Modeling and Bayesian Inference of Traceability in Food Supply Networks

Xue-Ting Zhang[1], **Jia-Hao Su**[1], **Ming Fang**[1], **Jiang Jiang**[1], **Ying-Wu Chen**[1], **and Xin Lu**[1,2,3,*]

[1]College of Information System and Management, National University of Defense Technology, Changsha, 410073, China.
[2]Department of Public Health Sciences, Karolinska Institutet, Stockholm, 17177, Sweden.
[3]Flowminder Foundation, Stockholm, 17177, Sweden.
[*]xin.lu@flowminder.org

ABSTRACT

The centralization of food production and globalization of trade have made foodborne disease a widespread public health problem in both developed and developing countries. Moreover, the complexity of Food Supply Networks (FSNs), together with the high cost of foodborne illness, demonstrates the acute need of innovative solutions to identify the outbreak source. By modeling the contamination procedure through a three-layer network, we construct the backward tracing matrix based on the transaction data with Bayesian Inference. Then, we propose Traceability Entropy (TE) as an estimator to measure the ability of supporting source identification. Additionally, we design a Tracing-back Algorithm to reduce the identification uncertainty where TE contribution and choosing strategies help ensure the best quarantine retailers. In the case study of Nanjing city from the Important Product Traceability System (IPTS) implemented by Commerce Department, we retrieve the quarantine sequence when one retailer is observed infected, and further verify the effectivity and generality of this algorithm by simulation of each retailer infected in the FSN. The statistical result reveals that optimistic strategy has better performance than pessimistic and hybrid strategy, where over 60% initial contaminated retailers narrow down the source identification to two farms within three steps.

Introduction

In many real world systems with spreading dynamics, such as internet virus transmission networks[1], human contact networks[2], food supply chains (FSCs)[3], it is critical to develop models and algorithms to improve the ability to trace the source of contaminated food or spreading diseases[4]. Traceability is of great significance for both public health management and emergency risk analysis, as damage can be prevented or reduced if the contamination source is identified efficiently[5–7].

Food as necessitiy of daily life affecting the national economy and human livelihoods, FSCs have long been the center of traceability studies. According to International Standard Organization (ISO 8402), the general definition of traceability is "the ability to trace the history, application or location of an entity by means of recorded identifications"[8]. However, due to the globalization of food circulation and the complexity of food production, the complexity of network[9] aggravates numerously, making the traceability problem both costly and computationally intensive[10].

In order to improve the food safety supervision, many countries have made efforts in both legislation and technology in the past years. Nowadays, mandatory food traceability laws are being enforced and the European Union has realized this by introducing the General Food Law (GFL)[11,12]. Accordingly, traceability data can be mandatory or optional, for which mandatory data include product ID, supplier ID, quantity, prize, buyer ID and optional data include supplier's information, receipt date, country of origin, date of pack, buyer's name, etc.[13] When it comes to the U.S., Food Safety Modernization Act (FSMA) was signed into law to expand the authority of the U.S . Food and Drug Administration[14,15], such as mandatory recall authority and the responsibility to develop a third-party audit system for the certification of imported foods, increasing the frequency of mandatory inspections, developing regulations for preventive control plans, establishing a product tracing system[16].

A variety of technologies of collection, storage, analysis and retrieval of transmission data empower the traceability of food items and ingredients among the distribution networks[17,18], such as alphanumerical codes, bar codes, and radio frequency identification (RFID)[19,20]. The food supply chain tracking information thus allows investigators to build early warning models for detecting abnormities of food safety issues[21], and tracing back contaminations as to support emergency management in food safety control[22,23]. These studies point out that developing effective and full chain Important Product Traceability

System (IPTS) is quite complex, which requires theoretical analysis of the uncertain information from methodological aspects[24]. These methods include mathematical models for identifying contamination sources[25], planning models for optimizing food distribution[26, 27], and simulation models for epidemic transmission[28].

In recent years, the PRC Ministry of Commerce constructed the IPTS in more than 13,500 corporations and 200,000 shops, involving meat, vegetables, Chinese medicinal crop and wine. More than 2 billion items of data have been collected so far, and the IPTS receives 3 million items of daily data. For a single product (e.g., pork), this system is an integrated supply information chain comprised of segments of household breeding, slaughter, wholesaler, retailer, etc.[29] Traditional researches focus on qualitative description of the traceability system or analyze the traceability problem from the linear supply chain view. In fact, for many IPTS with complicated connected relationships, it is easy to observe the symptoms of infection but difficult to identify the contamination source[30, 31].

In this paper, we propose a tracing back algorithm to ensure the potential source by quarantine process with the estimator (traceability entropy contribution) and several rational choosing strategies. Furthermore, Monte-Carlo simulation on the real pork supply network of Nanjing city during 2015 reveals that the tracing back algorithm reduces the traceability uncertainty to an acceptable range within three detection steps, and we find the best choosing strategy is optimistic strategy by comparison of stochastic contamination conditions. Overall, we elaborate how, and to what extent Bayesian Inference of TE identifies the outbreak source in FSN.

Results

Tracing Back Model with Bayesian Inference on Food Supply Network

Food supply chain is composed of a wide diversity of producers, processors, distributors and retailers[32], for which the spatial distribution makes inspection and quarantine work harder. Since the distribution processes of one product from production to consumption can vary considerably, without loss of generality[33], we take a sample model of pork by aggregating the underlying transaction network into the categories of Farms (F), Slaughters (S), and Retailers (R). For simplification, pork is produced by farms, transported to different slaughters, mixed with raw pigs from other farms, and then sold to customers at retailers (see Fig. 1). It is thereby straight forward to extend the current model to more layers[34].

