CHARACTERIZING EPIDEMICS IN METAPOPULATION CATTLE SYSTEMS THROUGH ANALYTIC MODELS AND ESTIMATION METHODS FOR DATA-DRIVEN MODEL INPUTS

by

PHILLIP RAYMOND BROOKE SCHUMM

B.S., Kansas State University, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

We have analytically discovered the existence of two global epidemic invasion thresholds in a directed meta-population network model of the United States cattle industry. The first threshold describes the outbreak of disease first within the core of the livestock system while the second threshold describes the invasion of the epidemic into a second class of locations where the disease would pose a risk for contamination of meat production. Both thresholds have been verified through extensive numerical simulations. We have further derived the relationship between the pair of thresholds and discovered a unique dependence on the network topology through the fractional compositions and the in-degree distributions of the transit and sink nodes.

We then addressed a major challenge for epidemiologists and their efforts to model disease outbreaks in cattle. There is a critical shortfall in the availability of large-scale livestock movement data for the United States. We meet this challenge by developing a method to estimate cattle movement parameters from publicly available data. Across 10 Central States of the US, we formulated a large, convex optimization problem to predict the cattle movement parameters which, having minimal assumptions, provide the best fit to the US Department of Agriculture's Census database and follow constraints defined by scientists and cattle experts. Our estimated parameters can produce distributions of cattle shipments by head which compare well with shipment distributions also provided by the US Department of Agriculture.

This dissertation concludes with a brief incorporation of the analytic models and the parameter estimation. We approximated the critical movement rates defined by the global invasion thresholds and compared them with the average estimated cattle movement rates to find a significant opportunity for epidemics to spread through US cattle populations.

CHARACTERIZING EPIDEMICS IN METAPOPULATION CATTLE SYSTEMS THROUGH ANALYTIC MODELS AND ESTIMATION METHODS FOR DATA-DRIVEN MODEL INPUTS

by

PHILLIP RAYMOND BROOKE SCHUMM

B.S., Kansas State University, 2009

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Caterina Maria Scoglio

Copyright

Abstract

We have analytically discovered the existence of two global epidemic invasion thresholds in a directed meta-population network model of the United States cattle industry. The first threshold describes the outbreak of disease first within the core of the livestock system while the second threshold describes the invasion of the epidemic into a second class of locations where the disease would pose a risk for contamination of meat production. Both thresholds have been verified through extensive numerical simulations. We have further derived the relationship between the pair of thresholds and discovered a unique dependence on the network topology through the fractional compositions and the in-degree distributions of the transit and sink nodes.

We then addressed a major challenge for epidemiologists and their efforts to model disease outbreaks in cattle. There is a critical shortfall in the availability of large-scale livestock movement data for the United States. We meet this challenge by developing a method to estimate cattle movement parameters from publicly available data. Across 10 Central States of the US, we formulated a large, convex optimization problem to predict the cattle movement parameters which, having minimal assumptions, provide the best fit to the US Department of Agriculture's Census database and follow constraints defined by scientists and cattle experts. Our estimated parameters can produce distributions of cattle shipments by head which compare well with shipment distributions also provided by the US Department of Agriculture.

This dissertation concludes with a brief incorporation of the analytic models and the parameter estimation. We approximated the critical movement rates defined by the global invasion thresholds and compared them with the average estimated cattle movement rates to find a significant opportunity for epidemics to spread through US cattle populations.

Table of Contents

List of Figures

List of Tables

Acknowledgments

My family has served in a huge way in the completion of this doctorate degree. Most significantly, Miss Ranjni J. Chand, my distinguished coworker and precious fiancèe has walked with me through all these days. When she gave me 15 kind minutes on the morning of September 20th, she again turned my world upside down. Let me not forget the rest of my family: Dad, Mom, Grandma, Ammuma, Amma, Jonathan, Allison, Chechi, Joshua, Zoë, Hank, Miriam, Deborah, Martha, Daniel, Diqing, Nicole, Alisa, Emmanuel, Jaq, Angel, 'Kiki, Ana, Ta'Dya, Tyrone, Jeremiah, Mercy, Isaiah, Kyrsten, Josiah, Timothy, Walter, Tim and Walter's little sibling, Mamaji and Tannaz Aunty, Kayannush, Uncle Brooke and Aunt Betty, Cousins Karl and Julie, Cousins Skip and Kari, Uncle Launie and Aunt Thelma, Uncle Larry and Aunt Gen, Maheli Uncle and Zarine Aunty, Andy, Kate, Nate, Marty, Lydia, Joe, Isaac, Heidi, Geoffrey, Cousin Gen, Cristalyn, Patrick, Nick, and Kathleen. Many thanks to you all. I would also like to thank those in my extended family: (in no particular order) Dr. Mina & Maro, Sakshi, Dr. Sydney & Netta, Dr. Li, Ben & Amber, Nikkie, Supriya, Sohini, Sweta, Nidhi, Amelia, Faryad & Ala, Dr. Xue, Anton, Hong, Lucy, Ms. Powdrill, Dr. Paul & Amy, Joel, Suresh Uncle and Dr. Mary, Mr. Tim and Janice, Lewis, Rachel, Caleb, Gabe, Joshua, Vijaya, Rashmi, Vijay, Venkat, Deepthi, Stuart, Yuba, Allen, Andy, Ms. Malve, Amit & Meera, Cheshta, David & Anita, Sam & Naomi, Brother Moore, Jairam & Lavanya, Ed & Lori, Wade & Shannon, Ms. Shah, Emma, Dr. Kabeer Jasuja, Abe, Kavitha, Natalie, Ryan, Barry, Keenan & Nada, Dr. Bob & Mary Taussig, Dr. and Mrs. Dunn, Dr. Muthukrishnan & Asha Aunty, Sankar and Uncle, Vijay A., Tasya, Annanya, Fiona, Anjali, Arjun, Zach B., Vishal, Josh, Russ, Ross, Brent, Todd, Navdeep, Saksham, Drs. Vikram & Neena, Rohit K., Lamuel & Anusha, Vaibhav & Sheelu, Ranjit, Sandeep, Neha, Purnima, Shiva, Rohit P., Bhushan, Jason, Jordan B., Chris, Dr. Thornley, the Hawkinson family, the Walter family, the Wolters family, and so so many more...

I would further like to thank a number of motivating and encouraging professors: Dr. Carpenter, Dr. Pahwa, Dr. Kuhn, Dr. Soldan, Dr. Natarajan, Dr. Harnett, Dr. Heier-Stamm, Dr. Poggi-Coradini, and Dr. Lee. I would like to especially recognize the professors who hosted me in visits to TU Delft, Indiana University, and the Institute for Scientific Interchange: Dr. Van Mieghem and Dr. Vespignani. These visits were also kindly facilitated by Dr. Wang and Dr. Balcan.

These past 6.5 years have been spent with Dr. Scoglio, my mentor and major professor. She has given me countless opportunities to collaborate, to travel, to make mistakes, to fail, and to try again. She has lead me in research all these years and suffered much frustration in this journey. Nevertheless, she has successfully brought me to the point of defending this dissertation before my advisory committee. Dr. Scott, Dr. Warren, Dr. Prakash, and the external chair Dr. Gonzalez have graciously donated their time and advice as my committee members. The committee began with one more member who has not been able to be with us to the end. I would like to thank my committee for their support in the unfortunate events that took Dr. Balcan from us. I would like to remember Dr. Duygu Balcan for her invaluable friendship and contributions to this work as an honest and close mentor. Very tragically, Dr. Balcan has recently passed away in an accident in August, 2013. I wish to convey my deepest condolences to all the lives she has touched.

