frontiers | Frontiers in Sustainable Food Systems

TYPE Methods
PUBLISHED 12 November 2024
DOI 10.3389/fsufs.2024.1491255

OPEN ACCESS

EDITED BY
Anindya Chanda,
Mycologics LLC, United States

REVIEWED BY
Patricia D. Millner,
Agricultural Research Service (USDA),
United States
Rahat Wadhwa Desai,
Syngenta, Canada

*CORRESPONDENCE
Jiayin Pei
✉ 2980559861@qq.com

RECEIVED 06 September 2024
ACCEPTED 23 October 2024
PUBLISHED 12 November 2024

CITATION
Song J and Pei J (2024) Mining text for causality: a new perspective on food safety crisis management.
Front. Sustain. Food Syst. 8:1491255.
doi: 10.3389/fsufs.2024.1491255

Mining text for causality: a new perspective on food safety crisis management

Jinyi Song and Jiayin Pei*

School of Business, Jiangnan University, Wuxi, China

The aim of the present study was to quantitatively analyze the importance of each risk factor in a food safety event, so as to fully elucidate the correlation between different risk factors and provide a reference for food safety governance. Text mining and complex network analysis methods were utilized to explore the causal mechanism of food safety incidents. By performing text mining on food safety event news reports, 15 major risk factors were identified based on high-frequency words. A causal network for food safety accidents was then constructed using strong association rules among these factors. Through network centrality analysis, the five core factors of food safety incidents and their associated sets were clarified. Based on text mining of 6,282 cases of food safety incidents reported by online media, 168 keywords related to food risk factors were extracted and further categorized into 15 types of food safety risk factors. Network analysis results revealed that microbial infection emerged as the most critical risk factor, with its associated sets including biotoxins and parasites, counterfeiting or fraud, processing process issues, and non-compliance with quality indicators.

KEYWORDS

food safety, risk factors, big data, text mining, complex networks

1 Introduction

With the rapid economic development and the continuous improvement in living standards, food safety has garnered increasing attention. This issue is crucial for public health and well-being, as the presence of food safety risk factors directly affects the efficacy of safety measures. These risk factors encompass various hazards and risks that may arise at different stages of the food supply chain, including production, processing, transportation, storage, and distribution. For instance, in the processing stage, to enhance the taste, color, or extend the shelf life of food products, businesses often use excessive additives which is the use of food additives beyond safety limits by national food safety standards. Some businesses go as far as using substandard raw materials in order to cut costs. Despite China's significant efforts and achievements in food safety governance, the ongoing presence of food safety risk factors remains a significant challenge. This issue is exacerbated by China's vast territory, diverse food types, and complex food supply chain (Yuntao et al., 2022).

Numerous studies have demonstrated that media participation in food safety governance, or broader social co-governance, is a crucial component of the social regulatory system (Yongqin et al., 2014; Kang et al., 2017; Rutsaert et al., 2013; Dillaway et al., 2011). Food safety incidents are frequently first highlighted by the media. As a result, using media-reported incidents as a data source has become a widely adopted approach in the field of sociology (Holtkamp et al., 2014). In the present study, food safety incidents exposed by domestic mainstream media were analyzed, extracting key food safety risk factors from a large number of news reports. The distribution characteristics of these risk factors were then examined, and the relationships between them were explored. The aim of the study was to provide a reference for decision-making in the prevention of food safety incidents.

Scholars have extensively researched the characteristics of food safety, risk factors, and supply chain links using data collected from media reports on food safety-related events. For instance, Ming et al. (2014) analyzed supermarket food safety incidents and summarized the causes at critical regulatory points based on food safety laws. Lan et al. (2013) examined the four elements of food safety risk—food type, the supply chain link where the risk occurs, the underlying cause of the risk, and the responsible entity. He conducted both qualitative and statistical descriptive analyses of 3,484 food safety events from 2001 to 2011. This research identified four primary sources of food safety risks in China. Using 9,314 food safety events from 2010 to 2019 as research samples, Hongxia (2021) employed Natural Language Processing and Information Retrieval (NLPIR), a big data semantic intelligence analysis platform, to extract high-frequency words, distill food safety risk factors, and analyze the causes and potential consequences of each factor. Mu et al. (2021) examined the resilience of food supply chains by analyzing the root causes of food safety incidents and explored strategies for building resilient food supply chains within the context of food safety. Goldberg et al. (2022) proposes a new approach that utilizes text mining and supervised machine learning to identify terms related to dangerous foods from text concerning food safety regulations in online media. Regarding accident causation mechanisms, Niu et al. (2021) explores the key factors driving food fraud and their interrelationships in China's emerging food market, employing a social co-governance perspective and a Decision Making Trial and Evaluation Laboratory (DEMATEL) based analytic network process. Long Peng et al. (2023) examines the causes and diagnoses of 420 accidents by constructing a Bayesian network (BN) model, combining Zagan's theory with Grey-DEMATEL and Interpretive Structural Modeling (ISM) techniques. Social network analysis has also been extensively used in accident causation studies. For instance, Huang et al. (2023) employ social network analysis (SNA) to examine the multidimensional structural characteristics and risk factors of nighttime tourism safety accidents, utilizing a dataset of 8,787 incidents from major nighttime tourism cities in China. Qiu et al. (2021) constructed a coal mine accident causation network through text mining and association rules, analyzing network centrality and paths to understand the relationships and influences among various factors. Xiumei et al. (2024) applied similar methods to general aviation accidents, proposing a causation model. Zhong et al. (2020) integrated deep learning and text mining techniques to identify and categorize accident hazards. Subsequently, the relationships between these hazards were analyzed using word co-occurrence networks. Identifying risk factors in food safety events is crucial for risk analysis and assessment. These factors provide critical reference points for developing food safety governance measures, ultimately enhancing overall food safety.

