New Scientometric-Based Knowledge Map of Food Science Research (2003 to 2014)

Jesús Blázquez-Ruiz, Vicente P. Guerrero-Bote, and Félix Moya-Anegón

Abstract: This article describes an analysis of keywords which was aimed at revealing publication patterns in the field of Food Science (FS) during the last decade, including the temporal evolution of its different research lines. To this end, the records of the specific subject area of FS were 1st retrieved from Scopus, and then their keywords were processed to resolve the obvious problems of synonymy and to limit the study to those most frequently used. These keywords were grouped into thematic clusters based on a scientometric technique known as co-word analysis. The structure of the clusters, their scientific impact, and their temporal evolution were then analyzed. This type of analysis is of great interest for all researchers in FS—for new researchers because they can form an objective vision of the subject based on the data from papers that have been published in the last decade, and for experienced researchers because they can contrast their own vision of the field with this objective overview. The results showed there to has been a clear increase in scientific production related to FS. This production had a structure corresponding to 5 major clusters which were themselves disaggregated into 18 2nd-level clusters. The cluster that had received most attention was that corresponding to antioxidants in food, being the cluster with the greatest scientific impact and the greatest growth in the period.

Keywords: co-word analysis, food science, food science structure, scientometrics, thematic analysis

Introduction

As shown in research by Guerrero-Bote and Moya-Anegón (2015), scientific production in Food Science (FS) has had a rapid growth in the last decade, even more rapid than the rest of world scientific production. This shows how important the field of FS is at present, and therefore that there is a need for research to study this phenomenon and the discipline of FS in general.

Given the importance that FS has already, an importance which will be even greater in the not too distant future, it is necessary to ensure that the organization and direction of future research in FS will not be completely dominated by the priorities of those countries carrying out most research today (Guerrero–Bote and others 2016b).

Once scientific papers have been published in a scientific journal they are listed together with their references in different bibliographic databases, thus facilitating their retrieval. In this way, these journals and databases become the protagonists of scientific communication (Miguel and others 2011). Another important aspect covered by these scientific databases is that of the keywords which are responsible for representing the content of the research in the most basic and concise way possible.

Scientometrics, 1st defined by Nalimov and Mulcjenko (1971) as "the quantitative methods of research on the development of

science as an informational process," is developed from these bibliographic databases.

However, we all know that not all published papers have the same value. To analyze the quality of research, scientometrics is based on the idea of scientific impact, in other words, the impact on the scientific community caused by a paper's publication. Its value is calculated using the citations that a paper receives, on the grounds that, despite there being different motivations for citing previous work (Bornmann and Daniel 2008), it does reflect recognition of that research (Moed 2005). Citation by 1 author of another provides links between people, ideas, publications, and institutions, and these links constitute a network that can be analyzed quantitatively.

This to a great degree had its origin in the work of Eugene Garfield. He identified its importance in creating the Science Citation Index as a bibliographic database that includes the references (Garfield 1955).

Price (1963, 1965), a historian, was one of the 1st to see the significance of the networks of scientific research and authors in beginning to analyze scientometric processes. This gave rise to the idea of "cumulative advantage" (Price 1976), also called "preferential connection," and is very similar to what Merton (1968, 1988) defined as the "Matthew effect." Price identified some key issues that scientometricians would have to solve: mapping the invisible colleges, the relationship between productivity and quality, and the different citation habits in different fields.

During this period, Science Politics began to use citation analysis. For example, the Institute for Scientific Information (ISI's) data were used in the 1972 Science Indicators Reports of the United

States and by the Organisation for Economic Co-operation and Development (OECD). And the Impact Factor was developed, which, despite all its weaknesses, is still used today (Garfield and Sher 1963).

Scientometrics currently plays an important, and often criticized, role in monitoring, collecting, and evaluating scientific activity based on bibliographic databases, as a response to the need for transparency and as an aid in decision making (Mingers and Leydesdorff 2015).

Paul Wouters (1999), in his doctoral thesis, describes how scientometricians, science politicians, science sociologists, or researchers in general see scientometrics because of this new role:

> "Scientometricians feel they are measuring science, either as 'scientists of science' or as sociologists. For science policy people, scientometrics is just one of many sources of policy instruments. Scholars in science studies tend to view scientometrics merely as a method without theory. Lastly, scientists tend to be divided into two groups: opponents and supporters. This is also true of researchers in the social sciences and the humanities. Adversaries raise all sorts of arguments against measuring science in general (e.g., the unmeasurable creative nature of scientific discovery) and citation analysis in particular (e.g., the lack of meaning of the citation). The proponents of citation analysis tend to see the scientometric scrutiny of the scientific process as a means of improving the quality of research, notwithstanding its limitations."

This role that scientometrics is playing means that, instead of just reflecting reality, it is actually transforming reality itself by changing the behavior of academics and researchers (Wouters 2014).

There has been little application of scientometrics to FS. Hinze and Grupp (1996) made thematic maps of biotechnology in FS by using the controlled terms of both patents and scientific publications (1985 to 1993). From their analysis, they concluded that the production of the less developed countries of the European Union (EU) had increased in this field.

Seetharam and Ravichandra (1999) compared the increase in production of FST (Food Science & Technology) in their institution (CFTRI) with their country overall (India) and the rest of the world. They used scientific publications, patents, PhD theses, and published standards for the period 1950 to 1990. Their finding was that, while there was growth, the rate of that growth was decreasing.

Alfaraz and Calviño (2004) studied the scientific production in FST of Ibero-American countries (including Spain and Portugal) between 1990 and 2000. They found that Spain accounted for more than half of the records and had a growth rate of 11% annually during that decade.

Zhou and others (2012) analyzed the changes undergone by the Chinese meat industry and the challenges and opportunities that lie ahead in the global market.

Muscio and Nardone (2012) addressed the relationship between industry and university regarding FS in Italy.

Guerrero-Bote and others (2016a) analyzed the FS research activities carried out in Spain and how these are reflected in international scientific journals. The study of Guerrero-Bote and Moya-Anegón (2015) found that Spain, China, and Italy were among the top 8 countries in FS scientific publication production.

Scientometrics also uses data from these large databases to analyze research thematically. One of the most used techniques in this regard is co-word analysis. This generates a network in which the different keywords are linked together and then weighted by the number of papers in which these co-words occur (Callon and others 1991). Procedures are applied to this network designed to detect groups of strongly related keywords. The result is a picture of the thematic structure of the research (Romo-Fernández and others 2013).

In the past, the intellectual structure of a discipline was known by senior researchers, usually when the discipline pertained to their own field of study. But this structure was neither formal nor registered on any support. It was a subjective structure that the researcher had formed mentally from his or her deep knowledge of the discipline. The result, therefore, suffered from conservatism, bias, and subjectivity (Irvine and others 1985; Bornmann 2011).

Therefore, the development of studies such as ours involves the disclosure of the structure of scientific fields in a more objective and easily assimilated way for both new and experienced researchers. The present study has as its main objective the establishment of the intellectual structure of FS on the basis of an analysis of the keywords present in the papers published in the field. The specific research questions are:

- How many subareas form the main structure of FS?
- How do they relate to each other?
- Which are the most centralized and which are the most specialized themes?
- What degree of internal cohesion do they show?
- What is the scientific impact of each theme and how has it evolved?
- Which are the keywords' bursty periods?

Materials and Methods

We used the Scopus database for this study. Created by Elsevier, this is one of the bibliographic databases, which includes the largest number of scientific journals. It has been used in numerous scientometric studies and has been the subject of many studies attempting to characterize and analyze it. One such case is that by Leydesdorff and others (2010) who compared it with the ISI database and concluded that they both do a good job in providing material for the mapping of science, and that the main differences between them are based on maturity and size. We decided to work with Scopus because it covers more international scientific journals in general, and, as stated in the study by López-Illescas and others (2009): "*Scopus coverage is more comprehensive, and citation impact of journals is less discriminative.*"

In the Scopus database, there are 2 types of keywords for each record: the "Author Keywords" are assigned by the author of the document; and the "Index Keywords" are added by professional indexers. Our study uses "Author Keywords." These are more numerous, and offer a more detailed description of the documents retrieved.

Contrary to what Romo-Fernández and others (2013) did in their study, namely, select a journal in the area and extract the keywords to analyze from that 1 journal, we downloaded all the keywords from all the citable documents published in all journals included in the specific Scopus subject area of FS during the period 2003 to 2014.

This gave a total of 184801 citable papers with 230007 different keywords. All of them were imported into a relational database created for the purpose. This allowed us to study them in a more

efficient manner, and also to subject them to various normalization processes.

As a novelty for the normalization of keywords, we used the Levenshtein distance and the Damerau-Levenshtein distance to identify similar keywords. The Levenshtein distance (Levenshtein 1966) is the minimum number of operations required to transform 1 string of characters into another. The definition of "operation" is either an insertion, deletion, or substitution of a character. The Damerau-Levenshtein distance (Damerau 1964) is an improvement on the latter with the difference that in addition to insertion, deletion, and substitution, it includes the transposition of 2 characters. In the Levenshtein distance, this last operation counts as 2 operations, 1 of deletion and another of insertion.

Specifically, the steps in the normalization process were the following:

(1) Eliminate punctuation lacking any meaning, such as quotation marks.
(2) Unify singular and plural.
(3) Group pairs of similar keywords, based on the Levenshtein and Damerau-Levenshtein distances.
    (a) Those with a small value of distance and a major part in common were automatically unified.
    (b) Those that did not meet the above condition were reviewed manually. We identified 7469 pairs with these characteristics.

Once normalized, the number of keywords decreased to 215409, and we proceeded to generate a network of co-words that allowed us to group them into clusters. Co-word networks consist of a set of nodes, which are the keywords, and a set of links that connect 2 keywords and which are weighted by the number of papers in which both keywords occur. Then, if a link between 2 keywords has a great weight, this means that those 2 keywords appear together in a large number of papers. As Neff and Corley (2009) state in their study, co-word analysis is based on the theory that research fields can be characterized and analyzed based on patterns of the keywords used in their publications.

To generate this network, we established a minimum threshold frequency of appearance in 300 papers, with a total of 297 keywords meeting this requirement.

There are many algorithms that can be used for mapping, each giving a different final layout. One of the most used is multidimensional scaling (Van Eck and others 2010), which tends to locate the items in an artificial circular structure that gives a final picture of the network's structure which is far from reality. There is also the Visualization of similarities (VOS) method (Van Eck and Waltman 2007) that locates the most recognized or best connected elements in the center of the map, leaving those least recognized on the periphery. This method does not impose any artificial structure, although, as Van Eck (2011) indicates, it can be a little disappointing at a local level. In contrast, a technique such as LinLog (Noack 2009) seems to give satisfactory results both globally and locally.

Similarly to cluster analysis, with networks one uses methods of detection of communities of nodes related more closely to each other than to the rest. The traditional method is to proceed by progressively eliminating the links with greatest "betweenness" (Newman and Girvan 2004). Waltman and others (2010) proposed a weighted and parametrized variant of modularity–based clustering that is implemented in the VOSviewer (Van Eck and Waltman 2010). This method allows the resolution of the commu-

nities to be varied by means of a parameter, yielding classifications of different granularity.

In the present study, we used the latest version of Van Eck and Waltman's VOSviewer software to find the layout. This version allowed us to select the attraction and repulsion parameters, and the method of normalization. The parameters we used were those of LinLog (Noack 2009).

On the maps generated, the labels' font size and the size of the circle vary depending on the number of documents associated with each keyword. The general clusters are differentiated by color, and the elementary clusters that make up each general cluster are differentiated by different shades of that color.

