept mapping task, focus groups were conducted to engage participants in sharing their thoughts on the use and attractiveness of specific plant-based product concepts. While respondents assessed the overall plant-based product idea in different communication framings in the concept mapping task, they received product descriptions in the focus groups. The product concepts were chosen to

"The first task for you today is to draw a concept map with everything that comes to your mind related to the central concept **"Plant based food products in which ingredients that previously were animal-based have been substituted with potato proteins"**. You can think of a concept map like a mind map with your associations with the central concept **"Plant based food products in which ingredients that previously were animal-based have been substituted with potato proteins"**

Researchers believe that our knowledge is stored in memory. The knowledge we have can be described through central concepts or associations and the relationship between them.

**"**In the center of the paper you see the "central concept" we are interested in. Please start thinking of anything that comes to your mind and write it down. Because people have a lot of different associations, we would like you to try and come up with anything whether it is positive or negative. To do so just think which associations you can link to the "central concept". You may also think of things that are only indirectly related. Just write anything down that comes to mind.

If you think something is closely related to the central concept (or to each other), try to position them very closely to each other. You may use the whole sheet of paper for all your associations.

**Important: Please number the words you write down accordingly!**

You can write down anything that comes to mind. It is all important! Once you can't think of anything anymore, please start to link the single words to each other.

**Connect all those words that are linked to each other in your opinion.**

If new words come to mind, please write them down as well. Also, if you have second thoughts about words or links just cross them out.

Please let the interviewer know that you are done by hand sign and wait for further instructions.

– *wait until done* –

"Great, thank you, finally, please indicate whether the associations are positive with a plus "+" or negative with a minus "-"."

**Fig. 1.** Verbal instructions to perform the concept mapping task in the substitution group Note: Instructions were developed based on Grebitus (2008) and Grebitus and Bruhn (2008).

**Table 2**
Questions to probe the discussion in the focus groups.

| | |
|---|---|
| 1 | What would you use these plant-based food products for? |
| 2 | What would you NOT use these plant-based food products for? |
| 3 | For whom would you use these plant-based food products? |
| 4 | Which product would you exchange these plant-based food products for? |
| 5 | Can you describe a person who you think would most likely buy these plant-based food products? |

span across utilitarian (sausage and protein drink) and hedonic (ice cream and candy) categories, where potato protein was used as a substitute to animal-based ingredients in all products. The focus groups were led by a trained interviewer who probed the discussion with the questions shown in Table 2. The main goal of the focus groups was to identify whether the communication framing for the broad concept of a plant-based diet evokes similar streams of discussion on the product-specific level and mirrored the findings of the associative network approach.

## 2.5. Analysis

Concept mapping allows analyses from a qualitative and quantitative perspective. To get a deeper understanding of the associations evoked by the various communication conditions, we apply both qualitative and quantitative techniques as well as network analysis. The different steps are described below. Coding and text mining techniques were applied to reduce the amount of data and understand its content in order to generate meaningful input for the network analysis, which serves to analyze the structure of the concept maps.

### 2.5.1. Counting

The first step in the analysis of the concept maps is to count the total number of associations, the number of positive, negative and neutral associations and the number of links in the network. The amount of associations is traditionally considered as a measure of complexity of a network, while the number of links reflect the degree of integration of knowledge structures (e.g., Joiner, 1998; Seitz & Roosen, 2015a). Networks with a high degree of integration are denser in terms of linkages between nodes and are characterized by a higher likelihood of fast information flow between nodes. While counting results in a quick overview of the networks on these parameters, it does not provide any qualitative information of the content of the networks. Before analyzing the networks as such, we employ two different techniques, coding and text mining, to filter the most relevant associations in the networks. The approaches are described below.

### 2.5.2. Coding

Qualitative data is characterized by its richness and fullness. We apply coding by one human rater as a systematic and still flexible and logic approach to analyzing qualitative data (e.g., Saunders, Lewis, & Thornhill, 2015). It serves as a transitional process between data collection and detailed analysis. Coding enables comprehension of a large amount of data, systematically organizes codes and forms categories of related coded data. The analysis is an interactive and repetitive process constructed to find patterns or themes occurring within the concept mapping data set.

The coding process involves multiple rounds of coding for the desired outcome. Firstly, it is essential to become familiar with the collected raw data. Therefore, in the first round of coding, each association is labelled with a code that symbolizes or summarizes the meaning of that association. The code as well as the association could be a single word or a short phrase. In the next, second round of coding, codes are grouped based on their relationship to reduce the number of codes into a few general categories or themes, which represent the related group of codes and are related to the research question. In the third round of coding, additional adjustments are made to elicit representative codes, which correspond to the themes. Moreover, this round of coding includes evaluation of codes in terms of meaningfulness, support of the theme and merging based on synonymy. A few examples from the

**Table 3**
Example codes from the coding procedure.

| 1st round (raw codes) | 2nd round (code) | 3rd round (theme) |
|---|---|---|
| Rainforest; Earth; Forest; Trees; Environment; Ecology | Nature | Environment |
| Less $CO_2$; Greenhouse gas; Climate change; GMO | Pollution | Environment |
| Health(y); Fit; Longer life; Better for the body; Energy | Healthy for life | Health |
| Easier to digest; Low fat; Feeling lighter; Less calories | Light food | Health |

Note: Examples of the coding procedure. Codes, such as "Rainforest" or "Earth" were extracted in the first round, organized into the code "Nature" in the second round and then grouped into the theme "Environment" in the third round.

procedure can be found in Table 3. Codes, which were mentioned by less than 10% of each experimental group, were eliminated from further analysis.

#### 2.5.3. Text mining

Text mining was applied as a novel method to analyze consumers' associative networks. With the aim to reduce the subjectivity of the coding analysis and shorten the time necessary for analyzing text data manually, we developed an approach that could work independently based on the implementation of some fuzzy logic. The associations can be considered quite messy in terms of sentence structure and syntax as many participants use neither single words nor full sentences as has been observed before, for example, with survey data (ten Kleij & Musters, 2003). In order to combine associations meaningfully, it is therefore necessary to define some rules which fit the data at hand. The steps we take are described below:

*Step 1 – cleaning associations:* First, each association (which could be a word or a short phrase) was stripped off any stop words using the Python toolkit for natural language processing (*NLTK*) (Bird, Klein, & Loper, 2009) and split into individual words (tokens). Those are standard procedures in computerized analysis of text data (e.g., Symoneaux, Galmarini, & Mehinagic, 2012; ten Kleij & Musters, 2003; Weiss et al., 2010).