Research on applying text mining technology to analyze food safety incidents reported by the media and identify associated risk factors has advanced significantly. Yet, earlier studies have mainly concentrated on identifying risk factors without quantifying their importance or elucidating the correlations between different factors. Building on existing research, news reports on food safety incidents from 2019 to 2023 were taken as a sample in the present study. Firstly, text mining technology was applied to perform word frequency analysis and extract high-frequency words, mapping them to existing food safety risk factors. Then, a causation network for food safety incidents was constructed to analyze these relationships. Finally, targeted countermeasures were proposed based on the established core causal model, offering insights and lessons for improving food safety risk management.

2 Materials and methods

2.1 Food safety incidents

The 2024 amendment to the Food Safety Law of the People's Republic of China defines food safety as food that is non-toxic, harmless, meets nutritional requirements, and poses no health hazards. Nonetheless, a consistent definition of food safety incidents remains elusive (Hong, 2011). Qing-guang et al. (2016) categorized these incidents narrowly as events directly linked to food that threaten human health, such as food poisoning and contamination. Broadly, he included incidents that also impact public opinion, potentially affecting consumer perceptions. Many media-reported incidents did not meet the legal criteria of food safety issues and thus were not considered hazardous to health. The focus of the present study was on food safety incidents related to food or food-contact materials containing physical, chemical, biological, or other hazards that could potentially harm human health.

2.2 Data set

To gather valuable information on food safety events, Word2Vec was first utilized to identify texts related to "food safety." The resulting keywords were then used to crawl food safety news reports from 2019 to 2023 across major domestic online media platforms, including WeChat, Weibo, and Today's Headlines, creating a food safety news database for the present study. The data in this database underwent text cleaning, which included removing duplicates and conducting additional manual cleaning to eliminate obviously irrelevant content. Subsequently, text categorization techniques were employed to identify food safety incidents within the database, resulting in 6,282 valid entries.

The following steps were performed to obtain key characteristics of food safety issues related to food safety incidents: Firstly, Python's third-party library Jieba 0.42.1[1](Van Rossum and Drake, 2009) was utilized to process the valid data. Secondly, words with no research significance were filtered out using the deactivated word lists from Harbin Institute of Technology and Baidu. Simultaneously, a food safety feature dictionary was constructed based on relevant national food safety standards to enhance the accuracy of word separation. Thirdly, the results were manually reviewed to select words related to food safety with a frequency >30. Ultimately, 692 high-frequency words were obtained.

2.3 Research framework

As illustrated in Figure 1, the research process involved the following steps. First, text mining technology was employed to extract high-frequency words from the text of news reports on food safety

1 Retrieved from https://github.com/fxsjy/jieba

events. Based on this, the main risk factors of the events were identified and categorized according to the key stages of the food supply chain. Next, the Apriori algorithm was used to extract strong association rules between the risk factors, laying the groundwork for constructing a causal network. Finally, a causal network of food safety events was constructed, and the core risk factors along with their closely related cause sets were identified through network centrality analysis.

3 Results and analysis

3.1 Food safety risk factor identification

Through manual screening, 168 high-frequency words were extracted from the initial set of 692 words related to food safety risk factors. Based on existing studies on the main risk factors involved in food safety incidents, these high-frequency words were classified into 15 categories using text clustering (Hongxia, 2021).

As shown in Table 1, food safety risk factors span the major stages of the food supply chain, involving production, processing, transportation, storage, and distribution. These stages are closely interconnected, and any issues at one stage can compromise the safety of the final product, potentially leading to food safety incidents.

3.2 Characterization of food safety risk factors and causal network construction

3.2.1 Characteristics of the distribution of food safety risk factors

Based on the extracted high-frequency words related to risk factors in food safety events, the risk factors associated with each incident in the food safety incident database were identified. The distribution of the main food safety risk factors is illustrated in Figure 2.

As shown, among the food safety risk factors, "raw material failure" accounted for the highest proportion, comprising 15.32% of the total. This was followed closely by "production date problems," which represented 12.40%. "Packaging failures" also had a significant share at 11.52%. The other notable risk factors included "counterfeiting or fraud," "unlicensed or unregistered manufacturing," "excessive use of additives" "residues of agricultural and veterinary drugs," and "substandard sanitary conditions."

3.2.2 Construction and analysis of a network of key risk factors for food safety

The distribution of food safety risk factors highlights the major contributors to food safety incidents, revealing the predominant issues involved in each case. Nevertheless, the occurrence of food safety incidents is rarely caused by a single risk factor. Instead, these incidents arise from a complex causation system where multiple risk factors interact. In this system, one or more risk factors can uncontrollably influence closely related factors, either directly or indirectly. This interconnected influence can lead to the spread of risk within the system and ultimately result in the incident.