Dedication

"*...To Him who is able to keep you from stumbling,*

and to make you stand in the presence

of His glory blameless with great joy,

to the only God our Savior,

through Jesus Christ our Lord,

be glory, majesty, dominion and authority,

before all time and now and forever.

Amen."

Jude 24,25

Preface

This dissertation is submitted as a requirement for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering at Kansas State University. The dissertation is entitled "Characterizing Epidemics in Metapopulation Cattle Systems Through Analytic Models and Estimation Methods for Data-driven Model Inputs" and has been performed under the supervision of Prof. Caterina M. Scoglio. This set of three journal papers/manuscripts comprises all the work in this dissertation:

- Schumm, P., Zhang, Q., Balcan, D., Gonçalves, B., Scoglio, C., Vespignani, A. **Global epidemic invasion threshold in directed subpopulation networks**, *Under final internal revision*, 2013.

- Schumm, P., Scoglio, C., Zhang, Q., [†]Balcan, D. **Global epidemic invasion thresholds in directed subpopulation networks having source, sink, and transit nodes**, *Under final internal revision*, 2013.

- Schumm, P., Scoglio, C., Scott, H. M. **"Dance of the Calves" An estimation of cattle movement parameters in the Central States of the US**, *Under final internal revision*, 2013.

[†] Dr. Duygu Balcan has passed away during the completion of this work. Dr. Balcan has co-designed the study and significantly contributed to the solutions of the quasi-equilibrium population levels. Furthermore, she has closely mentored QZ and PS through this project. PS and CS have finalized the derivations, experiments, and manuscript and are fully responsible for the content.

This dissertation focuses on the development of analytic and numerical models and the estimation of cattle movement parameters for these models for the objective of characterizing

epidemics in metapopulation cattle systems. This work provides two especially significant advancements for livestock disease modeling: the analytical results of chapters 2 and 3 provide novel insights into the analysis and control of potential epidemics in cattle, and the optimal estimation method of chapter 4 enables the extraction of cattle movement parameters for improved model inputs. I would like to extend my appreciation to the sponsors who have made these results possible.

- A portion of this material is based upon work supported by the US Department of Homeland Security under Grant Award Number 2010-ST061-AG0001 and a grant through the Kansas Biosciences Authority. The views and conclusions contained in this publication are those of the authors and should not be interpreted as necessarily representing the official policies, either explicit or implicit, of the US Department of Homeland Security and Kansas Biosciences Authority.

- A portion of this research was funded by the US Department of Agriculture - the Agricultural Research Service with Specific Cooperative Agreement 58-5430-1-0356.

- Partial funding support was provided by the National Agricultural Biosecurity Center at Kansas State University.

- A portion of this work was completed while Phillip Schumm was on sponsored visits to the Computational Epidemiology Laboratory at the Institute for Scientific Interchange, in Turin, Italy.

- The computing for the work of chapters 2 and 3 was performed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF grants CNS-1006860, EPS-1006860, and EPS-0919443.

I would further like to express my gratitude to Dr. Sanderson, Dr. Schroeder, and local farmers and stockers for helpful insights into the beef cattle industry. The experiments and

content of this document were compiled primarily by Phillip Schumm and should always be humbly open to further criticism, questions, and clarifications.

Chapter 1

Introduction

The cattle of the United States (US) number roughly 100 million head and supply careers, beef and dairy products, fertilizer, and more[1,2,3]. These cattle and the benefits they provide are constantly at risk for devastating epidemics, especially from foreign diseases[4,5]. The US cattle industry is large and pervasive. In 2007, cattle were reported in 3060 counties of the US[1]. To better understand the potential threats of new disease outbreaks, realistic and scalable epidemic models are needed. The current resolution of available data as well as the desire for scalability would suggest that population-level modeling would be the most suitable approach.

This dissertation describes a two-sided effort to advance the abilities of epidemiologists to model disease outbreaks in cattle. To model disease spread in any wide-spread population requires information describing the disease progression within an individual or a group of individuals and information characterizing the susceptible population, particularly how the individuals or groups interact[6,7,8]. Such information could be represented through an abstract model or a data-driven process. This dissertation focuses on the information describing the population demographics and movement parameters for cattle systems in the US. We approach this challenge through a) extending the definition of a global epidemic invasion threshold to a more realistic model of cattle systems and b) estimating unknown

parameters to describe the movements of cattle across the Central States of the US. The first portion steps from an abstract model towards reality by increasing the complexity of the model and analysis. The latter portion starts with a real database from the US Department of Agriculture and attempts to estimate cattle mobility parameters for use in an abstract, yet still more complicated, cattle population model. These two approaches have yet to reach each other except for a small example which we present in chapter 5. The remaining challenges are summarized for future work in subsection 5.1.1.

1.1 Broader impacts

The development of a modeling system capable of forecasting livestock disease and demographics on a national or global scale serves as the vision of this work[6,7,8]. In this dissertation, we focus first on deriving and validating analytic results for epidemic and movement models with increasing complexity and secondly on optimally estimating data-driven inputs for such a modeling system. Our contributions enable county-level resolution and an initial analytic understanding for the spread of influenza-like-illnesses (such as Foot and Mouth Disease). This spatial resolution is the highest reasonable level considering publicly available data. The adaptation of the models and parameter estimation to other species such as swine or goat is not a challenging extension. The consideration of diseases with greater complexities, such as diseases involving the environment, wildlife, or multiple species, would need further disease model expansions[5].

This modeling of epidemics in metapopulation livestock systems would serve as a platform for *in-silico* policy testing and implementation by animal health administrative agencies. This modeling system would improve scientists' understanding of the complex, spatial and temporal evolution of livestock disease outbreaks. Researchers could forecast the impacts of economic decisions, veterinary regulations, and control policies on the susceptibility of the United States cattle system to domestic and foreign diseases. With a national picture

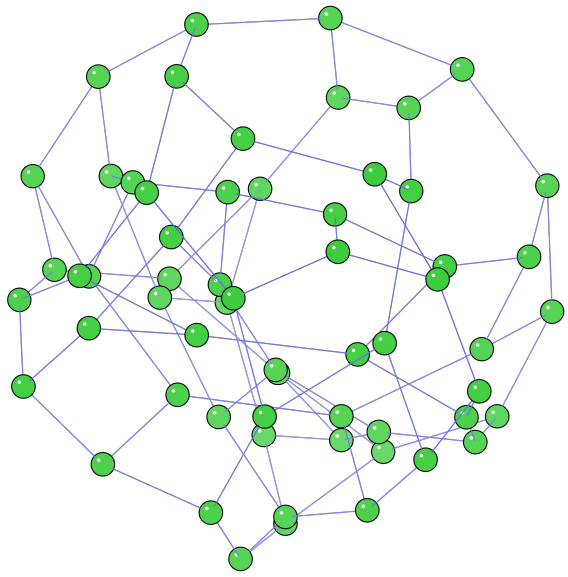

Figure 1.1: *A simple network composed of nodes shown as green circles which are connected through a set of arcs represented by blue lines.*

of the disease challenges, the potential arises for farmers, scientists, and administrators to control and exterminate threats to US livestock.

1.2 Basic terminology

The solutions to a problem as complex as characterizing the cattle systems of the US can be found through multidisciplinary efforts. Due to this, there may be basic vocabulary of one discipline that is unfamiliar to another discipline. This section aims to define a number of terms that may increase understanding of the dissertation content. The topics covered include abstract structures, such as the network of figure 1.1, and the cattle systems that provide familiar, everyday products.