*Step 2 – defining word groups:* The individual words have to be grouped based on their similarity to reduce the volume of data. Similar words often have the same meaning as their root is the same. Stemming or lemmatization are procedures that can identify roots and group words automatically based on pre-defined libraries (e.g., Weiss et al., 2010). These libraries can be standard packages or defined by the researcher (e.g, Spinelli et al., 2017). Since our data contains very specific vocabulary and we want the algorithm to run independently, we apply some fuzzy logic to define similarity. Similarity $S_w(w_j, w_k)$ between a word pair $w_j$, $w_k$ is determined based on an index of the amount of identical substrings between two words divided by the total number of letters of those two words:

$$S_w(w_j, w_k) = \frac{2L(w_j, w_k)}{|w_j| + |w_k|}$$

where $|w_j|$ is the number of letters in word $w_j$ and $L(w_j, w_k)$ is a function that returns the sum of lengths of the longest substrings of the two words $w_j$ and $w_k$. If this index $S_w$ was larger than 80%, those two words were considered as belonging to the same word group. As an example, consider the similarities between the words "nut", "nutrition" and "fruit":

$$S_w(nut, nutrition) = \frac{2 * 3}{3 + 9} = 0.5$$

$$S_w(nut, fruit) = \frac{2 * (1 + 1)}{3 + 5} = 0.5$$

$$S_w(fruit, nutrition) = \frac{2 * (1 + 1 + 2)}{5 + 9} \cong 0.57$$

In the first example, one word is a substring of the other word, so there is only one longest substring of length 3. Note that "ut" and "nu" are not counted in $L$("nut", "nutrition"), since they are substrings of a longer substring, and $L(w_j, w_k)$ only counts the lengths of the longest substrings. In the second example, there are two longest substrings of length 1, which are the letters "u" and "t". In the third example, there is one longest substring of length 2 ("it") and two longest substrings of length 1 ("r" and "u"). Based on the threshold criterion defined above, these three words would not have been matched into a word group.

*Step 3 – defining importance scores:* Before merging association based on matched word groups, we must define a rule to determine which associations are most similar, given the case that they contain the same amount of matched word groups. To this end, we assigned importance scores $I$ based on the frequency of occurrence of word groups. Those word groups, which were mentioned most often, were considered most important in our context, i.e. received the highest importance scores.

*Step 4 – combining associations:* Finally, associations were combined based on their similarity $S_a$ in terms of shared word groups and their importance scores. Similarity $S_a$ $(a_j, a_k)$ between a pair of associations $a_j, a_k$ was defined as half of the sum of importance scores $I(w)$ for the word groups found in both associations $a_j$ and $a_k$, divided individually by the sum of the importance scores for all word groups in each association $a_j$ and $a_k$:

$$S_a(a_j, a_k) = \frac{1}{2}\left(\frac{1}{\sum_{w \in a_j} I(w)} + \frac{1}{\sum_{w \in a_k} I(w)}\right) \sum_{w \in a_j \cap a_k} I(w)$$

here $I(w)$ is the importance score of the word group $w$ and the sums run over word groups that appear in both associations, i.e., in the intersection of the two sets ($w \in a_j \cap a_k$) or in each association individually ($w \in a_j$ or $w \in a_k$). If similarity $S_a$ was higher than 90%, associations were combined. This step was repeated until no more similarities between associations above 90% could be found. Associations which were mentioned by less than 10% of each experimental group, were eliminated from further analysis. *Step 4* was performed separately by valence rating of the association (positive, negative, neutral).

*Step 5 – naming of merged associations:* the final step involved researcher input to name the merged associations with a single term based on the word groups they contained (e.g. "nutritious, nutrition, nutritional" was named "nutritious").

#### 2.5.4. Network analysis

In social network analysis centrality measurements indicate the importance of nodes based on their location relative to other nodes in the network (Freeman, 1978; Wasserman & Faust, 1994). We follow Henderson et al. (1998) in applying these measures to semantic networks to identify nodes which are particularly influential in spreading information throughout the semantic network. Centrality in the context of product associations can be thought of as relevance to be targeted by marketing activities. The three most common centrality measurements, degree, closeness and betweenness centrality, will be further explained below.

*2.5.4.1. Degree centrality*. Degree centrality ($C_D$) is a measure for the number of direct links of one node to other nodes in the network (Freeman, 1978; Henderson et al., 1998). It is an indicator of activity in the network as activation of a node with a high degree centrality can activate many other associations quickly, following spreading

**Table 4**
Results of the counting approach.

| | Sustainability group (n = 27) Mean (SD) | Health group (n = 29) Mean (SD) | Substitution group (n = 31) Mean (SD) |
|---|---|---|---|
| Total associations | 15.0[a,b] (8.2) | 15.4[a] (5.5) | 11.03[b] (5.34) |
| Positive associations | 9.3 (8.2) | 9.5 (5.8) | 6.1 (4.5) |
| Negative association | 3.4 (3.2) | 3.9 (3.8) | 2.8 (2.5) |
| Neutral associations | 2.3 (3.7) | 2.1 (3.2) | 2.1 (4.2) |
| Number of links | 19.6[a,b] (10.2) | 19.2[a] (7.2) | 15.4[b] (8.3) |

Note: [a,b]Significant differences ($p < .05$) according to Tukey HSD post-hoc test are indicated with superscript letters.

activation theory (Collins & Loftus, 1975; Freeman, 1978).

Degree centrality for a node $p_k$ is defined as:

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k),$$

where $n$ equals the number of nodes in the network and $a(p_i, p_k) = 1$, if $p_i$ and $p_k$ are connected by a link, 0 otherwise.

*2.5.4.2. Closeness centrality*. Closeness centrality ($C_C$) in contrast to degree centrality measures all direct and indirect links of a node in a network (Henderson et al., 1998; Wasserman & Faust, 1994). It is an indicator of how close nodes are in the network in terms of the number of links connecting them. In this way closeness centrality indicates independence of control in a network (Henderson et al., 1998). With a high closeness centrality, a node is not dependent on activation by one specific node. When all nodes in a network are close, information flows quickly (Grebitus & Bruhn, 2008).

Closeness centrality for node $p_k$ is defined as:

$$C_C(p_k) = \left[\sum_{i=1}^{g} d(p_i, p_k)\right]^{-1}$$

where $d(p_i, p_k)$ describes the number of lines on the shortest path linking $p_i$ and $p_k$ (the geodesic).