Association rule analysis can uncover relationships between multiple items, with the Apriori algorithm being a classic method for mining such rules. In the present study, the Apriori algorithm was applied to identify strong association rules between the risk factors of food safety events, which were then used to construct a causal network for these events (Qiu et al., 2021). According to the requirements of the Apriori algorithm, the minimum support threshold was set to 0.002, the minimum confidence threshold to 0.1, the maximum number of antecedent terms to 3, and the minimum lift threshold to 1. A total of

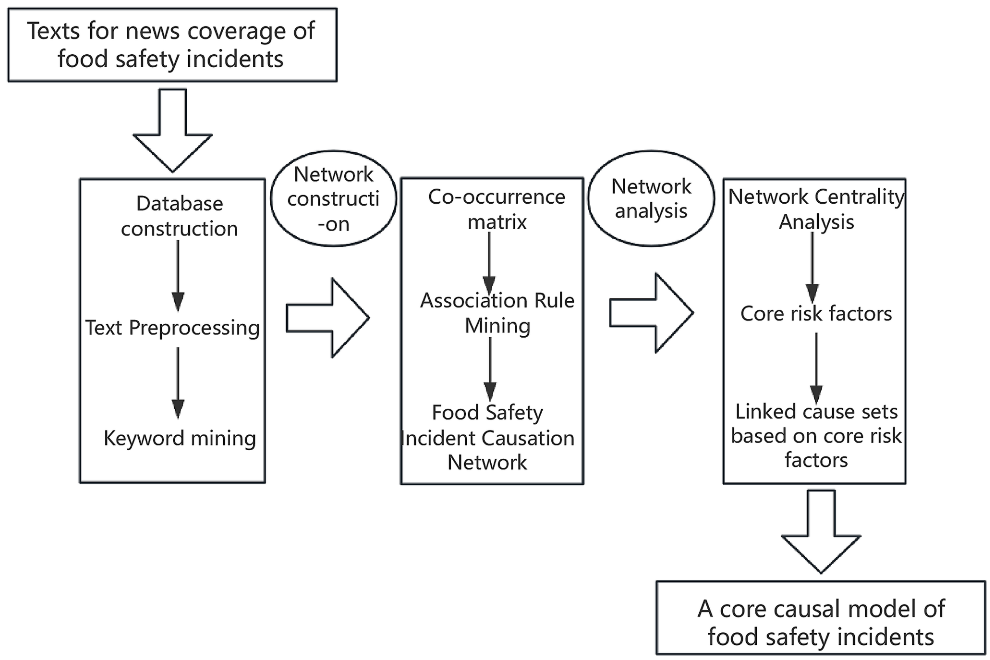

FIGURE 1
Process of analyzing the causal mechanism of food safety incidents.

TABLE 1 Main risk factors involved in food safety incidents.

Number	Risk factor	Selected high-frequency words
A0	Packaging failures	Packaging Box, Packaging Materials, Packing Crate, Product Packaging, Food Packaging, Packaging, Outer Packaging, Packaging Bag
A1	Production date problems	Expired, Inkjet Printer, Expiration or Due Date, Production Date, Expire, Date, Tampering, Shelf Life
A2	Excessive use of additives	Baking Powder, or Essence, Dehydroacetic Acid and Its Sodium Salt, Sodium Saccharin, Sodium Sulfonate, Pyrethroids, Sorbic Acid, Ethyl Maltol, Alum, Carmine, Benzoic Acid and Its Sodium Salt, Sorbic Acid and Its Potassium Salt, Food Additives, Cyclamate, Preservative, Additive, Dehydroacetic Acid, Nitrite, Benzoic Acid, Sulfur Dioxide
A3	Microbial contamination	Moldy, Mildew or Mold Contamination, Bacteria, Infested with Insects, Spoilage, Colony, Coliform Bacteria, Microorganism, Putrefaction or Decay, Mold or Fungus
A4	Counterfeiting or fraud	Violation, Counterfeiting, Counterfeit Alcohol, Unauthorized Use, Substandard and Counterfeit Products, Adulteration, Counterfeit and Substandard, Forgery, Selling Counterfeits, Fraud, Passing Off Inferior Goods as High-Quality, Fake and Inferior, Passing Off Fake Goods as Genuine, Infringement, Counterfeit, Intellectual Property Infringement, Impersonation, Knockoff or Shanzhai, Smuggling, Fabrication or Faking, Deception, Blending, Infringing Products, Counterfeit and Substandard Products
A5	Unlicensed or unregistered manufacturing	Without Permission or Unauthorized, Without a License, Without License or Business Permit, Operating Without a Permit or Unlicensed Business Operation, Permission, License, Business License, Business Registration Certificate, Production License, Operational Permit
A6	Physical contamination	Adulteration, Added, Foreign Object, Insect, Contaminant, Hair, Water Injection, Maggot, Incorporated
A7	Substandard sanitary conditions	Back Kitchen, Cockroach, Washing, Fly, Rat, Garbage, Garbage Bin, Fly Prevention, Cleaning, Dirty, Messy, and Poor, Disinfection
A8	Biotoxins and parasites	Toxic or Poisonous, Blue-Ringed Octopus, Puffer Fish, Poisonous Mushroom, Pufferfish, Wildlife, Accidental Ingestion, Deadly Poisonous, Poisoning, Toxin, Food Poisoning, Toxicity
A9	Residues of agricultural and veterinary drugs	Residue, Pesticide, Exceeding the Standard, Veterinary Drug, Procymidone, Clothianidin, Chlorobenzene, Chloramphenicol, Enrofloxacin
A10	Raw material failure	Soup Base, Mixed Storage, Leftovers, Gutter Oil, Discard, Raw and Auxiliary Materials, Quarantine, Raw Material, Raw Materials
A11	Processing issues	Fermentation, Brewing or Distillation, Pickling, High Temperature, Deep Frying, Handmade, Soaking, Craft, Thawing, Production Process, Simmering
A12	Harmful input Pollution	Poppy Husk, Industrial, Borax, Gold and Silver Foil, Clenbuterol, Morphine, Malachite Green, Harmful Substances, Aconite, Gold Leaf, Chemical Substances, Formaldehyde
A13	Non-conformity of Quality indicators	Product Quality, Quality Inspection, Food Quality, Quality Issue, Acid Value, Peroxide Value, Alcohol Content or Alcohol by Volume, Vitamin
A14	Improper storage	Preserve or Maintain, Store, Stock, Storage, Data Storage, Arrange, Position, Mixed Placement

512 strong association rules were mined, reflecting the close connections between the risk factors in food safety events. The top five association rules with the highest lift values are presented in Table 2.