Assume food in each node is mixed intensively, the probability of infection is then proportional to the amount of contaminated food. Given the amount of pork produced in the ith farm is f_i, $1 \leq i \leq I$, where I is the total number of farms, and the pork received at the jth slaughter is s_j, $1 \leq j \leq J$, where J is the total number of slaughters, the consumption amount of the kth retailer is r_k, $1 \leq k \leq K$, where K is the total number of retailers. Let F, S, R be the row vectors consist of the product amount, and let FS be the priori transition probability matrix where each element fs_{ij} denotes the proportion of pork transporting from farm i to slaughter j, then SR is the priori transition probability matrix where each element sr_{jk} denotes the proportion of pork transporting from slaughter j to retailer k. Then the calculation formulas are as follows: $F = [f_1, f_2, \cdots f_i, \cdots, f_I]$, $1 \leq i \leq I$; $S = F \times FS = [s_1, s_2, \cdots s_j, \cdots, s_J]$, $1 \leq j \leq J$; $R = F \times FS \times SR = [r_1, r_2, \cdots r_k, \cdots, r_K]$, $1 \leq k \leq K$, where

$$FS = \begin{bmatrix} fs_{11} & fs_{12} & \cdots & fs_{1J} \\ fs_{21} & fs_{22} & \cdots & fs_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ fs_{J1} & fs_{J2} & \cdots & fs_{IJ} \end{bmatrix}, \tag{1}$$

$$SR = \begin{bmatrix} sr_{11} & sr_{12} & \cdots & sr_{1K} \\ sr_{21} & sr_{22} & \cdots & sr_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ sr_{J1} & sr_{J2} & \cdots & sr_{JK} \end{bmatrix}. \tag{2}$$

Based on the above network model of supply chain, we can analyze the tracing back problem with Bayesian Inference[35, 36]. Given an observation of contamination at retailer layer and some knowledge of the topology and fluxes about the FSN, it is possible to induce the probability of each farm to be the contamination source. Bayesian Inference is a statistical approach to calculate the probability of one event based on conditions that might be related to it[37]. Known that FS is the priori probability matrix from F to S, then the posterior probability of SF can be calculated by the method of Bayes Inference. Also,

$$P(F_i | S_j) = \frac{P(F_i)P(S_j | F_i)}{P(S_j)} = \frac{F_i \times FS_{ij}}{S_j} = SF_{ji}, \tag{3}$$
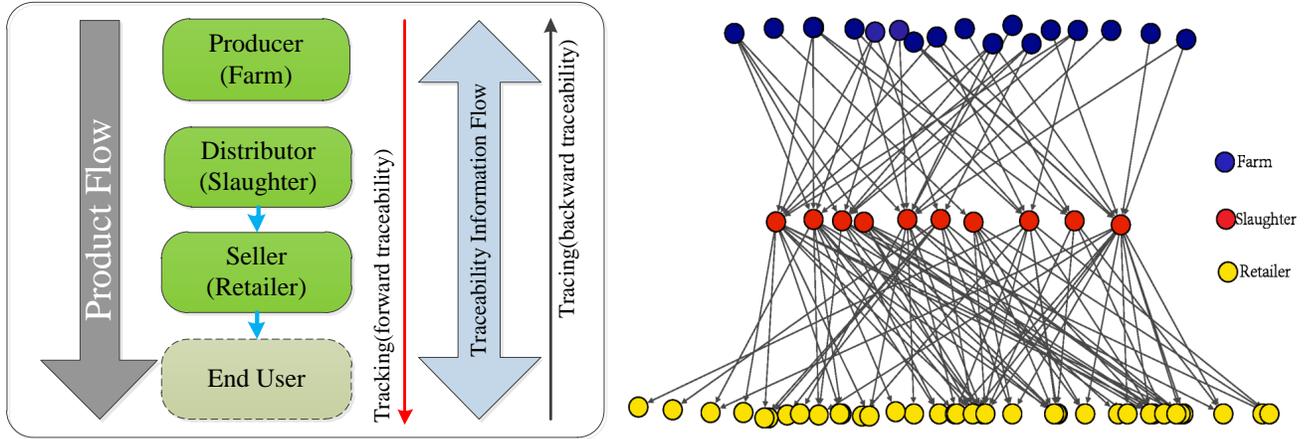
Figure 1. The Farm-Slaughter-Retailer Network and the schematic of product and traceability information flow. The traceability information is shared in two directions, and we focus on the backward tracing traceability for the source identification problem.

$$P(S_j|R_k) = \frac{P(S_j)P(S_j|R_k)}{P(R_k)} = \frac{S_j \times SR_{jk}}{R_k} = RS_{kj}. \tag{4}$$

Assume there is only one contamination source (F_i) in the pork supply network, then the posterior probability of R_k being the outbreak source is:

$$P(F_i|R_k) = \sum_{j=1}^{J} P(S_j|R_k) \times P(F_i|S_j) = \sum_{j=1}^{J} RS_{kj} \times SF_{ji} = (e_k{}^T \times RS) \times (SF \times e_i) = RF_{ki}. \tag{5}$$

Finally, the posterior probability calculation can be extended to the condition of any Retailer from any Farm by matrix calculation, as $RF = RS \times SF$.

Tracing Back Algorithm based on Traceability Entropy

Tracing Back Algorithm aims to solve the source identification problem by detecting the potential retailers. First, we shall figure out how the product and information spreads[38] in the traceability system (Fig. 1), after that we depict the tracing back procedure in mathematical model with Bayesian Inference, and then carry out the quarantine results to reduce the identity uncertainty with the least cost. In brief, there is one retailer detected as initial infected observation, then we shall detect the products in other retailers to affirm the real outbreak farm by tracing back algorithm. Traceability Entropy (TE) is proposed to evaluate the uncertainty during the source identification process[39](see Methods), and contribution (ΔTE) of each retailer demonstrates how well one more detection enhances the network traceability.

Known that R_{k_1} is infected, and the potential contamination set of R_{k_1} is $\Omega_{R_{k_1}}$, it is capable to find the potential contamination sub-network through tracing backward and tracking forward, and the process of which is presented as Fig. 2.