*2.5.4.3. Betweenness centrality*. Betweenness centrality ($C_B$) measures the probability that a node $p_k$ falls on the geodesic between two other nodes $i$ and $j$ (Freeman, 1978). It is thought of as a measure of control in a network (Henderson et al., 1998; Wasserman & Faust, 1994). A node with a high betweenness centrality falls on many geodesics and is therefore in control of spreading activation from one node to another. Betweenness centrality of a node $p_k$ is defined as:

$$C_B(p_k) = \sum_{i}^{n}\sum_{j}^{n} b_{ij}(p_k)$$

for all $(i < j) \neq k$, and where

$$b_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}}$$

where $g_{ij}$ is the number of geodesics from node $i$ to node $j$ and $g_{ij}(p_k)$ is the number of geodesics from node $i$ to node $j$ that include $p_k$. The probability that $p_k$ is included in a randomly selected geodesic from $i$ to $j$ is represented by $b_{ij}(p_k)$.

Analysis of centrality is conducted using the Python package *NetworkX* (Hagberg, Schult, & Swart, 2008). The results are displayed as normalized values to increase comparability. Network visualizations were created using the Python library *matplotlib* (Hunter, 2007).

*2.5.5. Analysis of focus groups*

Focus group recordings were transcribed and analyzed manually by a trained researcher. The level of analysis was not on the discussion of the specific products, but rather whether the communicational framing of a plant-based diet influences the discussion on the product level overall. Statements related to the overall themes were marked and taken out to support the main analyses of consumers' associations of plant-based food products.

## 3. Results

In this section, we first present the results from the counting approach to get an overview of the network solutions. Then we describe the outcome of the assessment of consumers' associations and move on to the results from the network analysis as it depends on the results from the assessment of consumers' evaluations.

### *3.1. Counting*

Without any reduction, 1186 associations were counted. The results show that the networks in the sustainability and the health groups are more complex and more integrated than those in the substitution group as can be seen in Table 4. Analysis of variance shows that the mean number of associations and the number of links differ significantly between groups (Total associations: $F_{(2,84)} = 4.20$, $p < .05$ and Number of links: $F_{(2,84)} = 3.83$, $p < .05$). The networks in all groups are characterized by more positive than negative associations and least neutral associations. Those differences were however not significant.

### *3.2. Consumers' associations*

The human coder process resulted in six themes, consisting of 21 codes across the three experimental groups. An overview can be found in Table 5. The frequency of the codes are displayed in Table 6 by experimental group and valence. The three codes with the highest number of associations are in bold print ("healthy for life", "animal origin substitutes" and "taste & quality"). The main findings from the frequency table can be interpreted the following way: Respondents' associations are predominantly positive, some are negative and least associations are neutral. The negative associations are mostly connected

**Table 5**
Overview of codes and merged associations from human and text mining analysis.

| Theme | Code from human coder | Merged association from text mining procedure |
|---|---|---|
| Environment | nature | sustainability |
| | pollution | |
| | environmentally friendly | environment |
| | resources | resources |
| | production & locality | organic |
| | ethics | animal welfare |
| Health | healthy for life | healthy |
| | diet & exercise | unhealthy |
| | light food | light food (fats) |
| | | unprocessed |
| | | processed |
| Food | food from nature | tofu |
| | animal origin substitutes | vegan products |
| | | soy |
| | | beans |
| | | fruit & vegetables |
| | | vegetables |
| Figures | protein | plant protein |
| | nutrition properties | nutritious |
| | chemicals | vitamins |
| Characteristics | taste & quality | taste |
| | selection variety | fake |
| | price | expensive |
| | difficulty | food preparation |
| People | vegan & vegetarian | veganism |
| | trend & culture | |
| | social responsibility | |

**Table 6**
Overview of the results from the coding procedure.

| Themes | codes | count | Positive | | | Negative | | | Neutral | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Sustain. | Health | Subst. | Sustain. | Health | Subst. | Sustain. | Health | Subst. |
| Environment | nature | 20 | 6 | **10** | 4 | | | | | | |
| | pollution | 22 | 6 | **12** | | | | | | 4 | |
| | environmentally friendly | 66 | 18 | 21 | **27** | | | | | | |
| | resources | 6 | **6** | | | | | | | | |
| | production & locality | 39 | **20** | 10 | 9 | | | | | | |
| | ethics | 52 | 11 | 16 | **21** | | | | | | 4 |
| Health | **healthy for life** | 99 | **34** | 22 | 25 | | 12 | 6 | | | |
| | light food | 25 | **12** | 8 | 5 | | | | | | |
| | diet & exercise | 14 | 5 | **9** | | | | | | | |
| Food | food from nature | 64 | 15 | **40** | | | | | 5 | 4 | |
| | **animal origin substitutes** | 74 | 9 | **16** | 10 | 8 | 5 | | 7 | 15 | 4 |
| Figures | protein | 9 | | 4 | | | | | | | **5** |
| | nutrition properties | 49 | 8 | **14** | | 4 | 11 | 6 | | | 6 |
| | chemicals | 15 | | 4 | | | **6** | | | 5 | |
| Characteristics | **taste & quality** | 82 | **16** | 13 | 14 | 5 | 5 | 12 | 7 | | 10 |
| | selection variety | 18 | **6** | 4 | | 4 | 4 | | | | |
| | price | 22 | | | | **8** | 6 | **8** | | | |
| | difficulty | 18 | 4 | | | **9** | 5 | | | | |
| People | vegan & vegetarian | 27 | 6 | **7** | 5 | | | | 4 | | 5 |
| | trend & culture | 37 | **9** | 7 | 6 | 5 | | 4 | 6 | | |
| | social responsibility | 17 | 4 | 4 | **9** | | | | | | |
| | *total sum of associations:* | *775* | *195* | *221* | *135* | *43* | *54* | *36* | *29* | *28* | *34* |

Note: Sustain. = Sustainability group, Subst. = Substitution group; the three codes with the highest number of associations are in bold. Numbers in bold indicate highest frequency values within each code (that is, horizontally).

to the themes “Characteristics” and “Figures”, thus related to product characteristics and to associations about the extent to which certain ingredients or nutrients are present in the product. The highest frequency of associations is found in the health and sustainability groups. The “Environment” theme has the highest number of codes and at the same time the highest number of associations among all themes. Consumers in the sustainability group seem mostly concerned with “healthy for life”, “production & locality” as well as “taste & quality” (positive valence). The health group comprises most associations concerning both of the codes in the “Food” theme (“food from nature” and “animal origin substitutes”). In addition, the health group stands out in that the codes “nutrition properties” and “chemicals” (negative valence) appear more often in this group than in the others. The substitution group contains most associations concerning the codes “environmentally friendly” and “ethics”. Overall, this group has more associations for the code “taste & quality” than the other two groups, spread out evenly across positive, negative and neutral associations.