1.2.1 Cattle systems

In the United States, cattle are counted by "head." With one head per individual, a count of 20 head of cattle is equivalent to 20 individuals[1]. Depending on the gender and individual life stage, cattle are known by a number of titles. Young offspring of cattle are "calves." A young female who has not given birth to a calf is called a "heifer." Once she has had her first calf, she is described as a "cow." A common inaccurate description of cattle is to refer to any and every individual as a cow. Although young male cattle have no title equivalent to heifer, they may be castrated and known as "steers," a common practice in beef production. If the male is kept whole, he is called a "bull" and will likely be used for breeding purposes. These cattle reside on premises or cattle "operations." Farms that focus on birthing and raising young calves are known as "cow-calf" operations[9]. In the production of beef, cattle may be fattened through grazing on spreads or ranches, or more efficiently through feeding yards called "feedlots." In feedlots, beef cattle are raised on a regulated diet of feed to prepare them for the end of their 15-24 month-long lives at a "slaughterhouse"[10]. Cows typically live on cow-calf farms or in "dairies" where both milk and calves are produced. Although some cattle are dedicated to become beef from the beginning of their lives, other cattle commonly find their way to a slaughterhouse as their usefulness expires in whichever purpose they were serving. The selection of cattle for slaughter based on the expiration of their utility is known as "culling." In chapter 4, we classify cattle into three types: "Preslaughter," representing cattle in a feed program with their next destination being a slaughterhouse; "Dairy," representing dairy cattle; and "Beef," representing a general collection of all remaining cattle.

1.2.2 Network science

Network Science studies systems through the perspective of networks[11,12]. A "network" can be defined as a set of elements connected through a set of relationships, such as the circles and lines of figure 1.1. Furthermore, a network often (always by mathematicians' defini-

tion) has a set of flows moving through the relationships. A similar concept, a "graph," is a network without any flows. The flows of a network can represent a diffusion dynamic, such as particles or rumors moving between nodes, or more complex dynamics, such as economic trading[12]. In a network, the elements are known as "nodes" or "vertices" and they can represent the components of nearly any system. Nodes have been used to represent individuals, geographical locations, computers, companies, economies, words, and much more[11]. The relationships that connect the nodes of a network are known by a variety of discipline-dependent terms, including "arcs," "links," "edges," and "lines." Once a system is represented as a network with nodes and arcs, one can study the static system of relationships or further a set of dynamic processes taking place in the network[12]. If the arcs between pairs of nodes or the flows taking place on them are not symmetrical, then the arcs and the network are referred to as "directed." One of the most basic parameters used to describe a network is the "node degree." The degree of a node is the count of the number of arcs it is connected to. For directed networks, nodes have both an "in-degree," the number of incoming or arriving arcs, and an "out-degree," the number of outgoing or departing arcs. Two special cases of directed node degrees can occur. When a node has an in-degree of zero, it is known as a "source" node because flows can only originate from it and not arrive to it. Similarly, a node with an out-degree of zero is called a "sink" node. In a directed network, if a node has both in- and out-degree nonzero, the node can be referred to as a "transit" node. Although these definitions are only a few with respect to the field of networks, they are sufficient for a majority of the network-related content of this document.

1.2.3 Metapopulation systems

A system composed of several "populations," namely several "subpopulations" of the system population, which are relatively separated geographically, is known as a "metapopulation" system[13,14,15]. If represented through a network such as in figure 1.1, a metapopulation system would suppose that each node (green circle) would have a population of individuals

residing within it. These nodes represent geographical locations, but as shown in figure 1.1, nodes may not always be visually presented in a geographical orientation. The relationships, represented by the arcs, represent pathways for interactions among the subpopulations. For the work of chapters 2 and 3, metapopulation networks are discussed in which the network and its nodes do not represent any concrete physical locations in the real world, but rather abstractions of such systems. Throughout this dissertation, the interactions that interface the subpopulations are the movements or shipments of cattle from location to location. If the subpopulations of a system are further divided by some criteria into different subparts, they may be referred to as "stratified" subpopulations. In chapter 4, the nodes of the considered metapopulation network represent 1034 counties in 10 States in the Central United States. The populations within these nodes (counties) are each stratified into nine subpopulations by the types of cattle mentioned in subsection 1.2.1 and the sizes of the operations.

1.2.4 Classical SIR disease model

Chapters 2 and 3 study epidemics in cattle populations through a compartmental disease model. The classical "Susceptible-Infected-Recovered" (SIR) model for disease sorts individuals into three disease states, susceptible (S), infected (I), and recovered (R)[12,16]. This model assumes that the considered population is "well-mixed," meaning that the individuals who compose the population are all likely to interact each with all others of the population. The classical SIR model is characterized by two parameters, the "infection rate" and the "recovery rate." The infection rate, β, represents the rate at which an infected individual produces new infections in each susceptible with which he is in contact with. The recovery rate, μ, is the rate at which each infected individual recovers and transitions to the recovered state R. Figure 1.2 presents the state transition diagram of the classical SIR model and an example of individuals with different disease states residing inside and moving into and out of a node.

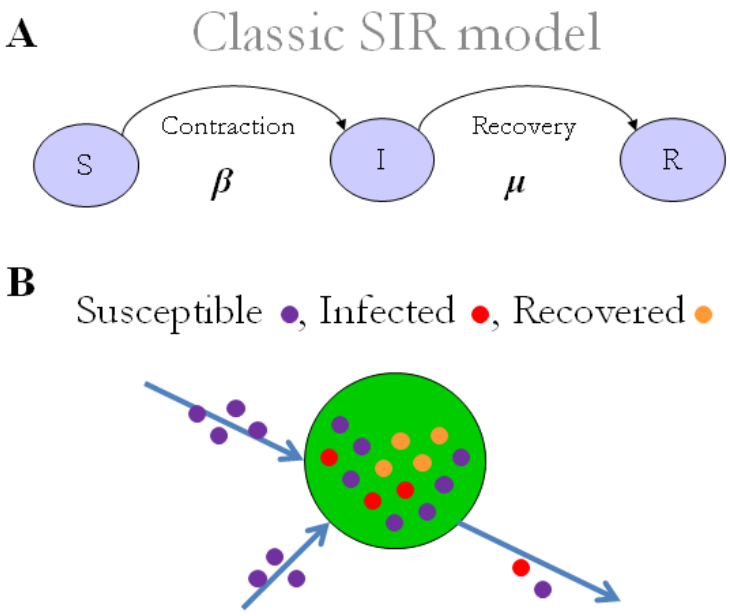

Figure 1.2: *(A) The state transition diagram of the classical SIR model, showing the transition from Susceptible to Infected driven by mixing with infected individuals and the infection rate β and the transition from Infected to Recovered driven by the recovery rate μ. (B) A node in a metapopulation system with susceptible (purple), infected (red), and recovered (orange) individuals residing inside and moving into and out of the node.*

Given the total size N_i of a population i and the initial states of the disease described by the number of individuals in each disease state, the progression of the SIR model can be followed either deterministically through differential equations or numerically through a discrete stochastic model of the processes[12,16]. These disease processes are modeled through binomial processes within each subpopulation of the metapopulation models of chapters 2 and 3. To characterize the disease outbreaks across the cattle systems, three basic metrics are used: the "epidemic duration," the "epidemic attack rate," and the "epidemic size"[17]. The duration of the epidemic measures the time period over which the disease persists, ending when the number of infected individuals reaches zero for all subpopulations of the metapopulation system. The epidemic attack rate is the fraction of the total system population which is infected during the outbreak. The third metric, the epidemic size, summarizes the geographical impact of system-wide outbreaks and is defined as the fraction of locations

(nodes and their respective subpopulations) which have experienced new infections.

