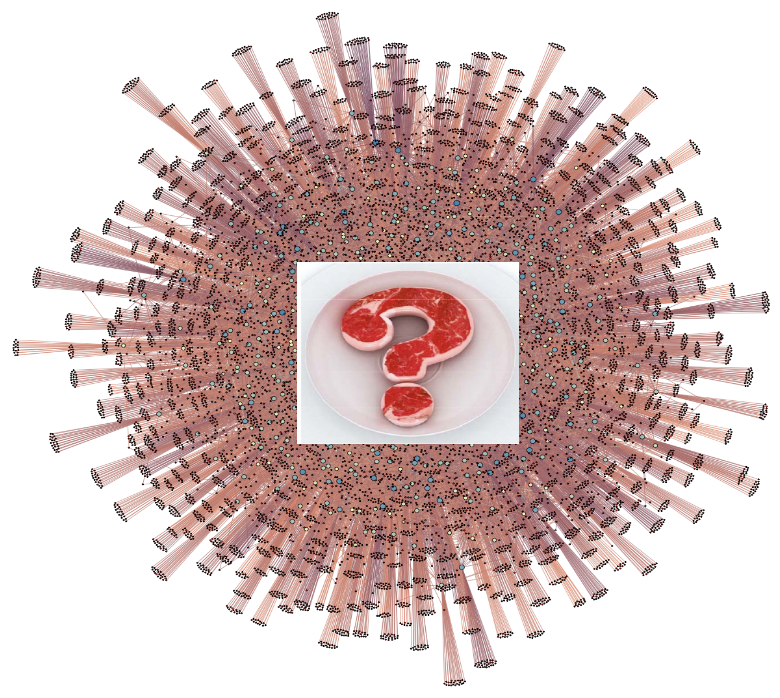
Modeling Supply Chain System Structure to Trace Sources of Food Contamination: Problem Framing Talk



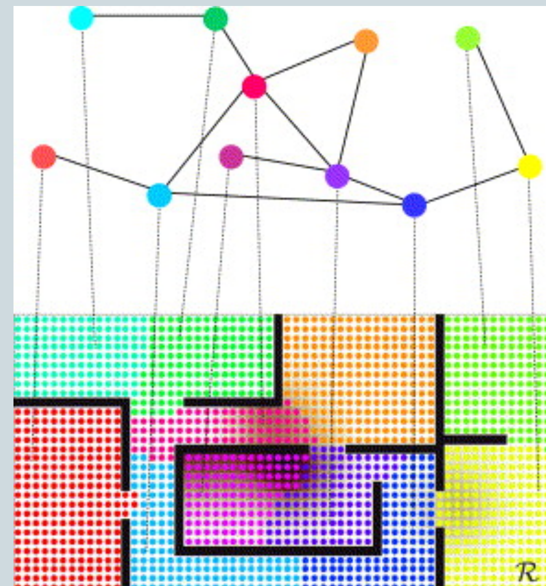
October, 2013

ABIGAIL HORN, PhD Candidate,
Engineering Systems Division,
MIT

<u>DOCTORAL COMMITTEE</u>:
PROF. RICHARD LARSON (Chair)
DR. STAN FINKELSTEIN
PROF. CÉSAR HIDALGO

Problem Framing: Optimal Search Theory

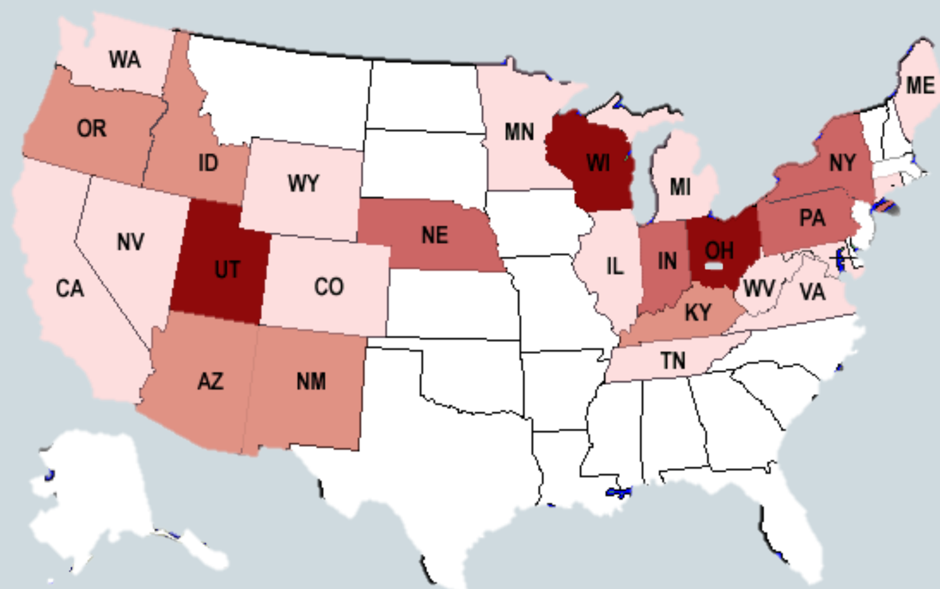
- Bernard Koopman's "Theory of Optimal Search" (Richardson, 1986).
- Anti-submarine warfare problem
  - Search over 3-D space
  - Prior probabilities
  - Bayesian updates
  - "Search effort" a highly nonlinear function of the updated probabilities
- Turned around the war in the North Atlantic (Nunn, 1981).



In 2006 there was an outbreak of E. coli O157:H7 in spinach in the US

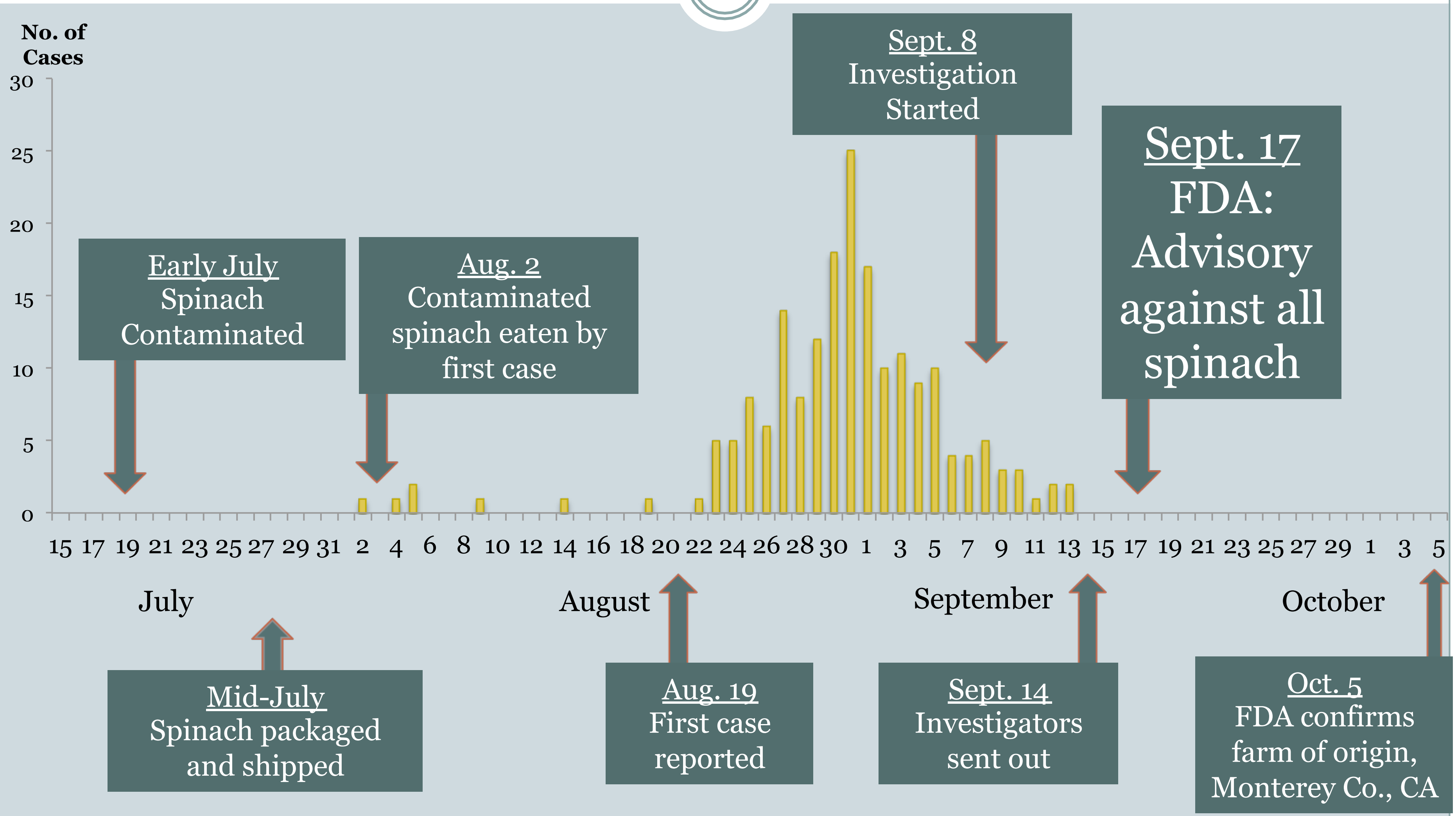
Known Impact of 2006 spinach outbreak:

- 276 Illnesses
- 102 Hospitalizations
- 31 People with Hemolytic-uremic syndrome (HUS)
- 3 Deaths
- 26 States with cases
- $350 million loss to spinach industry