The tracing back algorithm takes place in the following steps:

Step 1: Given the initial retailer R_{k_1} as the contamination observation;

Step 2: Acquire the potential contamination set $\Omega_{R_{k_1}}$, and the number of which is n_{k_1};

Step 3: Calculate TE and ΔTE for each non-detected retailers in $\Omega_{R_{k_1}}$;

Step 4: Choose the retailer with best contribution as next detected retailer by optimistic/ pessimistic/ hybrid strategy, retrieve the retailer number of which is k_t, save it in the quarantine sequence tree (see details in case study);

Step 5: Ensure the quarantine result as True/False, save the TE of the network and update the new probability matrix;

Step 6: Check whether it reaches any of three terminate conditions: the traceability entropy is less than 1; the traceability entropy does not decline anymore; or the number of iteration reaches the default value. If not, go back to Step 2. Otherwise go to Step 7;

Step 7: Output the quarantine sequence tree of R_{k_1}, the final network traceability entropy, and the potential source farms;
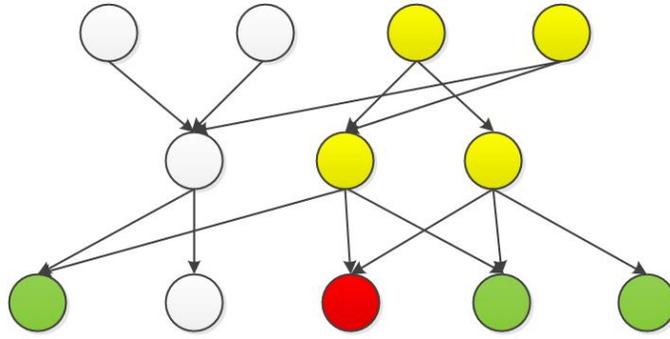
Figure 2. The tracing and tracking procedure of contaminated sub-network. Supposing the initial infected observation is red, then the potential contaminated sub-network is comprised of yellow and green nodes (we find yellow nodes by backward tracing and green retailers by forward tracking), here green nodes make up the potential contamination set $\Omega_{R_{k_1}}$.
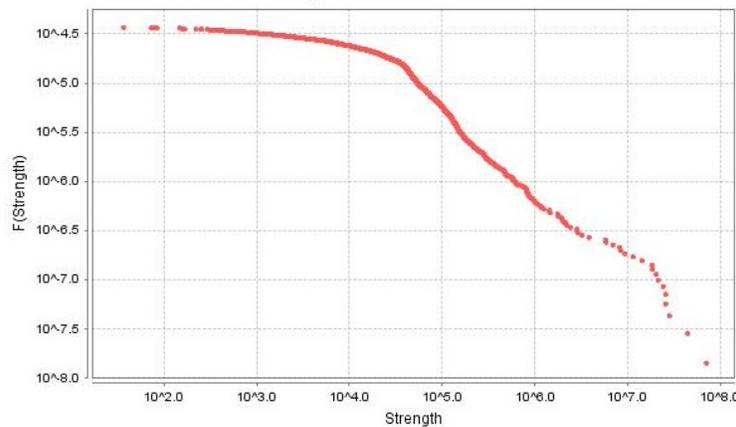
Figure 3. Weighted degree distribution of Pork Supply Network in Nanjing, which follows a long tail power law.

Case Study

Dataset

For illustration, we retrieve the farm, slaughter, distribution and product circulation data of pork in Nanjing in 2015 from the IPTS system (including 766,580 items), which allows us building the Food Supply Network, specifically speaking, the pork supply network. The weighted degree distribution of this FSN follows a heterogeneous long tail power law as shown in Fig. 3.

The complexity of tracing back algorithm is $OIJR^2$, and in this case the amount of retailers is 2350, the square of which is more than 5 million. In order to reduce the calculation complexity, the retailer nodes from and only from the same slaughter are classified as one cluster and the retailer with the largest degree represents the cluster, since all nodes in the cluster have the same topological features and only different transaction amount affects TE. The comparison of the network structure before and after node merging is shown in Fig. 4 , where the new retailers are numbered from 1 to 375.

Quarantine scheme with one infected retailer

Supposing that one retailer is observed contaminated in this real network, we carry out the tracing back algorithm to design an effective quarantine schemes to ensure the correct source farm based on traceability entropy contribution.

In this case, we calculate the TE of the initial infected retailer 227 as 4.0221, and obtain its potential contamination set as Ω_{227} of 338 retailers/retailer clusters. After the first contribution sorting of potential retailers in Ω_{227}, taking pessimistic strategy as an example, the next chosen retailer R_{k_2} is number 129, and the detection result is infected or not (True/False). If the detection result is True, the new TE is 2.5494, and the contribution is 1.4727; if the detection result is False, the new TE is 2.7858, and the contribution is 1.2363.

According to different quarantine results of R_{k_2}, we can calculate the number of next chosen retailer R_{k_3}. When the quarantine result of R_{k_2} is True, the next retailer R_{k_3} is number 8. If the detection result of R_8 is True, the new TE is 0, and the contribution is 2.5494; if the detection result of R_8 is False, the new TE is 2.5494, and the contribution is 0. On the contrary, if