1.2.5 Technical terms

In addition to the above terminology, a few more definitions are useful for comprehending this dissertation. In chapters 2 and 3, epidemic results are first derived and then verified through "stochastic" and numerical simulations. The stochastic portion of the simulations is the generation of pseudo-random numbers to model the evolution of the demographics and the disease through random processes. These processes create sample distributions of the metrics with which the results are captured. To present these distributions, statistics such as the sample median, average, and 95% "confidence interval" are employed. In these chapters, the 95% confidence interval represents the central 95% of each distribution, in which 2.5% has been removed from each end of the ordered distribution of samples. Similarly, chapter 4 describes a portion of results with 99% confidence intervals. In the same chapter, an "optimization problem" is formulated[18]. An optimization problem is composed of a set of "constraints" and usually an "objective function." It is also possible that the set of constraints is empty and only the objective is optimized. The constraints are a set of equalities and inequalities that are expressed through the problem variables and that define the feasible set of solutions to the problem, also known as the solution space. An objective function serves to quantify the quality of the solution, where the set of values of the problem variables in the solution space that produce the highest (lowest) feasible objective value represents the optimal solution of the maximization (minimization) problem. There exist many methods to determine solutions for different optimization problems, where the best method depends on the nature of the constraints and the objective function[18].

1.3 Contributions and overview

In recent years a global epidemic invasion threshold R_* has been introduced to characterize the system level threat of disease outbreaks for metapopulation systems on undirected networks[17,19]. This parameter defines the threshold between the situation where an epidemic is contained within one subpopulation and the situation where the epidemic spreads out across the system. In chapter 2, we derive this threshold for directed livestock networks without sink and source nodes. The critical movement rate p_c is derived, and simulations are used to validate the mean-field existence of epidemics only when the average movement rate p exceeds the critical movement rate.

Not satisfied to remain at this level of abstraction, we extended the livestock model of chapter 2 to include death, birth, and importation processes in chapter 3. With these additions, we modeled cattle systems on directed networks with source and sink nodes. For this second model, we derived the global epidemic invasion threshold R_*, discovered a second invasion threshold R_*^{TS}, determined the respective critical movement rates, p_c and p_c^{TS}, and validated these thresholds through extensive simulation. We found that the relationship between p_c and p_c^{TS} depends on the fractional composition and the in-degree distributions of the transit and sink nodes. The second threshold of R_*^{TS} and p_c^{TS} defines the criteria for an epidemic to move from the transit to the sink nodes of a network. The spread of diseased cattle into the sink nodes is of particular concern as these nodes represent the last feedlots before cattle are slaughtered and move into the human food chain.

Chapter 4 examines the characterization of epidemics in metapopulation cattle systems through an estimation of actual cattle movements. We developed a method to determine cattle movement parameters from publicly available databases of cattle demographics[1]. As described in chapter 4, we first collect and polish certain sets of parameters and then input them into the large, convex constrained optimal estimation problem we have formulated. This optimization maximizes the entropy of the out-going distributions of cattle movement parameters, permitting the minimal number of assumptions to remain in the optimal solu-

tion. We solve this problem for a set of 10 Central States in the US, yielding movement parameters and birth, culling, and slaughter rates for the region. We conclude chapter 4 with a proposal of a set of questions that would greatly benefit future efforts to estimate cattle movements without challenging the privacy of cattle operations.

The metapopulation cattle model and the movement estimation efforts have yet to converge into a single data-driven model of cattle demographics and disease. In chapter 5, we outline the remaining steps necessary for this convergence in a discussion of future work. The first contact between these two perspectives appears in subsection 5.1.2, where we compare the average movement rates of the 10 Central States with the critical movement rates p_c and p_c^{TS}, computed from the networks of estimated cattle movements. We close chapter 5 and this dissertation with a short summary of the content. The contributions of the work presented in this document are summarized as

1. The expression and numerical validation of the global epidemic invasion threshold R_* and the related critical movement rate p_c for directed metapopulation cattle movement networks with neither source nor sink nodes,

2. The expansion of a metapopulation model of cattle movements and epidemics for directed networks having source, sink, and transit nodes,

3. The expression and numerical validation of the global epidemic invasion threshold R_* and the related critical movement rate p_c for directed metapopulation cattle movement networks with both source and sink nodes,

4. The expression and numerical validation of a novel epidemic invasion threshold R_*^{TS} and the related critical movement rate p_c^{TS} between transit and sink nodes for directed metapopulation cattle movement networks with both source and sink nodes,

5. The derivation of the relationship between the pair of critical movement rates p_c and p_c^{TS},

6. The development of a flexible, scalable, and convex optimal estimation method to approximate the cattle movement parameters of the United States through publicly available data,

7. The estimation of demographic and movement parameters for cattle across 10 Central States, and

8. The comparison between the estimated cattle movement rates and the corresponding critical movement rates p_c and p_c^{TS} to determine the potential for epidemics to devastate the 10 Central States with the outbreak of an epidemic among their cattle populations.

1.4 List of symbols

Here we present two tables of the most commonly used symbols in this dissertation sorted by the two sides of our effort to characterize epidemics in cattle systems.

Symbol	**Definition**
V	Number of nodes (populations)
$N_i(t)$	Number of individuals in node i at time t
$\overline{N}$	Average node population
S	Susceptible individuals
I	Infected individuals
R	Recovered individuals
β	Infection rate
μ	Recovery rate
$\lambda_i(t)$	Force of infection on each individual in node i at time t
R_0	Basic reproduction number

Continued on next page

Table 1.1 – *Continued from previous page*

Symbol	**Definition**
p	Movement rate of each individual
d_{ij}	Per capita rate of flow of cattle from node i to node j
p_β	Fraction of recycled cattle born rather than imported
δ	Slaughter rate of each individual in a sink node
$\delta_{x,y}$	Kronecker Delta
k_i^{in}	In-degree of node i
k_i^{out}	Out-degree of node i
$\vec{k}_i = (k_i^{in}, k_i^{out})$	Joint node degree of node i
η_{in}	Fraction of nodes having no out-degree (sinks)
η_{out}	Fraction of nodes having no in-degree (sources)
$\vec{k}^{(1)}$	Joint node degree of a source node
$\vec{k}^{(2)}$	Joint node degree of a transit node
$\vec{k}^{(3)}$	Joint node degree of a sink node
$N_{\vec{k}}, d_{\vec{k}\vec{k}'}, V_{\vec{k}}$	Respective degree-block variables for N_i, d_{ij}, V
$D_{\vec{k}}^n$	Number of diseased nodes with degree $\vec{k}$ in generation n
R_*	Global epidemic invasion threshold
p_c	Critical movement rate of R_*
R_*^{TS}	Global transit-to-sink epidemic invasion threshold
p_c^{TS}	Critical movement rate of R_*^{TS}