We used the bursting algorithm developed by Jon Kleinberg (2003) to detect when certain terms come into fashion and when they fall out of fashion. In our study, it was used to detect the fashionable keyword trends during the period 2003 to 2014 in the thematic area of FS. The algorithm generated a table showing the "bursty" periods of the most frequent keywords, indicating the length, strength, and time interval in which the bursting occurred.

We also used some scientific impact indicators:

• Ndocc: Number of citable documents published in scientific journals that are included in the Scopus database. Documents of this type are those which really make a scientific contribution. Specifically, in Scopus, we consider citable documents to be articles, reviews, conference papers, and short surveys. Henceforward, we shall be referring to Ndocc when we speak of scientific production, and to %Ndocc with respect to the total Ndocc in FS when we speak of percentage production.
• Cites per Document: Average cites per document. Citations depend greatly on how much time the document has had to be cited; for this reason, we did not evaluate the evolution of this indicator.
• % Cited Documents: Percentage of documents cited. As with the previous indicator, this depends greatly on how much time the documents have had to be cited; so we did not evaluate the evolution of this indicator.
• Normalized Impact (NI): Average normalized citations received by each document. This is understood as the ratio between the citations received by the document and the average citations of documents of that same type, year, and category (Rehn and Kronman 2008). If the result of this average is 1, the NI is at the mean; if it is 1.2, then it is 20% above the mean; and if it is 0.7, it is 30% below the mean.
• % Excellence10 (%Exc): Percentage of documents that are among the 10% most cited of a given year, type, and category (Bornmann and others 2012). Clearly, the overall mean of this indicator is 10%, so that if one obtains a %Exc of 13% for a cluster, then this is 30% higher than the mean, and similarly if one obtains values lower than the mean.
• % Excellence1 (%Exc1): Percentage of documents that are among the 1% most cited of a given year, type, and category. The overall mean of this indicator is clearly 1%, so that if one obtains an %Exc of 1.3% for a cluster then this is 30% higher than the mean, and similarly if one obtains values lower than the mean.

Results and Discussion

Once we had determined this intellectual structure[1] that results from the clusters, we submitted it to the validation of an expert in

[1]As noted above, in our retrospective search in Scopus in the specific subject area of Food Science, we extracted 230007 different "Author Keywords" (with

Table 1–Cognitive structure of research in Food Science

L1 cluster	Ndocc	Citation	NI	%RCNI	Percentile	%Exc	%Exc1	%Cited	%	%RC%
1. Food composition and nutrition	33253	11.05	1.01	2.89	48.53	9.61	0.92	83.58	17.99	−22.3
1.1. Nutrition and metabolism	10677	11	1.07	5.95	49.26	10.56	1.48	81.34	5.78	18.1
1.2. Nutrients and quality	9930	10.4	0.91	3.52	49.97	7.61	0.52	83.87	5.37	−46.7
1.3. Food composition	10178	11.6	1.02	−10.10	47.69	9.86	0.67	84.52	5.51	−25.3
1.4. Composition and quality	6225	12.6	1.13	14.67	45.32	11.48	1.10	86.36	3.37	−35.2
2. Food processing and modification	31170	12.42	1.19	0.55	44.22	12.38	1.17	86.22	16.87	−1.4
2.1. Influence of food processing on sensory characteristics	15070	12.8	1.26	11.55	42.65	13.70	1.41	87.80	8.15	5.0
2.2. Methods of processing or treatment of food	9184	12.5	1.10	−7.13	45.02	10.48	0.74	86.18	4.97	−28.6
2.3. Beneficial microorganisms as starter cultures in the food processing	4680	13.6	1.31	−7.72	42.55	14.64	1.69	87.01	2.53	8.6
2.4. Modifications food during processing and how to determine them	5220	11.3	1.11	−16.97	46.92	11.87	1.09	81.96	2.82	46.7
3. Food security	31578	12.45	1.18	5.53	44.94	12.41	1.14	85.79	17.09	−10.2
3.1. Pathogenic microorganisms in milk and dairy products	11148	13.2	1.24	4.12	42.18	13.23	1.02	87.73	6.03	−24.3
3.2. Mycotoxins in cereals and methods of detection of mycotoxins in food	11796	11.0	1.07	12.79	47.33	11.00	0.61	83.95	6.38	−3.5
3.3. Other important food contaminants in food safety (heavy metals, pesticides)	6435	11.9	1.15	8.90	45.76	11.90	1.39	85.21	3.48	12.2
3.4. Antimicrobial agents used in food. Methods of determining these antimicrobial agents	4349	16.0	1.43	5.98	42.50	16.68	2.81	87.79	2.35	−1.7
4. Food preservation and shelf life	29154	11.65	1.06	−7.01	46.84	10.17	0.78	84.99	15.78	−18.8
4.1. Increased shelf life in plant foods during preservation. Methods and modifications of quality	14055	11.6	1.01	−6.26	47.87	9.82	0.67	84.78	7.61	−25.5
4.2. Fermentation process in wine and beer as a preservation method	10617	12.0	1.12	−13.62	45.12	10.99	0.65	86.33	5.75	−23.2
4.3. Parameters that influence food preservation	3979	11.1	0.99	−17.77	48.22	8.48	0.81	83.55	2.15	−33.2
4.4. Statistical analysis methods	3080	12.2	1.24	22.34	44.09	13.42	1.68	84.85	1.67	53.4
5. Antioxidants in food	20901	15.99	1.65	−12.01	37.69	21.26	2.85	88.05	11.31	110.4
5.1. Antioxidants and their effects	18047	16.7	1.74	−23.00	36.48	22.88	3.22	88.33	9.77	161.0
5.2. Plant antioxidants	3912	13.9	1.30	9.12	41.98	15.35	1.37	87.71	2.12	−12.8

Hierarchical structure composed of 5 level-1 (L1) clusters containing 18 level-2 (L2) clusters. Scientometric indicators summary: Ndocc, cites per document, normalized impact, % Excellence10, %Excellence1, %Cited Documents, %Ndocc, and the rates of change of NI and of %Ndocc.

the field, who also helped us to label all the clusters manually in accordance with the keywords that each contained. This thematic structure is presented in Table 1, including the level-2 clusters within the 5 clusters of level 1.

a total of 918588 occurrences) from a total of 184801 citable documents published in the period 2003 to 2014. After normalization, there remained 215409 keywords, of which only 297 met the requirement of appearing 300 times or more. The number of papers of the original data set which had 1 or more of these keywords was 110994, in other words, just more than 60%. To generate a hierarchical structure of clusters or communities, we tested several values of the resolution parameter. Of all the values tested, we chose 1.9 because it gave 18 level-2 (L2) clusters with quite acceptable sizes. Subsequently, the minimum cluster size parameter was set at 35, grouping the 18 clusters into 5 general or level-1 (L1) clusters. We could have used a lower value of the resolution parameter to find the L1 clusters, but in that case the new clusters might not include all of the L2 clusters. Therefore, with our choice we achieved a hierarchical structure.

Each of the level-1 clusters contains a certain number of keywords distributed among the set of its corresponding level-2 clusters. The Appendix presents the form in which they are distributed. In this study, we use blue to represent the L1 cluster nr 1, green for L1 cluster nr 2, gray for L1 cluster nr 3, red for L1 cluster nr 4, and yellow for L1 cluster nr 5.

Figure 1 shows the map of co-words with the 5 level-1 clusters differentiated by color. It is clear that level-1 cluster nr 5 (yellow) is that with the greatest internal cohesion of its keywords. They are well related with each other, occupying a well-defined zone of the map, and not overlapping with other clusters. On the other hand, knowing that the more centralized clusters are those found more in the center of the network, we cannot say that any particular cluster stands out for its centrality. Indeed, observing the keywords located closest to the center, one notes that they belong to various clusters. However, one does observe that the general clusters 3 and
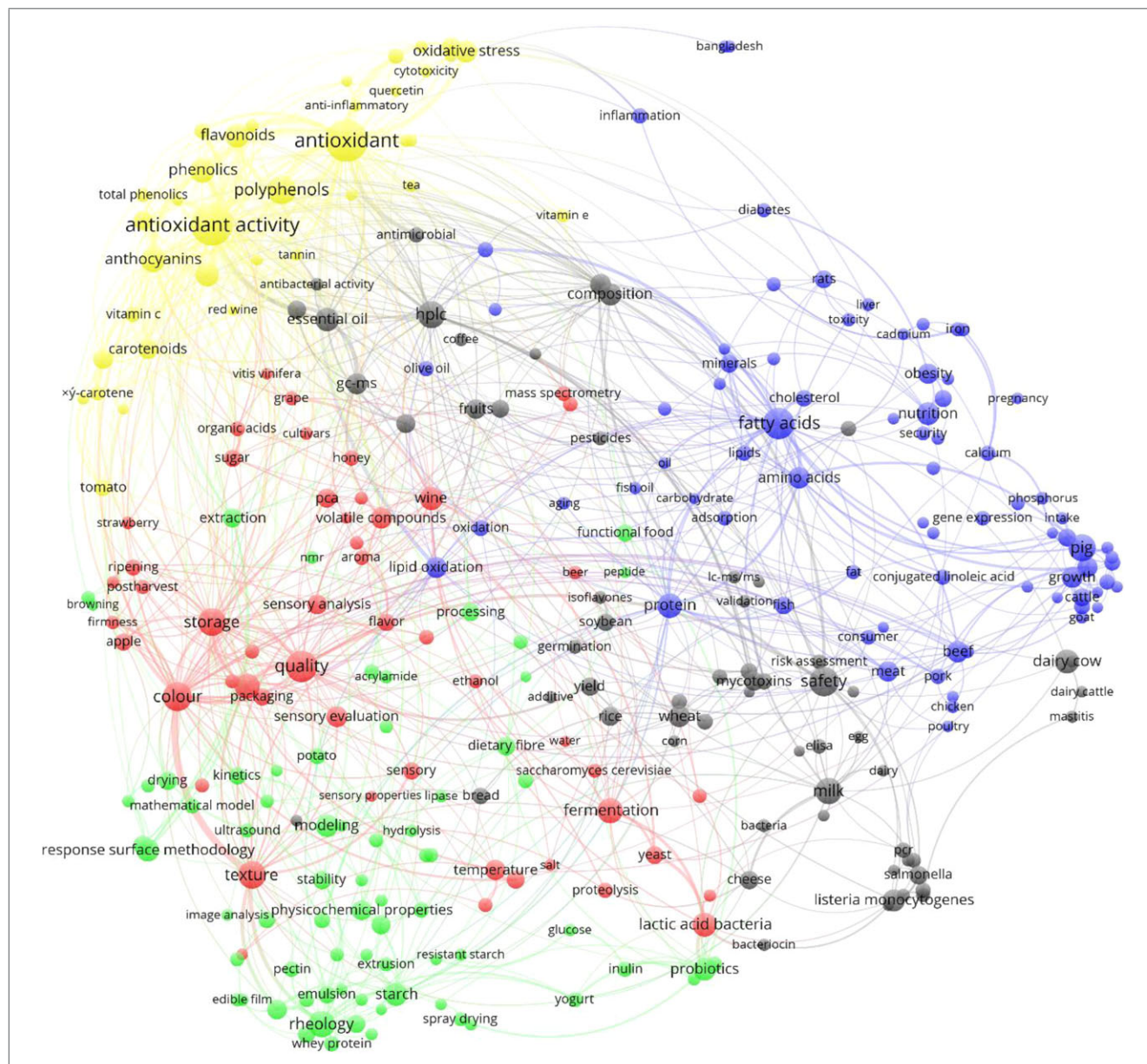
Figure 1–Overview of the co-word map, distinguishing 5 general clusters of level 1.

4 occupy less peripheral zones. For example, cluster nr 3 (gray) of Food Security starts from the bottom right area and crosses into the central area, touching or merging with other clusters. This is logical because Food Security cannot be an isolated field but has to be linked to all of FS.