(a)   **(b)**
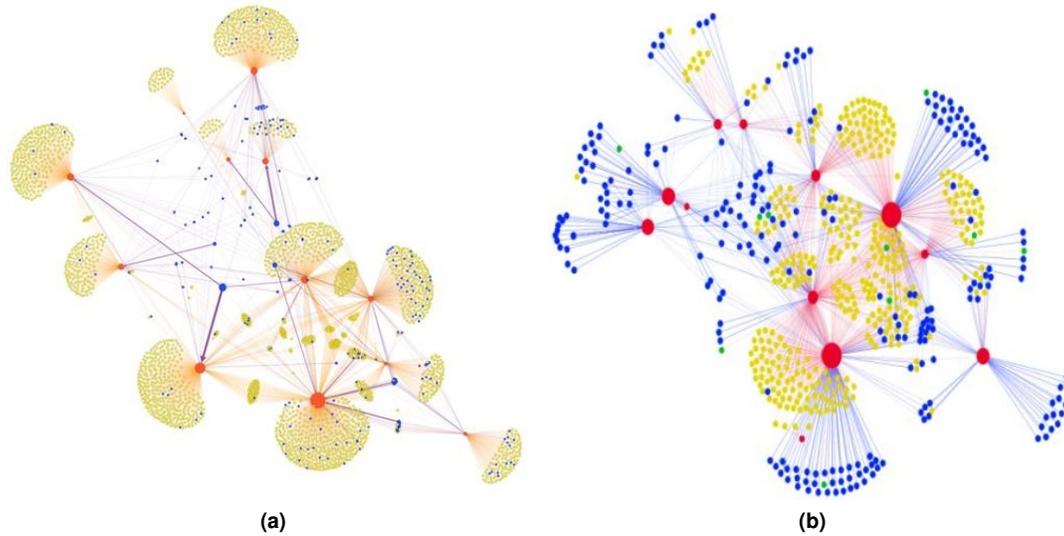
Figure 4. Comparison charts before and after node merging of Pork Supply Network in Nanjing, 2015. (a) The initial network is heterogeneous before node merging, where retailers are yellow, slaughter are red and farms are blue. (b) The network after retailers merging is less intensive and complicated.

the quarantine result of R_{k_2} is False, the next retailer R_{k_3} is number 129, which is a cluster with new value. If the detection result of R_{129} is True, the new TE is 2.5494, and the contribution is 0.2364; if the detection result of R_{129} is False, the new TE is 2.7838, and the contribution is 0.0020.

By analogy, we can get a quarantine sequence tree with pessimistic strategy, and also optimistic strategy and hybrid strategy separately as shown in Fig. 5 .

Comparison of different strategies with stochastic contamination

To further study the generality and robustness of tracing back algorithm, avoiding the specific observation influence, we carry out stochastic contamination simulation of all 375 retailers/retailer clusters, and calculate the quarantine results.

The process to generate the quarantine sequence tree is as follow: first, calculate the traceability entropy of the contamination retailer, then choose the next detected observation with maximum contribution; After that, calculate the entropy contribution according to the quarantine result (T/F), and verify the next detected observation. These steps are repeated until the termination conditions are met. Ultimately, the algorithm outputs the quarantine sequence tree (as shown above in Fig. 5) and the ultimate network traceability entropy. Fig. 6 depicts the heat map of pessimistic strategy, optimistic strategy and hybrid strategy after each detection step, respectively, where the result of optimistic strategy has the overall best result for both different conditions and different initial infected nodes, and the pessimistic result is the worst.

Further statistical analysis reveals that TE is a good index to depict the network traceability and the quarantine sequence tree result has realistic significance of source identification and epidemic controlling. After three-step quarantine process, we find the number and the proportion of initial infected retailers which is able to decrease TE less than 1 in different quarantine result conditions. The details of different strategies are shown in Tab. 1. Overall, it is obvious that optimistic strategy has better performance than pessimistic strategy, but all three strategies receive good traceability results.

It is notable that during the quarantine process, there is more contribution of truth detection result than false detection result, which means "truth verification is much easier than false verification". Specifically speaking, if one product is detected as infectious, this retailer is verified to be infected, then the same source slaughter of all infected retailers is deduced as infected; however, even if the products detected are not infectious, this retailer is still hard to be verified without any infected risk, which means the source slaughter cannot be ruled out.

Discussion

By analysis of the Food Supply Network (FSN) with real data in Nanjing (766580 items) from the IPTS, we find that Bayesian Inference on the 3-layer network model constitutes an effective tracing back algorithm of outbreak source identification, and it is possible to carry out rational quarantine tree schemes for stochastic contamination. Furthermore, different choosing strategies of next detected retailer are analyzed contrastively, where optimistic strategy reaches the best result that there is over 60%
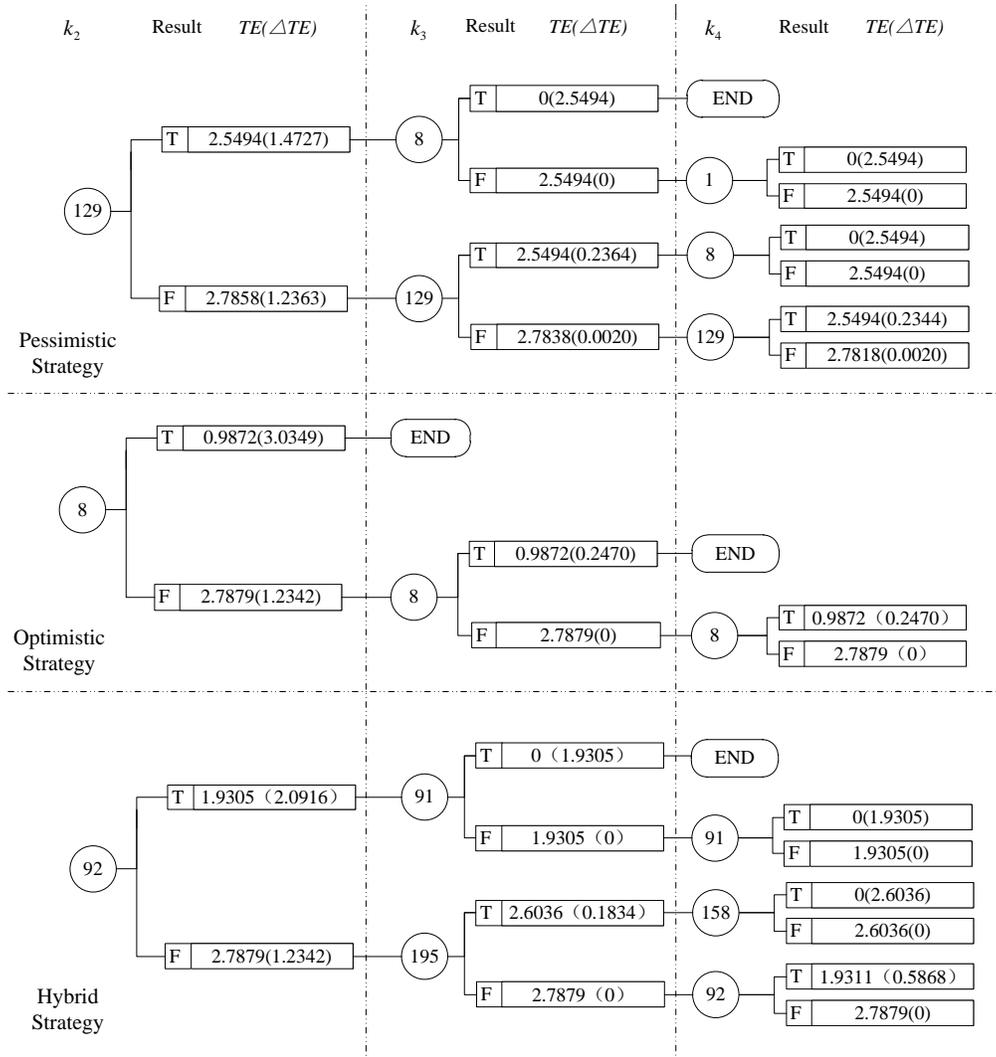
Figure 5. **Quarantine sequence tree with optimistic strategy, pessimistic strategy and hybrid strategy**

