



Article

A Novel Visual Analysis Method of Food Safety Risk Traceability Based on Blockchain

Zhihao Hao^{1,2,3}, Dianhui Mao^{1,3,*}, Bob Zhang^{2,*} , Min Zuo^{1,3} and Zhihua Zhao⁴

¹ National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing 100048, China; hao_zhihao@126.com (Z.H.); zuomin@th.btbu.edu.cn (M.Z.)

² PAMI Research Group, Department of Computer and Information Science, University of Macau, Taipa, Macau 999078, China

³ Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

⁴ The School of Law, Chinese University of Political Science and Law, Beijing 102249, China; zhaozhihua@cupl.edu.cn

* Correspondence: maodh@th.btbu.edu.cn (D.M.); bobzhang@um.edu.mo (B.Z.)

Received: 7 March 2020; Accepted: 26 March 2020; Published: 29 March 2020



Abstract: Current food traceability systems have a number of problems, such as data being easily tampered with and a lack of effective methods to intuitively analyze the causes of risks. Therefore, a novel method has been proposed that combines blockchain technology with visualization technology, which uses Hyperledger to build an information storage platform. Features such as distribution and tamper-resistance can guarantee the authenticity and validity of data. A data structure model is designed to implement the data storage of the blockchain. The food safety risks of unqualified detection data can be quantitatively analyzed, and a food safety risk assessment model is established according to failure rate and qualification deviation. Risk analysis used visual techniques, such as heat maps, to show the areas where unqualified products appeared, with a migration map and a force-directed graph used to trace these products. Moreover, the food sampling data were used as the experimental data set to test the validity of the method. Instead of difficult-to-understand and highly specialized food data sets, such as elements in food, food sampling data for the entire year of 2016 was used to analyze the risks of food incidents. A case study using aquatic products as an example was explored, where the results showed the risks intuitively. Furthermore, by analyzing the reasons and traceability processes effectively, it can be proven that the proposed method provides a basis to formulate a regulatory strategy for regions with risks.

Keywords: blockchain; visualization; risk; traceability; food safety

1. Introduction

With the improvement of living standards, people have higher requirements for food quality. Ensuring food quality safety has become a major problem for governments, business organizations, and merchants. However, due to the complexity of the food supply chain, there is no effective regulatory mechanism from farm to table, which has led to frequent food safety issues around the world over the past few decades. For example, according to The Sunday Times, on June 5, 2005, the British Food Standards Agency found the carcinogen “malachite green” in salmon sold in a well-known supermarket. Even in recent years, food safety issues have not been well resolved. In 2019, hundreds of African swine fever epidemics have occurred in many provinces of China [1]. These directly caused an increase of pork prices and seriously affected people’s daily lives.

At present, it is difficult to recall food quickly after it has entered the market. This is because the storage of food information is incomplete and the stored information can be easily forged. Countries

around the world have adopted different regulatory approaches to prevent food safety incidents. For example, the United States has utilized product traceability systems and product recall systems [2]. The European Union (EU) regulates the production, distribution, and processing of the entire food chain from farm to table [3]. China has used a segmented management safety oversight model [4]. This traditional regulatory model will lead to independent law enforcement by various departments. Each supervisory authority is responsible for its own affairs, or can shirk its responsibilities with other authorities, which is not conducive to supervision. In addition, due to the large number of food safety supervision departments and the lack of effective coordination departments and mechanisms, the regulatory information of a certain product cannot be effectively transmitted to the next stage. This may lead to repeated sampling and waste of manpower and resources. In addition to these issues, there are a number of problems with these models; for example, information can be easily tampered with, sources can be difficult to track, and so on.

In recent years, blockchain technology has received increasing attention due to features such as having a distributed system, decentralization, and generating data that cannot be tampered with [5]. Blockchain is a distributed database system containing multiple independent nodes. It is essentially a distributed ledger that is maintained jointly by the nodes in the network. Information can be recorded into the blockchain to ensure its security and credibility. It implements decentralized point-to-point transactions in distributed networks by using encryption algorithms, timestamps, consensus mechanisms, and reward mechanisms. The process is efficient and transparent. This technology solves the problems of poor reliability, low security, high cost, and low efficiency in the current centralized storage mode.

The introduction of blockchain technology into food safety supervision is becoming a trend. In October 2016, the Wal-Mart Food Safety Cooperation Center, International Business Machines Corporation (IBM), and Tsinghua University piloted a food traceability system based on blockchain technology in China. In March 2017, Alibaba cooperated with PricewaterhouseCoopers to launch a food supply blockchain test project in the food supply chain. In addition, Walmart has required its vegetable suppliers to use blockchain technology developed by IBM to track products in real time since September 2019. These practices have proven to the public that stakeholders in the global food supply chain consider food safety as being collaborative rather than competitive. Blockchain technology provides a new tool for traceability. By applying blockchain technology to food traceability systems, the authenticity of data can be guaranteed to solve trust issues. In addition, establishing a more transparent and traceable food system can ensure that every stakeholder in the blockchain, such as food producers, processors, retailers, and consumers, can benefit from it. For example, Tian [6] has built a traceability system for an agricultural product supply chain based on Radio Frequency Identification (RFID) and blockchain technology, which covers the entire process of data collection and information management in each link of the supply chain and realizes the monitoring, tracking, and traceability management of agricultural food quality and safety.

However, there are two main challenges. One is the lack of an intuitive display and analysis of the vast amounts of data. Visualization technologies can help people quickly identify and make relevant decisions. For example, Saura et al. [7] extracted useful knowledge from available user generated content (UGC) samples and visualized them. The results obtained are relevant to innovative education trends, and practitioners can use them to improve strategies and interventions in the education sector in the short-term future. Another challenge is that it is difficult to establish a good risk assessment model for food safety risks. Risk assessment refers to the use of the employed method to analyze existing data to assess the possibility of potential risks [8], and is applied to the food industry to evaluate food safety risks. Its basic content includes hazard identification, hazard feature description, exposure assessment, and risk feature description [9]. Therefore, visualization technologies and risk assessment models can be introduced into the food industry, as they help people formulate relevant strategies to reduce the risk of food safety incidents.

In order to meet these challenges, we propose a visual analysis method of food safety risk traceability based on blockchain. The theoretical framework and implementation process of this research were divided into three steps. First, we designed a data structure for users with different identity roles to update, view, and obtain information. Here, users can also upload relevant information to the network for verification. Food sampling data are used in this research for food safety risk assessment, where a method to quantitatively analyze food safety risks is proposed according to the food sampling data to facilitate traceability analysis. After the information is verified, it is uploaded to the blockchain. The consortium blockchain Hyperledger Fabric is used as a data hosting platform. Ordinary users and regulators are given different identity permissions in the system to perform different functions. Finally, visualization technologies are used to analyze food safety risk assessment results and food safety risk traceability processes based on spatial characteristics of the data in the blockchain. Heat maps are used to illustrate macro risks. Migration maps and force-directed graphs are also applied to show microscopic flow. By using a human's rapid recognition ability in visual mode, people can easily track and monitor the extent and impact of the dangerous spreading of food(s).

The rest of this research is organized as follows. Related work is described in Section 2. Section 3 presents the framework. Design and implementation are illustrated in Section 4. Section 5 shows experiments and analysis, and the conclusion is given in Section 6.