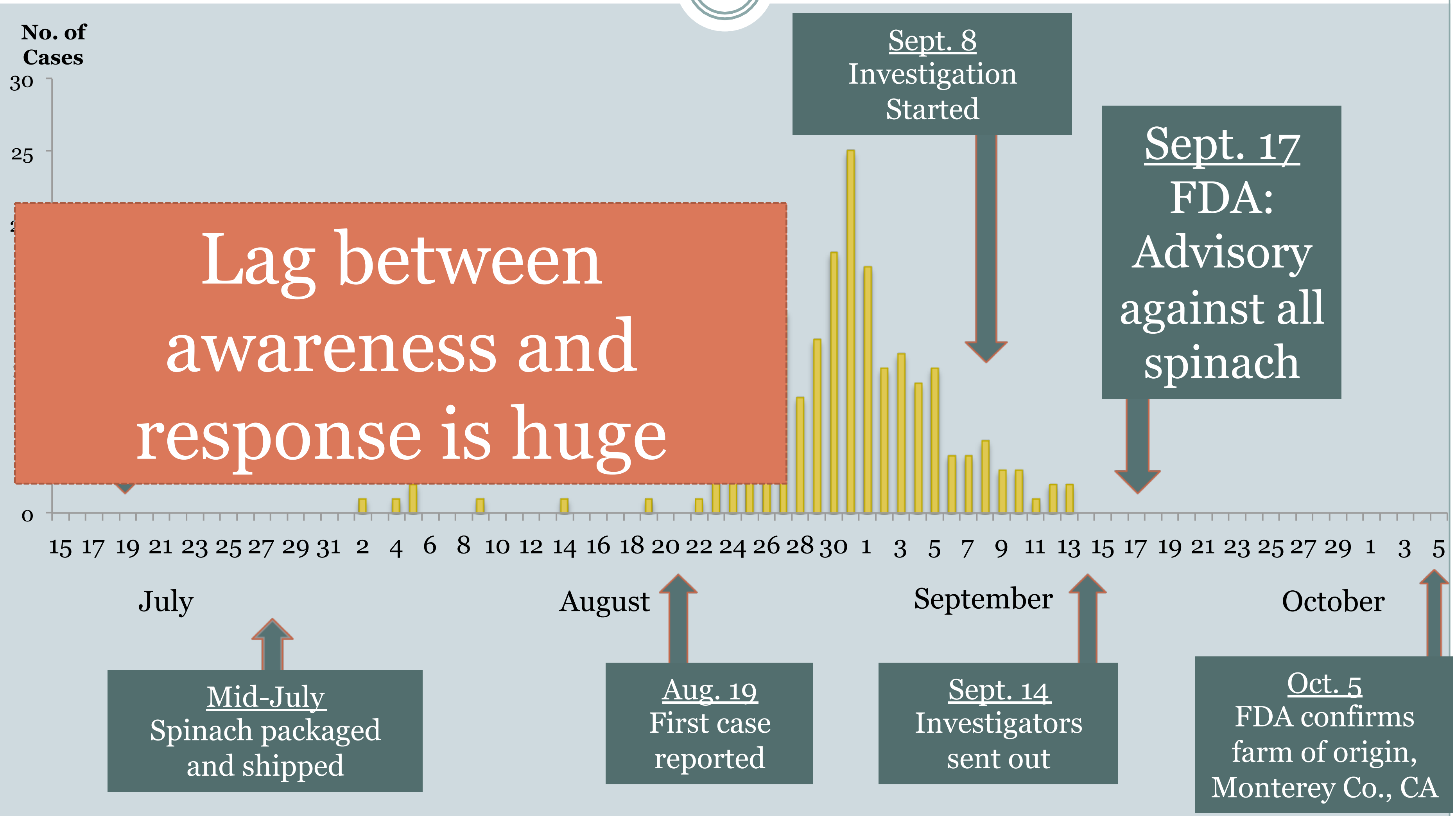
- Centers for Disease Control and Prevention (CDC) (2006a). Ongoing Multi-State Outbreak of Escherichia coli serotype O157:H7 Infections Associated with Consumption of Fresh Spinach. Morbidity and Mortality Weekly Report, 55(Dispatch); 1-2. September 26, 2006.
-California Department of Public Health (CDPH) (2007). Investigation of an Escherichia coli O157:H7 Outbreak Associated with Dole Pre-Packaged Spinach, Final Report prepared by the California Food Emergency Response Team. March 21, 2007. http://www.cdph.ca.gov

Could Damage Have Been Prevented?



California Department of Public Health (CDPH) (2007). Investigation of an Escherichia coli O157:H7 Outbreak Associated with Dole Pre-Packaged Spinach, Final Report prepared by the California Food Emergency Response Team. March 21, 2007. http://www.cdph.ca.gov.

Could Damage Have Been Prevented?



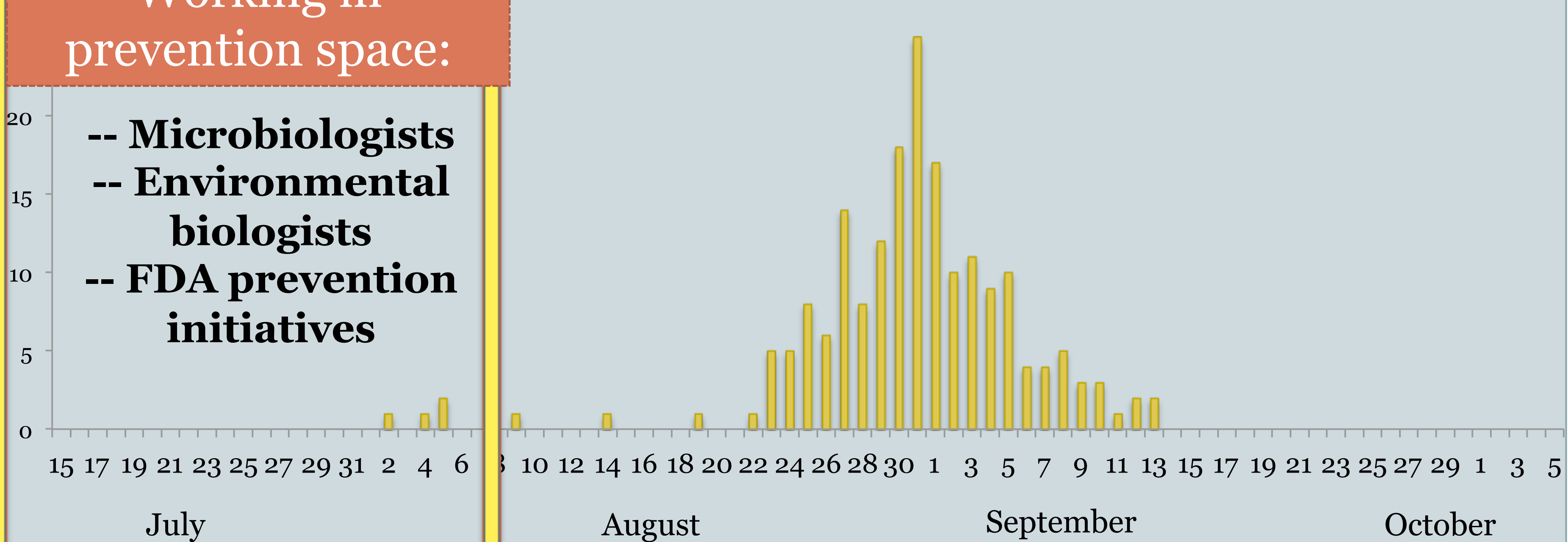
California Department of Public Health (CDPH) (2007). Investigation of an Escherichia coli O157:H7 Outbreak Associated with Dole Pre-Packaged Spinach, Final Report prepared by the California Food Emergency Response Team. March 21, 2007. http://www.cdph.ca.gov.

Outbreak Prevention

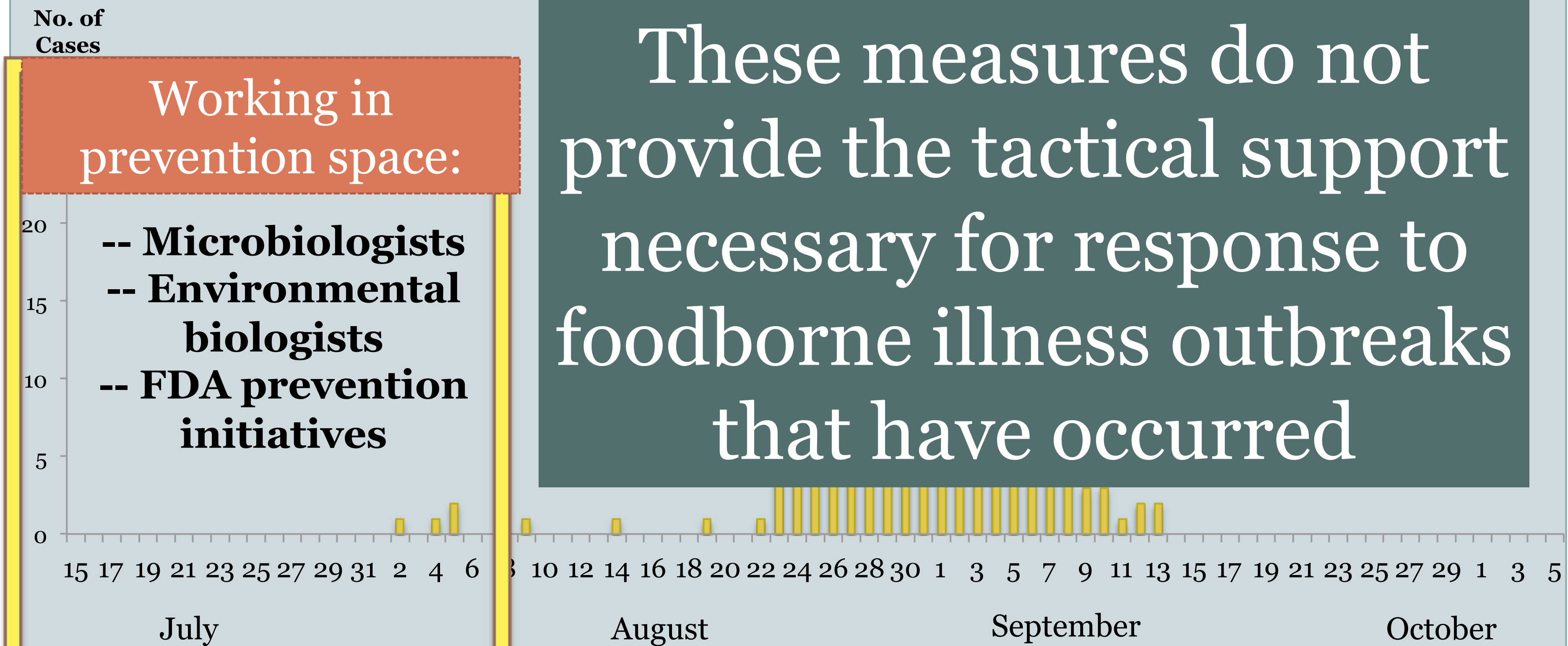
2011 Food Safety Modernization Act (FSMA)