Figure 2 corresponds to the density map. One clearly sees 11 zones highlighted in red. These correspond to a high intensity of keywords that link many documents and that are closely related to each other; in other words, they could be said to represent research fronts. This means that not all the L2 clusters will have a zone that stands out, and they thus may suffer from a lack of internal cohesion. Of these 11 zones, 4 of them especially stand out. The 1st of these outstanding zones is in L1 cluster nr 1, which mainly corresponds to livestock terms: pig, growth, cattle, and goat (subcluster 1.2). The 2nd is in the lower left part and corresponds to L1 cluster nr 2, namely the terms rheology, emulsion, whey protein, starch, pectin, and extrusion (subcluster 2.1). The other 2

zones of intense activity are in the L1 cluster nr 5. One is dedicated to antioxidant activity and the other to the keyword that links the most documents—antioxidant (subcluster 5.1). In line with what we noted in Figure 1 with respect to the center of the map, neither in this density map does one note anything remarkable.

Table 1 presents the number of documents contained in each cluster, the cites per document[2], the normalized citation and its percentage change, the average percentage at which the documents are located according to their citations (in each document type and publication year), the percentage of excellent documents in the 10% and 1% most cited, the total percentage of documents

---

[2]For the indicators that use citation (Cites per document, NI, Percentile, Exc10, Exc1, and Cited documents), the production of 2014 is not included as we considered that its citation data were not yet sufficiently stable.
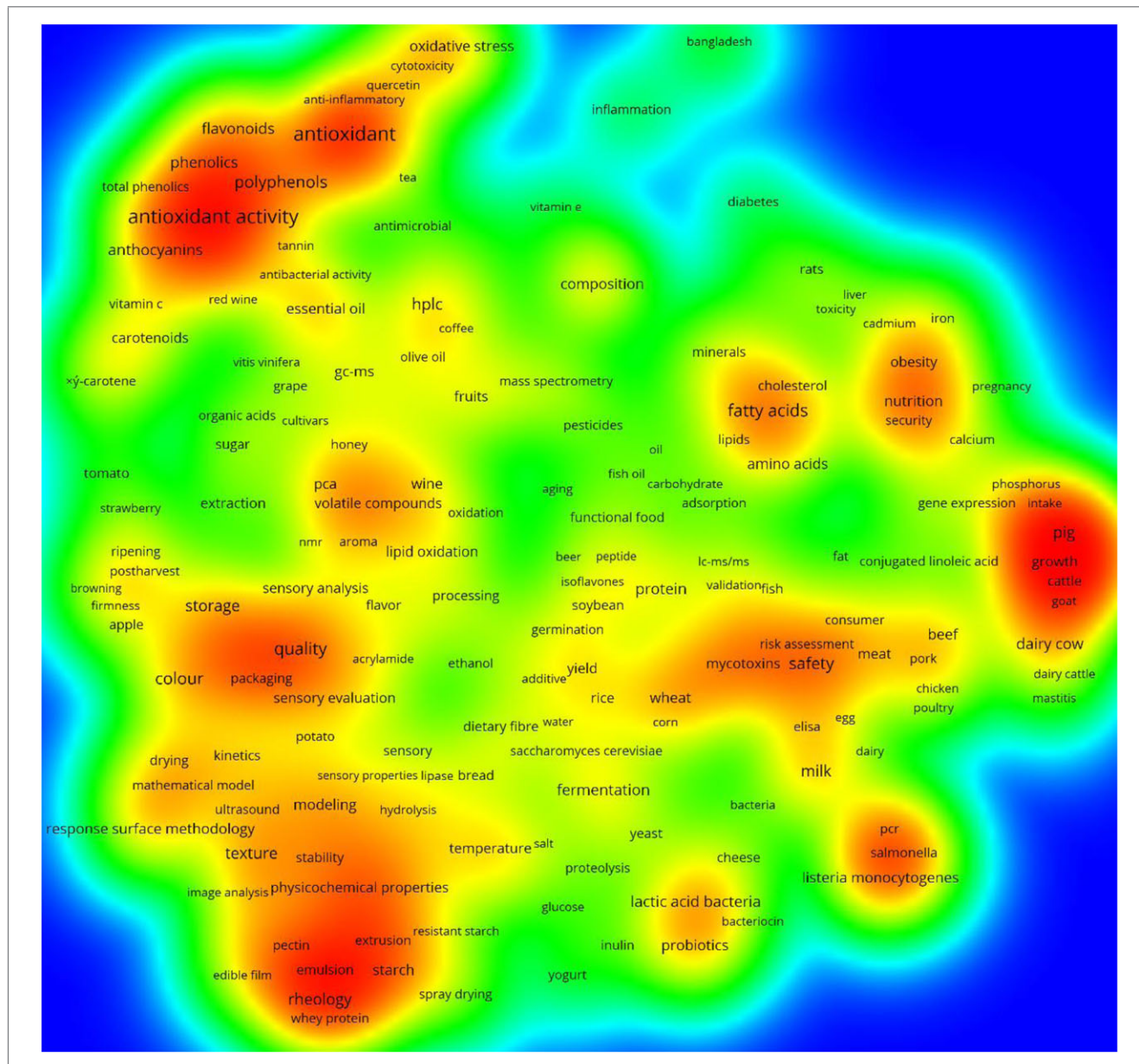
Figure 2–Density view of the general co-word map. In red, zones where there are many keywords that label many documents.

cited at least once, and the percentage of documents relative to the total in FS and its percentage change[3].

As can be seen, all the L1 clusters had a normalized citation above the mean (1)[4]. It is not surprising, therefore, that %Exc is also generally above the mean (10%), the exception being cluster nr 1 dedicated to Food Composition and Nutrition. The same is the case for %Exc1 (1%), although now cluster nr 4 is also an exception.

The documents in cluster nr 5 have the greatest scientific impact[5] of all the L1 clusters. One observes in Table 1 that it also has

the greatest percentage of excellent documents, double the mean of %Exc and almost triple the mean of %Exc1. This is also seen very clearly in Figure 3. All this is indicative of the specialization, cohesion, concreteness, and impact of this cluster.

At the other extreme is cluster nr 1 that has the lowest average normalized citation, the documents with least citations, the lowest percentage of documents in the top 10% of excellence, and the 2nd lowest of documents in the top 1% of excellence. In this last sense, it only surpasses cluster nr 4, which is clearly below the mean in terms of its documents among the top 1% cited, although it is above the mean in the normalized citation and %Exc[6].

[3]Percentage change from 2003 to 2014 in the percentage of documents relative to the total in FS production.

[4]This means that the selection of the most frequent keywords has led us to the scientific production with greatest scientific impact and, therefore, the mainstream of the scientific field.

[5]In this paper, we are using 3 scientometric indicators of scientific impact—NI, %Exc, and %Exc1.

[6]Level-1 cluster nr 5 peaked in normalized impact in 2006, which coincided with a subsequent major growth in the percentage of production in 2007. The decline that began in 2006 was interrupted by a slight upsurge in 2010, only for the fall to resume thereafter. Cluster nr 3, dedicated to Food Security, peaked in impact in 2004, also just before a notable increase in percentage production.
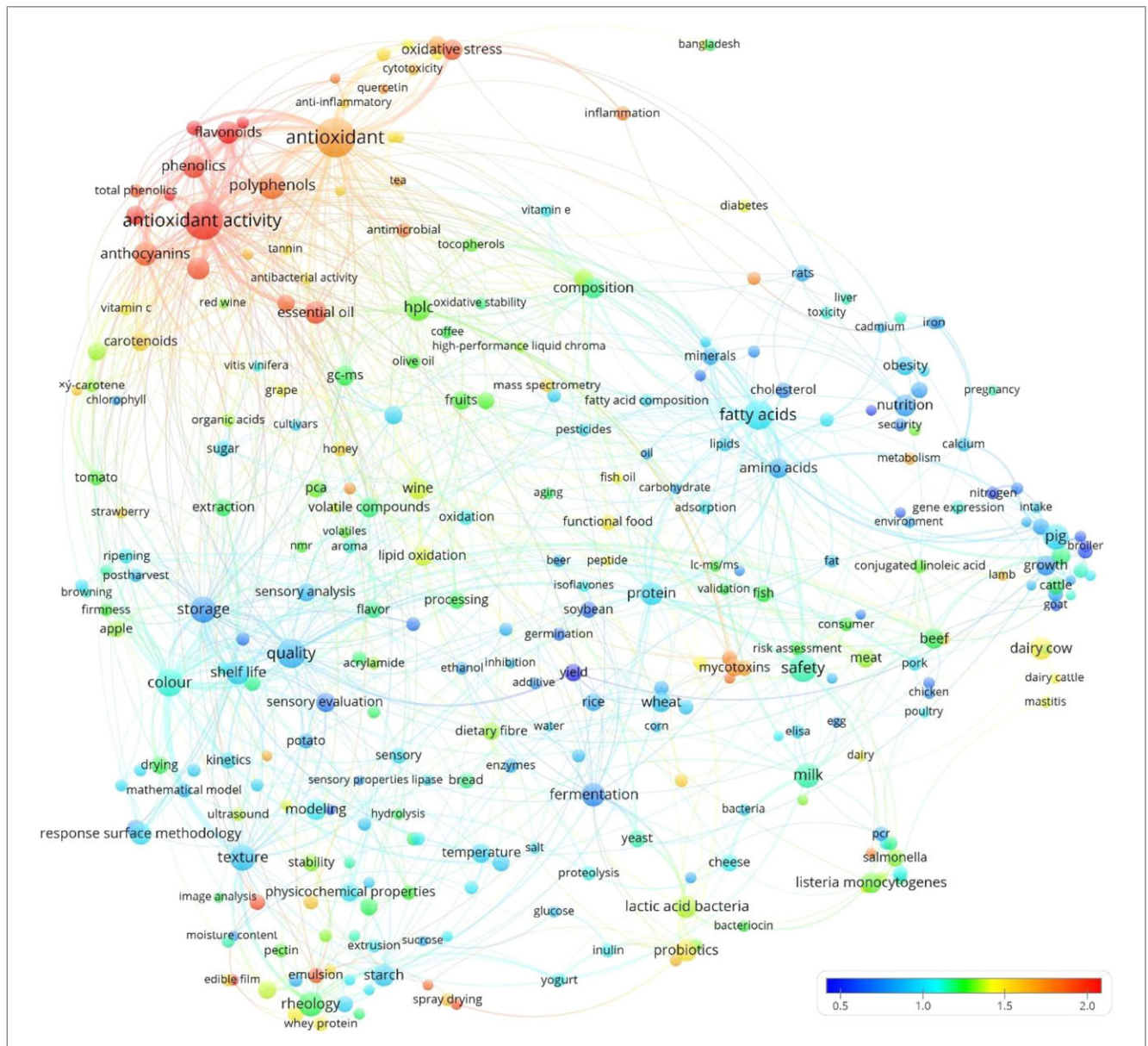
Figure 3–View of the co-word map in which colors mark the average normalized impact obtained by documents tagged with these keywords. The 2014 production is not taken into account considering that its citation data are insufficiently reliable.

All the L1 clusters except nr 5, which is dedicated to Antioxidants in Food, show a slight downward trend. This means that, during this period of time, Antioxidants in Food was attracting ever more attention. This trend, however, is in comparison to the total production in FS, because actually all the clusters showed increased production, and these increases were even at a higher rate than the overall global production in science. Nevertheless, as we indicated earlier, the scientific production in FS is growing at a faster rate than scientific production overall (Guerrero-Bote and Moya-Anegón 2015), thus leading to this decreasing trend being

seen in all the clusters except nr 5[7]. Figure 3, as already mentioned, represents the network of co-words, this time colored on the basis of the average normalized citation obtained by the documents labeled with the keywords that appear on the map. There stands out the zone corresponding to L1 cluster nr 5, this being the zone with the greatest percentage above the mean of cited documents. Table 1 gives the specific data of the average normalized citation of each cluster.

