Research and Analysis on the Field of Food Additive by Knowledge Graph Construction

Xinyue Zhang¹, Lu Yang¹²*, Limin Zheng¹² and Guodong Cheng¹

¹College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China
²Beijing Laboratory of Food Quality and Safety, China Agricultural University, Beijing, 100083, China
*Corresponding author’s e-mail: yanglumail@263.net

Abstract. The purpose of this study is to make people have a more intuitive understanding of relevant research in the field of food additives. In this paper, the bibliometrics and network analysis methods are used to study the literatures in the field of food additive research, which are included in the core collection database of web of science. With the help of citespace 5.2.R2 software, this paper analyzes and visualizes the age, institutions, research hotspots and research trends. In addition, it also explores the hot research direction of the field through the node centrality, prominence words and word frequency evolution.

1. Introduction
According to the provisions of the Food Sanitation Law of China, food additives are artificial synthetic or natural substances added to foods to improve the color, aroma, taste and other qualities of foods, as well as for the purpose of antiseptic and processing techniques. [1] Exploring the direction of food additive development has become a difficult issue. The burst terms detection algorithm is used to study words whose rate of word growth is different from normal. These words can reveal the hotspots and trends in the field. Keywords centrality can reflect the importance of the research direction represented by these words and can also explain the development opportunities of the research field. In order to provide reference for the research of food additives, this paper uses scientific metrology and network analysis methods to analyse the literature on food additives published in the web of science database from 1998 to 2018 by drawing knowledge graphs.

2. Data sources and research methods

2.1 Data sources
In order to present the trend of food additive research more objectively and effectively, the data used in this paper is derived from the high-level literature included in the web of science core collection database. The search term of the research data is “food additive”. The source of the literature is from 1998 to 2018. A total of 11875 valid documents are retrieved.

2.2 Research methods and tools
Scientific knowledge graphs are based on scientific knowledge. It is often used to analyze the literature of a subject area, to explore and discover the hidden phenomena and laws of the discipline.
Citespace 5.2.R2, developed by Dr. Chen from Drexel University in the United States\cite{2}, is a visualization software based on scientometrics and data visualization\cite{3}. Its main analytical principles are based on co-word analysis techniques, co-citation analysis techniques, and TF-IDF statistical methods.

3. Visual analysis of food additive research

3.1 Age analysis
The year distribution analysis of the literature can reflect the development trend and development speed of the food additive field. The number of documents in each year is compiled and plotted. By observing figure 1, it can be seen that from 1998 to 2006, the number of documents was small, with an average of less than 400 articles. Between 2007 and 2018, the number of documents had doubled and entered a period of rapid rise, indicating that food additives had gradually become a hot topic of research.

![Figure 1. Distribution of literature in the field of food additives.](image1)

3.2 Institutional analysis
The core academic institutions in the field of food additives can be analyzed by statistically publishing the relevant research institutions.

![Figure 2. Distribution of research institutions.](image2)

![Figure 3. Keywords distribution.](image3)

Through the analysis of research institutions, the key research institutions are related research institutes and major universities in the United States, but the cooperation of these institutions is not close enough, as shown in figure 2.

4. Research hotspot and trend analysis

4.1 Research hotspot analysis
The keywords are the refinement of the core content of the literature. Keywords with high frequency
and high centrality can be used to describe hot issues in the research field. The 50 objects with the highest frequency of occurrence are selected as the threshold, the node type is set to “keyword” and the network clipping method is the MST minimum tree method, as figure 3.

4.2 Intermediate centrality computing

The research perspective of social network analysis is usually divided into two types. The first one is the micro perspective. The nodes and edges in the network are used as research objects to analyze the relationship between these points or the relationship between points and edges. The second angle is the macroscopic perspective, taking the entire social network structure composed of nodes as the research object[4].

Node centrality can be used to measure the importance of nodes in the network, including degree centrality, local centrality, tight centrality, and intermediate centrality. The mediation centrality of a node calculates the ratio of the number of paths passing through the node to the total number of the shortest paths of any two points in the network. The large number indicates that the node is in a critical position in the network, and most of the information needs to pass through the node, which also shows that the node is more important in the whole network.