Working in prevention space:

-- Microbiologists
-- Environmental biologists
-- FDA prevention initiatives



Outbreak Prevention

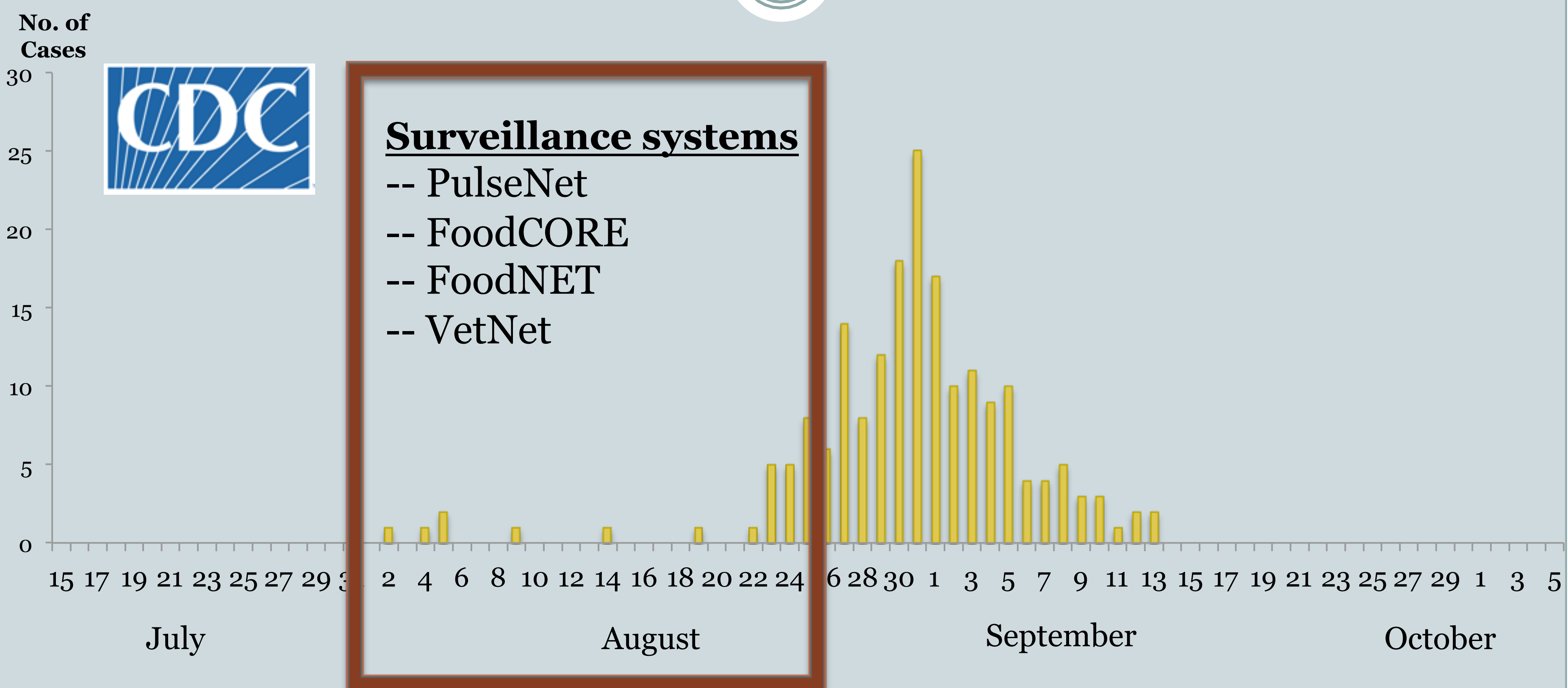


Working in prevention space:

-- Microbiologists
-- Environmental biologists
-- FDA prevention initiatives

These measures do not provide the tactical support necessary for response to foodborne illness outbreaks that have occurred

Outbreak Surveillance



Centers for Disease Control and Prevention (CDC) (2006b). Timeline for Reporting of E. coli Cases. September 19, 2006. http://www.cdc.gov/ecoli/reportingtimeline.htm

Toner ES, Nuzzo JB, Watson M, et al. Biosurveillance where it happens: state and local capabilities and needs. *Biosecur Bioterror*. 2011 Dec;9(4):321-330.

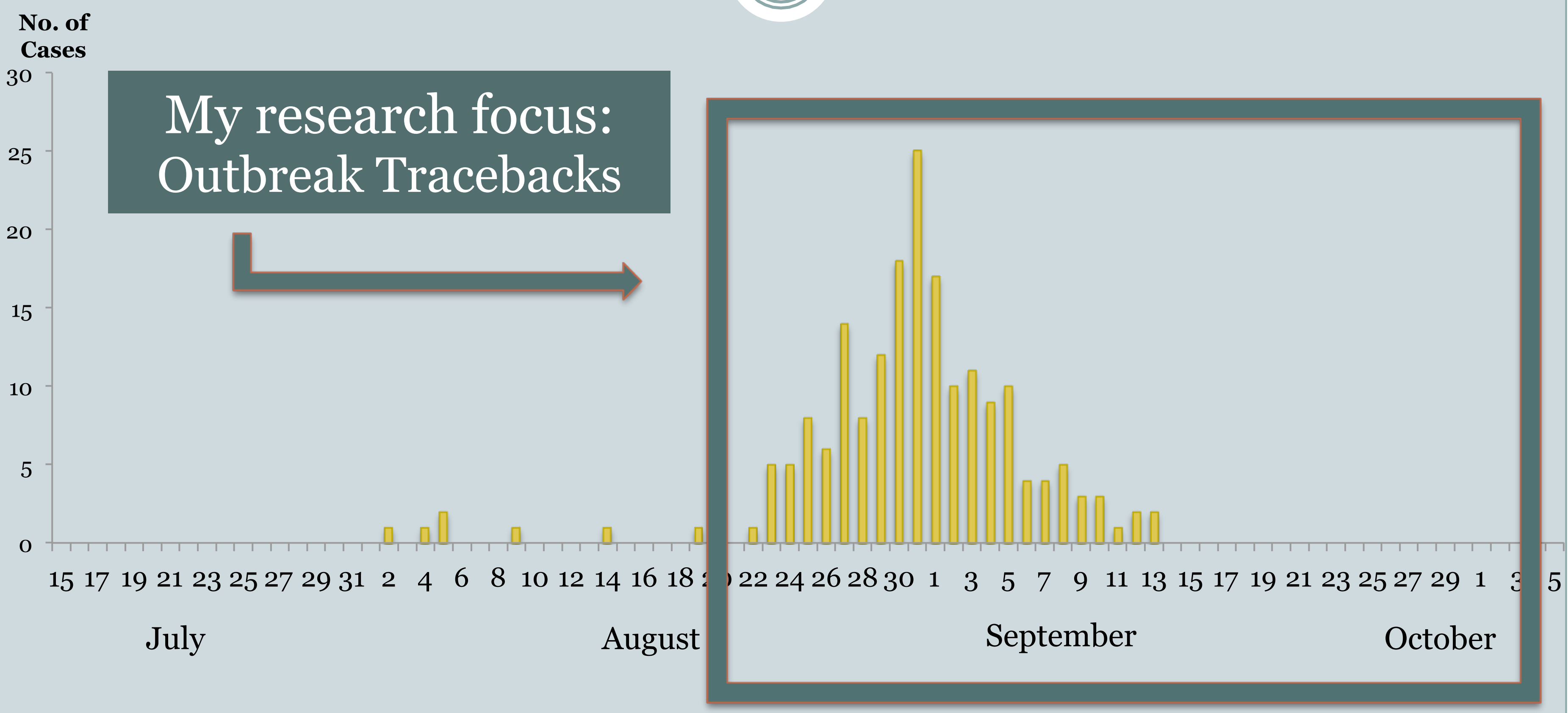
Annual Impact of Foodborne Disease Outbreaks

Despite efforts at prevention
the impact of foodborne disease outbreaks remains high:

- 48 million illnesses
- 128,000 hospitalizations
- 3000 deaths
- $77 billion in healthcare costs
- 55% - 65% of identified foodborne illness outbreaks UNSOLVED

- Osterholm, MT. Foodborne Disease in 2011 — The Rest of the Story. N Engl J Med 2011; 364:889-891, March 10, 2011.
- Scharff, R. (2009). Health-related costs from food borne illness in the United States. Retrieved from http://www.producesafetyproject.org
- Jennifer B. Nuzzo, Samuel B. Wollner, Ryan C. Morhard, Tara Kirk Sell, Anita J. Cicero, Thomas V. Inglesby. (2013). When Good Food Goes Bad: Strengthening the US Response to Foodborne Disease Outbreak. Final Report: Center for Biosecurity of UPMC.

Tactical Response to Outbreaks



Primary Research Question

How can the process of tracing the source of an outbreak be improved?