2. Related Work

Currently, many technologies have been used in the food industry to solve frequent food safety incidents, with one example being RFID (Radio Frequency Identification) [10]. For example, Zhang et al. [11] proposed a special food safety traceability network model based on RFID technology, which applies RFID technology to data acquisition of raw material procurement, production processing, warehousing management, logistics, and transportation. However, traditional models using RFID technology have problems such as low efficiency. In order to solve these problems, Alfian et al. [12] used Internet of Things (IoT) technology and machine learning methods to improve the efficiency of RFID-based perishable food traceability systems. In addition, Fan et al. [13] proposed a method to improve the continuous traceability of food by using barcode RFID two-way conversion equipment. In addition to RFID, IoT technology is also widely used in the food industry [14]. For example, Verdouw et al. [15] developed and applied a framework for the food industry, which is based on the Internet of Things system for modeling. However, these technologies still have some drawbacks. For example, the centralized storage of data increases the possibility of information loss and tampering. In addition, there are other problems, such as low transparency and easy leakage of information. For instance, the largest information leakage incident of South Korea occurred in 2013, where 104 million people's personal information was leaked [16].

Recently, blockchain technology has been applied to the food industry [17,18]. Tse et al. [19] proposed a method of applying blockchain technology to the food supply chain for ensured information security. Unlike the traditional food supply chain traceability system, it is transformed into a distributed storage platform based on the underlying protocol of the blockchain to ensure data security and traceability [20]. Blockchain technology is often combined with other technologies. For example, Hong et al. [21] implemented a traceability system based on the Internet of Things and blockchain technology. Tsang et al. [22] used the Internet of Things to achieve traceability through integrated consensus mechanisms. Through the application of blockchain, the environment of the food supply chain has been greatly improved [23].

However, current data about the food industry requires professional visualization methods to help quantitative analysis [24,25]. For example, ElMasry et al. [26] used near-infrared hyperspectral imaging to quantitatively analyze the prediction parameters of fresh beef. Cropotova et al. proposed a fluorimetric assay method to quantitatively analyze protein carbonyls [27]. Lohumi et al. [28] used Fourier transform infrared (FTIR) spectroscopy to quantitatively analyze Sudan dye adulteration for risk assessment. However, these results may cause great difficulties to the understanding of ordinary

people. Therefore, suitable data sets are needed to help users understand and avoid food safety risks. Therefore, we chose food sampling data as the data set for experimental testing. Blockchain technology is used to ensure the true validity of the data, and visual methods, such as heat maps and migration maps, are used to display and analyze risks.

3. Framework

The method uses Hyperledger Fabric as the underlying technology. Hyperledger is derived from the open source project led by the Linux Foundation in 2015. Hyperledger Fabric is its sub-project, which allows multiple parties to participate in the development, deployment, and operation of the consortium blockchain platform. It aims to create an extensible blockchain development framework that provides solutions for the development of enterprise-level blockchain applications.

In order to better understand Hyperledger Fabric, its architecture is introduced briefly here. Hyperledger Fabric includes multiple components: (1) Orderer. In Hyperledger Fabric, an ordering service is provided through multiple Orderers. They receive all transactions from the entire network and packs the transactions into blocks in order of time. It does not participate in the execution and verification of the transaction, so it does not care about the specific content of the transaction. The goal is to reach a consensus on the order in which the transactions occur, and then broadcast the results. (2) Client. The client is the access point between the user and the Hyperledger Fabric network, and deploys a proprietary Software Development Kit (SDK). Users can use the client to initiate a transaction request. (3) Endorser. When a client wants to initiate a transaction, it must first obtain a certain number of endorsements from Endorsers for the transaction, that is, signing to prove that the transaction has been processed by the endorsing nodes. (4) Committer. This type of node is the main body for maintaining the ledger in the network. Committers can receive packaged blocks and verify the validity of transactions in the blocks, and submit valid transactions to the ledger. Endorsers are special Committers. Their endorsement function is an additional function.

The framework is shown in Figure 1 and has three layers as follows.

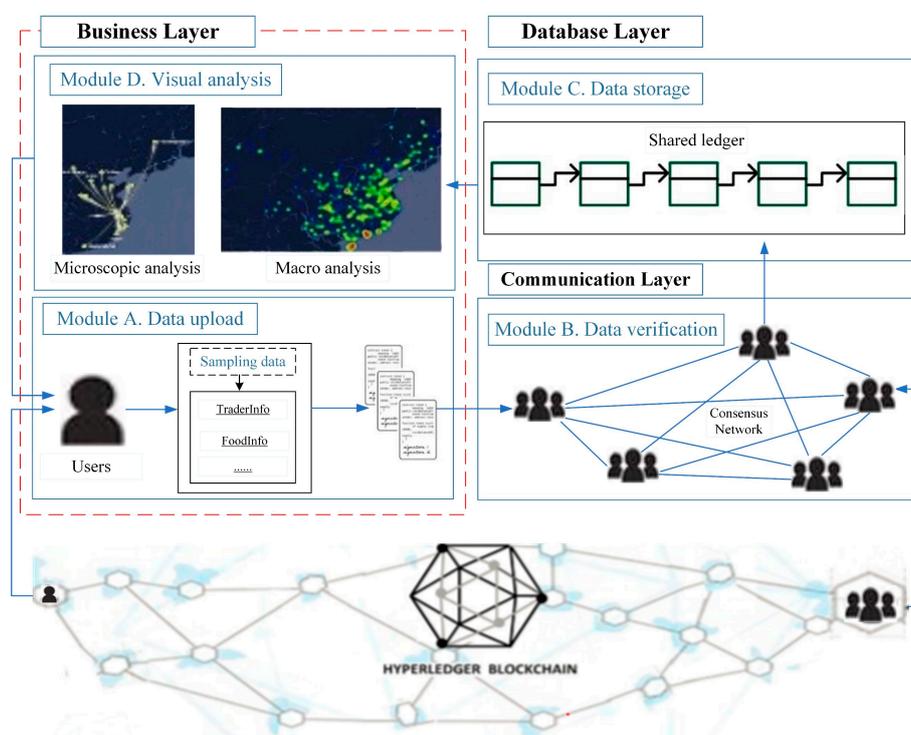


Figure 1. The framework of the method.

Business Layer. This layer supports user access and contains entry points for human-computer interaction. It consists of modules A and D in Figure 1. Module A is the operation of uploading data, and Module D is the visual display. Module A is mainly applicable to business developers who deploy smart contracts at the business level. Smart contracts can be understood as script code running on the blockchain, which provides programmable functions to support upper-layer applications. Users write their own smart contracts through the Application Programming Interface (API). Users can update or obtain information in the blockchain through smart contracts. Module D is the information visualization part. The results of this module can be fed back to users to help the user perform risk and traceability analysis. The heat map can show the risks in a macro view. The area shown by the heat map is an important attribute of food risk data and the basis of food safety traceability. Based on this, force-directed graphs and migration maps can be used to show the microscopic flow of products to analyze their traceability. These techniques can produce direct, simple, and high-quality results.