According to the definition of intermediary centrality, the calculation formula is as shown in equation (4.1):

\[ C(u) = \sum_{i \in V} \sum_{j \neq i \in V} \frac{\delta_{ij}(u)}{\delta_{ij}} \]  

\( \delta_{ij} \) indicates the number of shortest paths between two nodes i and j, \( \delta_{ij}(u) \) indicates the shortest path of the two nodes through node u. In an unprivileged network, the path length refers to the number of edges that pass. In a right network, the path length refers to the sum of the weights of the edges. Table 2 shows the top 50 high frequency keywords selected by the food additive keyword co-occurrence network map. At the same time, the calculation shows that their median centrality is greater than 0.1.

The key words in the table 1 represent research hotspots in the field of food additives, such as research category hotspots and research experiment methods.

Researchers always focus on two types of additives: preservatives and antioxidants. Preservatives and antioxidants can prevent food spoilage to a certain extent and prolong food shelf life, which is very important for food production and preservation[5]. Phenolic compounds are common antioxidants. In recent years, natural preservatives have become a research hotspot because of their unparalleled
advantages.

Due to the particularity of food additives, researchers must conduct long-term animal experiments to explore their different effects on organisms at different doses. Rats and humans are mammals, having organs and genes similar to humans and having small mass and fast reproductive cycle, so they are the first choice for experimental materials. As a food additive, titanium dioxide nanoparticles have been widely used in various products which are usually simultaneously high in sugar consumption. Chen, Zhang conducted an animal experiment to study the effects of oral titanium dioxide nanoparticles on glucose uptake and metabolism in rats of different masses and growth days \[6\]. The double-blind test means that during the test, nobody knows the group to which the test subject belongs (experimental group or control group), aiming at eliminating subjective subjects that may appear in the consciousness of the experimenters and the participants. Kuwano, Tetsuya et al. conducted a double-blind experiment to explore that the dietary supplement GDL has anti-inflammatory effects on the skin and can improve skin diseases caused by seasonal changes\[7\]. In addition, in vitro experiments are also indispensable experimental links due to fewer interference factors.

4.3 Burst terms detection

A burst term is defined as the sudden increase in the relative growth rate of a word (the instantaneous frequency at that moment relative to its own average frequency) over a short period of time. In addition, the frequency of the word changes is not smooth, but has a greater degree of migration.

According to the definition of the burst terms, the formula is calculated as equation (4.2):

$$p_i = \frac{N}{M} \left( p_i \leq 1 \right)$$

Where N is the number of documents in the target domain data that contain the target topic, and M is the number of documents in the target domain. s is called the scale parameter and reflects the severity of the burst state, which is used to indicate the degree of state resolution of the probability machine. Express the sequence of state flow occurrences as:

$$q = (q_{i1}, q_{i2}, ..., q_{in})$$

Where \( q_{in} \) indicates that the status of the nth batch of data is \( q_i \). Then, at the time of the kth batch of data, the loss function of the probability machine still in state q can be expressed as:

$$\sigma(i, r_k, d_k) = -\ln\left( d_k \right) p^r_k (1 - p_i) d_k - r_k$$

From the above equation (4.4), the batch of keywords in the text stream under state \( q_i \) obey the binomial distribution with probability p. Then the burst weight is expressed as:

$$\text{Strength} = \sum_{k=1}^{k_2} (\sigma(0, r_k, d_k) - \sigma(1, r_k, d_k))$$

The burst terms detection algorithm converts the problem into a cost-summation problem, indicating a state that is difficult to describe.

![Figure 4. Schematic diagram of high burst value keywords after 2010.](image-url)
The figure 4 is a high-burst keyword after 2010, which can be used to analyze the research frontiers in the field of food additives, including analytical methods in the field of analytical chemistry and substances that can be used as additives.

High performance liquid chromatography is a fast separation method with high accuracy and wide separation range. It is less destructive to the structure of the compound and is suitable for the separation of organic molecules and biomolecules. Mass spectrometry has unparalleled sensitivity to other analytical methods. It is very accurate for the structural analysis of unknown compounds, and the requirements for the corresponding standard samples are relatively low. Liquid chromatography-tandem mass spectrometry combines powerful separation analysis capabilities with sensitive identification and structural resolution capabilities to provide researchers with reliable, accurate relative molecular mass and structural information. Therefore, it has become one of the most important methods of separation and identification that researchers are now paying attention to and plays an important role in the field of analytical chemistry.