<u>My Approach:</u>

Model supply chain network structure and make predictions about the sites where contamination is likely to have taken place

Bayesian Updating Network Approach

Objective: Identify the highest probability source of contamination in the most efficient manner
Constraints: Limited available information

Investigations occur over the supply chain →
Leverage what is known about **network structure**

+

Information accruing continuously over time →
Method should allow for **dynamic updating**

+

Prior "hypotheses" from expert knowledge; history

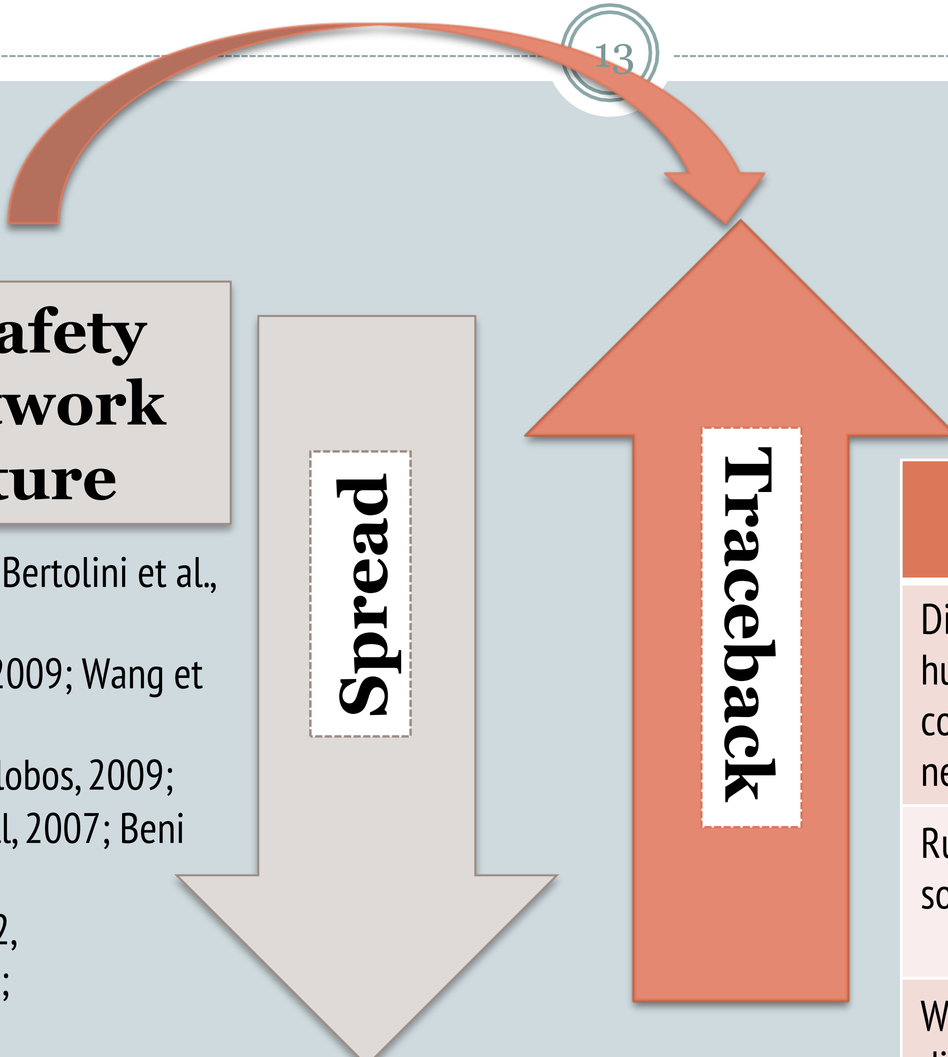
=

BAYESIAN UPDATING NETWORK APPROACH

Literature Gap

Food safety and network literature

Dupuy et al., 2005; Bertolini et al., 2006;
Fritz and Schiefer, 2009; Wang et al., 2009,
Ahumada and Villalobos, 2009;
Harlander and Sholl, 2007; Beni et al., 2011;
Hashemi et al. 2012,
Wein and Liu, 2005;
Pinior et al., 2012;
Conrad et al., 2012



Network literature only

Network domain	References
Disease in human contact networks	Liljeros et al., 2001; Newman, 2002; Pinto et al., 2012
Rumors in social media	Lloyd and May, 2001; Newman et al., 2002; Shah and Zaman, 2011
Water distribution	Tao et al., 2012; Preis and Ostfeld, 2007

Literature Gap

Food safety and network literature

Dupuy et al., 2005; Bertolini et al., 2006;
Fritz and Schiefer, 2009; Wang et al., 2009,
Ahumada and Villalobos, 2009;
Harlander and Sholl, 2007; Beni et al., 2011;
Hashemi et al. 2012,
Wein and Liu, 2005;
Pinior et al., 2012;
Conrad et al., 2012

Spread

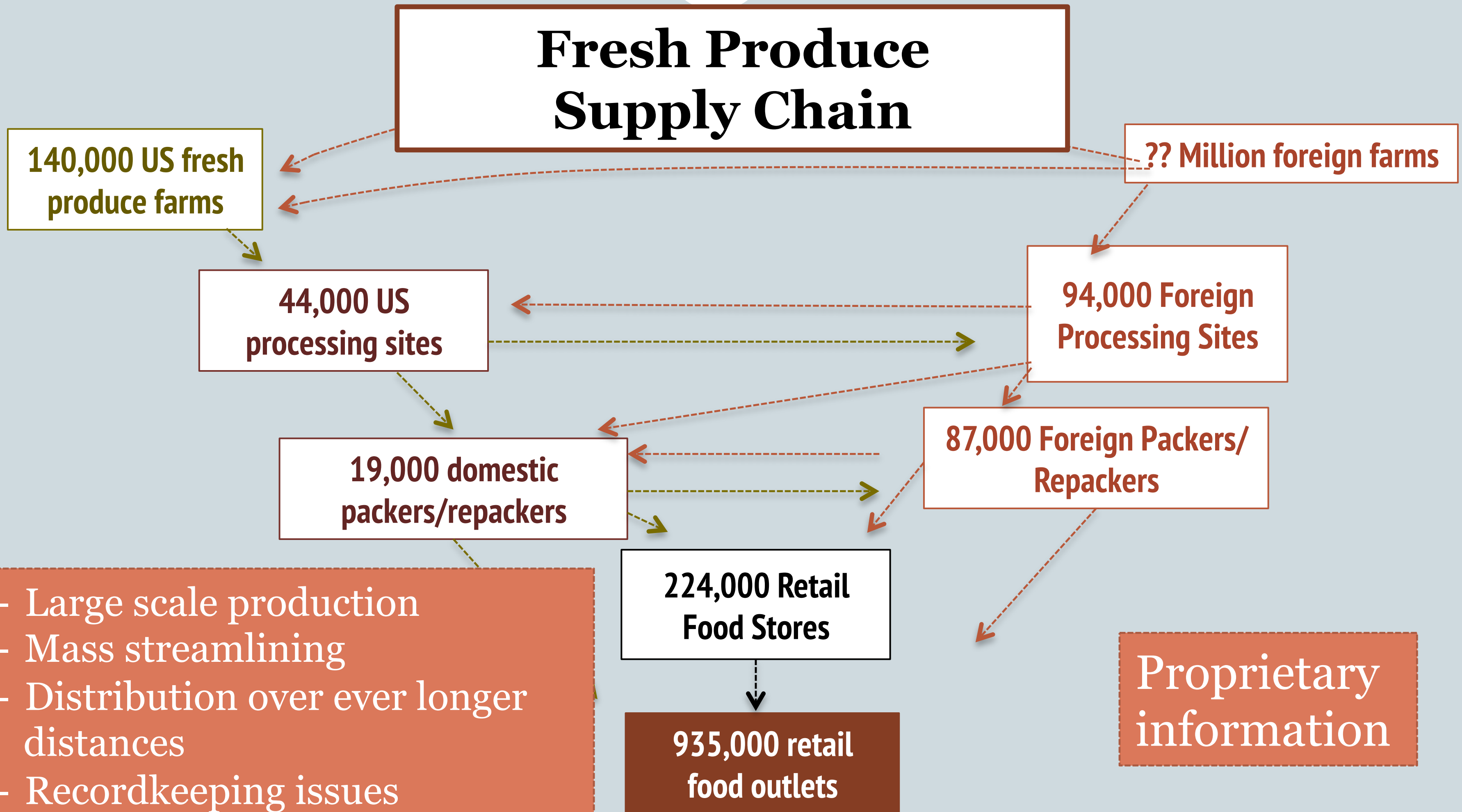
Tra

Network literature only

Network domain	References
	Shah and Zaman, 2011
Water distribution	Tao et al., 2012; Preis and Ostfeld, 2007

These models cannot be extended to food distribution

Complexity of Food Distribution



- Large scale production
- Mass streamlining
- Distribution over ever longer distances
- Recordkeeping issues

Johnson R. The Federal Food Safety System: A Primer. Washington, DC: Congressional Research Service; January 11, 2011. http://assets.opencrs.com/rpts/RS22600_20110111.pdf. Accessed February 4, 2013.

Research Goals and Applications

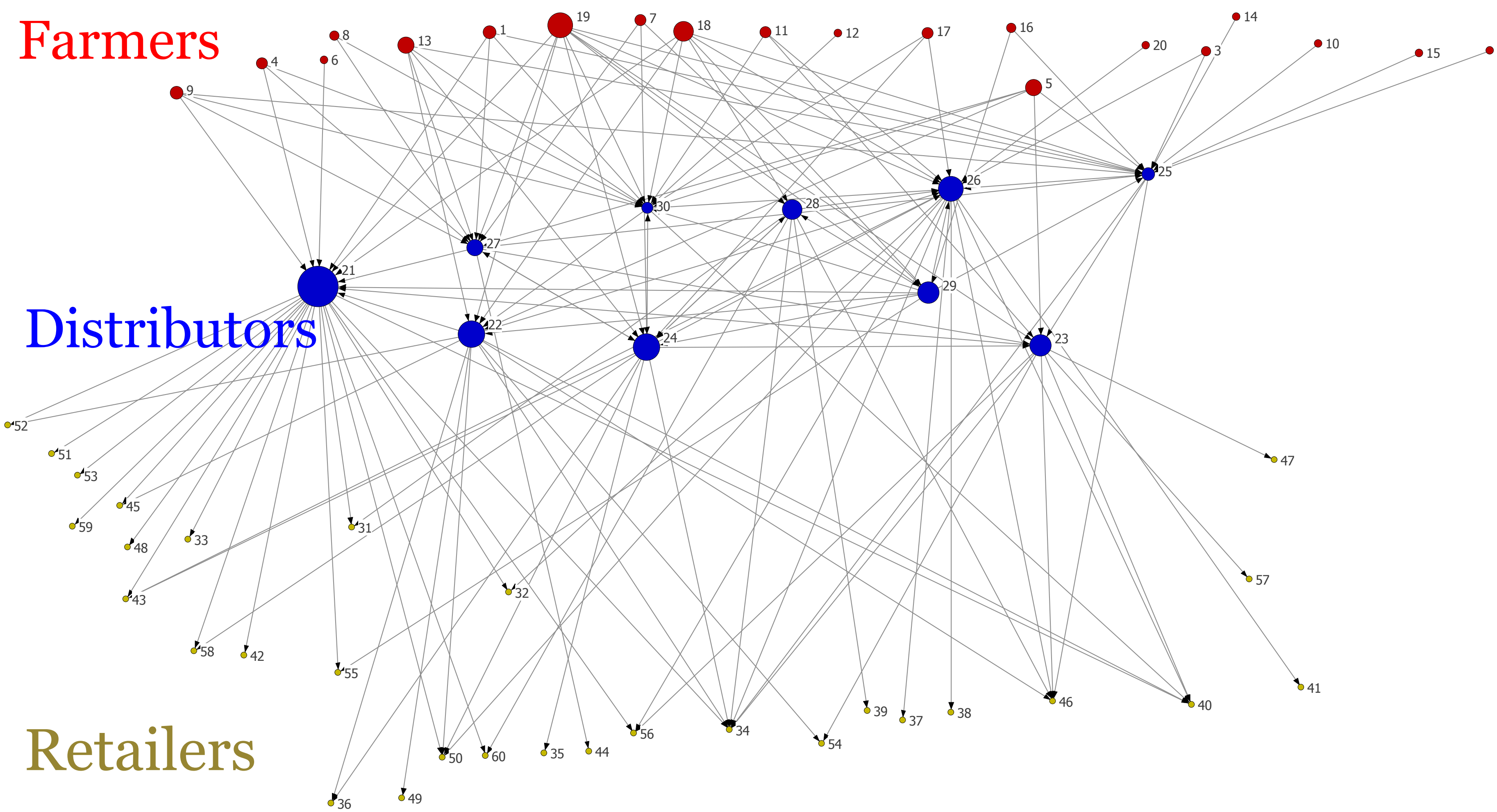
How can the process of tracing the source of an outbreak be improved?