Communication layer. This layer contains the network structure and protocols of the P2P network (Module B). It provides network services for the blockchain platform and uses the Gossip data communication protocol to achieve state synchronization and data distribution between nodes in the network. Due to the use of Hyperledger Fabric, nodes in the communication layer are assigned different roles to execute various services [29]. The communication layer not only deals directly with the business layer, but also connects with the database layer. The uploaded data is verified by the consensus algorithm and then uploaded to the blockchain to achieve consistency and correctness of the ledger data on different ledger nodes. Consensus algorithms are the foundation of blockchain technology. PBFT (Practical Byzantine Fault Tolerance) is used in Hyperledger Fabric. The main steps are: (1) The client (user) sends a request to activate the service operation of the master node (regulator). (2) After receiving the request, the master node broadcasts the request to each node. (3) The client waits for responses from different nodes. If more than half of the nodes (for example, 51%) have the same response, it is the result recorded in the blockchain.

Database layer. As can be seen from Module C, this layer consists of a shared ledger. Here, we create a data structure model to implement different types of data upload. The uploaded information is stored in blocks. Each block consists of a block header and a block body. The block header is divided into a few parts, such as version (the version number of the block header, used to track software/protocol updates), prevBlockHash (the hash address of the previous block), merkleRoot (the hash value of the Merkle tree root in the block), time (the creation timestamp of the block), and so on. The block body contains transaction information, which is the number of transactions and transaction details such as location, time, etc. The information recorded in the ledger is used as data for visual analysis.

4. System Design and Implements

4.1. Data Structure and Storage Process

A custom data model is created, as shown in Figure 2. The data model includes detailed parameters such as its identification and information for uploading. There are nine fields in total, which are ID, Name, Date, Location, FromLocation, ToLocation, PrincipalName, PrincipalPhone, and OtherInfo. Among these, ID is the classification of food categories, such as: milk and dairy products, fats, oils and emulsified fat products and cereals. Name refers to the sub-categories of food, such as high calcium milk. Date is the timestamp of data uploaded for each link. *Location* represents the geographical coordinates of the current link. The previous geographical coordinates are defined as FromLocation. ToLocation represents the next geographical coordinates. RunningTime is the transport time. PrincipalName the name of the principal in this link. PrincipalPhone represents phone calls of the principal. OtherInfo is the reserved field. Each link uploads a different type of data, but the ID and Name fields must be uploaded. Other fields can be filled out based on the specific link and role. The format of the key-value store sets the data format to JavaScript Object Notation (JSON). Therefore, these data models eventually need to be converted to JSON strings.

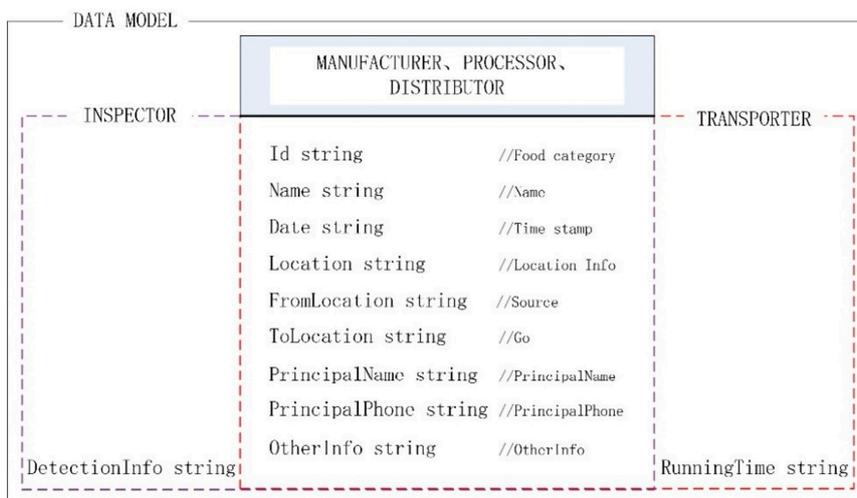


Figure 2. Data structure model for uploading to the blockchain.

In addition, there are five roles for data upload: manufacturer, processor, inspector, transporter, and distributor. Different roles have different responsibilities and parameters. The three roles of manufacturer, processor and distributor can use all of the above parameters. Different from these roles, inspector can use one more parameter, DetectionInfo, and transporter can use the other parameter RunningTime.

The process of data recording to the blockchain is shown in Figure 3. When a participant initiates a request, the system will call an embedded smart contract and then verify the data structure, signature integrity, and whether it is duplicated. After the verification is passed, the custom smart contract will be called to upload the data to the blockchain. Nodes in the blockchain network can access different information through different rights. In addition, digital signatures are used to ensure the integrity of the information. The digital signature in this research uses asymmetric encryption technology. It ensures that the information cannot be tampered with, and the identity information of the two nodes does not need to be disclosed. When the process is completed, the smart contract will automatically execute the content of the agreement. For example, during the transaction, both parties use asymmetric keys to encrypt and decrypt transaction information. It guarantees that the transaction information will not be tampered with maliciously, and solves the integrity problem in the transaction process.

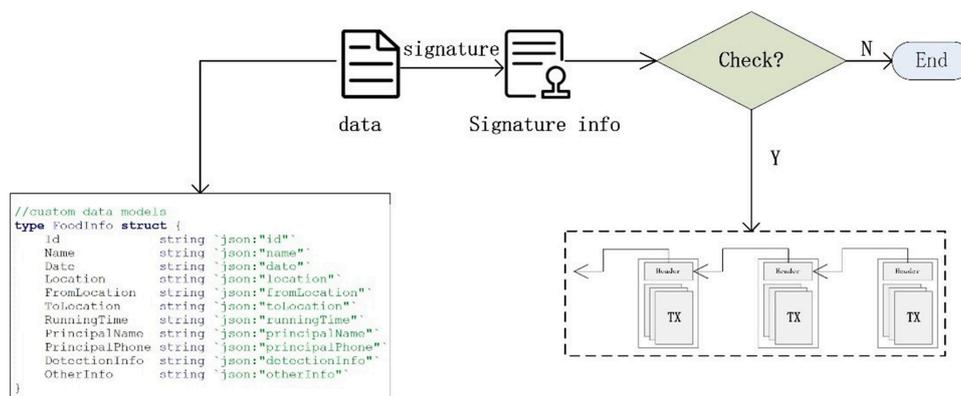


Figure 3. The process of recording data into the blockchain.

4.2. Quantitative Analysis of Safety Risks

Risk assessment refers to the consideration of all available relevant data, and on this basis to identify areas where risks may arise. The results of food safety risk assessments can provide a basis for regulators to formulate regulatory strategies. When constructing food safety risk indicators, we study

the degree of risk of food through the rate of non-conformity and the degree of deviation from conformity. The failure rate describes the frequency with which foods that do not meet the required standards [30] occur (unqualified products). The deviation rate describes how far these unqualified products deviate from their safety standards. They can be calculated using the following procedure.

The specific detection results of the product i in a region are defined as m_i^p , where p is the type of product. m_i^p needs to be compared with $[Min^p, Max^p]$, which is the scope of safety standards, and unqualified number count E_i^p ($E_i^p \in [0, 1]$) can be notified by following:

- If $Min^p \leq m_i^p \leq Max^p$, then the detected item meets safety standards, $E_i^p = 0$.
- If $m_i^p < Min^p$ or $m_i^p > Max^p$, then the detected item does not meet safety standards, $E_i^p = 1$.