		TTT	TTF	TFT	TFF	FTT	FTF	FFT	FFF
Pessimistic Strategy	number ($TE \leq 1$)	373	373	373	54	73	31	30	8
	proportion	99.47%	99.47%	99.47%	14.40%	19.47%	8.27%	8.00%	2.13%
Optimistic Strategy	number ($TE \leq 1$)	375	375	375	109	367	101	101	8
	proportion	100%	100%	100%	29.07%	97.87%	26.93%	26.93%	2.13%
Hybrid Strategy	number ($TE \leq 1$)	375	375	375	109	367	101	101	8
	proportion	100%	81.07%	81.07%	25.07%	94.67%	20.80%	22.93%	2.13%

Table 1. **Results of different quarantine choosing strategies.** The average proportion of acceptable $TE(TE \leq 1)$ for pessimistic strategy, optimistic strategy and hybrid strategy is 43.83%, 60.37% and 53.47%, respectively. And the average TE of the three strategies is 1.3383, 1.0631 and 1.2074, respectively.

probability to decline the traceability entropy less than 1 within 3 steps (meaning that the source is likely between two farms).

In this paper, the FSN is modeled as 3-layer Directed Weighted Acyclical Graph (DWAG), where the transaction data is applied to build the priority probability matrices, and Bayesian Inference assists generate the posterior probability matrices of tracing back information. For the purpose of this study, we propose Traceability Entropy to evaluate how the quarantine

result supports the identification of outbreak source, which is further proved that not affected by the specific observation. From a realistic perspective, once food crisis breaks out, it would be essential to quarantine the products on sale to ensure the contamination source[40, 41]. We henceforth propose the tracing back algorithm to implement quarantine scheme with the best traceability contribution and scientific choosing strategies adapted from Decision theory. Then, we carry out the quarantine sequence tree of specific contamination observation and the final TE of the network, providing a scientific method to identify the outbreak source with the least quarantine cost. Furthermore, the results of different strategies are contrastive analyzed with stochastic contamination, which is useful to find the best choosing strategy to find next detected object.

Our findings highlight the importance of taking the network structure into account in the study of food supply networks[42]. According to the real data from IPTS, further research could focus on more complex scenarios on different types of network, for example, uniform network, exponential network, scale-free network, homogeneous network, etc.[31] The joint effect of various network parameters (e.g., degree distribution, density, community) on the network traceability also enlarges possibilities for network optimization studies from another point of view. The IPTS data is day-by-day recorded, we are however limited to the ignorance of time influence, which is useful to narrow down the uncertainties of potential outbreak sources.

In summary, this paper is, to the best of our knowledge, the first attempt to propose Traceability Entropy as an estimator in FSN research, and carry out tracing back algorithm applied on a large real FSN data set. Traceability entropy is an effective estimator reflecting the traceability of the network itself without affected by specific observation. In addition, contribution and choosing strategies provide a feasible method to design quarantine scheme with high efficiency. Our results not only show the traceability of FSN can be improved by scientific tracing back algorithm, but also reveal that TE can apply to other traceability problems involving Directed Weighted Acyclical Graph (DWAG), such as disease contagion, virus infection, or rumor spreading.

Methods

Traceability Entropy

Entropy is a measure of disorder or uncertainty in the Information Theory[43], and Traceability Entropy (TE) is defined in this paper as the measure for the uncertainty in determining the outbreak source. In the tracing-back case, consider a game where we are allowed to ask True/False questions until we pinpoint the contamination source farm[44]. If there is less uncertainty about F_i, we obtain the value with fewer questions, meaning that the network has higher traceability. For this reason, we can see traceability entropy as the number of True/False questions we need to ask to identify the correct outbreak source. Turning this around, 2^E is the uncertainty of the source (the number of possible sources).