Table 1.1: Commonly used symbols in chapters 2 and 3

Symbol	Definition
P_t^{tot}	Total number of cattle of type t
$Type_A, Type_B$	Sets of cattle types
$Dairy(D)$	Dairy cows
$Preslaughter(P)$	Cattle on feed with slaughter as next destination
$All\ Cattle$	All cattle of any type
$Beef(B)$	All cattle except $Dairy$ and $Preslaughter$
$ShipType$	Set of shipment types
$All\ Movements$	Cattle movements or shipments of any type
$Slaughter$	Cattle movements or shipments for slaughter
$Size_A, Size_B$	Sets of size ranges for populations and shipments
^{x}	Notation to denote a decision variable
^{R}	Notation to denote a result of the data pre-processing
$Tc_{t,c}^x$	Number of cattle of type t in county c
$Tz_{t,i}^x$	Number of cattle of type t for size range i
$Pop_{t,c,i}^x$	Number of cattle of type t in county c of size range i
$Sales_{t,c,i}^x$	Number of cattle shipped of type t from county c for size range i
$Distance$	Set of distances in which cattle may be shipped
$p_{t_1,j_1,t_2,j_2,dist}^x$	Movement parameter from a subpopulation of type t_1 and size range j_1 to a subpopulation of type t_2 and size range j_2 over a distance $dist$
$f_{t_1,j_1,t_2,j_2,dist}$	Binary variable to implement industrial constraints
$bt_{c,t,j}^x$	Birth rate of subpopulation in county c of type t and size range j
$dt_{c,t,j}^x$	Culling rate of subpopulation in county c of type t

Continued on next page

Table 1.2 – *Continued from previous page*

Symbol	**Definition**
	and size range j
$sl^{x}_{c,t,j}$	Slaughter rate of subpopulation in county c of type t
	and size range j
R_C	Scaling factor for year-to-week time scale

Table 1.2: Commonly used symbols in chapter 4

Chapter 2

Global epidemic invasion threshold in directed subpopulation networks

Motivated by trade of livestock we have suggested a theoretical framework in order to define the conditions for global epidemic invasion in directed subpopulation networks. Our theoretical analyses are based on a metapopulation approach in which subpopulations correspond to geographically segregated social units coupled by exchanges of individuals. Inside each subpopulation we have considered the classical infectious-susceptible-recovered compartmental model in order to describe the spreading of the disease in the population via direct contacts. We have derived an analytical expression for the reproduction number at subpopulation level that determines the extinction or invasion of the disease in the metapopulation system. Our analytical calculations show the existence of a phase transition from invasion to extinction that is suppressed by heterogeneity in the architecture of subpopulation networks in the thermodynamic limit. Our analytical findings have been supported by stochastic, discrete time, and individual-based numerical simulations on synthetic subpopulation networks.

2.1 Introduction

Dynamical processes taking place on top of complex networks have attracted a lot of attention in recent years due to accumulation and availability of large-scale reliable data[12,20]. Opinion dynamics in social groups and consensus formation[21], propagation of failure in financial systems[22,23], spreading of infectious diseases in human communities[24,25] are among a vast number of examples where network science has substantially influenced and advanced our research approaches and understanding. The substrate network for geographical spread of infectious diseases corresponds to locations coupled by mobility fluxes of the host carrying and transmitting the disease. The host population in this case is structured in relatively isolated discrete patches or subpopulations where a contagion or reaction process takes place. These subpopulations are connected by a diffusion process or exchanges of individuals that enables the exportation of the contagion process into naive subpopulations. Reaction-diffusion processes have been integrated into a metapopulation modeling scheme where different subpopulations are coupled by mobility of individuals[13,14,15,26]. Metapopulation dynamics have been successfully applied to understand epidemics in spatially structured populations with well defined social units (e.g., families, villages, towns, cities, regions) connected though individuals' movements[27,28,29,30,31,32,33,34,35,36,37,38,39]. Metapopulation dynamics have generated a wealth of results and approaches that take into account the mobility of individuals explicitly[34,40,41,42,43,44,45] or implicitly by integrating the diffusion dynamics into an effective force of infection[32,35,46,47,48,49,50]. Metapopulation approaches have recently been implemented in data-driven computational models for the description of large-scale geographical spread of infectious diseases[6,39,51,52,53,54,55].

The metapopulation dynamics have been extended to complex settings by a particle-network description in that each node populated by a certain number of individuals is connected to a set of other nodes by mobility fluxes. The subpopulation network representation has enabled the study of contagion processes taking place on complex substrate network architectures and coupled by complex mobility schemes[19,56,57,58]. It has been shown

in particular that, along with the usual basic reproduction number, there is a second threshold parameter that depends on the parameters of the contagion process, the characteristics of the mobility scheme and the architecture of the subpopulation network. This global invasion threshold[7,17,19,57,58,59,60,61] defines whether a contagion process successfully spreading in a single subpopulation will lead to outbreaks in a nonzero fraction of subpopulations or will die out in a finite amount of time in the thermodynamic limit.

The metapopulation approach is based on detailed knowledge of the spatial structure of the population and the mobility scheme and network architecture[8,62,63,64,65,66,67,68]. In the case of human infectious disease epidemics, the links of the mobility networks are conveniently regarded undirected as the individuals get back to their original subpopulations within the timescale of the disease[8] or the number of trips between origin and destination are balanced[53]. However, the mobility or displacement of livestock is driven and strictly regulated by livestock trade industry. In the case of infectious diseases of livestock, the subpopulation network underlying the geographical spread is made of animal holdings or premises connected by displacements of livestock among them. Recent accumulation and availability of detailed data on livestock movements[69,70,71,72,73,74,75] have enabled the characterization of livestock movement networks in various European countries. In particular, the networks are highly unidirectional due to specializations of premises (e.g., diary farms, markets, slaughterhouses) and it is rather rare to observe an animal returning to its origin.

Conventional models for the geographical spreading of livestock infectious diseases employ mass-action laws and distance-based kernels in order to account for transmission of infection between premises[4,76]. Data-driven computational models informed by livestock movement databases have increasingly become available in the literature[77,78,79]. Except for several computational studies taking mechanistic approaches within the metapopulation framework[80,81], these models consider animal holdings as their basic units, missing the resolution at the level of individual animals. In this chapter, we propose a theoretical framework based on a metapopulation approach to identify the conditions for global epidemic invasion

of directed subpopulation networks motivated by livestock trade. In Section 2.2 we define the metapopulation model in which subpopulations corresponding to livestock communities are coupled by a Markovian mobility process. By introducing degree-block variables, we obtain the quasi-equilibrium of the system driven by this mobility scheme. In Section 2.3 we introduce the infection dynamics and obtain an analytical expression for the global epidemic invasion threshold parameter. This novel threshold parameter determines whether an infectious disease is going to spread to an appreciable fraction of the subpopulations or die out before reaching high proportions. We support our theoretical findings with numerical simulations in Section 2.4, followed by conclusions and discussions in Section 2.5.