Some keywords stand out more for their impact, although in a dispersed form, without it being possible to say that the entire

This pattern is understandable since the increase in production meant more citations, thus increasing the average impact of the previous years. However, in the following years, the citations these clusters received were distributed among more documents, which thus caused a decrease in the average impact.

[7]Although the L1 clusters had no clear "bursty" periods, cluster nr 1 had a percentage production which exceeded 20% during the early years, and which fell notably in 2006 and again at the end of 2012. Clusters 2, 3, and 4 showed falls in this parameter in 2004, followed by plateaux of different lengths and a generalized slight drop in the last part of the period studied. There stands out a major growth of cluster nr 2 in the last year of the period.

cluster to which they belong stands out. An example is the group of the mycotoxins (ochratoxin A, deoxynivalenol, cereals, aflatoxins) in the L2 subcluster 3.2. The other keywords that label documents with outstanding impact are more isolated. The level-1 cluster nr 1 was that with the lowest NI. Cluster nr 5 was followed in impact by clusters nr 2 (Food Processing and Modification) and nr 3 (Food Security), with the order of these 2 varying every year, and then cluster nr 4.

Figure 4 shows an enlarged view of the upper right zone of the co-word map, where level-1 cluster nr 1, dedicated to Food Composition and Nutrition, is located. The level-2 clusters are distinguished with different shades of blue. Although this is a cluster without any great cohesion, we can see that its 1st 3 level-2 clusters do maintain a certain cohesion, but not the 4th related to Composition and Quality. All this is clearly visible in the density map (Figure 2) where 3 zones of density stand out in the cluster— the zone already mentioned above on livestock terms, another on nutrition, obesity, and diet, and a final one on fatty acids. However, although there is a yellow zone about beef, pork, turkey, and chicken, it is not notable for its density. Except for subcluster 1.2, which covers a very limited zone, the other clusters have some keywords that are distant from the zone corresponding to the cluster itself and are mixed with other clusters. Such is the case of "inflammation" and "Bangladesh" in the case of subcluster 1.1 on Nutrition and Metabolism (which has a certain logic since these are terms that are not specific to FS), of "tocopherols," "oxidative stability," "olive oil," and "oxidation" in subcluster 1.3 dedicated to Food Composition which tend toward the zone of antioxidants, and of "lipid oxidation" in subcluster 1.4 dedicated to Composition and Quality. As can be seen in the Appendix, of the keywords in L1 cluster nr 1, the most frequent in occurrence is "fatty acids" (2331), and the least frequent "stress" (300). Of the 4 subclusters in this level-1 cluster nr 1, subcluster 1.1, dedicated to Nutrition and Metabolism, has the largest number of documents, and therefore the greatest percentage of documents with respect to the total. In the other indicators listed in Table 1, the best results of cluster 1 correspond to subcluster 1.4 dedicated to Composition and Quality, except for %Exc1 where cluster 1.1 is that with the greatest percentage[8]. This particularly stands out in cluster 1.1 because, despite it is not having particularly notable data in either NI or %Exc, it does exceed by 50% the average of %Exc1. This means that it has a very skewed distribution of impact, with a few high-impact documents (more than 50 documents with a NI of more than 10 times the average).

At the other extreme, namely, for the lowest results, subcluster 1.1 has the smallest percentage of documents cited at least once, not only with respect to the subclusters included in the general cluster 1, but also with respect to all the level-2 clusters. The case of subcluster 1.2, dedicated to Nutrients and Quality, is similar. Of all the level-2 clusters, it has the least average citations, the least average normalized citations, the most documents in the highest percentile of all the list, and the lowest percentage of excellent documents for both the 10% and the 1% cases. Subcluster 1.4 is, among the 4 level-2 clusters belonging to the general cluster 1, that with the fewest documents, and therefore also that with the lowest percentage of documents relative to the total. Subcluster

1.1 increases a little the percentage of papers in comparison to the total FS production, and subcluster 1.2 is the one with the greatest loss in this percentage between 2003 and 2014. Returning to Figure 3, one observes that the keywords that label documents with a NI greater than 1.5 were as follows: inflammation (1.75), bioavailability (1.74), and metabolism (1.62) of subcluster 1.1; lamb (1.56) of subcluster 1.2; and tenderness (1.52) of subcluster 1.4.

Finally, the data relative to the bursty period of the keywords are shown in Table 2. Subcluster 1.1 has 7 keywords, 4 of them with an intensity greater than 10 and with 3 similar bursty periods: "inflammation" from 2012 to the present, "security and sustainability" from 2013 to the present, and "obesity" from 2014 to the present. Subcluster 1.2 has 9 keywords, 6 of which have an intensity greater than 10. Of these 6, 5 started their bursty periods in 2003, and the 6th, "digestibility," began and ended in 2004. The 5 that started in 2003 finished their bursty periods in different years: 3 (sheep, beef cattle, carcass) in 2004, another (growth) in 2005, and the last (pig) in 2006. Subcluster 1.3 has 4 keywords, 2 of them with an intensity greater than 10 and with the same bursty period of 2003 to 2005. These 2 are "fat" and "conjugated linoleic acid." Finally, in subcluster 1.4 there are 4 keywords, 3 with the same bursty period of 2003 to 2004. These 3 are also those that exceed intensity 10: "beef," "tenderness," and "pork."

In Figure 5, in the lower right zone of the map there are areas of different shades of green which correspond to the level-2 clusters included in level-1 cluster nr 2, dedicated to Food Processing and Modification. In general, one observes that this cluster is separated from the rest of the map by the level-1 cluster nr 4 dedicated to Food Preservation and Shelf Life.

In the case of level-1 cluster nr 2, we observe that subcluster 2.1 appears to be strongly cohesive, whereas the rest are spread out. Indeed, going back to Figure 2, we can see a high-density zone including the keywords rheology, starch, viscosity, emulsion, and so on. In this figure, there is another lower density zone in subcluster 2.3 surrounding probiotics, prebiotics, and lactic acid bacteria (although this last corresponds to subcluster 2.2). Subclusters 2.2 and 2.4, dedicated to Food Processing or Treatment Methods, and Modifications of Food during Processing and How to Determine Them, respectively, merit special mention for making their way through cluster nr 4 to touch level-1 cluster nr 3. As can be seen in the Appendix, the most frequent keyword is "rheology" (1605) and "peptide" the least frequent (303).

We shall now look at the production, citations, and excellence data (Table 1). Subcluster 2.1, dedicated to Influence of Food Processing on Sensory Characteristics, is the subcluster with the most documents and therefore with the greatest percentage of documents relative to the total[9], as well as being the one with the greatest percentage of documents with 1 or more citations. Subcluster 2.3 is the one with the fewest documents, and also the one that includes the least proportion of documents relative to the total. It is also the subcluster with the 2nd greatest values of these 3 indicators among all the level-2 clusters. In the impact indicators, subcluster 2.3, dedicated to Use of Beneficial Microorganisms as Starter Cultures in Food Processing, has the best results, followed by subcluster 2.1 (although they all have an average NI greater than

---

[8]The only subcluster that exceeds the average in normalized impact over the entire period studied is subcluster 1.4, with a notable upward trend in 2014. Subcluster 1.2 is below the mean for nearly the whole period, only exceeding it in 2014. Subcluster 1.1 only fails to be above the mean in 2008, and subcluster 1.3, dedicated to Food Composition, is below the mean in the period 2009 to 2013, but recovers in 2014.

[9]Except for subcluster 2.2, the other 3 subclusters show increases in the percentages relative to the initial year of the period studied. Subclusters 2.3 and 2.4 have very similar percentage productions. They grow steadily, although they are still well below that of the other 2 subclusters in the general cluster 2, especially when compared with subcluster 2.1.
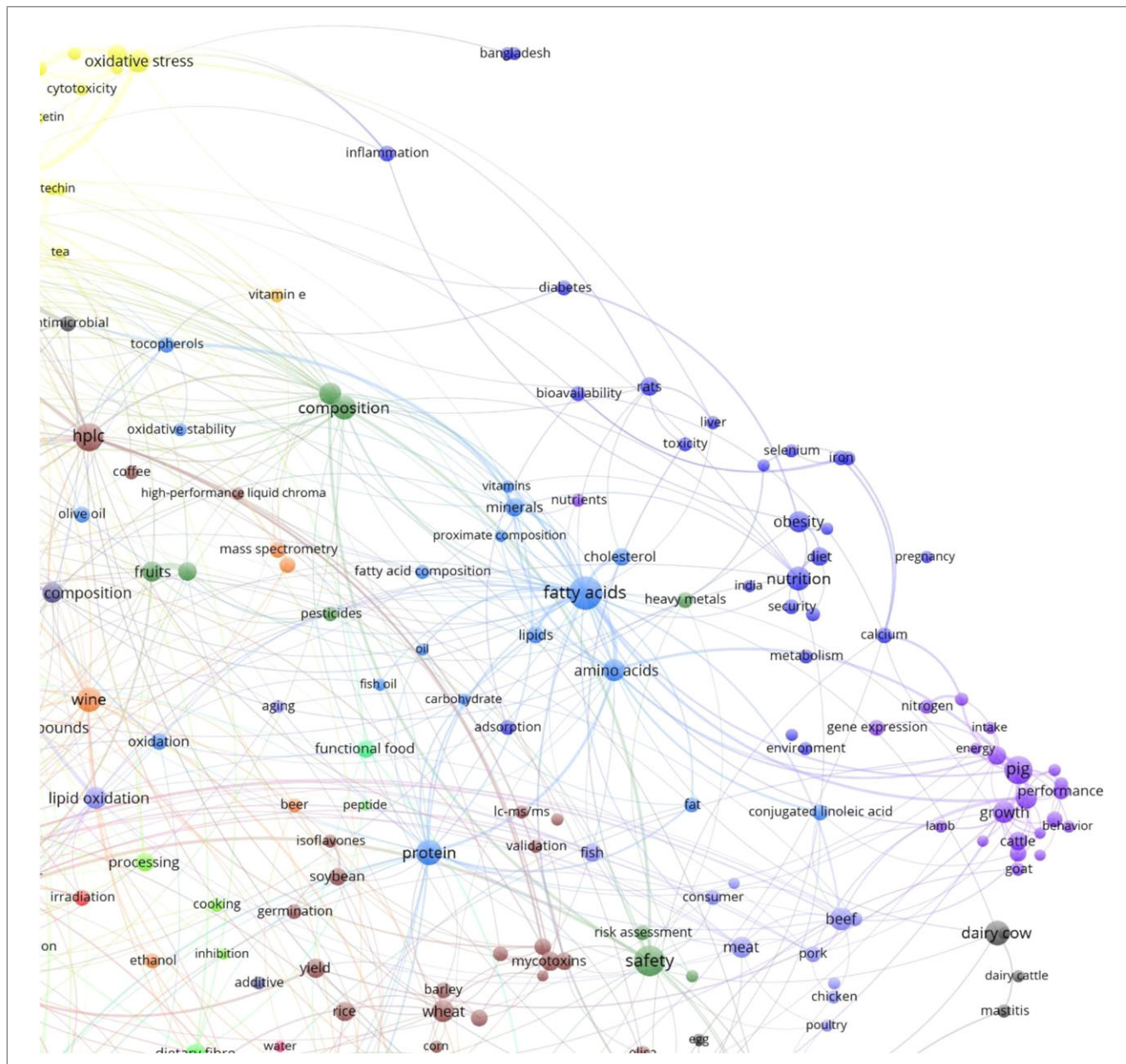
Figure 4–Zoom into the upper right zone of the co-word map where the level-1 (L1) cluster 1, dedicated to Food Composition and Nutrition, is located. Different shades of blue distinguish its level-2 (L2) clusters.

unity)[10]. As for the lowest results, they correspond to subcluster 2.2, dedicated to Food Processing or Treatment Methods, in the average NI and in the 10% and 1% of excellence. There stands out the small %Exc1 value of subcluster 2.2, which means that the distribution of impact is not very skewed.