The text mining procedure arrived at a total of 24 merged associations across the three experimental groups, which can also be found in Table 5. While these merged associations are not exactly the same as those from the human coder approach, they can be broadly classified into the six themes identified by the human coder procedure. Overall, it appears that it was possible to capture consumers’ associations with the text mining procedure in a similar fashion as with the human coder process, and we thus move on to the network analysis with these results.

### 3.3. Relevance of the associations in the networks

We start out by building associative networks for the aggregated solutions for each communication group. The networks are in line with the results from the counting approach in that they show networks of higher complexity for the sustainability (Fig. 2) and the health group (Fig. 3) compared to the substitution group (Fig. 4).

From the networks as well as the centrality measures (Tables 7 through 9), it can be seen that all associations are directly linked to the central concept, which was the communication about a plant-based diet, suggesting that betweenness centrality plays only a minor role for these networks. In all networks, the concept “healthy” (positive) scores very high in terms of degree centrality, suggesting that irrespective of the communication, this concept will be activated and has strong potential to activate other concepts directly. Most of the concepts with a high degree centrality are positive. Only in the substitution group, a negatively evaluated concept “expensive”, received a similarly high degree centrality score and has therefore a high ability to activate other concepts directly. The same pattern could be found for closeness centrality. It should be noted though that all closeness centrality scores are in a similar range, which suggests a high independence of control from other nodes in the network and an overall fast information flow. In line with that, betweenness centrality scores are rather low in all groups, with the exception of the central concepts as well as “healthy” in the sustainability group and “organic” in the substitution group. Organic here refers to associations about organic certification schemes. These concepts have a higher probability of controlling information flow in the networks. When communicating a plant-based product as being part of a sustainable diet, it seems that the association of “healthy” controls more information flow, while when communication the substitution ingredient, “organic” controls information flow. To secure fast information flow, both concepts should probably be activated to ensure thorough product communication.

The sustainability group scores highest in terms of closeness centrality suggesting a high degree of independence of associations evoked by sustainability communication. The substitution group scores highest in terms of degree centrality, suggesting potential faster information flow evoked by ingredient focused communication. The concepts in the health group show higher values in terms of betweenness centrality compared to the other networks, which means that some associations should be activated specifically in order to ensure activation of downstream associations. Contrary to the health and sustainability groups, there were more influential negative associations (e.g. “unhealthy” or “expensive”) in the aggregated network of the substitution group.

Overall, there are some noteworthy differences between the communication groups. It seems that the sustainability group associated more issues regarding the environment and the authenticity (e.g., “resources”, “fake”) of the given products. Communication from a health perspective seems to evoke more associations regarding the product

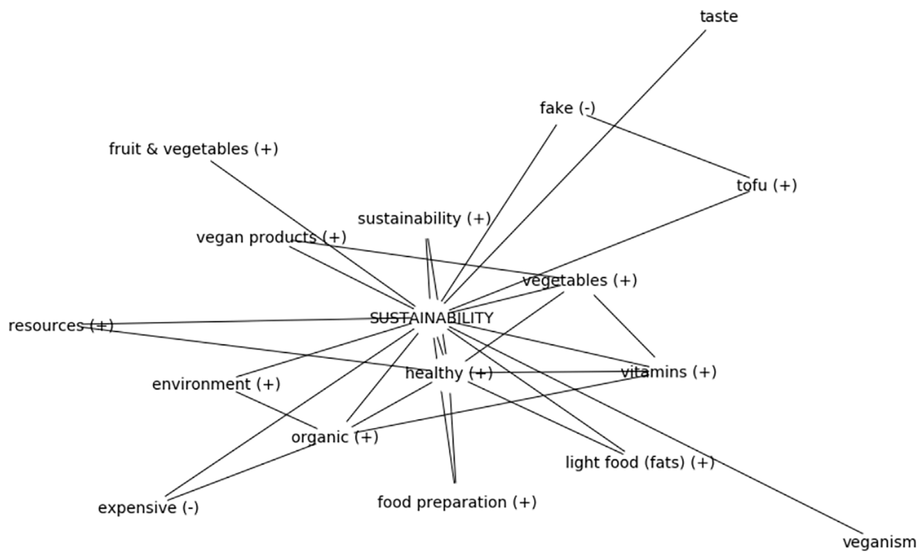


**Fig. 2.** Associative network for the sustainability group. Note: central concept is in capital letters, (+) denotes positive valence, (−) denotes negative valence, no sign indicates neutral valence.

itself and its intrinsic quality characteristics (e.g., "processed", "nutritious"). When the communication focuses on the substitution ingredient, this seems to foster associations concerned with taste (e.g., "taste" as positive, negative and neutral association). All groups associate expensiveness with a plant-based diet, but while it is an influential concept in the substitution group, it is more peripheral in the health group.

The data from the focus groups supports these findings by showing similar patterns. Selected quotes highlight how consumers expressed what we could observe as notable differences in the networks. For example, participants in the sustainability group made statements concerning the environment and authenticity, such as:

a) "The first thing I think about is the environment, and then the bonus is just that I would feel more healthy and so for me, the first thing, healthy, environment, reducing carbon footprint".
b) "I would avoid it. I am vegan myself so I don't use those substitutes for meat because you don't need them if there is a plant so just use 100% the plant, but not. How it's called not chemically but industrially made substitute for it".
c) "Well a person that were caring about environment or climate change… At least that's the way it's being marketed".
d) "I think that the marketing the product people should also be able to associate this with lessening impact on the environment so this could be done in a visual way".

Some selected quotes from the health group underlining how participants were concentrating more on the processing of the products, include:

a) "I think I would still prefer to use something like eggs or chickpeas or something that is less processed or made in a lab, a bit more natural. I think that would be more me".
b) "I am against all kind of plant based products. I think it's kind of everything unnatural even if they pretend to be natural".
c) "…But exactly if you can replace it with something that has preservatives I would rather have something natural but non-plant based that has something with preservatives in it but is plant-based