2.2 Metapopulation model of livestock industry

In order to describe the movement of livestock of a single species, let us consider a directed and weighted subpopulation network made of V nodes. Each node i corresponds to a subpopulation or a community which accommodates a number of animals $N_i(t)$ at time t. Each directed link $i \to j$ that will be called an arc represents the direction and the weight d_{ij} the per capital rate of the flow of animals from node i to node j. Each node i is connected to k_i^{in} other nodes via incoming arcs which originate from those nodes in the set υ_i^{in} and is connected to k_i^{out} other nodes via outgoing arcs which terminate at nodes in the set υ_i^{out}. Each node thus has a pair of degrees denoted by $\vec{k}_i \equiv (k_i^{in}, k_i^{out})$. In the continuous and deterministic limit, the temporal change of the animal population in node i due to livestock movement is determined by the rate equation:

$$\frac{dN_i(t)}{dt} = -N_i(t) \sum_{j \in \upsilon_i^{out}} d_{ij} + \sum_{\ell \in \upsilon_i^{in}} d_{\ell i} N_\ell(t) \quad . \tag{2.1}$$

The first expression accounts for the off-movements of livestock in subpopulation i to other subpopulations in the out-neighborhood, while the second expression corresponds to the

on-movements of livestock to subpopulation i from the in-neighborhood. The first question of interest then relates to the population sizes: How is the total population in the system distributed at the equilibrium of the above Markovian mobility process? In order to answer this question, we will rely on the assumption of the statistical equivalence of subpopulations with similar joint-degree $\vec{k}$. This is a mean-field approximation that considers all the subpopulations with same joint-degree as statistically equivalent, thus allowing the introduction of degree-block variables that depend only upon the subpopulation joint-degree.

While this is an obvious approximation to the system description, it has been successfully applied to many dynamical processes on undirected complex networks[7,17,19,56,57]. Imagine that the subpopulation network obeys the join-degree distribution $P_v(\vec{k})$. We will define vertex-based degree-block variables and assume that all the nodes within a joint-degree-class $\vec{k}$ are statistically equivalent. The average population size of nodes in degree-class $\vec{k}$ at time t is expressed as

$$N_{\vec{k}}(t) \equiv \frac{1}{V_{\vec{k}}} \sum_{i|\vec{k}_i=\vec{k}} N_i(t) \ , \tag{2.2}$$

where the sum is performed over all the nodes with joint-degree $\vec{k}$, and $V_{\vec{k}}$ is the total number of such nodes. Similarly, we can define arc-based degree-block variables. The average diffusion rate $d_{\vec{k}\vec{k'}}$ on arcs originating from nodes with joint-degree $\vec{k}$ and terminating at nodes with joint-degree $\vec{k'}$ is given by

$$d_{\vec{k}\vec{k'}} \equiv \frac{1}{E_{\vec{k}\vec{k'}}} \sum_{i|\vec{k}_i=\vec{k},j|\vec{k}_j=\vec{k'}} d_{ij} \ , \tag{2.3}$$

where the summation is performed over all the arcs whose origin and terminal end are occupied by nodes with degrees $\vec{k}$ and $\vec{k'}$, respectively, and $E_{\vec{k}\vec{k'}}$ is the number of such arcs. We can then write down the rate equations of population sizes in the mean-field description

by using the degree-block variables as

$$\partial_t N_{\vec{k}}(t) = -N_{\vec{k}}(t) k^{out} \sum_{\vec{k'}} P_a(\vec{k}, \vec{k'}|\vec{k}) d_{\vec{k}\vec{k'}} + k^{in} \sum_{\vec{k'}} P_a(\vec{k'}, \vec{k}|\vec{k}) N_{\vec{k'}}(t) d_{\vec{k'}\vec{k}} \ , \tag{2.4}$$

where $P_a(\vec{k}, \vec{k'}|\vec{k})$ is the conditional probability of finding a node with joint-degree $\vec{k'}$ in the out-neighborhood of nodes with joint-degree $\vec{k}$, and $P_a(\vec{k'}, \vec{k}|\vec{k})$ is the conditional probability of finding such a node in the in-neighborhood. While the sum in the first expression accounts for the off-movements of animals in subpopulation with join-degree $\vec{k}$ to any of the out-neighbors, the second sum corresponds to the on-movements to the subpopulation.

In order to proceed further with analytical calculations, we need to specify the network structure and the diffusion rates. The simplest case could be that the total per capita diffusion rate is the same p for all the subpopulations and that the diffusion rate from a subpopulation with joint-degree $\vec{k}$ to any one of the out-neighbors is equally weighted as

$$d_{\vec{k}\vec{k'}} = \frac{p}{k^{out}} \ , \ \forall k^{out} \neq 0 \ . \tag{2.5}$$

In this case the rate equation becomes

$$\partial_t N_{\vec{k}}(t) = -p N_{\vec{k}}(t)(1 - \delta_{k^{out},0}) + p k^{in} \sum_{\vec{k'}} P_a(\vec{k'}, \vec{k}|\vec{k}) \frac{N_{\vec{k'}}(t)}{k'^{out}} (1 - \delta_{k'^{out},0}) \ . \tag{2.6}$$

From the above equation, it is obvious that the nodes with zero in-degree will act as *sources* while those with zero out-degree will be *sinks*. We will consider that there are no sources and sinks in the network, i.e., $P_v(0, k^{out}) = 0$ and $P_v(k^{in}, 0) = 0$, corresponding to the fact that we are restricting our analysis to the (giant) strongly connected component of the network. This assumption yields

$$\partial_t N_{\vec{k}}(t) = -p N_{\vec{k}}(t) + p k^{in} \sum_{\vec{k'}} P_a(\vec{k'}, \vec{k}|\vec{k}) \frac{N_{\vec{k'}}(t)}{k'^{out}} \ . \tag{2.7}$$

If we assume that there are no correlations between the degrees of connected nodes, then we have that the conditional probability $P_a(\vec{k}', \vec{k}|\vec{k})$ is independent of $\vec{k}$ and is just equal to the probability of finding an arc originating from a node with joint-degree $\vec{k}'$ (see Eq. (A.2)). The rate equation for the average population size in subpopulations with degree $\vec{k}$ then becomes

$$\partial_t N_{\vec{k}}(t) = -pN_{\vec{k}}(t) + \frac{pk^{in}}{\langle k^{out} \rangle} \overline{N} \,, \tag{2.8}$$

where $\overline{N} = \sum_{\vec{k}'} P_v(\vec{k}') N_{\vec{k}'}(t)$ is the average node population and independent of time. It is worth remarking that we are ignoring birth and death processes, including importation and exportation of animals as well as their movements to slaughter houses. At the equilibrium of the diffusion process, i.e., $\partial_t N_{\vec{k}}(t) = 0$, the average population size of a node with joint-degree $\vec{k}$ is given by

$$N_{\vec{k}} = \frac{k^{in}}{\langle k^{out} \rangle} \overline{N} \,. \tag{2.9}$$

Once the system reaches this equilibrium, then the population size of each subpopulation will no longer change over time. However, this is a quasi-equilibrium in that the population continues moving between the subpopulations while keeping balanced the in- and out-flow of each location. In the following we will assume that the system is already in the equilibrium of the diffusion process and will introduce another dynamical process corresponding to the introduction of an infectious agent.