Based on strong association rules and Pajek 5.18[2] software (Batagelj and Mrvar, 2006), a causal network for food safety events was constructed. In this network, the antecedents and consequents of all the strong association rules were represented as nodes. The relationships between antecedents and consequents were depicted as edges, with the enhancement degree of the association rules serving as the weights of these edges. The resulting food safety incident causal network is shown in Figure 3. The network contains 102 nodes and 512 edges. The connection density around a node reflects the complexity of the region where the node is located. The closer the node is to the center of the network, the greater the influence of the node on other nodes. From Figure 3, an observation can be made that among the risk factors, the raw material failure and the excessive use of additives were closer to the center of the network (Table 3).

To identify the core risk factors of food safety incidents and propose targeted prevention strategies, the importance of each risk factor was quantified. Centrality analysis can measure the significance of key causal factors within a network. As such, Pajek 5.18 software was used to analyze the network's centrality. Newman (2008) proposed three types of centrality: degree centrality, closeness centrality, and betweenness centrality, each of which measures the importance of a node. Degree centrality assumes that a node's influence increases with the number of other nodes it can directly affect. Closeness centrality assumes that a node is more important if it has a shorter average path to all other nodes in the network. Betweenness centrality assumes that a node is more significant if it has a greater ability to mediate the shortest paths between different nodes. To comprehensively reflect the importance of the event risk factor nodes, the index values for degree centrality, closeness centrality, and betweenness centrality were normalized. These normalized values were then averaged to create a composite score for each node. Ultimately, the four event

2 Retrieved from http://mrvar.fdv.uni-lj.si/

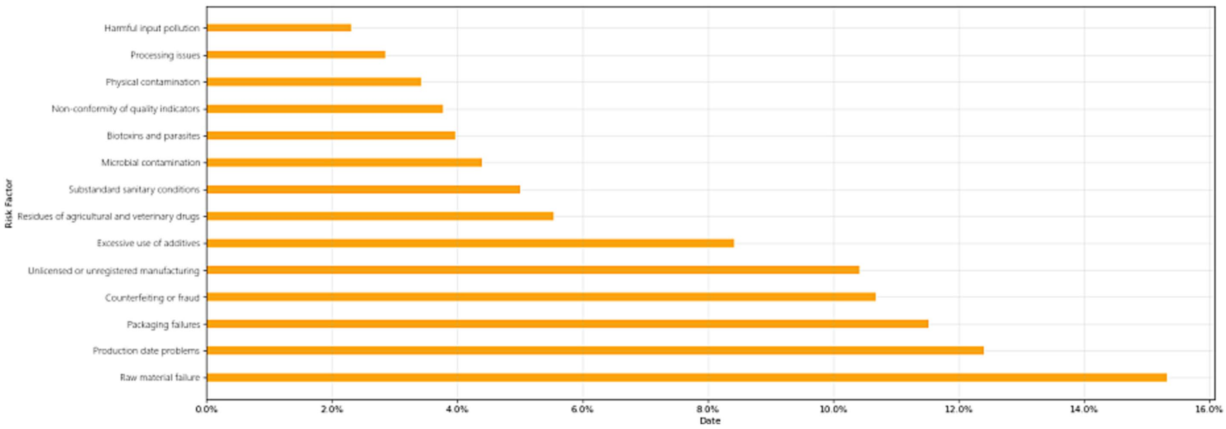

FIGURE 2
Distribution of the proportion of each risk factor.

TABLE 2 Top five association rules with the highest lift.

Antecedent	Consequent	Support (%)	Confidence (%)	Lift
A12,A6	A8	0.24	75.00	11.77875
A2,A8	A12	0.49	37.80	10.19271433
A12	A2,A8	0.49	13.30	10.19271433
A6,A8	A12	0.24	34.88	9.405130253
A9,A0	A3	0.41	61.90	8.798319328

risk factors with the highest composite scores were identified as the core causes of the incidents.

Based on the results of the composite scores, the core risk factors for food safety incidents were identified as Raw material failure, Excessive use of additives, Microbial contamination, and Packaging failures. From the results, the core risk factors derived from the network analysis were different from the statistical analysis.

Core risk factor nodes play a critical role in the event network. When a core risk factor is not controlled, the risk can gradually spread to closely related nodes. For instance, substandard food ingredients can lead to further issues such as processing problems, eventually culminating in food safety incidents. Therefore, effective prevention requires not only addressing the core risk factors directly but also implementing joint defenses for closely related cause sets. If controlling the core risk factor proves challenging, managing the associated cause sets becomes crucial to blocking the spread of risk within the network and preventing incidents.

To identify the associated cause sets for different core risk factors, Pajek 5.18 was used to perform network analysis on each core risk factor. Eigenvector centrality was employed to quantify the degree of connectivity between a node and a core node. Hence, the relevant associated cause sets were identified based on the eigenvector centrality value of each node in a single central network. For example, considering raw material failure (A10), all strong association rules containing A10 were first selected. Subsequently, the antecedents and consequents of these rules were treated as nodes in the network, with the relationships between them represented as edges, and the lift degree of the association rules used as the weight of each edge. Figure 4 illustrates a single central network where the size of each node is proportional to its eigenvector centrality value.