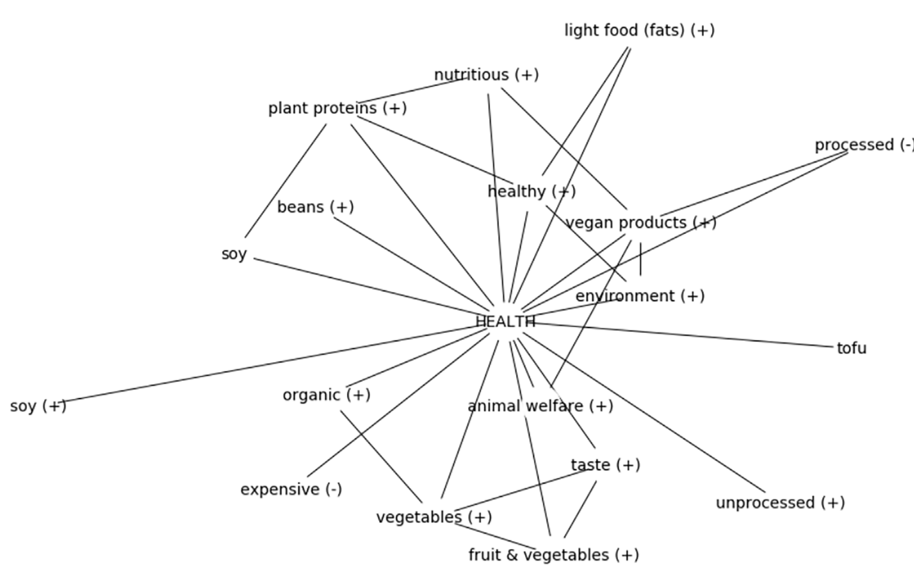


**Fig. 3.** Associative network for the health group. Note: central concept is in capital letters, (+) denotes positive valence, (−) denotes negative valence, no sign indicates neutral valence.

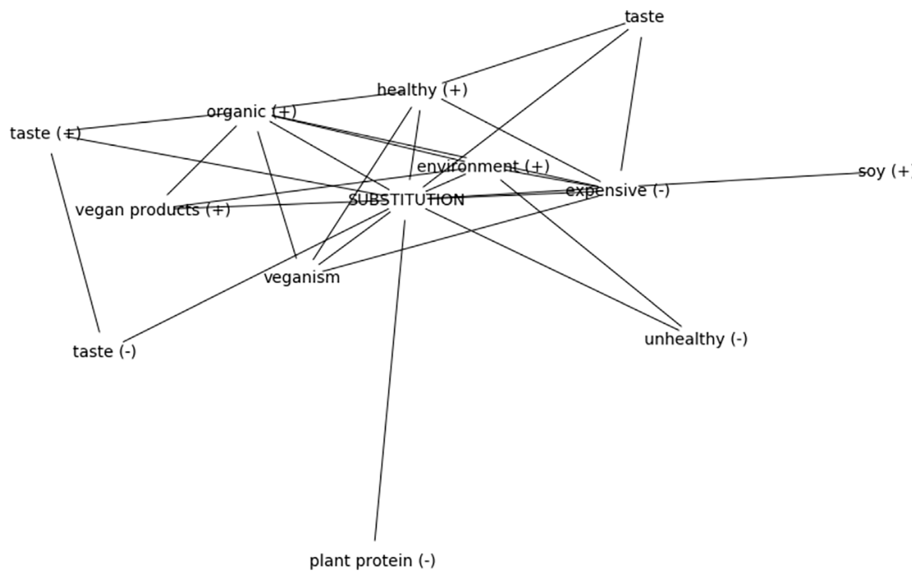


**Fig. 4.** Associative network for the substitution group. Note: central concept is in capital letters, (+) denotes positive valence, (−) denotes negative valence, no sign indicates neutral valence.

**Table 7**
Centrality measurements for the sustainability group.

| Descriptor | Centrality measures | | |
|---|---|---|---|
| | Degree | Closeness | Betweenness |
| Sustainability | 1.00 | 1.00 | 0.78 |
| healthy (+) | 0.50 | 0.67 | 0.08 |
| organic (+) | 0.31 | 0.59 | 0.02 |
| vegetables (+) | 0.25 | 0.57 | 0.01 |
| vitamins (+) | 0.25 | 0.57 | 0.00 |
| resources (+) | 0.25 | 0.53 | 0.00 |
| vegan products (+) | 0.25 | 0.53 | 0.00 |
| fake (−) | 0.13 | 0.53 | 0.00 |
| expensive (−) | 0.13 | 0.53 | 0.00 |
| sustainability (+) | 0.13 | 0.53 | 0.00 |
| environment (+) | 0.13 | 0.53 | 0.00 |
| tofu (+) | 0.13 | 0.53 | 0.00 |
| food preparation (+) | 0.13 | 0.53 | 0.00 |
| light food (fats) (+) | 0.13 | 0.53 | 0.00 |
| fruit & vegetables (+) | 0.06 | 0.52 | 0.00 |
| veganism | 0.06 | 0.52 | 0.00 |
| taste | 0.06 | 0.52 | 0.00 |

Note: Centrality scores are normalized to allow comparison across networks.

so its healthier I think to use something that has no preservatives and stuff".

d) "I think it would be perfect for someone who is willing to try new stuff and doesn't have a lot of habits. For me that can be all types of plant-based products. And also those who is into a more healthy lifestyle and really pays attention to exactly what he/she eats".

In the group focusing on the substitution ingredient, statements can be found which are more related to taste, such as:

a) "It really matters how all the ingredients match up with each other, how they are like – how to say – equal make the taste of it. That's what I think matters".
b) "I think that all substitutes maybe will have like a bit of a different taste and probably precisely as the meat substitutes or chicken substitutes, it won't be like perfectly substitutes. I think that will still have like a bit of a different taste".
c) "I think if you break the first barrier of people trying it, they going to realize that it tastes better than they expected and that a general

**Table 8**
Centrality measurements for health group.

| Descriptor | Centrality measurements | | |
|---|---|---|---|
| | Degree | Closeness | Betweenness |
| Health | 1.00 | 1.00 | 0.86 |
| vegan products (+) | 0.39 | 0.58 | 0.02 |
| healthy (+) | 0.33 | 0.56 | 0.01 |
| environment (+) | 0.28 | 0.55 | 0.00 |
| vegetables (+) | 0.22 | 0.56 | 0.01 |
| plant proteins (+) | 0.22 | 0.56 | 0.01 |
| fruit & vegetables (+) | 0.17 | 0.55 | 0.00 |
| soy (+) | 0.17 | 0.51 | 0.00 |
| taste (+) | 0.17 | 0.55 | 0.00 |
| nutritious (+) | 0.17 | 0.55 | 0.00 |
| animal welfare (+) | 0.11 | 0.53 | 0.00 |
| organic (+) | 0.11 | 0.53 | 0.00 |
| light food (fats) (+) | 0.11 | 0.53 | 0.00 |
| processed (−) | 0.11 | 0.53 | 0.00 |
| soy | 0.11 | 0.53 | 0.00 |
| expensive (−) | 0.06 | 0.51 | 0.00 |
| beans (+) | 0.06 | 0.51 | 0.00 |
| unprocessed (+) | 0.06 | 0.51 | 0.00 |
| tofu | 0.06 | 0.51 | 0.00 |

Note: Centrality scores are normalized to allow comparison across networks.