2.3 Global epidemic invasion of livestock industry

Now let us introduce a contagion process to one of the subpopulations and discuss the conditions for global invasion and extinction of the process. For this purpose we consider the classic susceptible-infected-recovered (SIR) model[16] in order to describe the spread of a biological virus in the livestock population. We will assume that the population is initially fully susceptible and also that livestock mix homogeneously in each subpopulation. Infec-

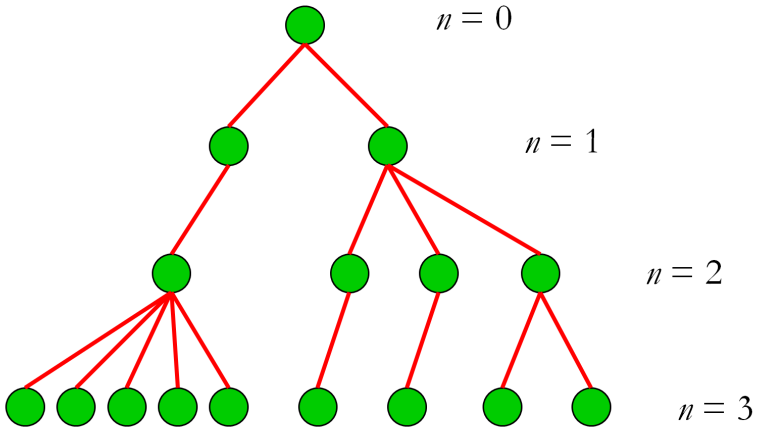

Figure 2.1: *A diagram of a branching process with labels denoting each "generation"* n *of the process. Each node produces a stochastic number of offspring nodes in the following generation.*

tion is transmitted from an infectious animal to a susceptible animal upon contact at rate β. The force of infection acting on each susceptible animal in a subpopulation at time t assumes a mass-action law that is proportional to the prevalence of infectious animals in the subpopulation, that is the rate at which each susceptible animal becomes infectious. Each infectious animal recovers permanently from the disease at rate μ. The model yields the basic reproduction number $R_0 = \beta/\mu$ at individual-animal-level that is the average number of secondary infections generated by a typical infectious animal in its entry infectious period[16]. In the thermodynamic limit (i.e., $N \to \infty$), the basic reproduction number acts as a threshold parameter in that the disease successfully spreads to a finite proportion of livestock in the subpopulation only if $R_0 > 1$. On the other hand, if $R_0 < 1$, then the disease fades out in a finite amount of time affecting only a tiny fraction of the animals (that is zero in the thermodynamic limit). In the early time of the SIR epidemic, we can consider the invasion dynamics as a branching process (see figure 2.1)[7,17,19,57,59,82,83] in which subpopulations experiencing the disease outbreak in the $n-1$th generation may export the contagion process to some of the subpopulations in their out-neighborhood. Letting $D^n_{\vec{k}}$ denote the number of diseased subpopulations with joint-degree $\vec{k}$ in the nth generation, we

may relate $D_{\vec{k}}^n$ and $\{D_{\vec{k}}^{n-1}\}$ by

$$D_{\vec{k}}^n = \sum_{\vec{j}} D_{\vec{j}}^{n-1} P_a(\vec{j}, \vec{k}|\vec{j}) j^{out} p(\vec{j}, \vec{k}) \prod_{m=0}^{n-1} \left(1 - \frac{D_{\vec{k}}^m}{V_{\vec{k}}}\right) . \tag{2.10}$$

The above equation reads as the following: Noting that we ignore bidirectional links, each one of the diseased subpopulations with degree $\vec{j}$ in the $n-1$th generation may export the disease to j^{out} other subpopulation in its out-neighborhood. The conditional probability that a node with degree $\vec{k}$ exists in such an out-neighborhood is given by $P_a(\vec{j}, \vec{k}|\vec{j})$. Given the existence of such a neighbor, the probability that the neighbor has not been infected in earlier generations is given by $\prod_{m=0}^{n-1}\left(1 - \frac{D_{\vec{k}}^m}{V_{\vec{k}}}\right)$. The probability $p(\vec{j}, \vec{k})$ that the diseased subpopulation $\vec{j}$ during the entire course of the epidemic will initiate the outbreak in the neighbor $\vec{k}$ is determined by the number of infected animals sent from $\vec{j}$ to $\vec{k}$ and the probability that the outbreak will occur. For the SIR epidemic model[84], the probability $p(\vec{j}, \vec{k})$ is given by

$$p(\vec{j}, \vec{k}) = 1 - R_0^{-\lambda_{\vec{j}\vec{k}}} , \tag{2.11}$$

where $\lambda_{\vec{j}\vec{k}}$ is the number of infectious seeds generated in subpopulation $\vec{j}$ and sent to subpopulation $\vec{k}$. Given the final size α of an SIR epidemic (i.e., the proportion of individuals who contract the disease)[16] and the average per capita infectious period μ^{-1}, the average number of exported seeds is

$$\lambda_{\vec{j}\vec{k}} = \alpha \mu^{-1} N_{\vec{j}} d_{\vec{j}\vec{k}} . \tag{2.12}$$

We can assume the term $\prod_{m=0}^{n-1}\left(1 - \frac{D_{\vec{k}}^m}{V_{\vec{k}}}\right) \simeq 1$ in the early time of the epidemic. If we are just above the local epidemic threshold, i.e., $R_0 \simeq 1$, then the probability $p(\vec{j}, \vec{k})$ can be approximated by $p(\vec{j}, \vec{k}) \simeq \lambda_{\vec{j}\vec{k}}(R_0 - 1)$ and α by $\alpha \simeq 2(R_0 - 1)R_0^{-2}$. Under these assumptions, the above recursion relation becomes

$$D_{\vec{k}}^n = 2(1 - R_0^{-1})^2 \mu^{-1} \sum_{\vec{j}} D_{\vec{j}}^{n-1} P_a(\vec{j}, \vec{k}|\vec{j}) j^{out} N_{\vec{j}} d_{\vec{j}\vec{k}} . \tag{2.13}$$

Plugging the diffusion rate Eq. (2.5) and the equilibrium population Eq. (2.9) into the above expression yields

$$D_{\vec{k}}^{n} = \frac{2\overline{N}p(1-R_0^{-1})^2\mu^{-1}}{\langle k^{out}\rangle}\sum_{\vec{j}} D_{\vec{j}}^{n-1} P_a(\vec{j},\vec{k}|\vec{j}) j^{in} \ . \tag{2.14}$$

If we assume that there are no correlations between the degrees of the nearest neighbors, then the conditional probability $P_a(\vec{j},\vec{k}|\vec{j})$ reduces to the probability of finding a node with joint-degree $\vec{k}$ at the terminal end of a randomly selected arc (see Eq. (A.3)), and the above equation yields

$$D_{\vec{k}}^{n} = \frac{2\overline{N}p(1-R_0^{-1})^2\mu^{-1}}{\langle k^{out}\rangle\langle k^{in}\rangle} k^{in} P_v(\vec{k}) \sum_{\vec{j}} D_{\vec{j}}^{n-1} j^{in} \ . \tag{2.15}$$

By defining auxiliary variable $\Theta^n \equiv \sum_{\vec{j}} D_{\vec{j}} j^{in}$ and noting that $\langle k^{out}\rangle = \langle k^{in}\rangle$, we obtain the following relationship:

$$\frac{\Theta^n}{\Theta^{n-1}} = 2\overline{N}p(1-R_0^{-1})^2\mu^{-1}\frac{\langle (k^{in})^2\rangle}{\langle k^{in}\rangle^2} \ . \tag{2.16}$$

The dynamical behavior of the system is determined by the right-hand-side of the above expression that we may call as the branching number or the basic reproduction number R_* of the epidemic at individual-subpopulation-level. This is equivalent to defining a subpopulation reproductive number[7,17,19,57,59,60,61] that, in structured metapopulation systems, is the average number of infected subpopulations generated by a typical infected subpopulation in a fully susceptible metapopulation system. If $R_* > 1$, then the epidemic is going to spread to a finite fraction of subpopulations in the thermodynamic limit, i.e., $V \to \infty$. On the other hand, if $R_* < 1$, then the epidemic is going to be confined to a small neighborhood and will die out in a finite amount of time. Thus, $R_* = 1$ sets the threshold values for the disease and mobility parameters as well as the properties of the connection matrices. The critical value p_c of the per capita diffusion rate above which the epidemic will spread globally in the subpopulation network is

$$p_c = \frac{1}{2\overline{N}(1-R_0^{-1})^2\mu^{-1}}\frac{\langle k^{in}\rangle^2}{\langle (k^{in})^2\rangle} \ . \tag{2.17}$$