Thus, the failure rate \widetilde{V}_i^p can be calculated by:

$$\widetilde{V}_i^p = \frac{\sum n_i^p E_i^p}{n_i^p} \quad (1)$$

where n_i^p is the number of sampled products. For unqualified products, the deviation rate needs to be calculated, as shown in Equation (2). It is represented by B_i^p ($0 < B_i^p \leq 1$).

$$B_i^p = \begin{cases} \frac{Min^p - m_i^p}{Min^p}, & 0 < m_i^p < Min^p \\ \frac{m_i^p - Max^p}{m_i^p}, & 0 < Max^p < m_i^p \end{cases} \quad (2)$$

The average deviation rate \widetilde{B}_i^p can be calculated by:

$$\widetilde{B}_i^p = \frac{\sum n_i^p B_i^p}{n_i^p} \quad (3)$$

For a region j , the risk indicator can be calculated according to this formula:

$$\psi_j = \sum_1^n (\widetilde{V}_i^p \times \widetilde{B}_i^p) \quad (4)$$

This obtains the risk indicators set $(\psi_1, \psi_2, \psi_3, \dots, \psi_n)$. After sorting, the region set with risks can be acquired according to the results in the order from large to small. Then, these can be analyzed using visualization techniques.

4.3. Visual Analysis Methods

Visual analysis consists of macro and micro analysis. The macro analysis method uses heat map technology to illustrate the risk distributions of regions. For a specific region, the migration map and force-directed graph technologies are used to demonstrate the reasons why these risks occur. The details of these methods are explained in the following.

4.3.1. Macro Analysis

The macro analysis of risks can be processed by heat maps [31]. These display areas of interest to users in colored highlights. The heat map generation process is roughly divided into three steps. First, the original data needs to be clustered to form clusters. Then, Gaussian fitting is performed according to the center points of the clusters to obtain each heat value in regards to its surrounding area. Finally, the heat map can be formed by coloring according to the heat value, which is combined with maps for intuitive display. The detailed process is as follows.

Clustering data is the basis of heat map generation. Clustering refers to clustering similar entities together to form a cluster [32]. The spatial distance between the data can be used as the basis for clustering. In a cluster obtained by clustering data with spatial-temporal characteristics (spatial points), the distance between any two spatial points must be smaller than the distance between any point in other clusters. From the aspect of density, the tight aggregation of points forms a cluster with high density and a cluster with low density represents points that are scattered. The operation of clustering is summarized as follows.

- 1. Data initialization. Data dimensions are reduced and certain characteristics are standardized.
- 2. Selection of data characteristics. The characteristics that can best distinguish the data are found, extracted, and stored.
- 3. Data clustering. Select or construct a certain clustering function to test the similarity of the data according to the characteristics, and perform clustering based on the test results.
- 4. Data cluster evaluation [33]. Perform evaluation of correlation and validity on clusters.

Processing the original data based on the grid clustering algorithm can preserve the important attributes of the data [34]. The efficiency of a grid cluster will not depend on the amount of raw data. It is determined by the number of elements in a one-dimensional case. It can improve data processing efficiency by flexibly processing the amount of data at different levels. The risk distribution can be shown by heat maps after the above operation of the risk regions (hotspots). These hotspots will produce a certain range of influence, and the strength of this range of influence can be calculated using a Gaussian function, as shown in Equation (5).

$$r(x) = \frac{k}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (5)$$

where x is the distance between sampling site and hotspot, σ is the scale factor of the function, and k is the influence factor. The larger of these two factors, the wider the range of influence. Equation (5) shows that its range of action is inversely proportional to distance. The influence of the center area is the largest, and the periphery is the smallest. When processing a certain area, it is necessary to add up all the heat values that can be generated in the surrounding area. Therefore, the final heat value of a certain region can be calculated by:

$$H = \sum_{i=1}^n r(x) \quad (6)$$

After H is obtained, the corresponding location information is required. We can process the discreteness of all sampling sites within the range into a matrix. Each sampling site corresponds to a specific position coordinate (longitude, latitude), and the heat value of the sampling site is calculated by Gaussian fitting. The area represented by the rectangular lattice consists of longitude (Lon_1, Lon_2), latitude (Lat_2, Lat_1), and the matrix size is $M \times M$. Then, the latitude and longitude corresponding to the hotspot of the row r and the column c can be calculated by Equations (7) and (8).

$$lat = r \times \frac{(Lat_2 - Lat_1)}{M} + Lat_1 \quad (7)$$

$$lon = c \times \frac{(Lon_2 - Lon_1)}{M} + Lon_1 \quad (8)$$

Combining the heat value with the corresponding coordinates, and then rendering the map with the corresponding chromaticity through the color table, we can further enhance the heat perception of the risk distribution. Therefore, the generation process is shown in Algorithm 1.

Algorithm 1 The Generation Process of Heat Maps for Food Safety Risks Evaluation**Require:** ChaincodeStubInterface, id, name**Ensure:** Heatmap Graph

```

1: begin:  $D = x_1, x_2, \dots, x_n$ 
2: for  $i = 1, 2, \dots, n$  do /* $i$  is the number of layers*/
3:   Calculate H and (lon,lat)
4:   Mark cell is relevant
5:   if  $i$  equal to the number of bottom layer then
6:     to  $n$ 
7:   else to  $i + 1$ 
8:   end if
9: end for
10: From the hierarchy to the next layer
11: calculateGrayValue()
12: renderColor()

```

4.3.2. Micro Analysis

For the risk distribution results, micro analysis is needed to discover the reasons for these risks occurring. Micro analysis is achieved through migration graphs and force-directed graphs. Migration diagrams can show relationships between regions with risks. Migration graph technology calculates, analyzes, and visualizes the data of a geographic location-based service (LBS) to dynamically, instantly, and intuitively display the trajectory and characteristics of data [35]. This technique is widely used in the analysis of movement. For example, Baidu launched a technology project called “Baidu Migration Map” during the Chinese Spring Festival Transport in 2014. It analyzed mobile phone user’s positioning information to map their tracks. This was used to observe the movement situation of China, as well as its provinces, cities, and districts in the current and earlier time periods, so as to intuitively determine the source and destination of the population. However, the migration graph can only mark the general flow direction, so the traceability of a specific product needs to be achieved through a force-directed graph.

Force-directed graphs can be used to illustrate specific flow directions, since they make relationships clear. Force-directed graphs are mainly used in the visualization of complex networks such as social networks [36]. They show the relationship between multiple nodes. Nodes are configured in two- or three-dimensional space, and the relationship between them is represented by lines. These lines are almost equal in length and as far as possible do not intersect. Both nodes and lines are subjected to forces (Coulomb repulsion, Hooke’s law, and damping attenuation also should be considered [37]), which can be calculated according to the relative positions. The motion trajectories of the nodes are determined according to the action of the forces, and their energy is continuously reduced, eventually reaching a stable and balanced state. Through predefined points, edge weights, and other information, the force-directed graph can reflect the data flow direction according to the real-time state.