To isolate the influence of other core risk factors, the eigenvector centrality values of nodes representing these factors were excluded when calculating the set of associated causes for A10. Consequently, A12, A11, A8, and A5 emerged as the four nodes with the highest eigenvector centrality values, specifically 0.357333, 0.230701, 0.210607, and 0.171216, respectively. This group of nodes was referred to as the associative cause set for raw material failure. Analysis results of food safety incidents indicate that raw material failure mainly arises when merchants, driven by profit motives, recycle leftovers and gutter oils to cut costs. This pursuit of higher profits may also lead to the use of harmful substances, such as poppy husks, to attract customers and gain undue financial benefits, significantly heightening food safety risks. The same analysis was applied to other core causes, and the associated cause sets for all identified core risk factors are summarized in Table 4.

From the perspective of the food supply chain, "unqualified raw materials" and "excessive quantity and scope of additives" mainly occur during food production and processing. These problems are often driven by merchants seeking to cut costs and maximize profits. Similarly, "microbial contamination" frequently occurs in the production and processing stages, primarily due to inadequate processing capacities and insufficient sterilization measures during production.

3.2.3 Key causal modeling of food safety incidents and recommendations for governance

Based on the distribution of food safety events across key links in the supply chain (Holtkamp et al., 2014) and their alignment with risk factors, conclusions from text mining and social network analysis

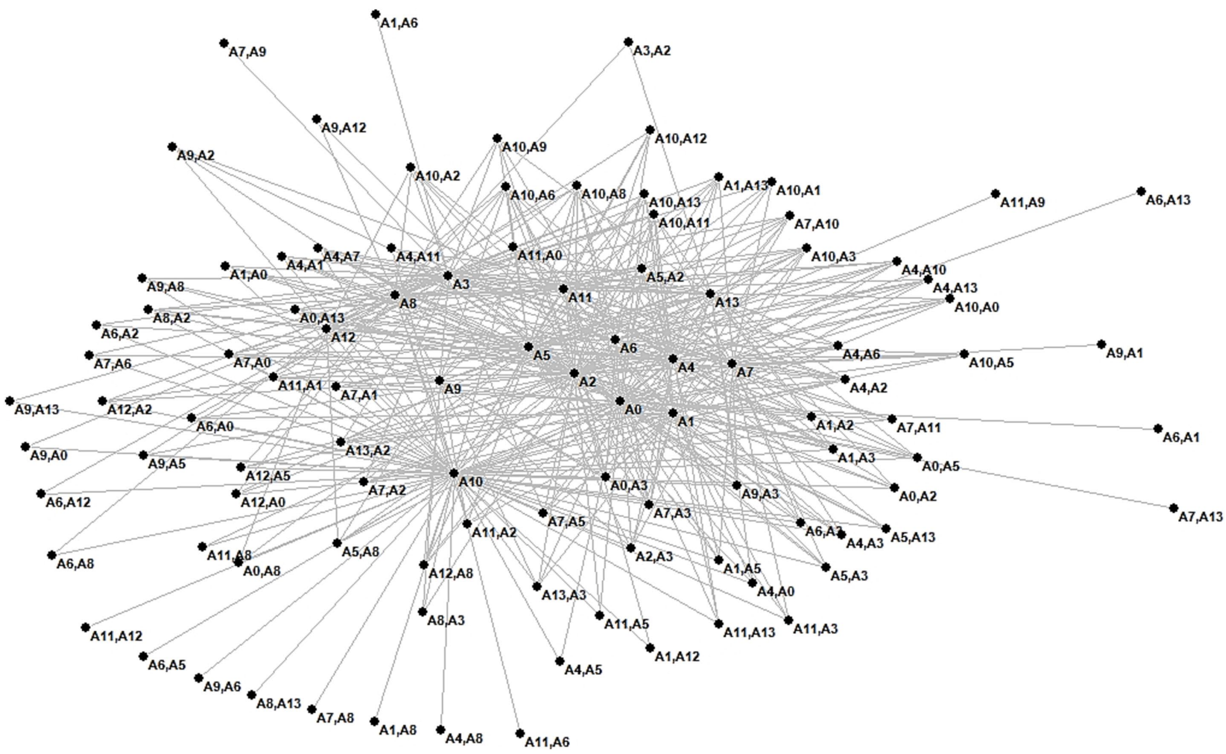

FIGURE 3
Food safety incident causation network.

TABLE 3 Core risk factors for food safety incidents.

Centrality analysis	Core risk factors for food safety incidents
Composite score	Raw material failure: 93.33943987
	Excessive use of additives: 73.7020027
	Microbial contamination: 63.2525949
	Packaging failures: 56.39036371
Degree centrality	Microbial contamination: 163.754548
	Excessive use of additives: 142.744054
	Raw material failure: 131.322699
	Biotoxins and parasites: 123.790569
Close centrality	Raw material failure: 0.759398
	Excessive use of additives: 0.701389
	Packaging failures: 0.668874
	Unlicensed or unregistered manufacturing: 0.63125
Betweenness centrality	Raw material failure: 0.354577
	Excessive use of additives: 0.171063
	Microbial contamination: 0.104025
	Packaging failures: 0.099467

were integrated to construct a core causal model of food safety events, as shown in Figure 5. The model construction process involved the following steps: First, 15 risk factors were identified through text mining of relevant keywords. Second, four core causal factors were determined using the composite scores derived from social network centrality analysis. Finally, significant connections between these core causal factors were established based on the associative relationships among the risk factors. To accurately represent the correlations between risk factors within each link and to exclude the influence of non-core factors, the study selected association rules with a single item and a lift value <2.