**Table 9**
Centrality measurements for the substitution group.

| Descriptor | Centrality measurements | | |
|---|---|---|---|
| | Degree | Closeness | Betweenness |
| Substitution | 1.00 | 1.00 | 0.63 |
| healthy (+) | 0.58 | 0.63 | 0.01 |
| organic (+) | 0.58 | 0.71 | 0.07 |
| expensive (−) | 0.50 | 0.67 | 0.03 |
| environment (+) | 0.42 | 0.63 | 0.03 |
| vegan products (+) | 0.42 | 0.57 | 0.00 |
| veganism | 0.33 | 0.60 | 0.00 |
| taste | 0.25 | 0.57 | 0.00 |
| taste (+) | 0.25 | 0.57 | 0.01 |
| unhealthy (−) | 0.17 | 0.55 | 0.00 |
| taste (−) | 0.17 | 0.55 | 0.00 |
| soy (+) | 0.08 | 0.52 | 0.00 |
| plant protein (−) | 0.08 | 0.52 | 0.00 |

Note: Centrality scores are normalized to allow comparison across networks.

thing".

## 4. Discussion and implications

Our research provided insight into how different communication framings of a plant-based diet can affect consumers' associations, and how text mining can add to analyzing the data instead of human coding alone. Consumers' associations are important for product evaluation and choice, because only associations, which are activated, can influence product evaluations. It is therefore valuable for producers to get an in-depth understanding of how their product communication affects consumers' associative networks before they decide on and invest in product packaging and market communication. To this end, we applied concept mapping to visualize consumers' cognitive structures for communicating a plant-based diet either as more sustainable, healthier or substituting animal-based protein with potato protein. While the first two types of communication refer to and reinforce the two most important consumer motives to choose more plant-based products (Verain et al., 2016), the latter taps into the current trend of offering products with a "clean label" that is transparent about the precise ingredient used to substitute animal-based ingredients (Asioli et al., 2017). In the following, we first discuss our findings regarding the different plant-based food product communication strategies, then we discuss methodological considerations in the light of our second research question and round off with an account of limitations.

### *4.1. Communication – plant-based food products*

A range of interesting findings emerge with regard to our research question on the differences in associative networks, depending on the communication presenting the plant-based food concepts (research question 1). However, before discussing the differences, it is worthwhile underlining the similarities as the baseline. Four important findings emerge. First, in terms of both the content and complexity of the associative networks, we find that health aspects are associated with a plant-based diet, irrespective of the communication implemented. The code "healthy" scores high in terms of all centrality measurements. Second, the majority of all associations are positive, thus showing that the selected target group is indeed mainly positive about the product concepts presented. Third, overall, codes concerned the issues of health, the environment, the plant-based characteristics and taste. These findings are well in line with research showing that there is a trend towards meat reduction and plant-based foods and interest in these groups' characteristics (Banovic et al., 2018; Verain et al., 2016). It is also in line with research that has repeatedly found that the underlying driver for choosing plant-based products is health, whether in a direct functional health or a holistic understanding of health (Aschemann-Witzel, 2015). As an implication, the findings suggest that either way, the aspect of health should be considered when communicating the benefits of a plant-based diet it being a central motive. The fourth observation is that in the case of any negative associations to be mentioned, they primarily relate to the product's characteristics and its content, ranging from taste, to price, to any ingredient perceived as 'chemical'. This finding mirrors the clean label trend (Asioli et al., 2017) and underlines that a major challenge of plant-based foods is found in the high degree of processing, and the negative perception of the ingredients needed for this (Bearth, Cousin, & Siegrist, 2014; Burdock & Wang, 2017; Dickson-Spillmann, Siegrist, & Keller, 2011). Further, it confirms that the product has to be right in terms of 'value for money', with price and taste typically being decisive for food product choice (Frewer & Van Trijp, 2006).

With regard to the communication frames, differences concerning the size and complexity of the networks emerged. When focusing on the substitution ingredients, the network was smaller and included less associations and connections compared to when the food was communicated focusing on health and sustainability aspects. This implies that producers choosing to highlight the substitution ingredient should expect that consumers are yet lacking associations to it. This has the disadvantage of potentially slower processing and less established beliefs and attitudes, which might be based on consumers' associations. However, it has the advantage and opportunity that producers might suggest new associations through the communication of their product. In addition and even more importantly, relatively more associations regarding the taste of products were found, positive, negative as well neutral. For producers choosing such an approach of underlining the new substitution ingredient, it might be worthwhile to also highlight the good taste of the product or make use of tasting samples at the point-of-sale to convince consumers of the product. Given that the association "expensive" (negative valence) was relatively high in degree and closeness centrality in the substitution group network, producers might have to be cautious when communicating solely based on the substitution ingredient and they should be able to justify a higher price meaningfully in that it is outweighed by the benefits of the product. The finding that the participants in the substitution group often referred to environmental friendliness and ethics, might suggest that these benefits could further be strengthened in product communication.

When communicating the healthiness of a plant-based diet, relatively more associations regarding the nutritional content and degree of processing of the potential product came up, which is line with previous results (Aschemann-Witzel et al., 2019). Products with a truly positive nutritional value and a low degree of processing might benefit from communicating the health benefits of a plant-based diet, consequently directing consumers to consider and value the health effect. If the product does not contain what consumers believe to be healthy, communicating health might however act as a boomerang (Bearth et al., 2014; Dickson-Spillmann et al., 2011). Therefore, products for which producers assume the health framing to be risky might benefit from communicating the sustainability benefits of a plant-based diet. This communication seems to evoke broader concepts related to the environment and the authenticity of products, or the production and locality of the ingredient source.

For the case of potato protein used as the substitution ingredient, the results imply that marketing outcomes depend on the protein's nutritional contribution and the product in question. Mentioning potato protein as the substitution ingredient likely triggers associations on the taste and price characteristics. The product's composition and marketing would need to provide a satisfactory response to address these concerns. Framing the communication towards health instead might be favorable if the ingredient is used in a product with a lower degree of processing, while framing communication towards sustainability could be used in processed, but environmentally favorable products. For potato protein, communicating both the ingredient and sustainability might be a valuable option, considering that the considerations of authenticity and local production triggered by the sustainability framing might go well with potatoes as a transparent ingredient of local sourcing.