In order to better understand the principle of its generation, we use physical methods as an analogy. First, the entire network can be considered as a virtual physical system. Each node of the network can be considered as a discharge particle in the system with a certain energy and, in the proposed method, represents a place (the start or end location of the unqualified products flow). The forces reflect the strength of these relationships among places. There is Coulomb repulsion between particles, which makes them repel each other. At the same time, some particles are implicated by some edges (lines), which can reflect the flows of the relationships in the method. These edges generate hooke’s gravitational force to keep the particles at both ends of the edge. Under the continuous action

of repulsion and gravity between particles, the particles are constantly displaced from the random and disordered initial state, and gradually tend to balance to form an orderly final state. At the same time, the energy of the entire physical system is also continuously consumed. After several iterations, relative displacements between the particles almost no longer occurs, and the entire system reaches a stable and balanced state. Based on this, the process of generating a force-directed graph is as follows.

- 1. The initial nodes positions are distributed randomly.
- 2. Calculate the unit displacement (edge/ line) caused by the repulsive force and gravitational forces between any two nodes in the area at each iteration.
- 3. Constantly adjust according to parameters such as the distance between nodes, the location of nodes, and the repulsive and gravitational coefficients.
- 4. Add up the unit displacements (edges/ lines) of all nodes.
- 5. Iterate n times until the desired effect is achieved.

Through the use of migration graphs and force-oriented graphs, regulators can more intuitively observe the causes of risks. Therefore, relevant rules can be formulated to reduce the occurrence of food safety incidents and ensure a food safety environment.

5. Experiments and Numerical Analysis

5.1. Experiment Platform

This research work is based on the Hyperledger Fabric v1.0 for development. The underlying environment is ubuntu16.04, Git client version 2.5.1, Node.js v8.11.2, NPM v5.6.0, Docker version 17.12.0-ce, Docker Compose-v1.8. The development languages are Python 3.6.0 and Go v1.9. LevelDB is used as the database, which is also the default database built into the Hyperledger. The front end is mainly developed through Hyper Text Markup Language (HTML), Javascript, JQuery, and Cascading Style Sheet (CSS). The background includes server deployment, information uploading to the blockchain, and information display of the blockchain. In the experiment, the CPU model of the microcomputer is Intel(R) Xeon(R) CPU e5-2603 v3@1.60ghz *6, and the memory is 8GB.

The description of the system interfaces can be divided into four parts. Part A the main interface of the system. It contains three main users of the system, which are inspectors, transporters, and users (manufacturers, processors and distributors). Here we take the transporter as an example. When the transporter completes the transportation of a batch of products, the detailed information of the products needs to be recorded and uploaded. Part B is the data upload interface, which contains detailed information of the transported products, such as name, ID, date, and so on. After the data upload operation is finished, the consensus of the nodes needs to be completed to ensure the validity of the data. Then, the valid data can be uploaded to the blockchain. Part C is the visual display interface. The upper part shows the relationship between the three roles, and the other part demonstrates the blockchain. It shows the basic information of blocks such as block number, hash value, and date. Part D shows the detailed information, which includes size, nonce, and so on. It can be used to analyze data traceability.

Hyperledger Fabric uses container technology such as Docker to host the “chaincode” (i.e., smart contracts) which include the system application logic. Smart contracts typically use the Domain-Specific Language (DSL) for development. These languages have specific restrictions for blockchain, like Ethereum. It is not only not conducive to the use of ordinary users, it also causes greater difficulties for developers. One of the features of Hyperledger Fabric is that it uses general-purpose programming languages such as Java, Node.js, and Go to develop smart contracts. It greatly reduces the learning costs for developers. In addition, it is the only channel for users to interact with Hyperledger Fabric and the only tool for performing transactions on the blockchain platform [38]. Based on these, we use smart contracts to implement the operations of recording data to the blockchain and downloading data, as shown in Figure 4.

```

func CreateFoodInfo(stub shim.ChaincodeStubInterface,
    args []string) ([]byte, error) {
    if len(args) < 2 {
        fmt.Println("Invalid number of args")
        return nil, errors.New("Expected at least two
            + " arguments for food info creation")
    }
    var foodId = args[0]
    var foodInfoInput = args[1]
    err := stub.PutState(foodId, []byte(foodInfoInput))
    if err != nil {
        fmt.Println("Could not save food info to ledger", err)
        return nil, err
    }
    fmt.Println("Successfully saved food info")
    return nil, nil
}

func GetFoodInfo(stub shim.A,
    args []string) ([]byte, error) {
    if len(args) < 1 {
        fmt.Println("Invalid number of arguments")
        return nil, errors.New("Missing food ID")
    }
    var foodId = args[0]
    bytes, err := stub.GetState(foodId)
    if err != nil {
        fmt.Println("Could not fetch food with id "
            + foodId + " from ledger", err)
        return nil, err
    }
    return bytes, nil
}
    
```

(a)

(b)

Figure 4. (a) Upload the data to the blockchain. (b) Retrieve the data from the blockchain.

Since the sampling results are stored on the blockchain, we collected the sampling data for the whole year of 2016. In order to protect the privacy of the merchant, we performed address hiding and other processing on the data, and the results are shown in Figure 5.

transactionHash	blockHash	blockNumber	from	to	timestamp
0xaf0a0627d7...	0xa773763b1...	33	0x7ebfdffdf...	0xebcd4...	2016-08-04...
0xc5c5db83fd...	0x3d48efcb0...	88	0x7ebfdffdf...	0xebcd4...	2016-08-06...
0x7f74410da8...	0xc339ee7f3...	165	0x7ebfdffdf...	0x18b9a...	2016-08-07...
0x3bd672e529...	0x578596954...	436	0x7ebfdffdf...	0x7913e...	2016-08-07...
0xf4a0f82612...	0xa800aeeef...	465	0x7ebfdffdf...	0x918b9...	2016-08-07...
0x402f093c46...	0xeaaaa318f...	491	0x7ebfdffdf...	0x8b9ab...	2016-08-08...
0xb44062b0a4...	0xe1081b1ce...	596	0x7ebfdffdf...	0xbb3a3...	2016-08-09...
0x1c17e51d79...	0x9abd4ba59...	768	0x7ebfdffdf...	0x3e8ac...	2016-08-09...

Figure 5. Data in the blockchain.

As shown in Figure 5, using the blockchain platform to store the sampling data can meet the following security requirements: (1) Privacy protection. The merchant information stored in the block is hash encrypted (from and to). It does not disclose any private information about merchants, thus achieving privacy protection and improving security. (2) Data integrity. After the information is record into the block, it will correspond to a specific block number (blockNumber), and the block will be encrypted by the hash function to generate a unique hash value (blockHash). (3) Information security. Digital signatures are used to prove ownership, and the hash function is also used to encrypt transaction information (transactionHash). In addition, the use of digital signatures can prevent duplicate transactions. This ensures the security to a certain extent.