Task: Create methodology for tactical, real-time outbreak response
Application: Enable public health and emergency preparedness officials to make informed traceback policy decisions

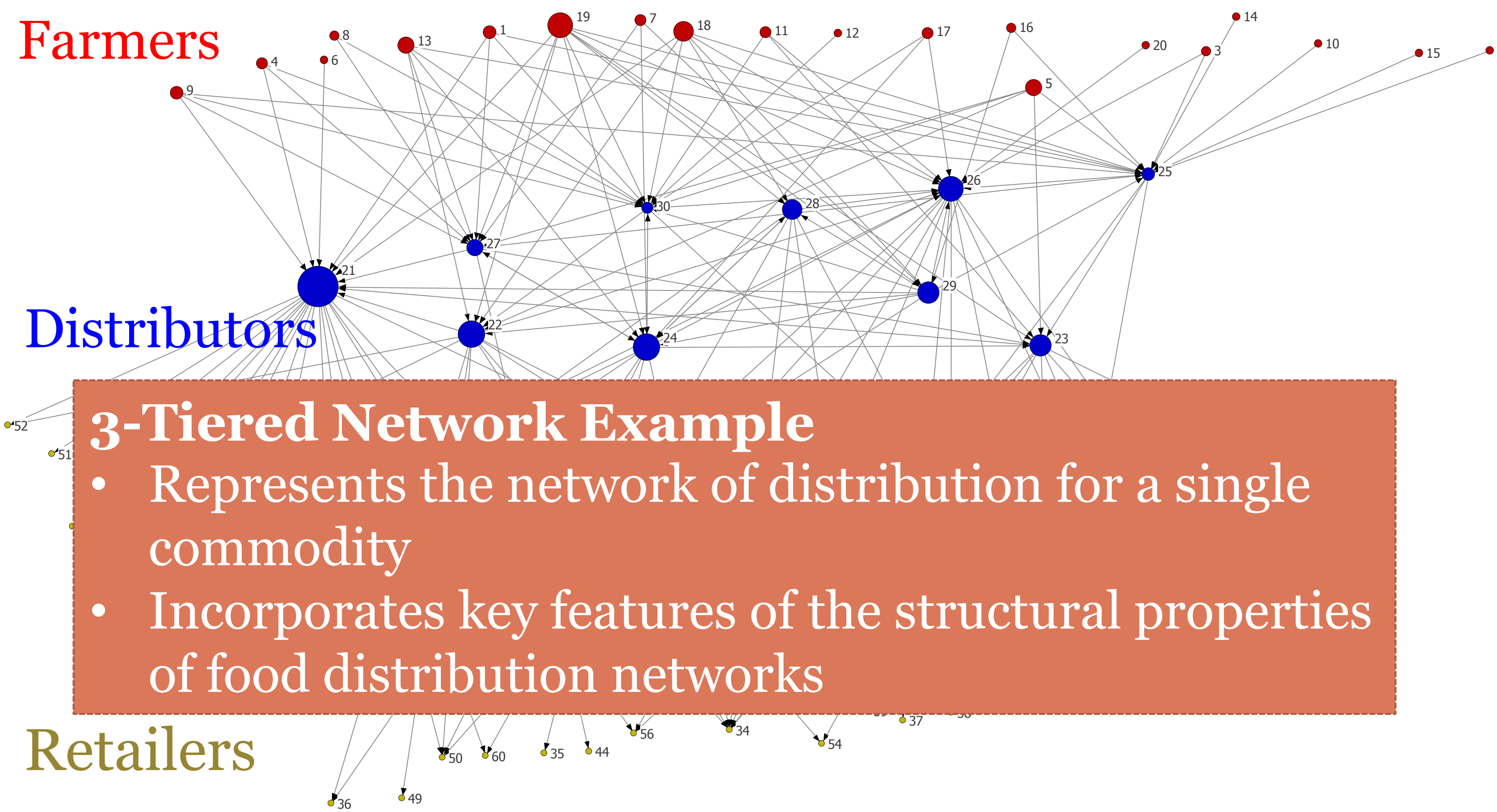
Task: Determine how much information about the supply chain is necessary to achieve accurate traceback.
Application: Make recommendations for what recordkeeping data the FDA should collect systematically.

I can explore both of these questions using the following framework:

Modeling Framework



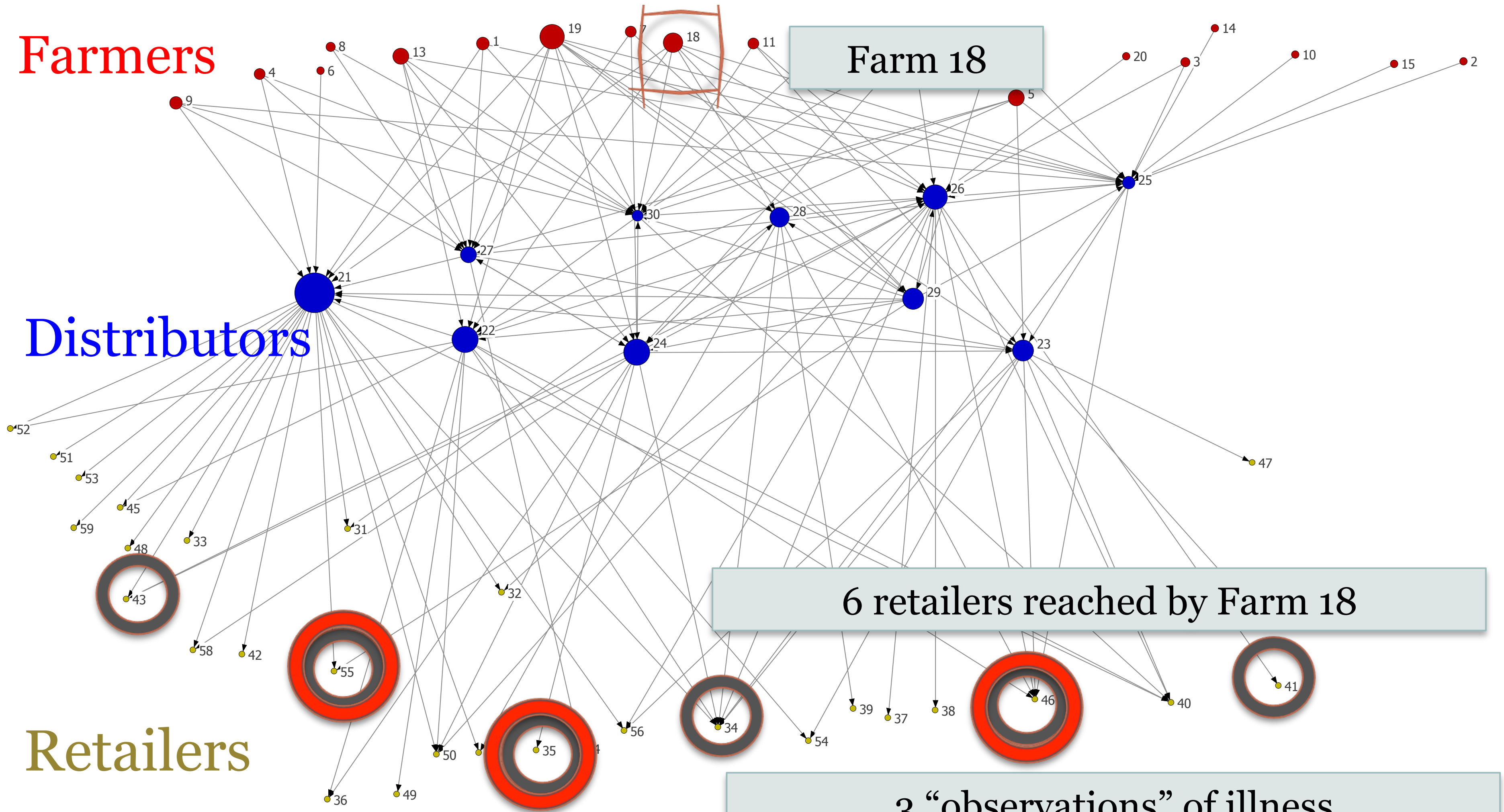
Modeling Framework



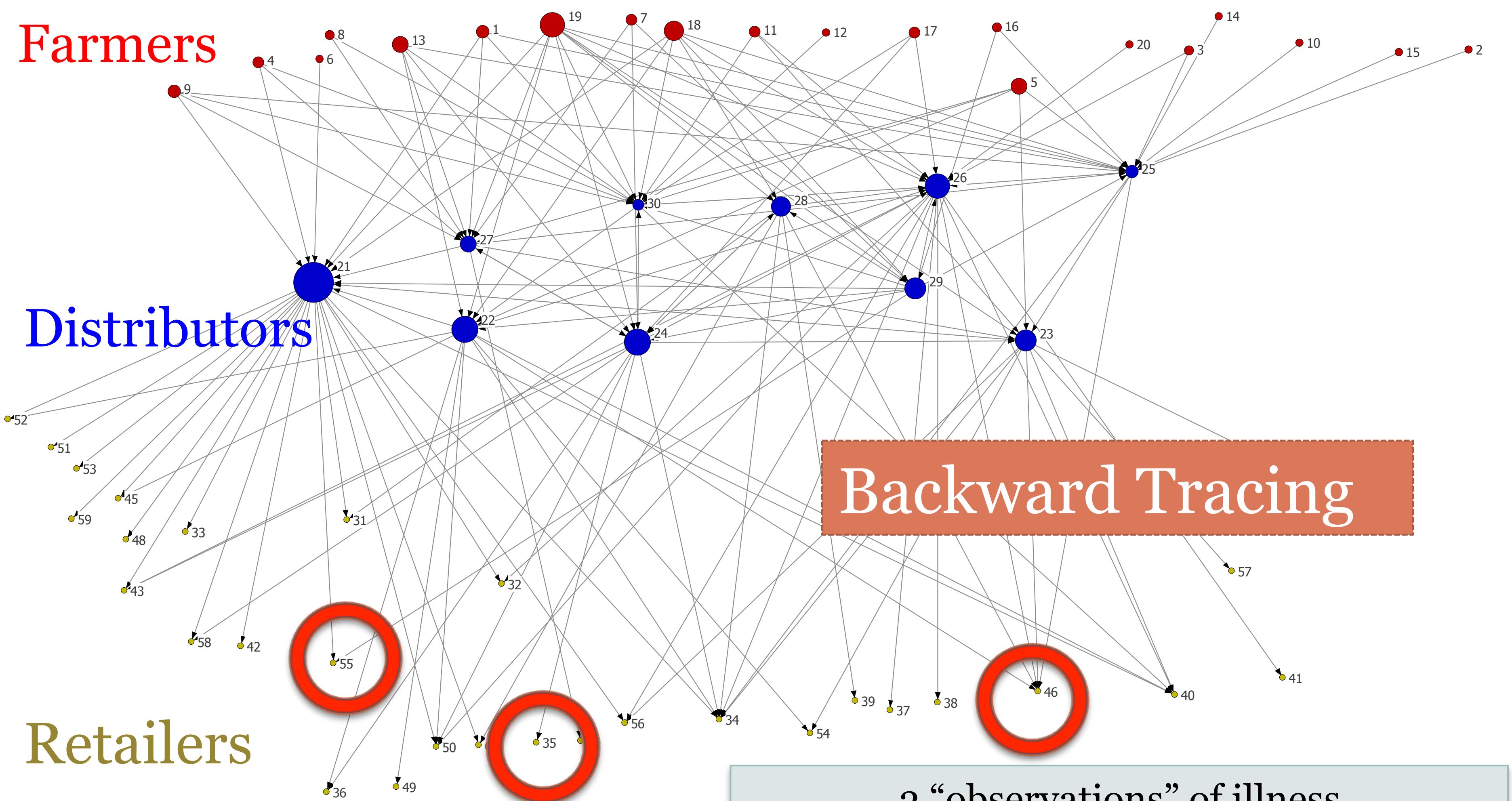
3-Tiered Network Example

- Represents the network of distribution for a single commodity
- Incorporates key features of the structural properties of food distribution networks

Modeling Framework



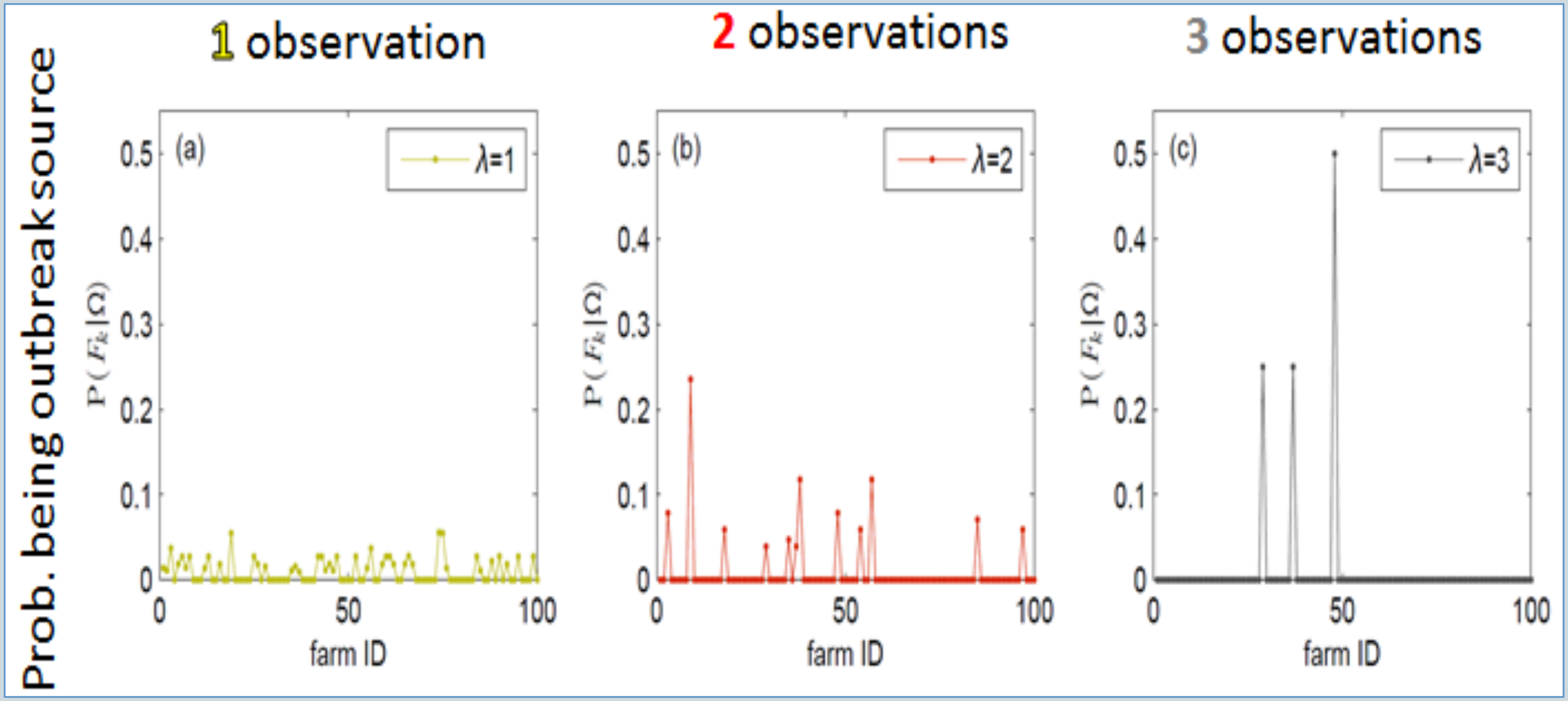
Modeling Framework



3 "observations" of illness

Probability Distribution Example

Probability of being the outbreak source given…



Defining Probability Space

$FD:$ Distribution matrix for links going from Farms to Distributors $(|F| \times |D|)$

$DR:$ Distribution matrix for links going from Distributors to Retailers $(|D| \times |R|)$

$\Omega:$ Set of retailers reporting infection

$$FR = F \times FD \times DR = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_F \end{bmatrix} \times \begin{bmatrix} fd_{11} & fd_{12} & \cdots & fd_{1D} \\ fd_{21} & fd_{22} & \cdots & fd_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ fd_{F1} & fd_{F2} & \cdots & fd_{FD} \end{bmatrix} \times \begin{bmatrix} dr_{11} & dr_{12} & \cdots & dr_{1R} \\ dr_{21} & dr_{22} & \cdots & dr_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ dr_{D1} & dr_{D2} & \cdots & dr_{DR} \end{bmatrix} = \begin{bmatrix} fr_{11} & fr_{12} & \cdots & fr_{1R} \\ fr_{21} & fr_{22} & \cdots & fr_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ fr_{F1} & fr_{F2} & \cdots & fr_{FR} \end{bmatrix}$$

Inference: Determine conditional probabilities

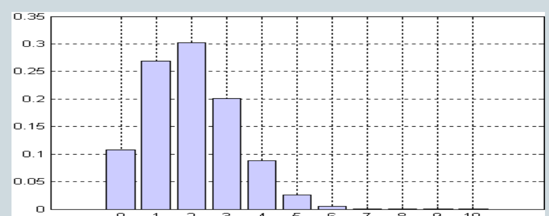
$$P\left(R_j | F_i\right) = \frac{f_i r_j}{\sum_j f_i r_j}$$

$$P\left(F_k \mid \Omega\right) \equiv P(F_k \mid \bigcap_{i \in \Omega} R_i) = \frac{P(F_k \bigcap_{i \in \Omega} R_i)}{\bigcap_{i \in \Omega} R_i}$$

$$P(F_k \bigcap_{i \in \Omega} R_i) = P\left(F_k\right) P\left(R_1 \mid F_k\right) P\left(R_2 \mid F_k \cap R_1\right) \ldots$$

$$P\left(\bigcap_{i \in \Omega} R_i\right) = \sum_{k \in F} P\left(F_k\right) P\left(R_1 \mid F_k\right) P\left(R_2 \mid F_k \cap R_1\right) \ldots$$

$P(F_i)$



Retailers

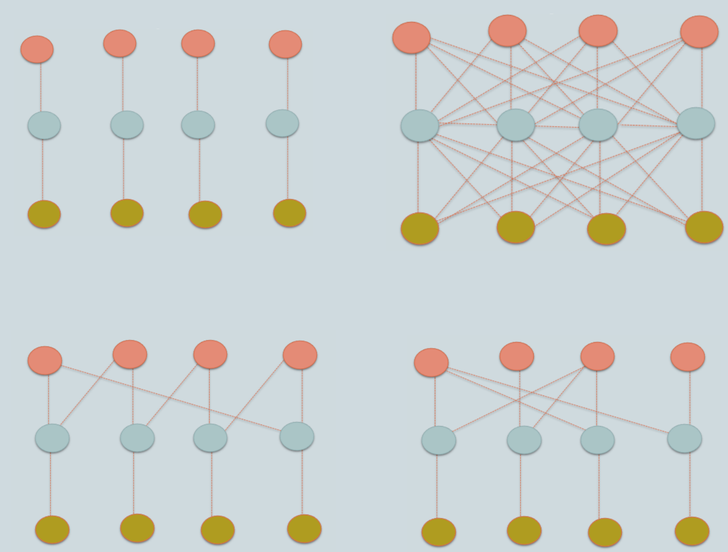
Methodology: Metrics, Limiting Cases

METRICS

- Network Traceability Accuracy
  - Expected accuracy as a function of $|\Omega|$ observations and m "guesses"
  - Expected accuracy with **incomplete knowledge about network**?
    → How much knowledge is necessary for traceback within bounds of accuracy?