The core causation model for food safety incidents identified 15 risk factors (depicted as solid rectangles in Figure 5), including 4 core risk factors (shown as solid grey rectangles). These risk factors correspond to the five stages of the food supply chain: production, processing, distribution, sales, and consumption. Notably, the processing stage included 6 risk factors—Raw material failure, Excessive use of additives, Counterfeiting or fraud, Microbial contamination, Packaging failures, and Processing issues —which constituted 40% of all identified risk factors and encompassed all core risk factors. The production stage included 3 risk factors: Residues of agricultural and veterinary drugs, Harmful input pollution, and Non-conformity of quality indicators. The sales stage also had 3 risk factors: Unlicensed or unregistered manufacturing, Physical contamination, and Production date problems. The consumption stage involved 2 risk factors: Biotoxins and parasites, and Substandard sanitary conditions. Finally, the distribution stage included one key risk factor: Improper storage.

As evidenced, food safety incidents are most likely to occur during the processing stage. The relationships between core risk factors (represented by solid lines in Figure 5) indicate that the connections between various stages of the food supply chain are complex and interdependent. Specifically, the links between processing, production, and consumption are the most robust. The connection between processing and consumption primarily relates to microbial

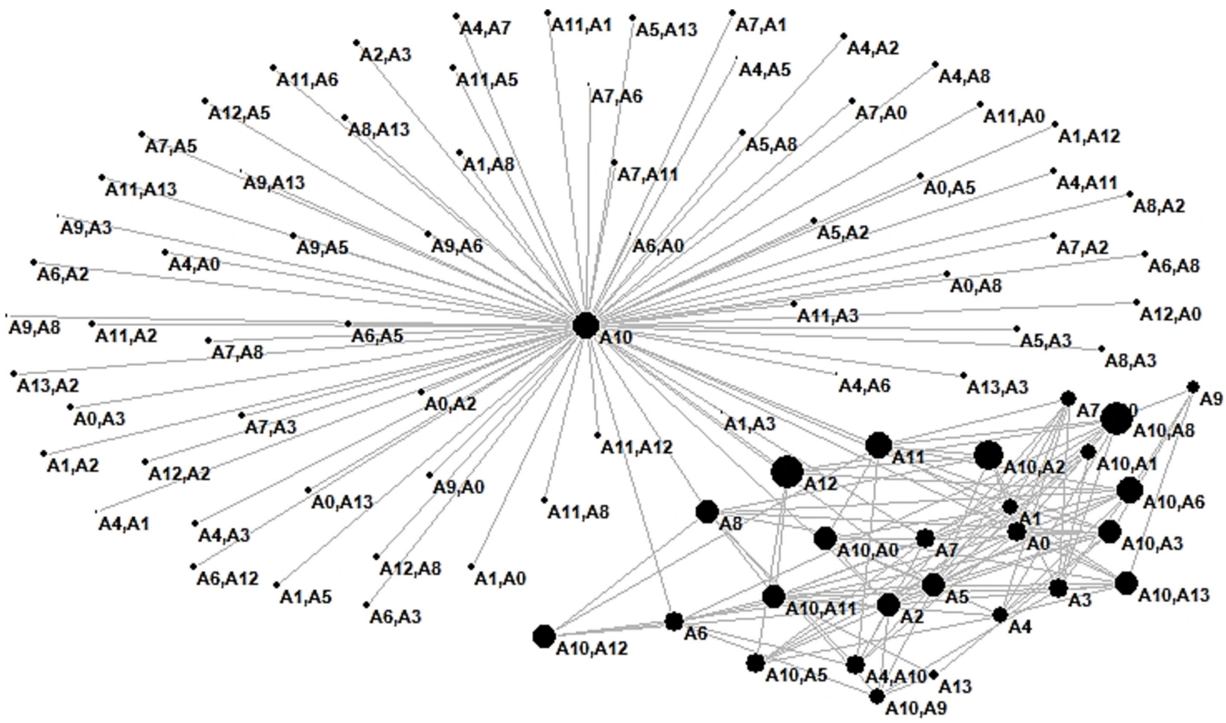

FIGURE 4
Single center network based on raw material failure (A10).

TABLE 4 Set of associated causes for all core risk factors.

Core risk factors	Associated-cause set based on core cause
Raw material failure	Harmful input pollution; Processing issues; Biotoxins and parasites; Unlicensed or unregistered manufacturing; Substandard sanitary conditions
Excessive use of additives	Harmful input pollution; Processing issues; Residues of agricultural and veterinary drugs; Unlicensed or unregistered manufacturing; Production date problems
Microbial contamination	Physical contamination; Substandard sanitary conditions; Processing issues; Residues of agricultural and veterinary drugs; Non-conformity of quality indicators
Packaging failures	Unlicensed or unregistered manufacturing; Production date problems; Processing issues; Counterfeiting or fraud; Substandard sanitary conditions

contamination, processing issues, and biotoxins and parasites. The link between processing and production is mainly associated with the quality of raw materials, hazardous input contamination, and processing issues. The connection between consumption and production primarily concerns biotoxins, parasites, hazardous input contamination, and pesticide and veterinary drug residues. Overall, the findings indicate that the food safety risks are most closely linked in the production, processing, and consumption stages, with the processing stage being identified as the most critical point of vulnerability.

4 Policy recommendations

Based on the present findings, the primary food safety risk factors include unqualified raw materials and issues related to production dates. Network analysis results further reveal the core risk factors, such as unqualified raw materials and the excessive or inappropriate use of additives. At present, food safety risks in China are still predominantly driven by human factors (Hongxia, 2021). China's food safety governance faces significant challenges, including the complexity of the food safety industry chain, the absence of a robust social integrity system, and an imperfect regulatory framework. In response to these challenges, several recommendations were proposed to enhance food safety risk governance.