In term of R_k, the calculation of Traceability Entropy is as follow,

$$TE(k) = -\sum_{i=1}^{I} P(F_i|R_k)\log(P(F_i|R_k)),\tag{6}$$

where $P(F_i|R_k)$ is the posterior probability when R_k is the outbreak source and F_i is the contamination observation.
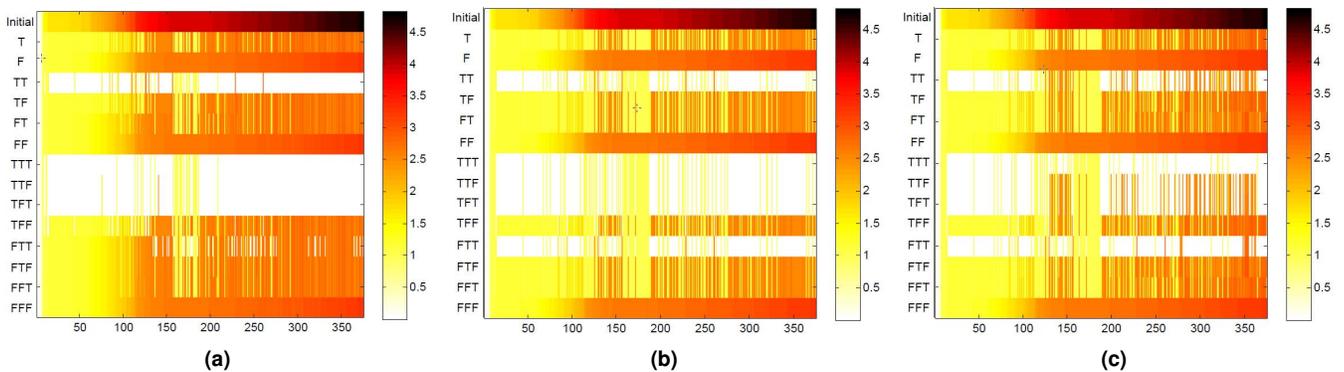
Figure 6. Heat map of different choosing strategies. (a) Pessimistic strategy has best performance when the detection results are most true, but not robust in other conditions. (b) Optimistic strategy has the overall beast effect for both different conditions and different initial infected nodes. (c) Hybrid strategy fits different conditions better than pessimistic strategy, and more uniform for different retailers.

Apparently, when the supply network forms a tree structure, each retailer only receives product from the same farm, $TE = 0$; but when the number of links between each layer is maximized and products are equally produced and distributed, each retailer will receive the same product amount from all farms, $TE = \log I$. For any supply network, given one infected retailer R_k, the uncertainty in identifying F_i as the outbreak source (assuming the contaminant is produced in only one farm) is similar to picking one out of $2^{TE(k)}$ farms randomly. Consequently, TE provides a universal measurement for network traceability, and may be used to guide the construction of food supply infrastructures[45]. Once given TE, the system is predictable as choosing one randomly among 2TE sources, thus the expected prediction rate is $1/2^{TE}$. To further prove the effectiveness of the notification of TE without the influence of specific observations, here is a simple example.

For example, given distributions A $= \{0.5, 0.49, 0.01\}$ and B $= \{0.5, 0.25, 0.25\}$, the TE for A is $TE_A = -(0.5 \times log(0.5) + 0.49 \times log(0.49) + 0.01 \times log(0.01)) = 0.7422$, and the expected prediction rate is $1/2^{TE_A} = 0.5798$; the TE for B is $TE_B = -(0.5 \times log(0.5) + 0.25 \times log(0.25) + 0.25 \times log(0.25)) = 1.0397$, and the expected prediction rate is $1/2^{TE_B} = 0.4864$.

We can see that the entropy and predictability distinguish the differences of traceability uncertainty in A and B properly[46]. Assuming we use a simple predictability algorithm to predict: always predict the source to be in the top X farms with highest probability, then when $X = 1$, prediction rate on both A and B will be 0.5; when $X = 2$, prediction rate becomes 0.99 and 0.75, separately.

The above example demonstrates the accuracy difference between TE and traditional prediction algorithm[47], as distribution A is more predictable than B, which can be measured from TE but not the prediction algorithm when $X = 1$.

Traceability Entropy Contribution

Additional contamination observations could rule out farms that are unable to reach some contamination retailers, which reduces the uncertainties in identifying the outbreak source and enhance traceability. However, more quarantine objects aggrandize the cost so that we need better traceability with less observation amount. Consequently, it is essential to define the traceability gain of one more observation as Contribution.

If R_{k_1} has already been infected, and we find that $R_{k_2}(k_2 \neq k_1)$ is also infected, the probability of S_j being infected is

$$P(S_j|R_{k_1},R_{k_2}) = \frac{P(R_{k_1}|S_j)P(S_j)P(R_{k_2}|S_j,S_{k_1})}{P(S_{k_1})P(R_{k_2}|R_{k_1})} = \frac{P(R_{k_1}|S_j)P(S_j)}{P(R_{k_1})} \times \frac{P(R_{k_2}|S_j)}{P(R_{k_2})} = \frac{sr_{jk_1} \times s_j}{r_{k_1}} \times \frac{sr_{jk_2}}{r_{k_2}} = rs_{k_1j} \times \frac{sr_{jk_2}}{r_{k_2}} = RS^{k_1}_{k_2j} , \quad (7)$$

where RS^{k_1} is the RS matrix of the k_1th retailer, and $RS^{k_1}_{k_2j}$ is the element at the k_2th row and the jth column of RS^{k_1} when R_{k_2} is detected as infected.

Similarly, the probability of F_i being the outbreak source is:

$$P(F_i|R_{k_1},R_{k_2}) = \sum_{j=1}^{J} P(S_j|R_{k_1},R_{k_2}) \times P(F_i|S_j) = \sum_{j=1}^{J} RS^{k_1}_{k_2j} \times sf_{ji} = (e_{k_2}{}^T \times RS^{k_1}) \times (SF \times e_i) . \quad (8)$$

Consequently, the corresponding Traceability Entropy of network $TE(k_1,k_2) = -\sum_{i=1}^{I} P(F_i|R_{k_1},R_{k_2})\log(P(F_i|R_{k_1},R_{k_2}))$.
And the contribution gained by R_{k_2} is defined as $\Delta TE(k_2) = TE(k_1,k_2) - TE(k_1)$.