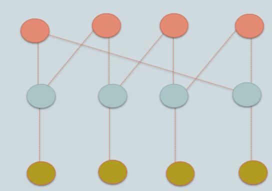
- Expected # of farms tested until source identified
  - Also allows computation of the expected value of the cost of the investigation.

Limiting Cases

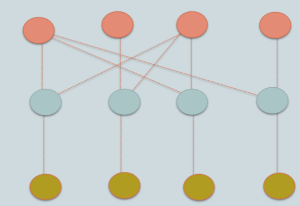


Numerical Example: Number of Farms to Test

- 100 farms
- Each retailer reached by 20 farms

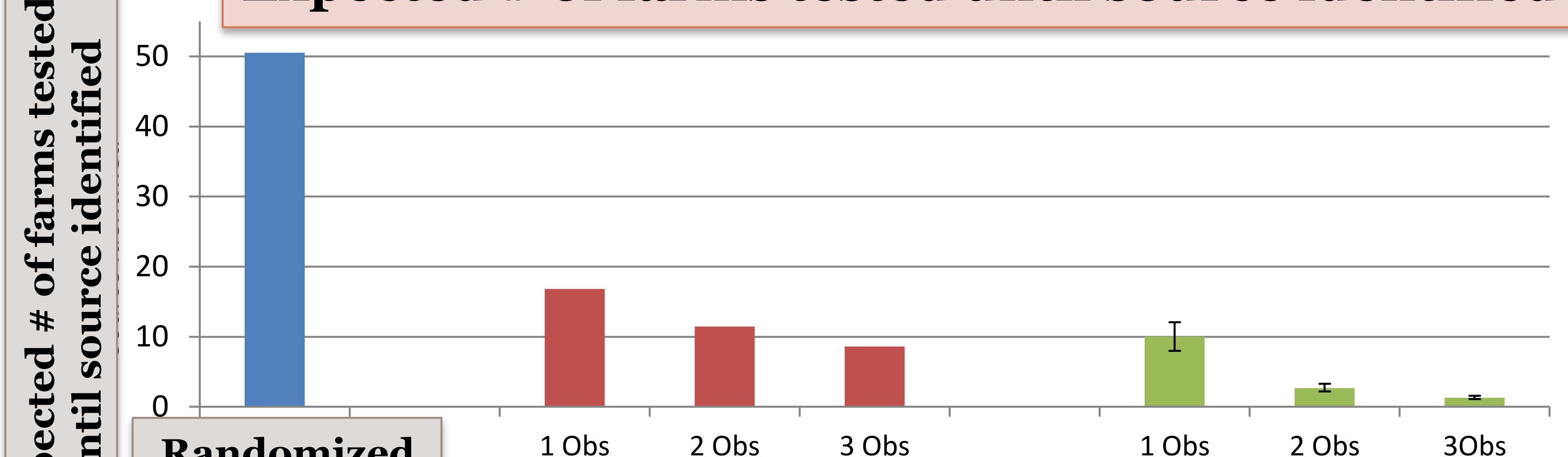


Case A: Network with highest uncertainty



Case B: Random network structure

Expected # of farms tested until source identified



Randomized Testing (no knowledge of network)

Case A

Case B

Methodology: Three Classes of Networks

- Closed form analytical solutions
  - Analytically tractable results require considerable restrictions on modeling framework
  - Lower bound on accuracy set by statistically identical connectivity patterns, which are most tractable analytically

- Algorithmic approaches using Bayesian Network framework
  - Investigate networks with more layers and complicated connectivity patterns

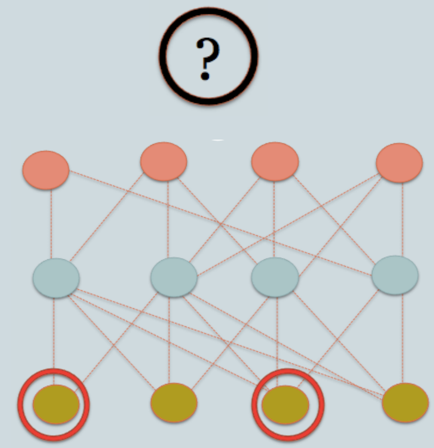
- Monte Carlo simulation
  - For situations not tractable with algorithm (e.g. cycles)

How far can each be pushed before resorting to the next?

Applications: Tactical, real-time outbreak response

GUIDANCE PROVIDED:

- Identify highest probability sources of contamination
- **Search theory**: Determine how search effort should be allocated
  - Automatically produces estimates of the cost of a given success probability
- Guide targeted messaging by localizing outbreak



Comparison to current methods

Make comparisons to examples of past outbreaks

- Include prior probabilities derived from temporal, historical, environmental, and scientific evidence

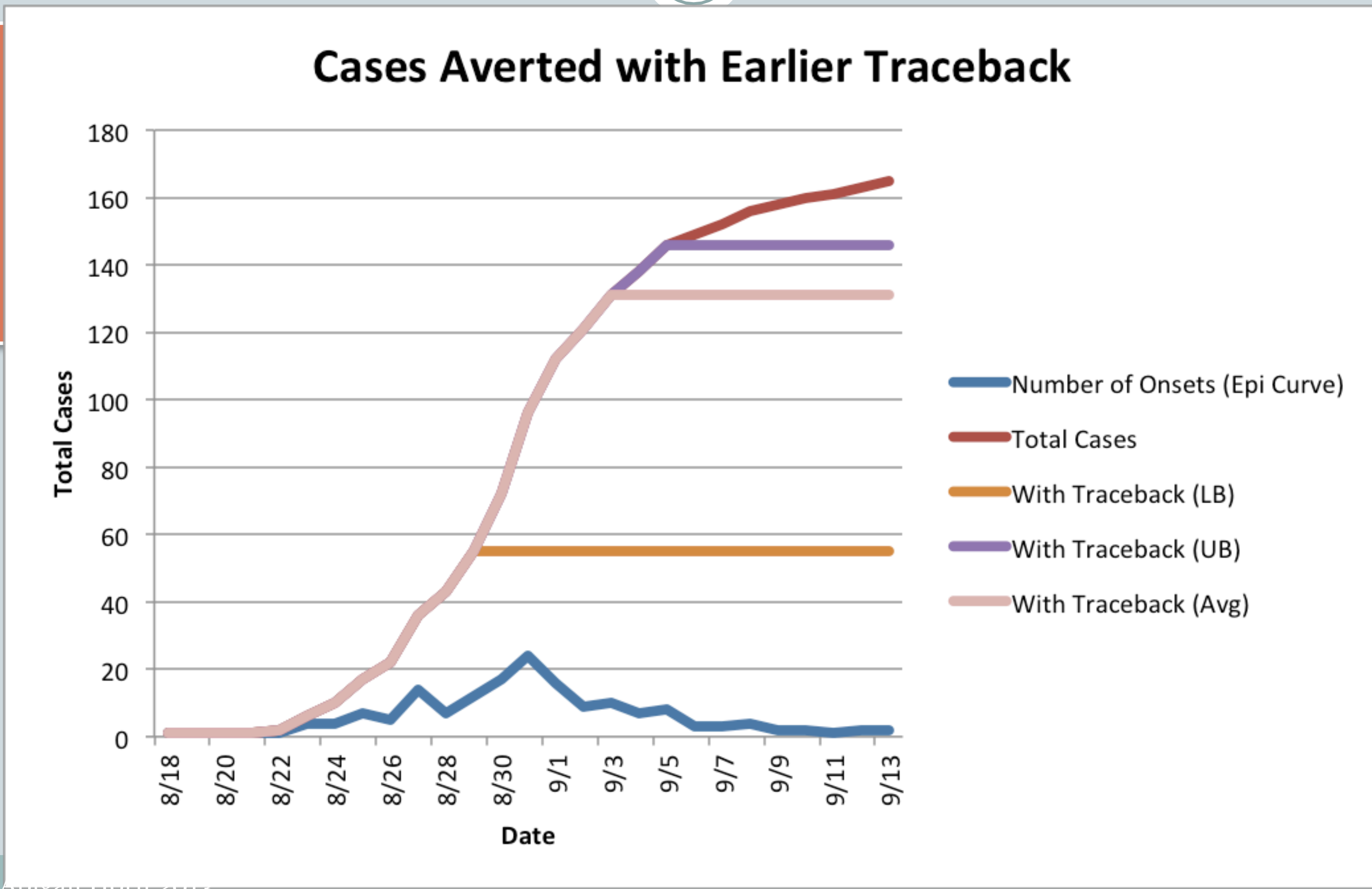
COMPARISON METRICS

Accuracy improvement at increasing levels of "available information"

Time reduced = illnesses averted

Size reduction: Narrow down the firms involved or the geographical area containing the origin of the outbreak

Comparison to current methods



Contributions

Academic and practical contribution is to improve traceability

Contributions

Academic: to develop implementable algorithms that identify the highest probability sources of contamination in the most efficient manner;

To develop a theory of optimal search effort on distribution networks

Practical: Work with stakeholders to develop a scientifically sound, *implementable* methodology to establish hypotheses early to guide investigation and control measures

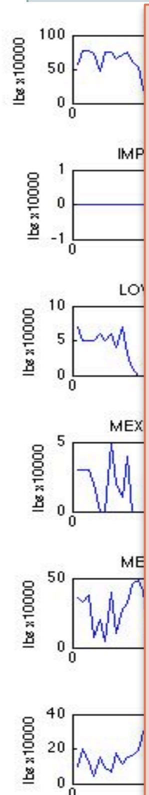
Practical: Make recommendations on what "recordkeeping" data the FDA should systematically collect.