5.2. An Example of Using Aquatic Products Data

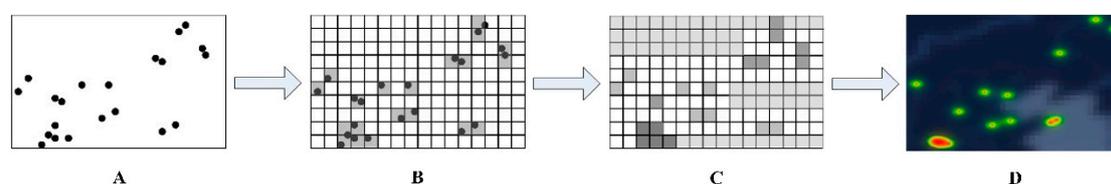
The data set used in this research is the Food Sampling Data of the full year of 2016. The data set contains 341 types of food with a total of 1,048,575 items. The detailed parameters are shown in Table 1. Here we have selected the data related to 19,578 aquatic products as the experimental data set, which includes fish, shrimps, crabs, shells, and mollusks. Among these, a total of 19,142 items were marked with qualified items. There were 436 items marked as unqualified items, question items, and non-determined items. In order to protect the privacy of the merchants, we blurred the information in the data set and processed the detection results in proportion to ensure the validity of the results.

Table 1. An example of the dataset.

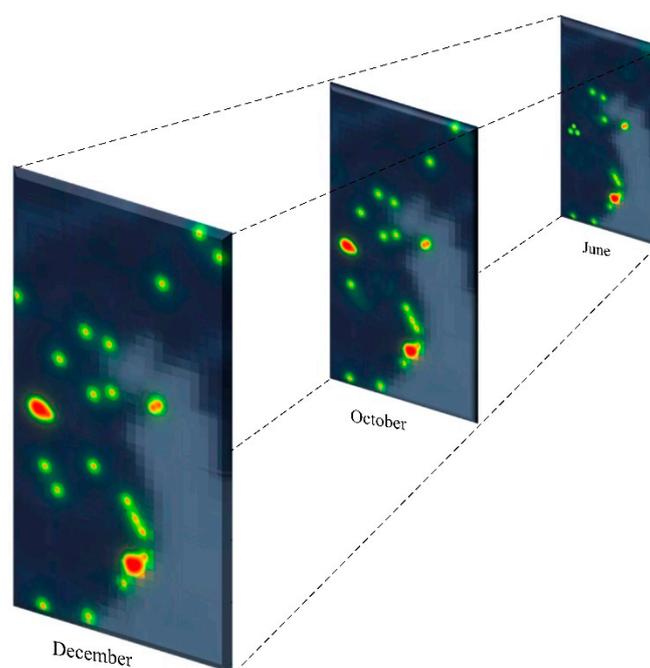
ID	Product ID	Product Name	Place of Production	Place of Sold	Food Category ID	Food Category	Substance ID	Substance Name	Result	Judgement	Date
1	93074	Razor clam	C ₁	A ₂	745	Shell	136	Tetracycline	0	Qualified	1/1/2016

5.2.1. Quantitative Risk Analysis Based on Heat Map

In order to visualize the areas where unqualified products appear, we use a heat map here. The generation results are shown in Figure 6. (A) and (B), where we first conduct grid clustering based on unqualified products data. (A) shows relatively rough clustering results, and (B) is a graph representing the results with accurate clustering. After the data is processed, the gray value, which reflects the color depth of the points in the black and white image, is superimposed and displayed as shown in (C). Then, the value is mapped using a color-band with 256 colors to finally form a colored heat map, as shown in (D).

**Figure 6.** Schematic diagram of the generation process.

The heat map will show different results over time. In order to measure this, we built a 3D model similar to [39,40]. Figure 7 shows the results of the heat map changes from June to December 2016. In order to better present the results, we selected the results in June, October and December. From Figure 7, it can be seen that most of the areas where incidents appear are initially near the water. Incidents are concentrated in some areas. As time goes on, incidents have gradually emerged in inland areas and some areas have a tendency to concentrate. From June to October, in particular, this trend is obvious. For the areas where incidents occur, we use the results from December for analysis (see Figure 8).

**Figure 7.** Heat maps over time.

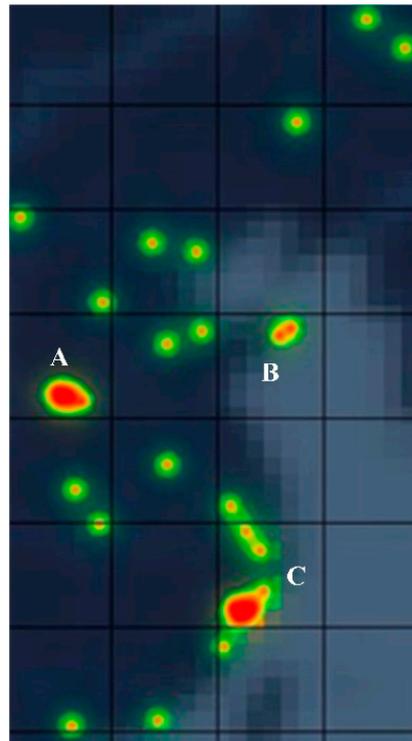


Figure 8. Part of the heat map.

For better analysis, the pictures were divided into 28 different grids. It can be seen from Figure 8 that large red areas appear in grids A, B, and C, which indicates that a large number of unqualified items occurred in these three grids. In particular, continuous red areas appeared in grid C, which indicates that the unqualified items occurred more frequently in grid C. Therefore, we recommend finding solutions for all grids where incidents occur. For grids A, B, and C, a high degree of attention must be paid to the resolution process to reduce the risk of the incident recurring. In this way, special solutions can be formulated for different causes of problems in different places, greatly improving the efficiency of problem solving.

5.2.2. Traceability Analysis Based on Migration Map and Force-Directed Graph

In order to provide effective evidence to facilitate the development of solutions, we need to conduct a traceability analysis based on the above results. We select a_1 in Grid A for further analysis. As shown in Figure 9, the migration map and force-directed graph realizes the traceability analysis of unqualified products. These product parameters are shown in Table 2. However, the results presented in Figure 9 are too complex and greatly increase the difficulty of analysis. Thus, we simplified the original image for better analysis and produced Figure 10. As can be seen from Figure 10, the unqualified products appearing in a_1 come from eight locations in grids B, C, D, E, F, and G, which are b_1 , c_1 , c_2 , c_3 , d_1 , e_1 , f_1 , and g_1 . These results reflect the flow of unqualified products. Analyzing these flow processes can help regulators in these regions perform better management. From Table 2 and Figure 10 it can be observed that the unqualified products appear in regions near the water. The regulations in these areas can be strengthened to reduce the occurrence of unqualified products. For example, regulators need to formulate stricter regulations, such as increased frequency of spot checks on products, and increased penalties for manufacturers of unqualified products [41].



Figure 9. (a) Migration map and (b) force-directed graph illustrating that they can realize traceability analysis of the unqualified products.

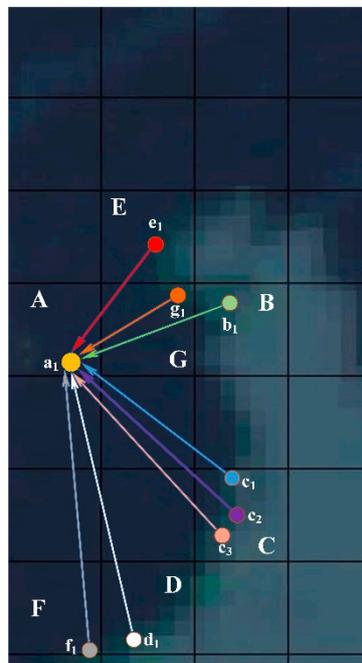


Figure 10. Traceability analysis graph of unqualified products.