Figure 3 shows that the keywords that exceed 1.5 in NI are the following: edible film (2.17), chitosan (1.9), emulsion (1.85), encapsulation (1.83), polysaccharide (1.57), and mechanical properties (1.5) from subcluster 2.1; microencapsulation (1.88), prebiotics (1.61), and spray drying (1.57) from subcluster 2.3; and

[10]Subcluster 2.1 is the only one to show increases in NI. In most years, the 4 subclusters exceeded the mean of 1, but subclusters 2.2 and 2.4 fell below it in 2011, subcluster 2.2 in 2014, and subcluster 2.4 in 2004 and from 2013 onwards. In 2006, subcluster 2.3 obtained the greatest impact of all the subclusters of this chart, surpassing the mean by almost 7 tenths.

pulsed electric field (1.66) from subcluster 2.2. Subcluster 2.4, dedicated to Modifications of Food during Processing and How to Determine Them, is the one with the fewest citations, with its documents in the highest percentile, and with the lowest percentage of documents cited at least once. Only 2 of its level-2 clusters have keywords with an intensity greater than 10 (Table 2). This is the case of subclusters 2.1 and 2.2, with 7 keywords each. In the 1st, the keywords that exceed the intensity value 10 are: cyclodextrin, with a bursty period from 2006 to 2007, and encapsulation, with a bursty period from 2013 to the present. In the 2nd they are: mathematical model, with a bursty period from 2014 to the present, and high pressure, with a bursty period from 2003 to 2006. Subcluster 2.3 only includes 1 keyword, glucose, with an intensity of 6.38 and a bursty period from 2003 to 2004. Subcluster 2.4 has 4 keywords. One of them, enzymes, has a bursty period
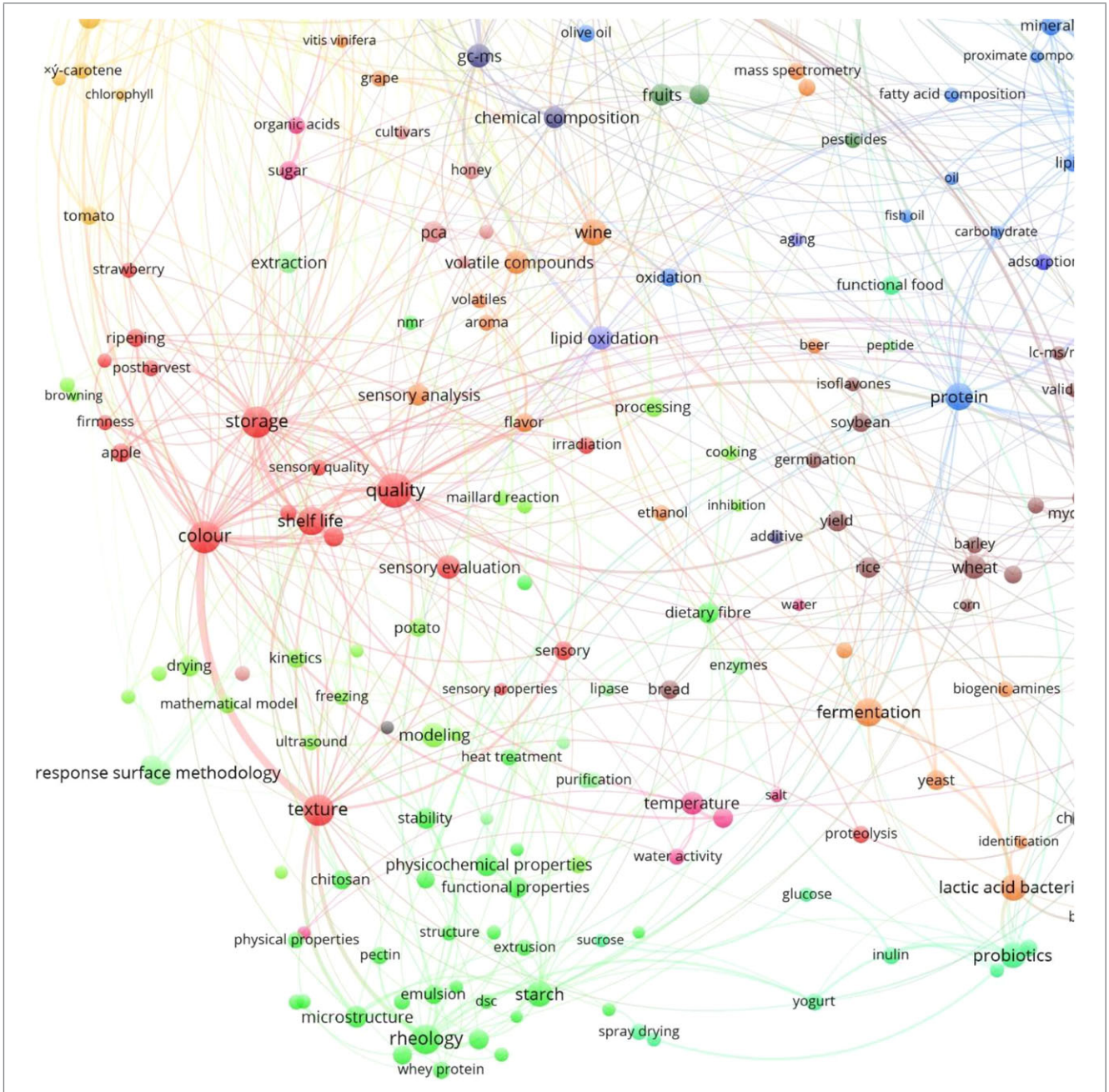
Figure 5–Zoom into the lower left zone of the co-word map where the level-1 (L1) clusters 2 and 4 (green and red), dedicated to Food Processing and Modification and Food Preservation and Shelf Life, respectively, are located. Different shades of the 2 colors distinguish their level-2 (L2) clusters.

from 2003 to 2004, and has an intensity of almost 10, reaching 9.54.

Figure 6 shows the central zone of the co-word map. We can observe in different mixes of gray with colors the 4 level-2 clusters in the level-1 cluster nr 3, dedicated to Food Security. This cluster extends diagonally across the map and touches most of the other clusters. As it is not located in a particular zone, we cannot say that it has any great coherence as a whole. However, returning to Figure 2, we see that it includes 2 high-density zones. The 1st surrounds the keywords "salmonella," "*Listeria monocytogenes*," and Polymerase Chain Reaction ("PCR.") The other, which is more extensive but less intense, surrounds "safety" and "mycotoxins" and spreads out on 1 side to "cereals: wheat, maize, corn, or bar-

ley," to "dairy products and eggs," and to "meat: meat, pork, beef, turkey" which belong to the L1 cluster nr 1. The most frequent keyword is "safety" (2055), and "allergy" is the least frequent (301). Subcluster 3.1, dedicated to Pathogenic Microorganisms in Milk and Dairy Products, has its documents in the lowest percentile. Subcluster 3.2, dedicated to Mycotoxins in Cereals and Methods of Detecting Mycotoxins in Food, has the greatest number of documents and, logically, contributes the greatest percentage of documents to the total[11]. With respect to the number of citations, the number of documents in the top 10% and 1% of excellence,

---

[11]Except for subcluster 3.3, whose percentage increases, the percentages of the other 3 subclusters decline, although only slightly.
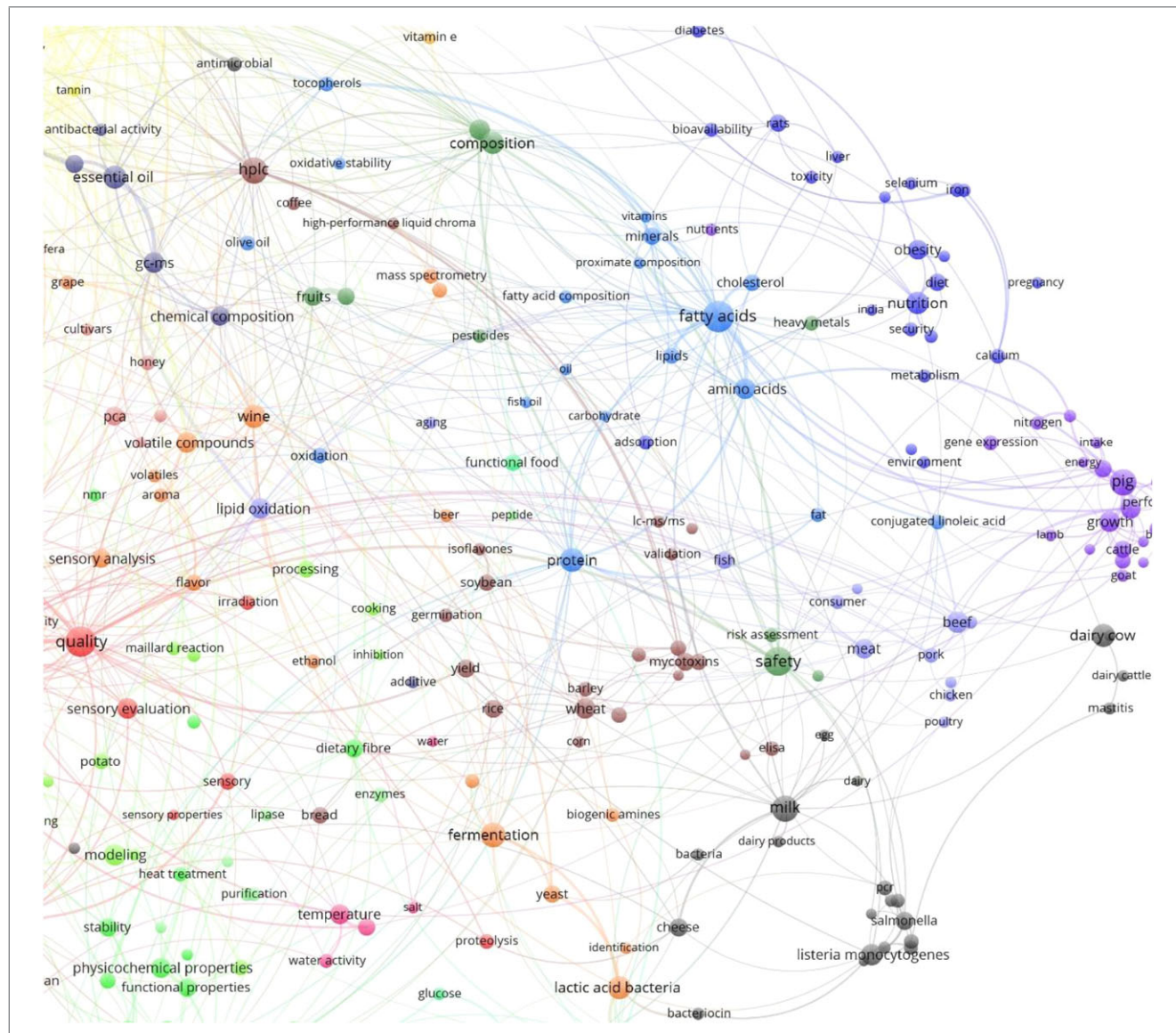
Figure 6–Zoom into the central zone of the co-word map where the level-1 (L1) cluster 3, dedicated to Food Security, is located. Different shades of gray distinguish its level-2 (L2) clusters.

and the percentage of documents cited at least once, it is subcluster 3.4, dedicated to Antimicrobial Agents Used in Food, which has the best results in addition to having the best average NI[12] in the general cluster 3 and the 2nd best score of all the level–2 clusters. It stands out in this regard that, although subcluster 3.1 is 2nd in impact and in %Exc, it is clearly surpassed by subcluster 3.3 in %Exc1, which means that the distribution of impact is more skewed in the latter. However, subcluster 3.2, which is last in impact, despite being above 1 and 10 in NI and %Exc, respectively, falls to 0.60 in %Exc1, which means that it contains very few outstanding documents.