### *4.2. Methodological considerations*

In the study, concept mapping was used as a tool to elicit consumers' associations and their interlinkages. Concept mapping has the advantage of triggering spontaneous and cognitively less controlled associations. Our sample size was about average having 90 consumers. Previous research in the area of food marketing reports sample sizes between 29 (Donaldson et al. 2019) and 314 (across four countries, Seitz & Roosen, 2015b) participants. As mentioned before, we chose a free elicitation approach, which is expected to result in a higher number of concepts as sample size increases than an approach that uses predefined concepts. We found 1186 concepts, which could be reduced to about 20 depending on the extraction method. On average, we found 15 associations per concept map. This number is higher than in, for example, Seitz and Roosen (2015b), who found between four and ten associations on average and opted to analyze the top ten associations per country. Seitz and Roosen (2015a) found a similar total number of associations, however with a sample twice the size of ours. Grebitus and Bruhn (2008, 2011) found similar values to ours, with about half of the associations being freely elicited and the other based on pre-determined concepts. Hasimu et al. (2017) only report that they arrive at 37 concepts to be used for the network analysis after including only those concepts mentioned by at least 50% of the sample. The other studies did not report on the raw counts of the concept maps or used a different extraction method. These findings show, however, that there is great variation, depending on the central concept to be evaluated. Nevertheless, our results seem to concur with previous findings, which strengthens the validity of our approach. For those studies reporting relative centrality measures, we can compare the results to see if they are in the same range. We find that closeness centrality is relatively high across studies and networks (Grebitus & Bruhn, 2008, Seitz & Roosen, 2015b). This might be because only fractions of the total number of concepts tend to be used in the analysis, so all concepts are relatively close in the aggregated network. For betweenness centrality, we do not find as dominant concepts as in Grebitus and Bruhn (2008) but rather compare our networks to Seitz and Roosen's (2015b) solution. It seems that the central concept is more dominant in our case and triggers associations directly.

With regard to our research goal of exploring differences between text mining and a human coder approach (research question 2), we find that the pattern of the elicited associations is comparable across the human coder and the text mining approach, even though some differences arise. This is well in line with Christensen et al. (2018) thus reinforcing the value of text mining methods. In our specific case, we could observe some minor differences between the solution of the human coder and the text mining procedure, while the overall pattern was comparable. While this was sufficient for the purpose of our study, other studies might have more specific requirements. In these cases, it might be necessary to train the algorithm with a greater degree of human input, as it was done, for example, in Spinelli et al. (2017). Thus, human control of the algorithm or triangulation of data analysis appears to be important to consider, depending on the research goal in question.

### 4.3. Limitations

We would like to acknowledge the limitations of our presented research. As with all qualitative research, generalization is not the main purpose but rather a deeper understanding of a certain phenomenon. For generalization, future research might validate our findings on a larger scale. Currently, our findings are limited to a Danish context and to consumers who are already interested in consuming a more plant-based diet. Other cultural contexts might lead to different results (Ares, 2018) as well as studies focusing on other segments of the population. In addition, we used one specific ingredient – potato protein – for substitution, and we did so in order to be able to communicate the ingredient transparently as it is done in 'clean label' products (Asioli et al., 2017). Other plant-based protein sources might evoke other reactions, such as soy, for example, due to its potential allergens and the fact that it is not locally grown in the country (Lazzarini, Zimmermann, Visschers, & Siegrist, 2016). However, there are a range of other new alternative ingredients and proteins in particular, which are applied in new food products, such as peas, lupine, or even protein extracted from grass that leave room for further investigation. We encourage future research to address these shortcomings.

Furthermore, we tailored our text mining procedure to the given study characteristics at hand. The results thus relate to the given process. Using a different approach might lead to a different outcome, which could deviate from the human coder procedure to a greater or lesser extent than in the results presented. To account for that, we were as transparent as possible regarding our methodological decisions. We encourage future research to replicate our procedure in a different context and to be similarly transparent about the choices made with regard to the text mining procedure.

## 5. Conclusions

Findings of this study on consumers' associative networks in reaction to different communicational framings of plant-based food products allow to conclude that there are differences found through both human coder and text mining-based analysis of consumers' concept map data. A major conclusion is that health and sustainability framing leads to richer and more complex associations than communicating the substitution ingredient, but that while health leads to product-centered associations on nutritional quality and chemicals, sustainability leads to broader associations on environment and authenticity. Further, both human coder and text mining can lead to similar results and thus have great value in analyzing consumers' associative networks.

## Acknowledgments

The research conducted in this study was supported by funding from Innovation Fund Denmark, in the scope of the project 'ProPotato – Potato proteins – Challenges and Industrial Possibilities', large scale project, grant no. 5158-00001B.

Thanks to Viktorija Kulikovskaja, Anna Kock, and Christian Kræmer Aagaard for their help in the tedious work of data collection and processing, and to Dan Mønster for insightful comments on the manuscript.

## References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261–295.
Ares, G. (2018). Methodological issues in cross-cultural sensory and consumer research. *Food Quality and Preference, 64*, 253–263.
Aschemann-Witzel, J. (2015). Consumer perception and trends about health and sustainability: trade-offs and synergies of two pivotal issues. *Current Opinion in Food Science, 3*, 6–10.
Aschemann-Witzel, J., Varela, P., & Peschel, A. O. (2019). Consumers' categorization of food ingredients: Do consumers perceive them as 'clean label' producers expect? An exploration with projective mapping. *Food Quality and Preference, 71*, 117–128.
Asioli, D., Aschemann-Witzel, J., Caputo, V., Vecchio, R., Annunziata, A., Næs, T., et al. (2017). Making sense of the "clean label" trends: A review of consumer food choice behavior and discussion of industry implications. *Food Research International, 99*, 58–71.
Askelson, N. M., Golembiewski, E. H., DePriest, A. M., O'Neill, P., Delger, P. J., & Scheidel, C. A. (2015). The answer isn't always a poster: Using Social marketing principles and concept mapping with high school students to improve participation in school breakfast. *Social Marketing Quarterly, 21*(3), 119–134.
Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *American Political Science Review, 67*(4), 1248–1266.
Banovic, M., Arvola, A., Pennanen, K., Duta, D. E., Brückner-Gühmann, M., Lähteenmäki, L., et al. (2018). Foods with increased protein content: A qualitative study on European consumer preferences and perceptions. *Appetite, 125*, 233–243.
Bearth, A., Cousin, M.-E., & Siegrist, M. (2014). The consumer's perception of artificial food additives: Influences on acceptance, risk and benefit perceptions. *Food Quality and Preference, 38*, 14–23.
Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit.* O'Reilly Media, Inc.
Burdock, G. A., & Wang, W. (2017). Our unrequited love for natural ingredients. *Food and Chemical Toxicology, 107*, 37–46.
Chae, B. K. (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.
Christensen, K., Liland, K. H., Kvaal, K., Risvik, E., Biancolillo, A., Scholderer, J., et al. (2017). Mining online community data: The nature of ideas in online communities. *Food Quality and Preference, 62*, 246–256.
Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In search of new product ideas: Identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management, 26*(1), 17–30.
Christensen, K., Scholderer, J., Hersleth, S. A., Næs, T., Kvaal, K., Mollestad, T., et al. (2018). How good are ideas identified by an automatic idea detection system? *Creativity and Innovation Management, 27*(1), 23–31.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.
Cowley, E., & Mitchell, A. A. (2003). The moderating effect of product knowledge on the learning and organization of product information. *Journal of Consumer Research,*