Firstly, the monitoring of core risk factors should be strengthened. The analysis reveals that some risk factors are intricately interconnected within the overall food safety risk landscape. It is essential to improve the management of these core risk factors to prevent their amplification or activation of additional risks, thereby decreasing the probability of food safety incidents.

Secondly, particular emphasis should be placed on controlling the processing stage. The core causal model of food safety events reveals that all four core risk factors are concentrated in the food

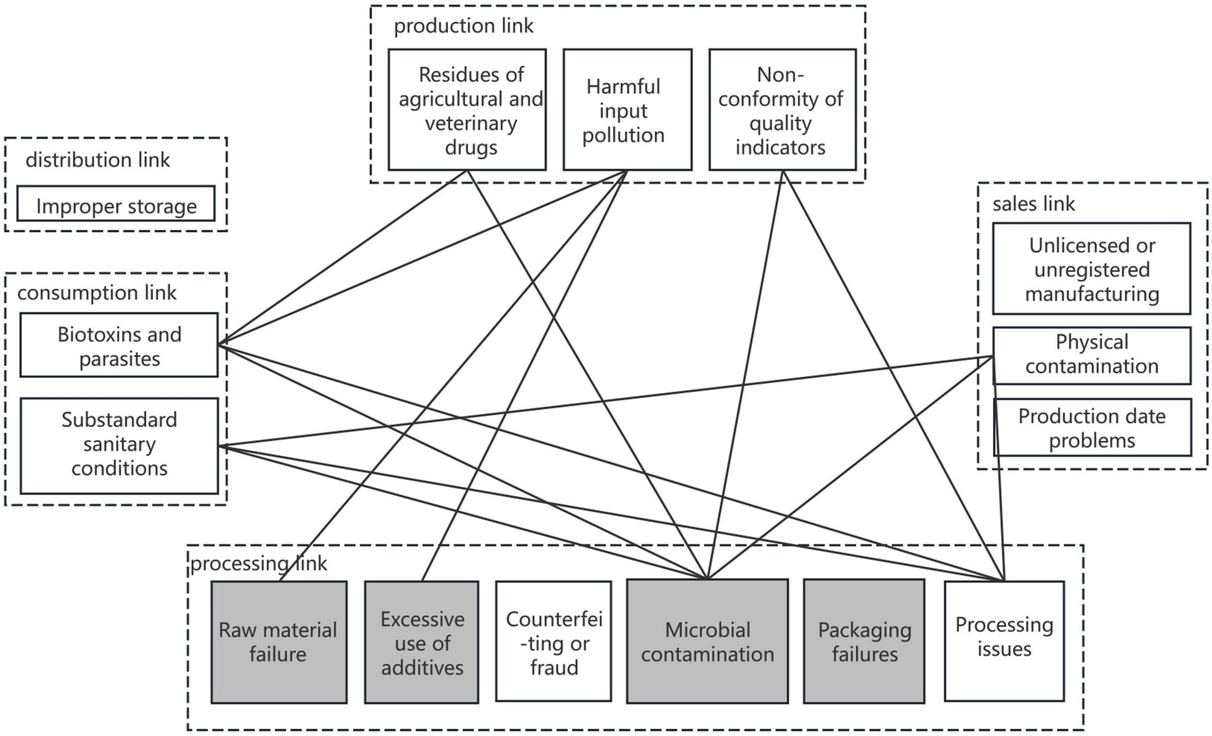

FIGURE 5
Core causation model for food safety incidents.

processing stage, which is also closely linked to risk factors in other stages. As the most critical phase in the food supply chain, processing involves significant human factors that directly influence food safety. Thus, it is advisable to enhance regulatory oversight of the processing stage by increasing transparency of the production process to consumers and improving the sampling and testing of raw materials.

In addition, it is essential to refine traceability mechanisms and clearly define responsibilities. The food supply chain involves multiple stakeholders, and establishing a comprehensive traceability system is crucial for identifying those responsible when food safety incidents occur. Simultaneously, the roles and responsibilities of all parties involved in supervision must be clearly defined. Presently, the State Quality Supervision Bureau, the Ministry of Health, and the Ministry of Agriculture all participate in food safety management (Liwei, 2024). Despite such participation, inconsistencies in management standards and ambiguous delineation of authority have impeded effective oversight. It is recommended to standardize regulatory standards, centralize management, and improve coordination among departments to ensure clear and effective regulatory responsibilities.

Finally, it is imperative to enhance relevant laws and regulations and strengthen enforcement. Food safety incidents often arise from interest-driven behaviors by stakeholders within the supply chain. To address these root causes, it is crucial to bolster penalties. Relevant authorities should improve and update legislation governing food production and processing, ensure transparent enforcement mechanisms, and effectively safeguard consumer rights. While increasing penalties, it is also crucial to build public trust in the regulatory system.

5 Conclusion

In the present study, 15 food safety risk factors were identified through text mining techniques applied to food safety incidents. These factors included Packaging failures, Production date problems, Excessive use of additives, Microbial contamination, Counterfeiting or fraud, Unlicensed or unregistered manufacturing, Physical contamination, Substandard sanitary conditions, Biotoxins and parasites, Residues of agricultural and veterinary drugs, Raw material failure, Processing issues, Harmful input pollution, Non-conformity of quality indicators and Improper storage. These factors represent the primary risks associated with food safety incidents.

To further explore the relationships among these individual risk factors, a causal network of food safety events was constructed and analyzed. Four core risk factors and their associated causal sets were identified: raw material failure, excessive use of additives, microbiological contamination, and packaging failures. This analysis offers valuable insights for the prevention and control of food safety incidents.