The keywords that label the documents with greatest average NI are: essential oil (1.91) and antimicrobial activity (1.85) of subcluster 3.4; deoxynivalenol (1.8), ochratoxin A (1.75), mycotoxins (1.69), and aflatoxins (1.65) of subcluster 3.2; and foodborne pathogen (1.77) and antimicrobial (1.76) of subcluster 3.1. Subcluster 3.4 does not include any keyword with a bursty period (Table 2). Subcluster 3.1 has 8 keywords, and 2 of them exceed an intensity of 10. These are "salmonella" and "milk production" with bursty periods in 2012 for the 1st and from 2003 to 2004 for the 2nd. Subcluster 3.2 has 4 keywords, and only 1 of them exceeds an intensity of 10. This is liquid chromatography ("LC-) mass spectrometry (MS)/MS" with a bursty period from 2013 to the present. Finally, in subcluster 3.3 there are "composition" and "analysis," with bursty periods from 2009 to 2011 for the 1st, and from 2010 to 2011 for the 2nd, both exceeding the intensity of 10.

The map of the level-1 cluster nr 4, dedicated to Food Preservation and Shelf Life, is shown in Figure 5 together with cluster

---

[12]All the clusters exceed the mean of 1 throughout the period studied except for subcluster 3.2 which falls below that value in 2003 and 2011, but in 2014 again rises above it. In 2004, subcluster 3.4 has a peak in impact, the greatest impact of all the subclusters in this L1 cluster, exceeding the mean by more than 1 point. It also has another peak in 2008. Interestingly, there are other clusters with peaks in those years.

Table 2–Bursty periods of the keywords, ordered by the level-2 clusters

Word	Length	Strength	Start	L2 cluster
Inflammation	4	29.93	2012	1.1
Security	3	27.20	2013	1.1
Obesity	2	18.05	2014	1.1
Sustainability	3	13.36	2013	1.1
Pregnancy	1	4.60	2003	1.1
Children	1	3.10	2005	1.1
Cadmium	1	2.61	2004	1.1
Pig	4	81.08	2003	1.2
Sheep	2	29.16	2003	1.2
Growth	3	25.77	2003	1.2
Beef cattle	2	12.84	2003	1.2
Digestibility	1	12.50	2004	1.2
Carcass	2	11.98	2003	1.2
Goat	2	7.14	2004	1.2
Lamb	3	5.48	2003	1.2
Cattle	2	4.18	2003	1.2
Fat	3	16.73	2003	1.3
Conjugated linoleic acid	3	14.48	2003	1.3
Carbohydrate	1	4.57	2003	1.3
Olive oil	1	3.11	2004	1.3
Beef	2	28.72	2003	1.4
Tenderness	2	13.36	2003	1.4
Pork	2	12.42	2003	1.4
Chicken	2	2.76	2005	1.4
Cyclodextrin	2	30.94	2006	2.1
Encapsulation	3	14.31	2013	2.1
β-lactoglobulin	2	5.14	2003	2.1
Viscosity	1	4.25	2006	2.1
Gelatinization	1	3.83	2003	2.1
Ultrafiltration	1	2.95	2003	2.1
Gelatinization	1	6.74	2008	2.1
Mathematical model	2	17.21	2014	2.2
High pressure	4	10.76	2003	2.2
Kinetics	1	7.22	2003	2.2
Osmotic dehydration	1	3.89	2007	2.2
Modeling	1	3.61	2006	2.2
Ultrasound	2	3.05	2014	2.2
Inhibition	1	2.60	2006	2.2
Glucose	2	6.38	2003	2.3
Enzymes	2	9.54	2003	2.4
Purification	2	6.23	2014	2.4
Response surface methodology	2	3.33	2014	2.4
Lipase	1	3.01	2003	2.4
Salmonella	1	10.92	2012	3.1
Milk production	2	10.18	2003	3.1
Dairy cow	1	9.85	2004	3.1
Dairy cattle	1	6.35	2004	3.1
Mastitis	3	5.38	2003	3.1
Escherichia coli	1	3.37	2003	3.1
Dairy	3	3.36	2013	3.1
Cheese	1	3.01	2003	3.1
LC-MS/MS	3	24.44	2013	3.2
Ochratoxin A	2	7.32	2005	3.2
Soybean	2	3.96	2003	3.2
Residues	1	2.79	2006	3.2
Composition	3	29.19	2009	3.3
Analysis	2	22.19	2010	3.3
Proteolysis	2	26.60	2003	4.1
Irradiation	2	15.43	2003	4.1
Proteolysis	1	38.12	2006	4.1
Storage	2	9.12	2003	4.1
Ripening	2	8.77	2005	4.1
Firmness	1	7.08	2006	4.1
Irradiation	1	20.58	2007	4.1
Apple	1	5.01	2003	4.1
Aroma	1	7.14	2003	4.2
Grape	1	5.14	2006	4.2
Volatiles	1	4.08	2003	4.2
Yeast	1	2.58	2003	4.2
pH	1	14.76	2003	4.3
Moisture content	1	3.64	2007	4.3
Water	1	3.08	2004	4.3
PCA	2	5.81	2014	4.4
Oxidative stress	3	37.81	2013	5.1
Bioactive compounds	3	24.07	2013	5.1
Anti-inflammatory	4	14.70	2012	5.1
Tannin	1	5.86	2004	5.1
Tannin	1	9.70	2006	5.1
Vitamin E	2	6.67	2003	5.2

nr 2. The keyword with the greatest frequency is "quality" (2272), and that with the lowest frequency is "water" (304).

The L2 subclusters included in this cluster do not have very sharply confined zones, but tend to merge considerably with other clusters. As was mentioned above, it acts as a barrier between cluster nr 2 and the rest of the map, although merging substantially into the latter. However, in the density map (Figure 2) one sees a pair of high-density zones. The 1st is in the central zone of subcluster 4.1, around the keywords "quality," "shelf life," "sensory quality," and "packing." The 2nd is around "volatile compounds" of subcluster 4.3. There stands out the existence of a subcluster, 4.4, formed by keywords alluding to numerical analysis. Subcluster 4.1, dedicated to Increased Shelf Life in Plant Foods during Preservation, and Methods and Modifications of Quality, has the greatest number of documents, and contributes the greatest number of documents to the total (Table 1)[13]. Subcluster 4.2, dedicated to Fermentation Process in Wine and Beer as a Preservation Method, has the greatest percentage of documents cited at least once. In the impact indicators, the best situated subcluster is 4.4 dedicated to Statistical Analysis Methods, which suggests that the works using these methods have a greater average scientific impact[14].

The keywords that label the documents with greatest scientific impact are: chemometrics (1.72) and honey (1.52) from subcluster 4.4; biogenic amines (1.53) and MS (1.51) from subcluster 4.2; and strawberry (1.51) from subcluster 4.1.

The worst results in this series of indicators are: subcluster 4.2 for the number of excellent documents in the 1% most cited; and subcluster 4.3, dedicated to Parameters that Influence Food Preservation, for the number of citations, average normalized citations, percentile location, number of excellent documents in the 10% most cited, and the percentage of documents cited at least once. In the case of the number of documents and percentage of documents relative to the total, subcluster 4.4 is the worst located both in the set of subclusters belonging to the general cluster 4 and in the total set of 18 level–2 clusters analyzed in this study. Turning to the bursting of this general cluster 4 (Table 2), we observe that subcluster 4.1 includes 8 keywords, 2 of which have an intensity greater than 10 and the peculiarity of having 2 different bursty periods. The 1st keyword, "proteolysis," has a period from 2003 to 2004 and another that begins and ends in 2006. The 2nd, "irradiation," also has a 1st bursty period from 2003 to 2004 and the 2nd begins and ends in 2007. Subcluster 4.2 includes 4 keywords, none with an intensity greater than 10. The case is similar for subcluster 4.4, which has only 1 keyword and neither does it surpass the intensity barrier of 10. Finally, subcluster 4.3 has 3 keywords, with "pH" being the only one with an intensity greater than 10 (14.76). Its bursty period is in 2003.

Figure 7 shows the upper right zone of the map where shades of yellow distinguish the 2 level-2 clusters in the level–1 cluster nr 5 dedicated to Antioxidants in Food. Documents labeled with the keywords of this cluster are those with the greatest scientific impact. Moreover, returning to Table 1, we see that they

[13]Subcluster 4.4 is the only one showing an increase in percentage. The other 3 show decreases, especially subcluster 4.1. But this last also has the greatest production, followed by subcluster 4.2, with subclusters 4.4 and 4.3 much farther behind.

[14]Subclusters 4.2 and 4.4 are above the mean at all times. However, this is not so with subclusters 4.1 and 4.3, which before 2008 exceeded the mean of 1, but from that year onwards began to fall, with a slight recovery in 2010, but again going below the mean in 2011. Subcluster 4.1 does not recover, and is still at present below the mean. In 2007, subcluster 4.4 obtains the greatest impact of all the subclusters of this L1 cluster.
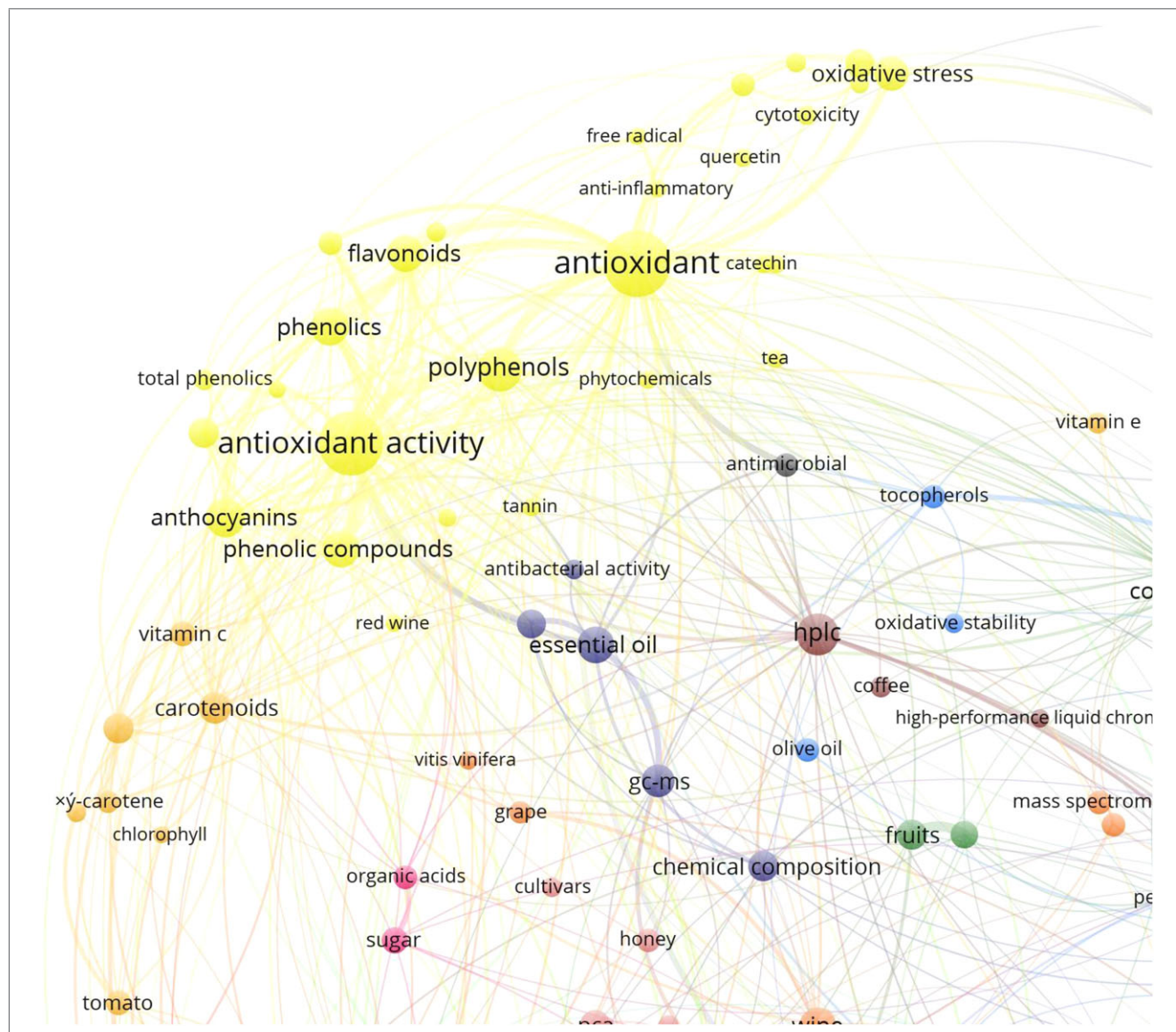
Figure 7–Zoom into the upper left zone of the co-word map where the level-1 (L1) cluster 5, dedicated to Antioxidants in Food, is located. Different shades of yellow distinguish its level-2 (L2) clusters.