Chitosan, also known as chitosan, is obtained by deacetylation of chitin, which is widely found in nature. Chitosan and its derivatives have good antibacterial activity and can inhibit the growth and reproduction of some fungi, bacteria, and viruses. In recent years, many researchers have proposed that chitosan can achieve antibacterial purposes by inducing pathogenesis-related proteins, accumulating secondary metabolites and signaling. As a high-efficiency, wide-ranging, high-molecular substance with excellent properties such as bio-functionality and compatibility, blood compatibility, safety and microbial degradability, chitosan has attracted wide attention of researchers and becomes a research hotspot today.

4.4 Word frequency evolution analysis
The frequency of occurrence of keywords in the literature is one of the key factors to measure the importance of nodes. The analysis of the frequency of occurrence of keywords with the year can be used to learn the development trend and heat change of the research direction represented by the keywords. If the frequency and centrality of the keyword increase with time, then the keyword represents a research direction that is hot and has good development potential, and vice versa.

All documents are grouped by year, and the word frequency and centrality of each selected keyword are calculated. After constructing the word frequency evolution table and the central evolution table of the keywords, and drawing them into graphs, you can see the evolution trend of the keywords. The figure is a plot of the word frequency evolution trend of selected keywords from 1998 to 2018.

[Figure 5. Keywords frequency evolution diagram.]

As you can see from figure 5, research on preservatives and antioxidants has been hot, and a large number of research scholars have devoted themselves to it and obtained excellent research results. Food migration refers to the potential ability of food to migrate from packaging to food and the
presence or absence of migrating substances. In recent years, the amount of papers in this area has shown a downward trend.

4.5 Research trend analysis
By analyzing the keywords in the field of food additives, we can grasp the research hotspots and development directions in this field. Select the Pathfinder algorithm to tailor the knowledge network for each time slice and the merged network, and select the Time zone view time zone view display method during visualization. Set the minimum visible cluster size to 9 and make appropriate debugging. By analyzing the keyword changes, we can grasp the evolution trend of food additive research hotspots.

From 1998 to 2004, the keywords in the field of food additives were mainly “food additives”, “in vitro studies”, “rats”, “aspartame”, “high performance liquid chromatography”, "Acid", "phenolic compounds" and so on. Aspartame may cause brain damage as well as lymphoma, so its use restrictions were not completely abolished until 1996 and caused a wave of research. After the 1990s, the rapid development of bioengineering and life sciences has brought more and more new topics of separation, purification and preparation for high performance liquid chromatography. Before a new food additive is marketed, applicants must perform high-dose long-term animal experiments to demonstrate that human consumption does not cause harmful effects[8]. This requires an in-vivo experiment, which is also the reason for the high frequency of "mouse" keywords.

From 2005 to 2008, the keywords in the field of food additives were “double blind”, “diet”, “water”, “children”, “hyperactivity disorder”, “citric acid” and “preservative” and so on. The impact of food additives on children has become a research hotspot at this stage. For example, due to the large intake of aspartame in beverages, it has been reported that aspartame is one of the causes of ADHD in children. However, controlled trials have shown that the intake of aspartame does not cause significant adverse behavior or affects cognitive function in children[9]. Since many experiments at this stage are related to the human body, double-blind experiments have been taken as an experimental method that achieves a high degree of rigor and is suitable for use on the human body.

From 2009 to 2013, there were a large number of keywords in the field of food additives, such as “identification”, “protein”, “Listeria monocytogenes”, “metabolism”, “oxidative stress”, “Genotoxicity, "cells", etc. It can be seen from the keywords that at this stage, researchers pay more attention to research at the cellular level, study the stress response of cells and the toxicity of substances to cells. For example, Picone, Gianfranco evaluated the effect of carvacrol on the E. coli 555 metabolome by using H-1-NMR spectroscopy[10]. Jayasena, Dinesh D. et al. discuss that plant-derived essential oils have significant antimicrobial efficacy against spoilage and pathogenic microorganisms in meat and meat products[11].