Further, a comprehensive analysis of the causes of food safety incidents was conducted from the perspective of the food supply chain, resulting in the development of a core causal model. This model identified the processing stage as the most susceptible to food safety incidents and highlighted its close interconnection with both the production and consumption stages. It suggests that food safety incidents often arise from the combined effects of raw material failure, microbial contamination and biotoxins and parasites.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JS: Methodology, Writing – original draft, Writing – review & editing. JP: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Social Science Foundation of China (Grant no.: 20FTQB018), Jiangsu Province Social Science Foundation (Grant no.: 20TQB001), and the Fundamental Research Funds for the Central Universities (Grant no.: 2019JDZD06).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Batagelj, V., and Mrvar, A. (2006). Pajek-program for large network analysis. *Connections* 21, 47–57.

Dillaway, R., Messer, K. D., Bernard, J. C., and Kaiser, H. M. (2011). Do consumer responses to media food safety information last? *Appl. Econ. Perspect. Policy* 33, 363–383. doi: 10.1093/aepp/ppr019

Goldberg, D. M., Khan, S., Zaman, N., Gruss, R. J., and Abrahams, A. S. (2022). Text mining approaches for postmarket food safety surveillance using online media. *Risk Anal.* 42, 1749–1768. doi: 10.1111/risa.13651

Holtkamp, N., Liu, P., and McGuire, W. (2014). Regional patterns of food safety in China: what can we learn from media data? *China Econ. Rev.* 30, 459–468. doi: 10.1016/j.chieco.2014.07.003

Hong, S. (2011). Analysis of food safety. *Food Ind.* 5, 89–91. doi: 10.1007/978-1-4614-6205-7_8

Hongxia, Z. (2021). Identification and distribution characteristics of food safety risk factors——an empirical analysis based on 9314 food safety incidents. *Contemp. Econ. Manage.* 43, 66–71. doi: 10.13253/j.cnki.ddjjgl.2021.04.009

Huang, R., Xie, C., Lai, F., Li, X., Wu, G., and Phau, I. (2023). Analysis of the characteristics and causes of night tourism accidents in China based on SNA and QAP methods. *Int. J. Environ. Res. Public Health* 20:2584. doi: 10.3390/ijerph20032584

Kang, X., Yi, L., and Xin, Z. (2017). The condition and strategy of media participating in food safety social co-governance. *Manag. Rev.* 29, 192–204. doi: 10.14120/j.cnki.cn11-5057/f.2017.05.018

Lan, L., Yu-fa, A., Chuan, G., and Yang, L. (2013). Analysis of sources of risk and regulatory strategy of Chinese food safety. *J. Food Sci. Technol.* 31, 77–82.

Liwei, P. (2024). Construction and optimization of food safety risk control system. *China Food Safety Magazine.* 10, 18–30. doi: 10.16043/j.cnki.cfs.2024.10.002

Long Peng, J., Liu, X., Peng, C., and Shao, Y. (2023). Comprehensive factor analysis and risk quantification study of fall from height accidents. *Heliyon* 9:e22167. doi: 10.1016/j.heliyon.2023.e22167

Ming, M., Yu-fa, A., and Zhong-wei, H. (2014). An analysis of key regulatory points and control measures of supermarket food safety——based on 359 supermarket food safety incidents. *Theor. Prac. Financ. Econ.* 35, 137–140.

Mu, W., Van Asselt, E. D., and Van der Fels-Klerx, H. J. (2021). Towards a resilient food supply chain in the context of food safety. *Food Control* 125:107953. doi: 10.1016/j.foodcont.2021.107953

Newman, M. E. J. (2008). The mathematics of networks. *New Palgrave Encycl. Econ.* 2, 1–12. doi: 10.1057/978-1-349-95121-5_2565-1

Niu, L., Chen, M., Chen, X., Wu, L., and Tsai, F. S. (2021). Enterprise food fraud in China: key factors identification from social co-governance perspective. *Front. Public Health* 9:752112. doi: 10.3389/fpubh.2021.752112

Qing-guang, L., Lin-hai, W., and Xiao-li, W. (2016). Review of Chinese food safety incidents research progress. *Food Ind.* 37, 219–224.

Qiu, Z., Liu, Q., Li, X., Zhang, J., and Zhang, Y. (2021). Construction and analysis of a coal mine accident causation network based on text mining. *Process Saf. Environ. Prot.* 153, 320–328. doi: 10.1016/j.psep.2021.07.032

Rutsaert, P., Regan, Á., Pieniak, Z., McConnon, Á., Moss, A., Wall, P., et al. (2013). The use of social media in food risk and benefit communication. *Trends Food Sci. Technol.* 30, 84–91. doi: 10.1016/j.tifs.2012.10.006

Van Rossum, G., and Drake, F. L. (2009). Python language reference. Wilmington, DE: Python Software Foundation.

Xiumei, Z., Xiaokang, L., Qing, L., and Beibei, G. (2024). Text mining and social network analysis of general aviation accident causes. *J. Saf. Environ.* 24, 602–609. doi: 10.13637/j.issn.1009-6094.2023.0041

Yongqin, W., Siyuan, L., and Julan, D. (2014). Contagion effects vs. competitive effects in credence goods markets: theory and event study on China's food markets. *Econ. Res. J.* 49, 141–154.

Yuntao, S., Kai, Z., Shuqin, L., Meng, Z., and Weichuan, L. (2022). Heterogeneous graph attention network for food safety risk prediction. *J. Food Eng.* 323:111005. doi: 10.1016/j.jfoodeng.2022.111005

Zhong, B., Pan, X., Love, P. E., Sun, J., and Tao, C. (2020). Hazard analysis: a deep learning and text mining framework for accident prevention. *Adv. Eng. Inform.* 46:101152. doi: 10.1016/j.aei.2020.101152