"Any measure that will help to determine where we should focus our attention and give leads on the investigation would have a lot of application and utility for public health. Messaging could be more targeted because we would be able to narrow down more quickly where the product is not coming from…This could really make a difference early on!"

–– S. McGarry, Foodborne Outbreak Coordinator at FDA Headquarters, Personal communication, December 20, 2012

Backup Slides

Weight Networks with Prior Information



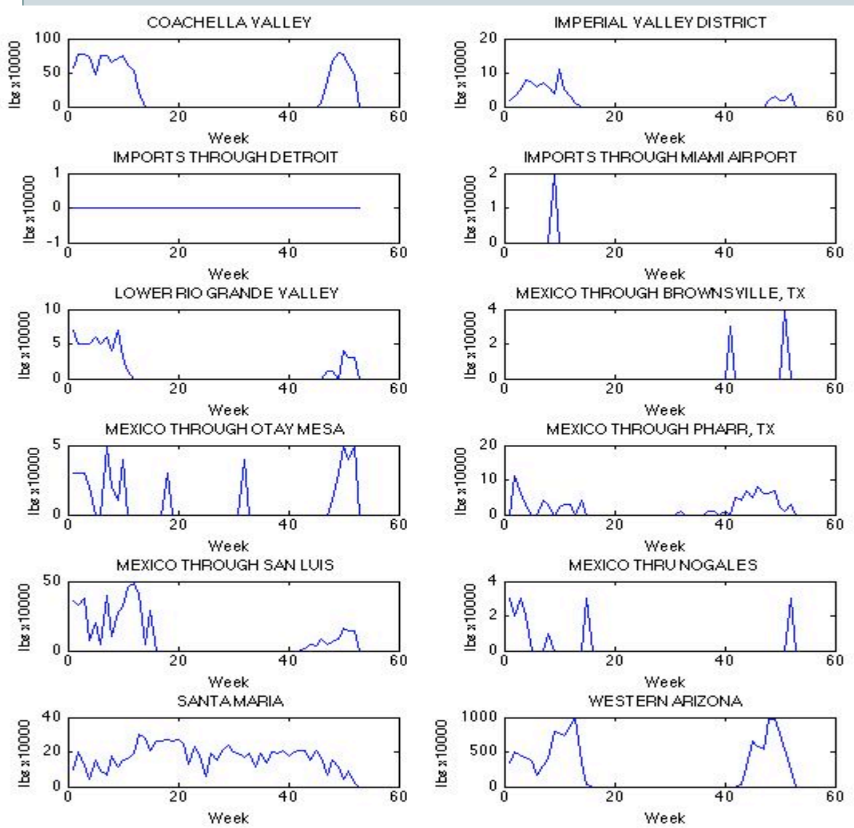
Incorporate as "prior probabilities" information about risk-related factors such as:

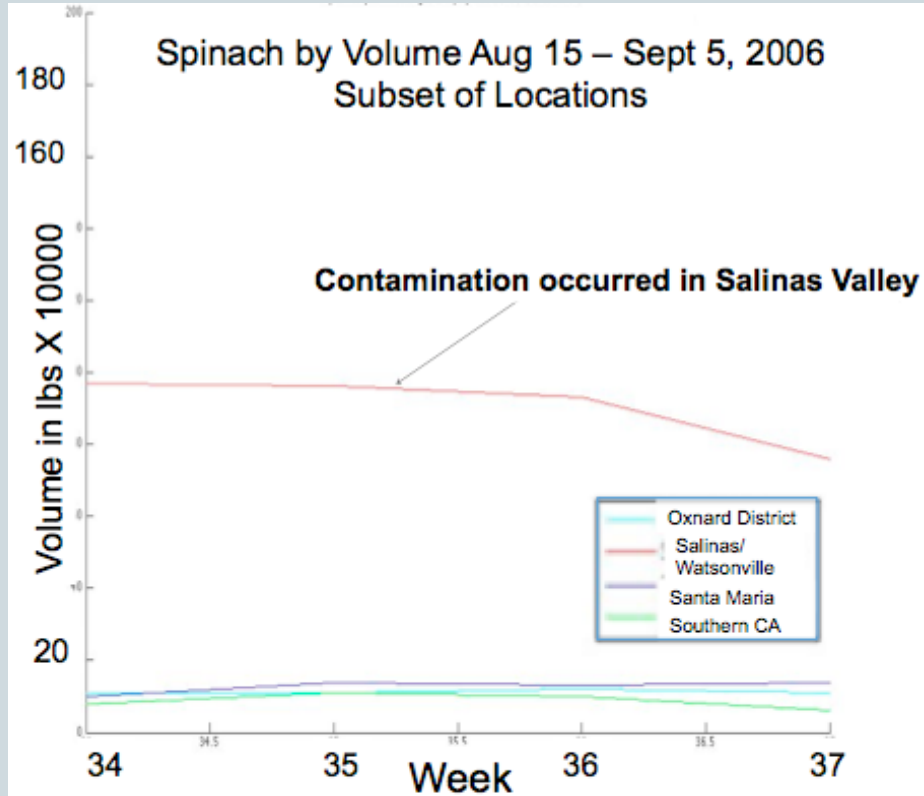
- **Temporal information**
- **Historical data**: results from inspections, statistical associations between particular regions and commodities
- **Environmental factors:** weather events, movement of wild animals
- **Pathogen-commodity risk models** (Reviewed in: Anderson et al., 2011)

Shipping records for movement of spinach over 2012 (volume)
(from USDA Agricultural Marketing Service)

Movement of spinach for ***possible*** origin districts during August 15 – September 6th, 2006.

Weight Networks with Prior Information





Shipping records for movement of spinach over 2012 (volume)
(from USDA Agricultural Marketing Service)

Movement of spinach for ***possible*** origin districts during August 15 – September 6th, 2006.

Model Building and Data Sources

Type A: Build high-level models from ground up

Product-specific point of view must be taken

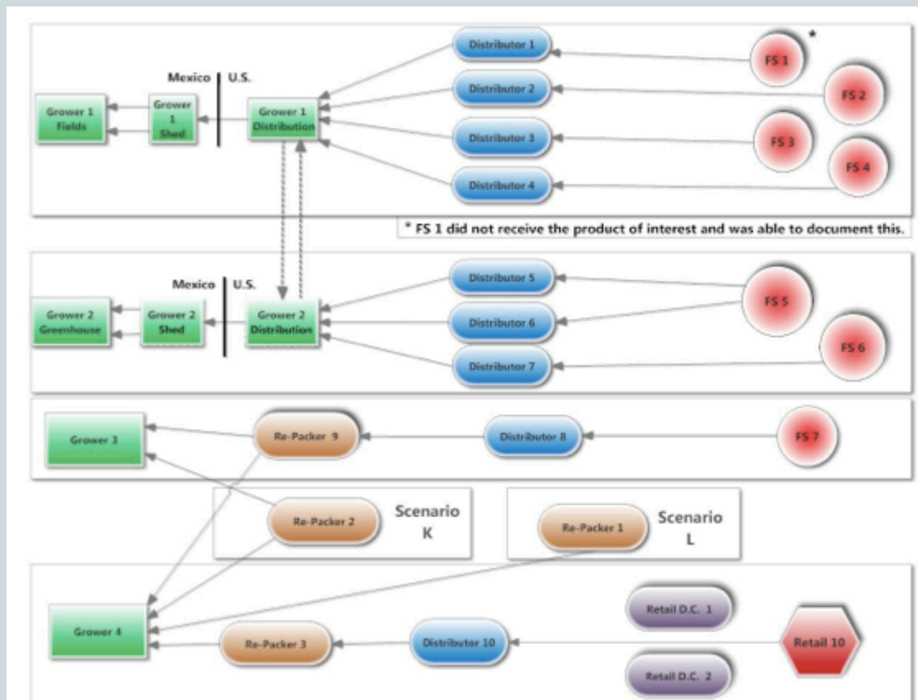
Model Building and Data Sources

Type A: Build high-level models from ground up

Data Types	Data Sources
NODES: Locations of growing regions, locations of distribution centers, brokers, wholesalers, and retail warehouses	• Secondary data collection of shipping records and expert elicitation compiled by BTSafety, LLC, for their Consequence Management System • Expert elicitation with state agriculture and commerce departments, marketing associations/trade organizations, and cooperative extension centers
LINKS: Supply and demand data for traders	• National Agricultural Statistics Service (NASS) • Gravity models to fill in (Pinior et al., 2012)
WEIGHTS: Weekly shipment and border crossing information, commodity seasonality	• Agricultural Marketing Service • FDA pathogen-commodity risk models • Expert elicitation as above
OUTPUT: Location of reported cases	• Marler Clark Litigation Firm

Model Building and Data Sources

Type B: Use supply chain models from FDA, IFT (Institute for Food Technologists)



Pilot Projects for Improving Product Tracing along the Food Supply System – Final Report. Institute of Food Technologists – prepared for the FDA. August 2012.

Model Corroboration

Models will be corroborated through

– **Conversations with experts to approve inputs:** Conversations with experts at the FDA and in industry who have comprehensive knowledge of supply chain structure for the three chosen commodities to ensure the structural and temporal parameters of the models are reasonable

– **Using data on past outbreaks to evaluate outputs**: Because the models are representations rather than exact characterizations of supply chains, we cannot directly project the exact data from past outbreaks reported in FDA and state health department post-incident outbreak investigation reports. We will make use of these cases by basing parameter values on the actual values from cases in selected cases to recreate the story line in our models, represent an analogous progression of the contamination event, and check whether the resulting statistics of the outbreak traceback correspond qualitatively with those reported.

Promise of Technology-Enabled Traceability

Technology exists to have fully traceable food supply system

But along with logistical difficulties of tracing loose produce…

current lack of will to implement full traceability due to:

Mandated by Government

- No meaningful purposed legislation
- Unfavorable legislative environment
- Only can go in after "reasonable cause"

Adopted by Industry

- **Full compliance a distant reality**
- Not incentivized to create a system that tracks food once it has been sold and consumed
- Failure to supply adequate traceability systems for basic food safety control
- Firms find value in anonymity (Golan et al., 2004)