There are not many new and frequently occurring keywords from 2014 to 2018, including “safety”, “antioxidant activity” and “nanoparticles”. Because the time span is not big enough, the related research is not enough, which also shows that the food additive field has entered a new stage of tackling. It can be seen from the keywords that field research is paying more and more attention to nature and health and the granularity is getting finer and finer. For example, Moreira, Manuela M. et al. have found that as a natural resource for biologically active compounds, Portuguese cane waste produced during vine pruning can be used as a natural source of antioxidants readily available in the food or pharmaceutical industry[12].

5. Conclusion
In this paper, the visual analysis software citespace 5.2.R2 is used to create knowledge graphs. The age, organization, research hotspots and research trends in the field of food additives are analyzed by using the mediation centrality, the burst word detection and the word frequency evolution. Due to technical feasibility constraints, the scope of the study is limited to the relevant literature in the web of science core collection database. The threshold of the keywords is also limited, which may exclude some important information.
At present, there are various researches in the field of food additives, including research on antioxidants and preservatives, research on the use of additives in children and patients, research on the method of determination of additive content, research on additive effects and toxicity and consumer perception of additives. In the future, the utility of natural additives and additives at the cellular level will be the focus of research.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Project No. 2017YFC1601800) and National System for Layer Production Technology (Project No. CARS-40-K27).

References

[1] Chen Y. (2016) Overview of Food Additives. China Food Safety Magazine, 3:76-77.
[2] Chen C. (2009) CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the Association for Information Science & Technology, 57(3):359-377.
[3] Li J, Chen C. (2016) Citesspace: technology text mining and visualization. Capital University of Economics and Business Press,Beijing.
[4] Dorogovtsev, Sergei N., José FF Mendes. (2013) Evolution of networks: from biological nets to the internet and www. Physics Today, 57:10-15.
[5] Zhang H, Jia J, Wang W, Chen X. (2016) Research Progress and Development Trend of Domestic Food Additives. Journal of Food Science and Biotechnology, 35(3):225-233.
[6] Chen Z, Wang Y, Wang X, et al. (2018) Effect of Titanium Dioxide Nanoparticles On Glucose Homeostasis After Oral Administration. JOURNAL OF APPLIED TOXICOLOGY, 38 (6): 810-823.
[7] Kuwano T, Kawano S, Kagawa D, et al. (2018) Dietary Intake of Glucono-Delta-Lactone Attenuates Skin Inflammation and Contributes to Maintaining Skin Condition. FOOD & FUNCTION, 9 (3): 1524-1531.
[8] Picone G, Laghi L, Gardini F, et al. (2013) Evaluation of the Effect of Carvacrol On the Escherichia Coli 555 Metabolome by Using H-1-Nmr Spectroscopy. Food Chemistry, 141 (4): 4367-4374.
[9] Guo X, Zhang H, Huang J. (2007) Overview of domestic and foreign food additive management regulations and standards. Transactions of the Chinese Society of Agricultural Engineering, 6:12-18.
[10] Ernest T B, Elder D P, Martini L G, et al. (2007) Developing paediatric medicines: identifying the needs and recognizing the challenges. Journal of Pharmacy& Pharmacology, 59(8):1043-1058.
[11] Teng Y. (2017) The Analysis of Hotspot and Detection of Research Front of Supply Chain Management Domestic and Abroad Based on Knowledge Map. http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201702&filename=1017139047.nh&uid=WEEvREdxOWJmbC9oM1NjYkZCbDdrdTViZINLUTg1ZlZTb1dezJjaHdsNlY=5RlyZ0H6jyaaoen3RxVUd8dfiHi7XMXMD07mtKT6mSmEvTuk1112gFA!!&v=MDY0NjJDVVJMT2ZaT1J2Rn16bFdyN1BWRjJ2RJN0Y5EixSkViUEISOGVYMUx1e5ITN0RoMVZcVRYV00xRnl=.
[12] Moreira M M, Fatima Barroso M, Porto J V, et al. (2018) Potential of Portuguese Vine Shoot Wastes as Natural Resources of Bioactive Compounds. Science of the Total Environment, 634: 831-842.