are also those with the greatest percentage of excellence, doubling the mean of %Exc and almost tripling the mean of %Exc1.

In general terms, cluster nr 5 occupies a very limited zone of the map, which is indicative of its great internal cohesion. In the density map of Figure 2, we see that there are 2 of the zones of greatest density in this cluster, one around the keywords "antioxidant," "anti-inflammatory," "quercetin," and another around "antioxidant activity," "phenolics," "polyphenols," and so on. However, if we go down a level to that of the subclusters, we see that while all this is true about subcluster 5.1, it is not about subcluster 5.2. Subcluster 5.2 occupies a broader zone, with some keywords being more remote, and with no high-density zone. The keyword "antioxidant" is the one with the greatest frequency (4227). This is so not only for this cluster of keywords alone, but also for all the 297 keywords used in this study. The keyword with the lowest frequency is "chlorophyll" with 305 appearances.

The subcluster with the greatest number of documents is 5.1, dedicated to Antioxidants and Their Effects, with a total of 15060

documents[15]. In addition, this same subcluster received the greatest number of citations, and also has the greatest average normalized citations, the most excellent documents in the 10% and 1% most cited[16]. As well as this, its remaining documents are in the lowest percentile relative to the other level-2 clusters, the greatest percentage of documents cited at least once, and the greatest percentage of documents relative to the total. Namely, subcluster 5.1 has the best values in all the categories of Table 1, and in general the best of all the 18 level-2 clusters. The only other level-2

[15]We found that subcluster 5.1 increases its percentage of production, evidence of the great interest generated in this field. This subcluster peaked in 2011, with a noticeable drop in 2012. Subcluster 5.2 hardly varies. It loses 12% in 2014 in comparison to 2003, but remained very stable.

[16]The 5.1 subcluster decreased in normalized impact with respect to 2003 by 23%. However, both of the subclusters in this general cluster remain above the mean in all the years of the period studied, 2003 to 2014. The peaks of impact occur in 2006, with the peak being more pronounced in subcluster 5.2, although it is still below the general impact of subcluster 5.1.
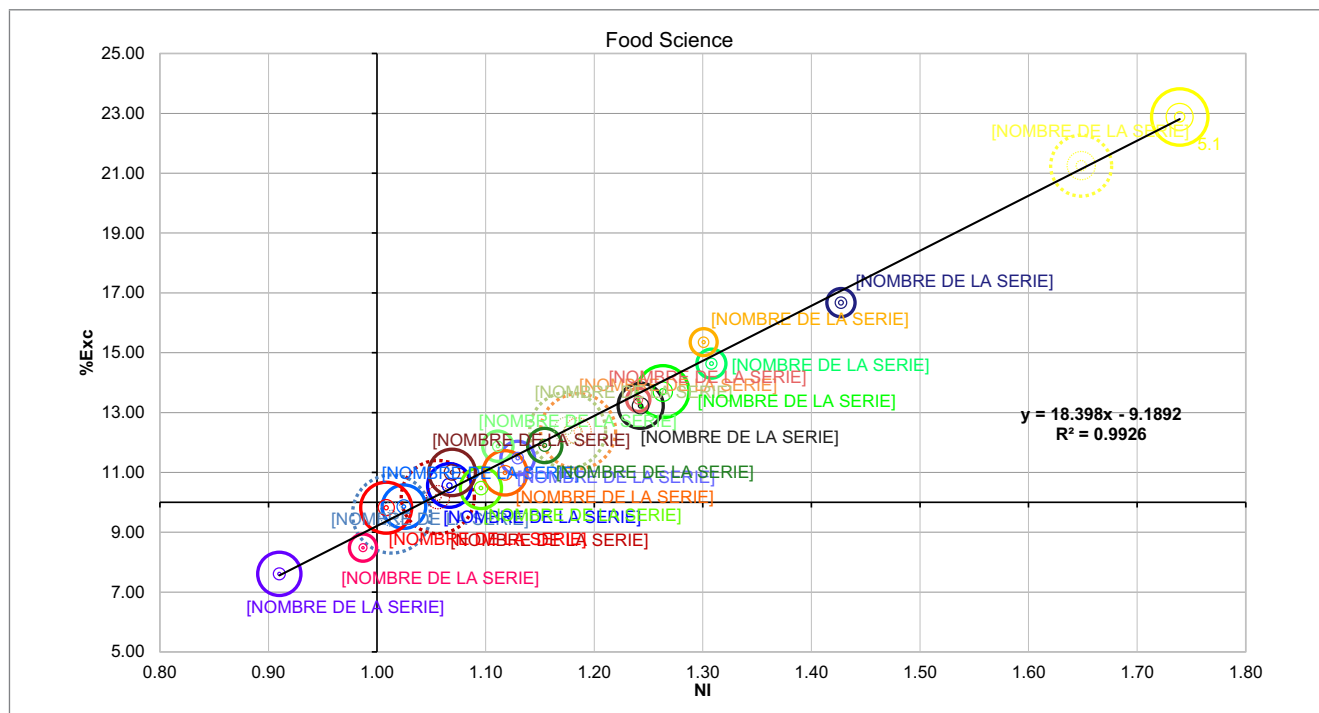
Figure 8–Percentage of top 10% excellent documents as Research Guarantor compared with the normalized impact of the 5 level-1 (L1) clusters (dashed circles) and the 18 level-2 clusters. The 3 concentric circles correspond, respectively, to Ndocc (citable scientific production), Exc (scientific production of excellence, among the 10% most cited), and Exc1 (scientific production among the 1% most cited), respectively.

cluster is subcluster 5.2 dedicated to Plant Antioxidants. While this has the worst values in Table 1, its impact results are still good in relation to the other 16 L2 clusters, and it is actually the 2nd after subcluster 5.1 in citations and the location of its documents in a low percentile.

Virtually all of the keywords in subcluster 5.1 are above the average NI of 1.5. Indeed, the keyword with the lowest average is "red wine" with 1.33. The keywords with greatest values of average NI are: 2,2-diphenyl-1-picrylhydrazyl (DPPH (2.32)), total phenolic content (2.28), phenolic acids (2.19), and flavonoids (2.17). The only ones in subcluster 5.2 that exceed the NI of 1.5 are: lycopene (1.55) and carotenoids (1.52). Finally, only the 1st level-2 cluster of general cluster nr 5 has keywords with intensities greater than 10 (Table 2). These are: "oxidative stress" with a bursty period from 2013 to the present; "bioactive compounds," also with a bursty period from 2013 to the present; and "anti-inflammatory" with a bursty period from 2012 to the present. This subcluster also has the keyword "tannin" with 2 different bursty periods, although neither reaches an intensity exceeding 10: one in 2004, and another in 2006 which fails by three tenths to reach an intensity of 10. Subcluster 5.2 has only one keyword, "vitamin E," with a bursty period from 2003 to 2004, and it does not surpass an intensity of 10.

Figure 8 shows the scatter plot comparing the percentage of documents of excellence 10 (%Exc) with the NI of the 5 level-1 clusters and the 18 level-2 clusters[17].

[17]At a glance, one observes that there is a strong positive correlation between the 2 variables since, 1st, all the clusters are concentrated around the trend line, and 2nd, as the value of the Y-axis increases so does the value of the X-axis, and the other way round. In other words, the greater the normalized impact of a cluster, the greater the number of excellent documents in the 10% most cited, and the other way round. Besides, this correlation is more formally confirmed

The concentric circles represent 3 of the parameters studied: the number of documents in each cluster, the number of excellent documents in the 10% most cited, and the number of excellent documents in the 1% most cited, respectively. Moreover, as has been mentioned above, one sees clearly that the great majority of clusters surpass the mean NI of 1, with only subclusters 1.2 and 4.3 being below this value. One also observes, as so many other times before, that there stand out the level-1 cluster nr 5 and its 1st level-2 cluster.

Conclusions

The co-word analysis and the generation of the corresponding maps for the specific subject area of FS have made clearly observable a structural division of this area into 5 major clusters, conforming the 5 general areas of FS. At a finer level of resolution, there was a division into 18 subclusters, which correspond to the 18 subareas of FS.

Right from the beginning of the study, subcluster 5.1, dedicated to Antioxidants and their Effects, appeared to begin to break away from the rest. The more data that were collected, the more this supposition was confirmed. The theme corresponding to this subcluster 5.1 had the best results in all the indicators studied— evidence of its solidity, specialization, and high internal cohesion.

Acknowledgments

by the value of the square of the Pearson's correlation coefficient (R^2) which is equal to 0.9926, and is close to being a perfect positive correlation with total direct dependence between the 2 variables. This would have been the case if R^2 had been equal to 1.

Author Contributions

Jesús Blázquez-Ruiz collected test data and drafted the manuscript. Vicente P. Guerrero-Bote designed the study and interpreted the results. Félix Moya-Anegón interpreted the results and revised the manuscript.

References

Alfaraz PH, Calviño AM. 2004. Bibliometric study on food science and technology: scientific production in Iberian-American countries (1991-2000). Scientometrics 61(1):89–102.

Bornmann L. 2011. Scientific peer review. Annu Rev Inform Sci Technol 45(1):197–245.

Bornmann L, Daniel HD. 2008. What do citation counts measure? A review of studies on citing behavior. J Document 64(1):45–80.

Bornmann L, Moya-Anegón F, Leydesdorff L. 2012. The new excellence indicator in the World Report of the SCImago Institutions Rankings 2011. J Informetr 6(2):333–5.

Callon M, Courtial JP, Laville F. 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics 22(1):155–205.

Clauset A, Newman M, Moore C. 2004. Finding community structure in very large networks. Phys Rev E 70(6):1–6.

Damerau FJ. 1964. A technique for computer detection and correction of spelling errors. Commun ACM 7(3):171–6.

Garfield E. 1955. Citation indexes for science: a new dimension in documentation through association of ideas. Science 122(3159):108–11.