Table 2. Parameters of the unqualified products.

ID	Product ID	Product Name	Place of Production	Place of Sold	Food Category ID	Food Category	Substance ID	Substance Name	Result ¹	Judgement	Color
1	60813	Sea crab	b ₁	a ₁	738	Crab	41	Cadmium	1.481364	Unqualified	Green
2	4708	Croaker	c ₁	a ₁	737	Fish	184	AOZ	42.804198	Unqualified	Blue
3	96064	Scylla serrata	c ₂	a ₁	738	Crab	41	Cadmium	3.7944408	Unqualified	Purple
4	93568	White shrimp	c ₃	a ₁	736	Shrimp	1027	AOZ	1.9691520	Unqualified	Pink
5	11859	Weever	d ₁	a ₁	737	Fish	1027	AMOZ	3.344956	Unqualified	White
6	9290	Turbot	e ₁	a ₁	736	Fish	182	SEM	56.294333	Unqualified	Red
7	93109	Pomfret	f ₁	a ₁	737	Fish	1027	AOZ	1.1054634	Unqualified	Gray
8	9415	Mantis Shrimp	g ₁	a ₁	736	Shrimp	123	Chloramp-henicol	0.304360	Unqualified	Orange

¹ This refers to detection results. Due to the length limitation here, only six digits after the decimal point are retained.

6. Conclusions

Most current research only uses blockchain technology to ensure the authenticity and validity of the data. This research not only proposes a method based on blockchain technology to realize the storage and management of food sampling data, but also introduced visualization methods to intuitively show risks and help the traceability analysis of food. This research can expand the current research system from the following aspects.

First, unlike the current data storage and management methods, the designed data structure can meet the needs of different roles and normalize the recording rules. In addition, the blockchain technology is used to store data. Since the blockchain features are distributed, the stored data cannot be tampered with maliciously. Therefore, the reliability and integrity of data can be ensured. Second, most current methods for risk analysis are qualitative, that is, the risk is classified into several levels. Based on food sampling data, a quantitative analysis method has been proposed for food safety risk assessment. We have proven that this method can provide a scientific basis for management, thereby reducing the occurrence of risks and protecting people's health. Finally, unlike many current analysis methods that only focus on the results, a holistic and detailed analysis approach has been adopted to obtain results and the reasons for their occurrence, respectively. The experimental results visualize risks through heat maps, and the traceability can be analyzed through migration and force-oriented graphs. By using visualization methods to display information, people can easily mine some important data. In practical implications, this research can help regulators (such as the Food and Drug Administration) through using these results to develop more scientific and reasonable regulatory strategies for reducing the occurrence of food safety incidents. In addition, it can also facilitate the management and control of the regions with risks.

However, there are still deficiencies that need to be improved. Blockchain technology has issues related to the speed and scalability of generating blocks. These problems will greatly reduce the efficiency of data processing. In addition, the visual analysis is also affected by the quality of the sampling data, and problems such as sampling errors can greatly affect the validity of the results. In the future, we should consider effective methods to solve these problems in order to obtain more accurate results.

Author Contributions: Conceptualization, D.M. and Z.H.; methodology, Z.H.; software, Z.H.; validation, Z.H., D.M. and B.Z.; formal analysis, Z.H.; investigation, Z.H.; resources, M.Z.; data curation, D.M. and B.Z.; writing—original draft preparation, Z.H.; writing—review and editing, Z.H.; visualization, D.M. and B.Z.; supervision, D.M., B.Z. and Z.Z.; project administration, D.M.; funding acquisition, D.M. and B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Beijing Municipal Philosophy and Social Science Foundation (19GLC091), the National Social Science Fund of China (18BGL202), National Key Technology R&D Program of China (2016YFD0401205, 2019YFC1605306), the Social Science and Humanity on Young Fund of the Ministry of Education (17YJCZH127, 17YJCZH007, 20YJCZH229), Special Subject of Innovation Method Work of the Ministry of Science and Technology (2018IM020200), Beijing Natural Science Foundation (4202014) and the University of Macau (MYRG2019-00006-FST).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vilanova, E.; Tovar, A.M.; Mourão, P.A. Imminent risk of a global shortage of heparin caused by the african swine fever afflicting the chinese pig herd. *J. Thromb. Haemost.* **2019**, *17*, 254–256. [[CrossRef](#)] [[PubMed](#)]
2. Xu, X.-H.; Fan, B.-B.; Du, D.-Q. The enlightenment from the “food safety modernization act” of the united states to the food safety supervision of china. *Hubei Agric. Sci.* **2011**, *21*.
3. Zeng, W.-G.; Xiao, F. The “convenience” characteristics of european food safety supervision legislation and its implications for china. *J. Dalian Univ. Technol. Soc. Sci.* **2013**, *3*.
4. Wang, S.-B.; Wang, J.-N.; Ma, B.-C.; Wu, X.-Q. Study on performance evaluation of food safety supervision in Jiangsu province. *China Condiment* **2017**, *37*.