*30*(3), 443–454.
Dickson-Spillmann, M., Siegrist, M., & Keller, C. (2011). Attitudes toward chemicals are associated with preference for natural food. *Food Quality and Preference, 22*(1), 149–156.
Donaldson, A., Reimers, J. L., Brophy, K. T., & Nicholson, M. (2019). Barriers to rejecting junk food sponsorship in sport—A formative evaluation using concept mapping. *Public Health, 166*, 1–9.
Erasmus, A. C., Bishoff, E., & Rousseau, G. (2002). The potential of using script theory in consumer behaviour research. *Journal of Consumer Sciences, 30*(1).
Evans, J. A., & Aceves, P. (2016). Machine translation: mining text for social theory. *Annual Review of Sociology,* 42.
Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.
Frewer, L., & Van Trijp, H. (2006). *Understanding consumers of food products.* Woodhead Publishing.
Gidlöf, K., Anikin, A., Lingonblad, M., & Wallin, A. (2017). Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite, 116*, 29–38.
Grebitus, C. (2008). *Food quality from the consumer's perspective: An empirical analysis of perceived pork quality.* Doctoral ThesisGoettingen, Germany: Department of Agricultural Economics, Kiel University, Cuvillier Verlag.
Grebitus, C., & Bruhn, M. (2008). Analyzing semantic networks of pork quality by means of concept mapping. *Food Quality and Preference, 19*(1), 86–96.
Grebitus, C., & Bruhn, M. (2011). A way to more effective marketing strategies: Analyzing dimensionality of cognitive structures quantitatively. *SAGE Open, 1*(2) 2158244011418380.
Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *The Journal of Marketing,* 60–72.
Hagberg, A., Schult, D., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In: Los Alamos National Lab. (LANL), Los Alamos, NM (United States).
Hasimu, H., Marchesini, S., & Canavari, M. (2017). A concept mapping study on organic food consumers in Shanghai, China. *Appetite, 108*, 191–202.
Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research, 111*(2), 306–327.
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering, 9*(3), 90–95.
Joiner, C. (1998). Concept mapping in marketing: A research tool for uncovering consumers' knowledge structure associations. *ACR North American Advances*.
Ju, D., Mu, T.-H., & Sun, H.-N. (2017). Sweet potato and potato residual flours as potential nutritional and healthy food material. *Journal of Integrative Agriculture, 16*(11), 2632–2645.
Lazzarini, G. A., Visschers, V. H., & Siegrist, M. (2017). Our own country is best: Factors influencing consumers' sustainability perceptions of plant-based foods. *Food Quality and Preference, 60*, 165–177.
Lazzarini, G. A., Zimmermann, J., Visschers, V. H., & Siegrist, M. (2016). Does environmental friendliness equal healthiness? Swiss consumers' perception of protein products. *Appetite, 105*, 663–673.
Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them.
Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* Cambridge University Press.
Peschel, A. O., Orquin, J. L., & Mueller Loose, S. (2019). Increasing consumers' attention capture and food choice through bottom-up effects. *Appetite, 132*, 1–7.
Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research, 28*(1), 11–31.
Roininen, K., Arvola, A., & Lähteenmäki, L. (2006). Exploring consumers' perceptions of local food with two different qualitative techniques: Laddering and word association. *Food Quality and Preference, 17*(1–2), 20–30.
Roman, S., Sanchez-Siles, L. M., & Siegrist, M. (2017). The importance of food naturalness for consumers: Results of a systematic review. *Trends in Food Science & Technology, 67*, 44–57.
Russo, J. E., & Johnson, E. J. (1980). What do consumers know about familiar products? *ACR North American Advances*.
Saunders, M., Lewis, P., & Thornhill, A. (2015). *Research methods for business students.* Pearson Education579–587.
Seitz, C. C., & Roosen, J. (2015a). Does consumer ethnocentrism influence product knowledge? *Food Quality and Preference, 43*, 113–121.
Seitz, C. C., & Roosen, J. (2015b). An empirical analysis of international consumers' associations with Bavarian food products by means of concept mapping. *British Food Journal, 117*(3), 987–1006.
Spinelli, S., Dinnella, C., Masi, C., Zoboli, G. P., Prescott, J., & Monteleone, E. (2017). Investigating preferred coffee consumption contexts using open-ended questions. *Food Quality and Preference, 61*, 63–73.
Stewart-Knox, B., & Mitchell, P. (2003). What separates the winners from the losers in new food product development? *Trends in Food Science & Technology, 14*(1–2), 58–64.
Symoneaux, R., Galmarini, M., & Mehinagic, E. (2012). Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality and Preference, 24*(1), 59–66.
ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: a complementary method to preference mapping. *Food Quality and Preference, 14*(1), 43–52.
Thøgersen, J. (2011). Green shopping: for selfish reasons or the common good? *American Behavioral Scientist, 55*(8), 1052–1076.
Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory, 1*, 381–403.
Vanga, S. K., & Raghavan, V. (2017). How well do plant based alternatives fare nutritionally compared to cow's milk? *Journal of Food Science and Technology,* 1–11.
Verain, M. C. D., Sijtsema, S. J., & Antonides, G. (2016). Consumer segmentation based on food-category attribute importance: The relation with healthiness and sustainability perceptions. *Food Quality and Preference, 48*, 99–106.
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.
Waglay, A., & Karboune, S. (2016). *Potato Proteins: Functional Food IngredientsAdvances in potato chemistry and technology* (second ed.). Elsevier.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press.
Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information.* Springer Science & Business Media.
Wood, K., Carragher, J., & Davis, R. (2017). Australian consumers' insights into potatoes-Nutritional knowledge, perceptions and beliefs. *Appetite, 114*, 169–174.