Garfield E, Sher IH. 1963. New factors in the evaluation of scientific literature through citation indexing. Am Document 14(3):195–201.

Guerrero-Bote VP, Moya-Anegón F. 2015. Analysis of scientific production in food science from 2003 to 2013. J Food Sci 80(12):R2619–26.

Guerrero-Bote VP, Olmeda-Gómez C, Moya-Anegón F. 2016a. La ciencia de los alimentos georreferenciada. aproximación bibliométrica a nivel institucional. El profesional de la información 25(1):25–34.

Guerrero-Bote VP, Olmeda-Gómez C, Moya-Anegón F. 2016b. Atlas of scientific institutions in food science (Scopus, 2003-2013). LWT - Food Sci Technol 67:133–42.

Hinze S, Grupp H. 1996. Mapping of R&D structures in transdisciplinary areas: new biotechnology in food sciences. Scientometrics 37(2):313–35.

Irvine J, Martin B, Peacock T, Turner R. 1985. Charting the decline of British science. Nature 316:587–90.

Kleinberg J. 2003. Bursty and hierarchichal structure in streams. Data Min Knowl Discov 7(4):373–97.

Levenshtein VI. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Phys Doklady 10(8):707–10.

Leydesdorff L, Moya-Anegón F, Guerrero-Bote VP. 2010. Journal maps on the basis of Scopus data: a comparison with the journal citation reports of the ISI. J Am Soc Inform Sci Technol 61(2):352–69.

López-Illescas C, Moya-Anegón F, Moed HF. 2009. Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus: the effect of poorly cited journals in oncology. J Inform Sci 35(2):244–56.

Merton RK. 1968. The Matthew effect in science. Science 159(3810):56–63.

Merton RK. 1988. The Matthew effect in science II: cumulative advantage and the symbolism of intellectual property. ISIS 79:606–23.

Miguel S, Chinchilla-Rodríguez Z, Moya-Anegón F. 2011. Open access and Scopus: a new approach to scientific visibility from the standpoint of access. J Am Soc Inform Sci Technol 62(6):1130–45.

Mingers J, Leydesdorff L. 2015. A review of theory and practice in scientometrics. Eur J Operat Res 246(1):1–19.

Moed HF. 2005. Citation analysis in research evaluation. Dordrecht: Springer.

Muscio A, Nardone G. 2012. The determinants of university-industry collaboration in food science in Italy. Food Policy 37(6):710–8.

Nalimov V, Mulcjenko B. 1971. Measurement of science: study of the development of science as an information process. Washington, D.C.: Foreign Technology Div., US Air Force System Command.

Neff MW, Corley EA. 2009. 35 Years and 160,000 articles: a bibliometric exploration of the evolution of ecology. Scientometrics 80(3):657–82.

Newman M, Girvan M. 2004. Finding and evaluating community structure in networks. Phys Rev E 69(2). DOI: 10.1103/PhysRevE.69.026113.

Noack A. 2007. Energy models for graph clustering. J Graph Algorithms Appl 11(2):453–80.

Noack A. 2009. Modularity clustering is force-directed layout. Phys Rev E Stat Nonlin Soft Matter Phys 79(2). DOI: 10.1103/PhysRevE.79.026102.

Price DJdS. 1963. Little science, big science. New York, N.Y.: Columbia Univ. Press.

Price DJdS. 1965. Networks of scientific papers. Science 149(3683):510–5.

Price DJdS. 1976. A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inform Sci 27:292–306.

Rehn C, Kronman U. 2008. Bibliometric handbook for Karolinska Institutet. Huddinge: Karolinska Institutet.

Romo-Fernández LM, Guerrero-Bote VP, Moya-Anegón F. 2013. Co-word based thematic analysis of renewable energy (1990-2010). Scientometrics 97(3):743–65.

Seetharam G, Ravichandra Rao IK. 1999. Growth of food science and technology literature: a comparison of CFTRI, India and the world. Scientometrics 44(1):59–79.

Van-Eck NJ, Waltman L. 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–38.

Van-Eck NJ. 2011. Methodological advances in bibliometric mapping of science. [PhD Series Research in Management 247]. Rotterdam: Erasmus Research Institute of Management. ISBN: 978 90 5892 291 5.

Van Eck NJ, Waltman L. 2007. VOS: a new method for visualizing similarities between objects. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-540-70981-7_34.

Van Eck NJ, Waltman L, Dekker R, van den Berg J. 2010. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. J Am Soc Inform Sci Technol 61(12):2405–16.

Waltman L, Van Eck NJ, Noyons ECM. 2010. A unified approach to mapping and clustering of bibliometric networks. J Informetr 4(4):629–35.

Wouters P. 1999. The citation culture [Unpublished PhD thesis]. Univ. of Amsterdam, Amsterdam.

Wouters P. 2014. The citation: from culture to infrastructure. In: Cronin B, Sugimoto C, editors. Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. London: MIT Press. p 47–66.

Zhou G, Zhang W, Xu X. 2012. China's meat industry revolution: challenges and opportunities for the future. Meat Sci 92(3):188–96.

Appendix: The Distribution by Cluster of the Keywords and their Ndocc Values

1. Food composition and nutrition	
1.1. Nutrition and metabolism	Nutrition (1237), obesity (948), rats (734), diet (708), calcium (571), inflammation (559), adsorption (506), iron (489), bioavailability (486), diabetes (470), security (464), zinc (440), toxicity (430), metabolism (429), health (417), selenium (387), liver (384), Bangladesh (381), medicinal plants (374), environment (366), children (355), cadmium (330), pregnancy (324), sustainability (322), India (321)
1.2. Nutrients and quality	Pig (1744), growth (1035), meat quality (950), cattle (773), digestibility (731), performance (622), sheep (621), beef cattle (560), gene expression (520), nitrogen (450), nutrients (404), phosphorus (397), goat (395), broiler (394), intake (348), growth performance (347), lamb (342), behavior (319), carcass (304), heritability (303), energy (303), stress (300)
1.3. Food composition	Fatty acids (2331), protein (1425), amino acids (1078), minerals (773), cholesterol (647), oxidation (619), lipids (585), conjugated linoleic acid (570), tocopherols (557), olive oil (552), fat (486), fatty acid composition (441), oxidative stability (407), fish oil (349), proximate composition (347), vitamins (330), oil (321), carbohydrate (308)
1.4. Composition and quality	Beef (1086), lipid oxidation (1003), meat (945), fish (752), pork (573), consumer (513), tenderness (417), chicken (409), aging (407), turkey (343), poultry (338), consumer behavior (303)
2. Food processing and modification	
2.1. Influence of food processing on sensory characteristics	Rheology (1605), starch (1243), physicochemical properties (936), microstructure (900), functional properties (837), dietary fiber (812), stability (797), viscosity (740), gelatin (704), emulsion (689), chitosan (655), polysaccharide (606), whey protein (562), pectin (544), rheological properties (515), physical properties (491), heat treatment (476), extrusion (468), mechanical properties (451), structure (441), cyclodextrin (426), nmr (406), edible film (365), particle size (363), solubility (363), DSC (differential scanning calorimetery) (362), ultrafiltration (347), Î²-lactoglobulin (343), encapsulation (325), resistant starch (314), gelatinization (309)
2.2. Food processing or treatment methods	Modeling (1139), drying (821), processing (761), kinetics (696), potato (618), high pressure (584), mathematical model (551), Maillard reaction (513), ultrasound (490), cooking (463), microwave (455), acrylamide (446), polyphenol oxidase (434), freezing (432), osmotic dehydration (381), pulsed electric field (330), image analysis (326), simulation (315), inhibition (309), browning (305)
2.3. Use of beneficial microorganisms as starter cultures in food processing	Probiotics (1249), functional food (690), inulin (546), lactobacillus (546), yogurt (534), spray drying (473), prebiotics (405), microencapsulation (370), glucose (369), sucrose (326)
2.4. Food modifications during processing and how to determine them	Response surface methodology (1231), extraction (850), optimization (847), enzymes (574), purification (487), characterization (477), lipase (379), enzymatic hydrolysis (350), hydrolysis (348), peptide (303)
3. Food security	
3.1. Pathogenic microorganisms in milk and dairy products	Milk (1642), dairy cow (1319), *Listeria monocytogenes* (1157), salmonella (820), cheese (797), *Escherichia coli* (571), antimicrobial (569), PCR (568), real-time PCR (508), *Staphylococcus aureus* (491), detection (399), bacteria (393), *Escherichia coli* o157:h7 (389), mastitis (379), bacteriocin (376), milk production (333), high hydrostatic pressure (329), dairy cattle (324), dairy products (322), dairy (313), biofilm (308), egg (308), foodborne pathogen (305)
3.2. Mycotoxins in cereals and methods of detection of mycotoxins in food	HPLC (high-performance liquid chromatography) (1712), wheat (1079), mycotoxins (847), rice (838), yield (822), soybean (695), bread (661), aflatoxins (655), ochratoxin A (639), maize (597), barley (575), ELISA (530), cereals (514), germination (490), coffee (425), LC-MS/MS (406), isoflavones (401), validation (378), HPLC (345), residues (341), corn (326), deoxynivalenol (310), allergy (301)
3.3. Other important food contaminants in food safety (heavy metals, pesticides)	Safety (2055), composition (1178), analysis (1032), fruits (913), vegetables (757), heavy metals (543), risk assessment (507), pesticides (469), contamination (316)
3.4. Antimicrobial agents used in food. Methods of determining these antimicrobial agents	Essential oil (1303), GC-MS (gas chromatography-mass spectrometry) (1054), chemical composition (868), antimicrobial activity (804), antibacterial activity (393), additive (382)
4. Food preservation and shelf life	
4.1. Increased shelf life in plant foods during preservation. Methods and modifications of quality	Quality (2272), color (1930), storage (1787), texture (1772), shelf life (1491), sensory evaluation (1016), packaging (720), apple (699), sensory (682), ripening (606), proteolysis (570), irradiation (530), sensory quality (480), modified atmosphere packaging (475), postharvest (471), firmness (445), strawberry (413), ethylene (374), sensory properties (306)
4.2. Fermentation process in wine and beer as a preservation method	Fermentation (1474), lactic acid bacteria (1344), wine (1260), volatile compounds (1022), sensory analysis (896), yeast (774), flavor (701), MS (540), GC (523), aroma (514), biogenic amines (483), *Saccharomyces cerevisiae* (482), ethanol (476), grape (466), volatiles (437), beer (424), identification (347), *Vitis vinifera* (334)
4.3. Parameters that influence food preservation	Temperature (987), pH (766), sugar (693), organic acids (536), water activity (491), salt (343), moisture content (329), water (304)
4.4. Statistical analysis methods	PCA (867), honey (556), artificial neural network (463), cultivars (405), chemometrics (393), multivariate analysis (314), cluster analysis (311)
5. Antioxidants in food	
5.1. Antioxidants and their effects	Antioxidant (4227), antioxidant activity (3974), polyphenols (1915), anthocyanins (1454), phenolics (1419), flavonoids (1327), phenolic compounds (1317), oxidative stress (1144), antioxidant capacity (912), apoptosis (816), DPPH (575), lipid peroxidation (544), total phenolics (459), phenolic acids (397), cytotoxicity (387), reactive oxygen species (363), catechin (359), tea (348), tannin (335), bioactive compounds (333), phytochemicals (331), green tea (329), quercetin (327), antioxidant enzymes (320), anti-inflammatory (313), total phenolic content (312), red wine (311), free radical (303)
5.2. Plant antioxidants	Carotenoids (942), ascorbic acid (889), tomato (641), vitamin C (580), Î²-carotene (484), vitamin E (430), lycopene (374), chlorophyll (305)