5. Yuan, Y.; Wang, F.Y. Blockchain and cryptocurrencies: Model, techniques, and applications. *IEEE Trans. Syst. Man and Cybern. Syst.* **2018**, *48*, 1421–1428. [[CrossRef](#)]
6. Saura, J.R.; Reyes-Menendez, A.; Bennett, D.R. How to extract meaningful insights from ugc: A knowledge-based method applied to education. *Appl. Sci.* **2019**, *9*, 4603. [[CrossRef](#)]
7. Tian, F. An agri-food supply chain traceability system for China based on RFID & blockchain technology. In Proceedings of the 2016 13th International Conference on Service Systems and Service Management (ICSSSM), Kunming, China, 24–26 June 2016; pp. 1–6.
8. Fan, G.; Chen, G. Research on Food Safety Risk Assessment System in China. *Grain Distrib. Technol.* **2018**, *98*, 83–84.
9. Lu, J.; Sun, Y.; Geng, N.; Xu, C.; Zhang, X. Research on the Establishment of Food Safety Issues and Supervision Model in China. *Food Sci.* **2010**, *31*, 319–324.
10. Aung, M.M.; Chang, Y.S. Traceability in a food supply chain: Safety and quality perspectives. *Food Control* **2014**, *39*, 172–184. [[CrossRef](#)]
11. Zhang, C.; Li, S.; Qu, J. Safety traceability system of characteristic food based on rfid and epc internet of things. *Int. J. Online Biomed. Eng.* **2019**, *15*, 119–126. [[CrossRef](#)]
12. Alfian, G.; Syafrudin, M.; Farooq, U.; Ma'arif, M.R.; Syaekhoni, M.A.; Fitriyani, N.L.; Lee, J.; Rhee, J. Improving efficiency of rfid-based traceability system for perishable food by utilizing iot sensors and machine learning model. *Food Control* **2020**, *110*, 107016. [[CrossRef](#)]
13. Fan, B.; Qian, J.; Wu, X.; Du, X.; Li, W.; Ji, Z.; Xin, X. Improving continuous traceability of food stuff by using barcode-rfid bidirectional transformation equipment: Two field experiments. *Food Control* **2019**, *98*, 449–456. [[CrossRef](#)]
14. Da Xu, L.; He, W.; Li, S. Internet of things in industries: A survey. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2233–2243.
15. Verdouw, C.; Sundmaeker, H.; Tekinerdogan, B.; Conzon, D.; Montanaro, T. Architecture framework of IoT-based food and farm systems: A multiple case study. *Comput. Electron. Agric.* **2019**, *165*, 104939. [[CrossRef](#)]
16. Song, D.H.; Son, C.Y. Mismanagement of personally identifiable information and the reaction of interested parties to safeguarding privacy in South Korea. *Inf. Res.* **2017**, *22*, 770.
17. Mao, D.; Hao, Z.; Wang, F.; Li, H. Innovative blockchain-based approach for sustainable and credible environment in food trade: A case study in shandong province, china. *Sustainability* **2018**, *10*, 3149. [[CrossRef](#)]
18. Mao, D.; Hao, Z.; Wang, F.; Li, H. Novel automatic food trading system using consortium blockchain. *Arab. J. Sci. Eng.* **2019**, *44*, 3439–3455. [[CrossRef](#)]
19. Tse, D.; Zhang, B.; Yang, Y.; Cheng, C.; Mu, H. Blockchain application in food supply information security. In Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; pp. 1357–1361.
20. Mondal, S.; Wijewardena, K.P.; Karuppuswami, S.; Kriti, N.; Kumar, D.; Chahal, P. Blockchain inspired rfid-based information architecture for food supply chain. *IEEE Internet Things J.* **2019**, *6*, 5803–5813. [[CrossRef](#)]
21. Hong, W.; Cai, Y.; Yu, Z.; Yu, X. An agri-product traceability system based on iot and blockchain technology. In Proceedings of the 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, China, 15–17 August 2018; pp. 254–255.
22. Tsang, Y.P.; Choy, K.L.; Wu, C.H.; Ho, G.T.S.; Lam, H.Y. Blockchain-driven iot for food traceability with an integrated consensus mechanism. *IEEE Access* **2019**, *7*, 129000–129017. [[CrossRef](#)]
23. Saberi, S.; Kouhizadeh, M.; Sarkis, J.; Shen, L. Blockchain technology and its relationships to sustainable supply chain management. *Int. J. Prod. Res.* **2019**, *57*, 2117–2135. [[CrossRef](#)]
24. Pérez-Rodríguez, F.; Skandamis, P.; Valdramidis, V. Quantitative methods for food safety and quality in the vegetable industry. In *Quantitative Methods for Food Safety and Quality in the Vegetable Industry*; Springer: Basel, Switzerland, 2018; pp. 1–9.
25. Ha, T.M.; Shakur, S.; Do, K.H.P. Consumer concern about food safety in hanoi, vietnam. *Food Control* **2019**, *98*, 238–244. [[CrossRef](#)]
26. ElMasry, G.; Sun, D.-W.; Allen, P. Near-infrared hyperspectral imaging for predicting colour, ph and tenderness of fresh beef. *J. Food Eng.* **2012**, *110*, 127–140. [[CrossRef](#)]

27. Cropotova, J.; Rustad, T. A novel fluorimetric assay for visualization and quantification of protein carbonyls in muscle foods. *Food Chem.* **2019**, *297*, 125006. [[CrossRef](#)] [[PubMed](#)]
28. Lohumi, S.; Joshi, R.; Kandpal, L.M.; Lee, H.; Kim, M.S.; Cho, H.; Mo, C.; Seo, Y.-W.; Rahman, A.; Cho, B.-K. Quantitative analysis of sudan dye adulteration in paprika powder using ftr spectroscopy part a chemistry, analysis, control, exposure & risk assessment. *Food Addit. Contam.* **2017**, *34*, 678–686.
29. Androulaki, E.; Barger, A.; Bortnikov, V.; Cachin, C.; Christidis, K.; De Caro, A.; Enyeart, D.; Ferris, C.; Laventman, G.; Manevich, Y. Hyperledger fabric: A distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference, Porto, Portugal, 23–26 April 2018; pp. 1–15.
30. Zhang, P.; Liang, A.; Wang, D.; Yang, C.; Wu, X.; Yin, W. Gb 5009.268-2016 Verification of the determination method of sodium and lead in food. *Mod. Prev. Med.* **2017**, *44*, 4256–4262.
31. Bojko, A.A. Informative or misleading? Heatmaps deconstructed. In Proceedings of the International Conference on Human-Computer Interaction, San Diego, CA, USA, 19–24 July 2009; Springer: Berlin, Germany, 2009; pp. 30–39.
32. Pedersen, T.; Purandare, A.; Kulkarni, A. Name discrimination by clustering similar contexts. In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, Mexico, 13–19 February 2005; Springer: Berlin, Germany, 2005; pp. 226–237.
33. Kranen, P.; Kremer, H.; Jansen, T.; Seidl, T.; Bifet, A.; Holmes, G.; Pfahringer, B. Clustering performance on evolving data streams: Assessing algorithms and evaluation measures within moa. In Proceedings of the 2010 IEEE International Conference on Data Mining Workshops, Sydney, Australia, 13 December 2010; pp. 1400–1403.
34. Aslam, N.; Xia, K.; Haider, M.T.; Hadi, M.U. Energy-aware adaptive weighted grid clustering algorithm for renewable wireless sensor networks. *Future Internet* **2017**, *9*, 54. [[CrossRef](#)]
35. Chen, Y.; Li, F.; Chen, J.; Du, B.; Choo, K.-K.R.; Hassan, H. Epls: A novel feature extraction method for migration data clustering. *J. Parallel Distrib. Comput.* **2017**, *103*, 96–103. [[CrossRef](#)]
36. Bannister, M.J.; Eppstein, D.; Goodrich, M.T.; Trott, L. Force-directed graph drawing using social gravity and scaling. In Proceedings of the International Symposium on Graph Drawing, Redmond, WA, USA, 19–21 September 2012; Springer: Berlin, Germany, 2012; pp. 414–425.
37. Walshaw, C. A multilevel algorithm for force-directed graph drawing. In Proceedings of the International Symposium on Graph Drawing, Williamsburg, VA, USA, 20–23 September 2000; Springer: Berlin, Germany, 2000; pp. 171–182.
38. Wang, S.; Yuan, Y.; Wang, X.; Li, J.; Qin, R.; Wang, F.-Y. An overview of smart contract: Architecture, applications, and future trends. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 108–113.
39. Mao, D.; Hao, Z.; Wang, Y.; Fu, S. A novel dynamic dispatching method for bicycle-sharing system. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 117. [[CrossRef](#)]
40. Mao, D.; Hao, Z. A novel sketch-based three-dimensional shape retrieval method using multi-view convolutional neural network. *Symmetry* **2019**, *11*, 703. [[CrossRef](#)]
41. Kaufmann, W. *Going by the Book: The Problem of Regulatory Unreasonableness*; Routledge: New York, NY, USA, 2017; pp. 119–212.