On the contrary, if we find that R_{k_2} is not infected, the probability of S_j being infected is:

$$P(S_j|R_{k_1},\overline{R_{k_2}}) = \frac{P(R_{k_1}|S_j)P(S_j)}{P(R_{k_1})} \times \frac{P(\overline{R_{k_2}}|S_j)}{P(\overline{R_{k_2}})} = \frac{P(R_{k_1}|S_j)P(S_j)}{P(R_{k_1})} \times \frac{1-P(R_{k_2}|S_j)}{1-P(R_{k_2})} = \frac{sr_{jk_1} \times s_j}{r_{k_1}} \times \frac{1-sr_{jk_2}}{1-r_{k_2}} = rs_{k_1j} \times \frac{1-sr_{jk_2}}{1-r_{k_2}} = RS^{k_1}_{\overline{k_2},j} , \quad (9)$$

where RS^{k_1} is the RS matrix of the k_1th retailer, and $RS^{k_1}_{\overline{k_2},j}$ is the element at the k_2th row and the jth column of RS^{k_1} when R_{k_2} is detected as not infected.

Then, the probability of the ith Farm being the outbreak source is:

$$P(F_i|R_{k_1},\overline{R_{k_2}}) = \sum_{j=1}^{J} P(S_j|R_{k_1},\overline{R_{k_2}}) \times P(F_i|S_j) = \sum_{j=1}^{J} RS^{k_1}_{\overline{k_2},j} \times sf_{ji} = (e_{k_2}{}^T \times RS^{k_1}_{\overline{k_2}}) \times (SF \times e_i) . \quad (10)$$

Consequently, the corresponding Traceability Entropy of network $TE(k_1,\overline{k_2}) = -\sum_{i=1}^{I} P(F_i|R_{k_1},\overline{R_{k_2}})\log(P(F_i|R_{k_1},\overline{R_{k_2}}))$.
And the contribution gained by R_{k_2} is $\Delta TE(\overline{k_2}) = TE(k_1) - TE(k_1,\overline{k_2})$.

Quarantine Strategy

In consideration of two possible situations (True/False) after detecting one potential infected retailer, we apply three classic strategies adapted from Decision theory, to find the optimum quarantine solution: pessimistic strategy , optimistic strategy, and hybrid strategy[48].

Pessimistic strategy always chooses the maximum contribution from the test sample $k_2, k_3, ..., k_r \in \Omega_{k_1}$ with the minimum possible value of detection result, ensuring it is optimal event though the result of True/False detection is worse. Consequently, the corresponding retailer $R_{k_t}(k_t \in \Omega_{k_1})$ is the next detection target, when

$$k_t = \arg\max\min\{\Delta TE(k_t), \Delta TE(\overline{k_t})\}, k_t \in \Omega_{k_1}. \tag{11}$$

Similarly, optimistic strategy always chooses the maximum value from contribution from the test sample within the maximum possible value of detection result, thus the next detection retailer R_{k_t} is

$$k_t = \arg\max\max\{\Delta TE(k_t), \Delta TE(\overline{k_t})\}, k_t \in \Omega_{k_1}. \tag{12}$$

Hybrid strategy always takes into account of both possible T/F results, and choose the maximum weighted traceability contributions from the test sample, when

$$k_t = \arg\max(P(k_t)\Delta TE(k_t) + (1 - P(k_t))\Delta TE(\overline{k_t})), k_t \in \Omega_{k_1}. \tag{13}$$

References

1. Li, T., Wang, Y. & Guan, Z. H. Spreading dynamics of a siqrs epidemic model on scale-free networks. *Communications in Nonlinear Science & Numerical Simulation* **19**, 686–692 (2014).

2. Mi, B. *et al.* Cooperative and competitive spreading dynamics on the human connectome. *Neuron* **86**, 1518–29 (2015).

3. Dabbene, F., Gay, P. & Tortia, C. Traceability issues in food supply chain management: A review. *Biosystems Engineering* **120**, 65–80 (2014).

4. Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z. & Baranyi, J. Complexity of the international agro-food trade network and its impact on food safety. *PLoS One* **7**, 810–837 (2012).

5. Canavari, M., Centonze, R., Hingley, M. & Spadoni, R. Traceability as part of competitive strategy in the fruit supply chain. *British Food Journal* **112**, 171–186 (2010).

6. Parreño-Marchante, A., Alvarez-Melcon, A., Trebar, M. & Filippin, P. Advanced traceability system in aquaculture supply chain. *Journal of Food Engineering* **122**, 99–109 (2014).

7. Larraín, M. A., Díaz, N. F., Lamas, C., Uribe, C. & Araneda, C. Traceability of mussel (mytilus chilensis) in southern chile using microsatellite molecular markers and assignment algorithms. exploratory survey. *Food Research International* **62**, 104–110 (2014).

8. ISO. Quality management and quality assurance; vocabulary. *International Standard Iso* **8402**, 74–75 (1994).

9. Standish, R. K. Complexity of networks. *Complexity* **17**, 50–61 (2005).

10. Bender, S. F., Valadares, R. B. D. S. & Taudiere, A. Mycorrhizas dynamic and complex networks of power and influence. *New Phytologist* **204**, 15–18 (2014).

11. V. Kher, S. *et al.* Experts' perspectives on the implementation of traceability in europe. *British Food Journal* **112**, 261–274 (2010).

12. Bukeviciute, L., Dierx, A. & Ilzkovitz, F. *The functioning of the food supply chain and its effect on food prices in the European Union*. 47 (Office for infrastructures and logistics of the European Communities, 2009).

13. Folinas, D., Manikas, I. & Manos, B. Traceability data management for food chains. *British Food Journal* **108**, 622–633 (2006).

14. Kumar, S., Heustis, D. & Graham, J. M. The future of traceability within the us food industry supply chain: a business case. *International Journal of Productivity and Performance Management* **64**, 129–146 (2015).

15. Strauss, D. M. An analysis of the fda food safety modernization act: protection for consumers and boon for business. *Food and drug law journal* **66**, 353–376 (2011).

16. Regattieri, A., Gamberi, M. & Manzini, R. Traceability of food products: General framework and experimental evidence. *Journal of food engineering* **81**, 347–356 (2007).

17. Yao, Y., Palmer, J. & Dresner, M. An interorganizational perspective on the use of electronically enabled supply chains. *Decision support systems* **43**, 884–896 (2007).

18. Costa, C. *et al.* A review on agrifood supply chain traceability by means of rfid technology. *Food and Bioprocess Technology* **6**, 353–366 (2013).

19. Liang, W., Cao, J., Fan, Y., Zhu, K. & Dai, Q. Modeling and implementation of cattle or beef supply chain traceability using a distributed rfid-based framework in china. *PloS one* **10**, 39–58 (2015).

20. Piramuthu, S., Farahani, P. & Grunow, M. Rfid-generated traceability for contaminated product recall in perishable food supply networks. *European Journal of Operational Research* **225**, 253–262 (2013).

21. Brandenburg, M., Govindan, K., Sarkis, J. & Seuring, S. Quantitative models for sustainable supply chain management: Developments and directions. *European Journal of Operational Research* **233**, 299–312 (2014).

22. Resende-Filho, M. A. & Hurley, T. M. Information asymmetry and traceability incentives for food safety. *International Journal of Production Economics* **139**, 596–603 (2012).

23. Olsen, P. & Aschan, M. Reference method for analyzing material flow, information flow and information loss in food supply chains. *Trends in food science & technology* **21**, 313–320 (2010).

24. Hasani, A., Zegordi, S. H. & Nikbakhsh, E. Robust closed-loop supply chain network design for perishable goods in agile manufacturing under uncertainty. *International Journal of Production Research* **50**, 4649–4669 (2012).

25. Buscema, M., Grossi, E., Breda, M. & Jefferson, T. Outbreaks source: A new mathematical approach to identify their possible location. *Physica A: Statistical Mechanics and its Applications* **388**, 4736–4762 (2009).

26. Wang, X., Li, D., O brien, C. & Li, Y. A production planning model to reduce risk and improve operations management. *International Journal of Production Economics* **124**, 463–474 (2010).

27. Ahumada, O. & Villalobos, J. R. Application of planning models in the agri-food supply chain: A review. *European journal of Operational research* **196**, 1–20 (2009).

28. Chao, D., Halloran, M., Obenchain, V. & Jr, L. I. Flute, a publicly available stochastic influenza epidemic simulation model. *Plos Computational Biology* **6**, 656–672 (2010).

29. Wu, L. *et al.* Simulation of market demand for traceable pork with different levels of safety information: A case study in chinese consumers. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* **63**, 513–537 (2015).

30. Aung, M. M. & Chang, Y. S. Traceability in a food supply chain: Safety and quality perspectives. *Food control* **39**, 172–184 (2014).

31. Yu, M. & Nagurney, A. Competitive food supply chain networks with application to fresh produce. *European Journal of Operational Research* **224**, 273–282 (2013).

32. Manzini, R. & Accorsi, R. The new conceptual framework for food supply chain assessment. *Journal of Food Engineering* **115**, 251–263 (2013).

33. Eskandarpour, M., Dejax, P., Miemczyk, J. & Péton, O. Sustainable supply chain network design: an optimization-oriented review. *Omega* **54**, 11–32 (2015).

34. Soysal, M., Bloemhof-Ruwaard, J. & Van der Vorst, J. Modelling food logistics networks with emission considerations: The case of an international beef supply chain. *International Journal of Production Economics* **152**, 57–70 (2014).

35. Odbert, H. & Aspinall, W. Updating parameters for volcanic hazard assessment using multi-parameter monitoring data streams and bayesian belief networks. In *EGU General Assembly Conference Abstracts*, vol. 16, 56–65 (2014).

36. Shankar, J. *et al.* Using bayesian modelling to investigate factors governing antibiotic-induced candida albicans colonization of the gi tract. *Scientific reports* **5** (2015).

37. Park, J. & Yook, S.-H. Bayesian inference of natural rankings in incomplete competition networks. *Scientific reports* **4** (2014).

38. Bosona, T. & Gebresenbet, G. Food traceability as an integral part of logistics management in food and agricultural supply chain. *Food Control* **33**, 32–48 (2013).

39. Saikouk, T., Badraoui, I. & Spalanzani, A. The forest supply chain management: An entropic perspective. *Next Generation Supply Chains: Trends and Opportunities* **18**, 487 (2014).

40. Septiani, W., Marimin, M., Herdiyeni, Y. & Haditjaroko, L. Method and approach mapping for agri-food supply chain risk management: A literature review. *International Journal of Supply Chain Management* **5**, 51–64 (2016).

41. Beske, P., Land, A. & Seuring, S. Sustainable supply chain management practices and dynamic capabilities in the food industry: A critical analysis of the literature. *International Journal of Production Economics* **152**, 131–143 (2014).

42. Devika, K., Jafarian, A. & Nourbakhsh, V. Designing a sustainable closed-loop supply chain network based on triple bottom line approach: A comparison of metaheuristics hybridization techniques. *European Journal of Operational Research* **235**, 594–615 (2014).

43. Rothstein, J. Information and entropy. *Science* **114**, 171–175 (1951).

44. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences* **109**, 11576–11581 (2012).

45. West, J., Bianconi, G., Severini, S. & Teschendorff, A. E. Differential network entropy reveals cancer system hallmarks. *Scientific reports* **2** (2012).

46. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).

47. Newton, P. K. *et al.* Entropy, complexity, and markov diagrams for random walk cancer models. *Scientific reports* **4**, 58–75 (2014).

48. Isen, A. M. & Means, B. The influence of positive affect on decision-making strategy. *Social Cognition* **2**, 18–31 (2011).

Acknowledgements

Author contributions statement

X.L, J.J and Y.-W.C designed the research plan; X.-T.Z performed the research and wrote the paper; J.-H.S and M.F. conducted the experiments; X.L analyzed the results and proposed the improvement suggestions. All authors reviewed the manuscript.

Additional information