One interesting feature of the above expression we like to highlight is the explicit dependence of p_c on the topology of the subpopulation networks through the moments $\langle k^{in} \rangle$ and $\langle (k^{in})^2 \rangle$ of in-degree distribution. As has already been shown for several dynamical processes on undirected subpopulation networks, heterogeneity in the network connections favors the spreading of infectious diseases by lowering the threshold value. Interestingly, however, in the case of directed subpopulation networks, this striking effect is caused by the heterogeneity of in-degrees. Indeed, for heavy-tailed in-degree distributions $P_v^{in}(k) \sim k^{-\gamma}$ with $\gamma > 1$ and $k_{\min}^{in} \leq k \leq k_{\max}^{in}$, the second moment scales with the maximum in-degree in case of $\gamma < 3$. This means that the second moment of the in-degree distribution tends to diverge at the infinite size limit of the network, as in this limit $k_{\max}^{in} \to \infty$, virtually reducing the threshold to zero.

It is worth noting that the above analyses are restricted to epidemics of animal infectious diseases initiated in the giant strongly connected component of a synthetic livestock industry network in which death and birth processes have also been ignored. We have relied our analyses on the statistical equivalence of degree classes. It is also worth remarking that we have considered that there are no correlations between the joint-degrees of connected subpopulations. We have also made a strong simplifying assumption that the total off-movement rates are the same across all the subpopulations and are independent of time. We have considered only a single species of livestock and a single type of livestock subpopulation.

2.4 Stochastic simulations

In order to validate our analytical findings, we have performed Monte Carlo simulations. We adopt mechanistic numerical simulations in which each individual-animal was tracked in time during both the diffusion and the infection dynamics.

2.4.1 Synthetic subpopulation networks and dynamical processes

In order to compare with theoretical calculations, the subpopulation network structure has been generated by wiring the subpopulations according to three different random graph topologies. We have assumed that the in-degree and out-degree of each single subpopulation is independent from each other, meaning that there are no single-vertex degree correlations, i.e., $P_v(\vec{k}) = P_v^{in}(k^{in})P_v^{out}(k^{out})$. The random network topologies obey the following degree distributions:

- Heterogeneous in-degree & homogeneous out-degree distribution.
- Heterogeneous in-degree & heterogeneous out-degree distribution.
- Homogeneous in-degree & heterogeneous out-degree distribution.

All the synthetic networks have been generated by an uncorrelated configuration model[85,86]. The heterogeneous in(out)-degree distribution assumes a power-law function $P_v^{in(out)}(k) \sim k^{-\gamma}$ confined to the interval $k_{min} \leq k \leq k_{max}$. The homogeneous out(in)-degree distribution, on the other hand, follows a Poisson function whose mean value is set by the mean value of the power-law in(out)-degree distribution for the sake of comparison. In all the network topologies, we have set the minimum in(out)-degree to $k_{min} = 1$, while the maximum in(out)-degree is fixed at $k_{max} = V^{1/(5-\gamma)}$ in order to avoid topological two-vertex degree correlations[87]. The exponent of the power-law in(out)-degree distribution has been fixed at $\gamma = 2.1$.

We construct the metapopulation system of V subpopulations from a pool of $\overline{N}V$ animals. Initial population size N_i of each subpopulation i is chosen at random from a multinomial distribution with probability proportional to k_i^{in}, which ensures that the metapopulation system obeys Eq. (2.9). Each individual animal is subject to discrete and stochastic processes defined by the mobility and infection dynamics as detailed in appendix section A.2. The rate d_{ij} at which each animal in subpopulation i moves to a neighboring subpopulation

$j \in v_i^{out}$ assumes Eq. (2.5):

$$d_{ij} = \frac{p}{k_i^{out}} \quad . \tag{2.18}$$

Each animal in subpopulation i at time t leaves its current subpopulation to join subpopulation j in the out-neighborhood with probability $d_{ij}\Delta t$, where Δt is the time interval considered. Inside each subpopulation, we consider an SIR epidemic model in which each animal is classified by one of the discrete disease states at any point in time. The rate at which a susceptible animal in subpopulation i acquires the infection, the so-called *force of infection* λ_i, is determined by interactions with infectious livestock in the subpopulation. The force of infection $\lambda_i(t)$ acting on each susceptible animal in subpopulation i at time t has been assumed to follow the mass-action principle

$$\lambda_i(t) = \beta \frac{I_i(t)}{N_i(t)} \quad , \tag{2.19}$$

where β is the transmission rate of infection and $I_i(t)/N_i(t)$ is the prevalence of infectious livestock in the subpopulation. Given the force of infection $\lambda_i(t)$ in subpopulation i, each animal in the susceptible compartment (S) contracts the infection with probability $\lambda_i(t)\Delta t$ and enters the infectious compartment (I). Each infectious animal permanently recovers with probability $\mu\Delta t$ and enters the recovered compartment (R).

2.4.2 Numerical results

The global epidemic invasion is determined by the disease parameters as well as the architecture of the trade networks and the rate at which the animals diffuse on them. In the following, we present numerical simulations focusing on the interplay between network topology and diffusion rate, and verify the analytical result of Eq. (2.17).

In all the simulations, we have set the size of the subpopulation networks to $V = 10^5$ and the average number of animals per subpopulation to $\overline{N} = 10^4$, which yields a total of 10^9 livestock. Note that each subpopulation with joint-degree $\vec{k}$ is occupied by a number of

animals that is determined by the in-degree of the subpopulation given in Eq. (2.9). Simulations have been initialized with $I(0) = 10$ infectious animals seeded in a single subpopulation chosen at random with minimum out-degree, while the rest of the livestock population is assumed to be susceptible to infection. Since the networks are directed, we have chosen the seed subpopulation in the giant strongly connected component of the network. As our focus is on the global invasion threshold, we have let the metapopulation system progress until the infection dies out. Since we aim at uncovering the interplay between network topology and diffusion rate on the global epidemic invasion, we have fixed the disease parameters for all the simulations, setting the basic reproduction number to $R_0 = 1.2$, the infectious period to $\mu^{-1} = 7\ day$, and the unitary time scale of the simulations to $\Delta t = 1\ day$. In the results we present here, all the realizations resulting in the outbreak of initially seeded subpopulation (requiring an attack rate of 1%) have contributed to the statistical analysis. For each set of parameters, we have generated at least $1,000$ system realizations. Since the subpopulation networks and dynamical processes on them are subject to fluctuations, for each set we have sampled 10 network realizations and 100 dynamical realizations on each of them resulting in successful epidemics (referring to observation of outbreak in the initially seeded subpopulation).

The expression Eq. (2.17) for the critical value of diffusion rate displays dependence on the topology of the subpopulation networks only through the in-degree distribution. In Fig. 2.2 we compare the results of numerical simulations performed on subpopulation networks with the same heterogeneous in-degree distribution while varying the out-degree distribution from heterogeneous to homogeneous. On the same plots we also show the critical value p_c of diffusion rate obtained from analytical calculations above which the metapopulation is invaded by the disease. We clearly see that the out-degree distribution does not alter the critical value of the diffusion rate, that is in agreement with theoretical findings.

We now turn our attention to the impact of varying in-degree distribution. Fig. 2.3

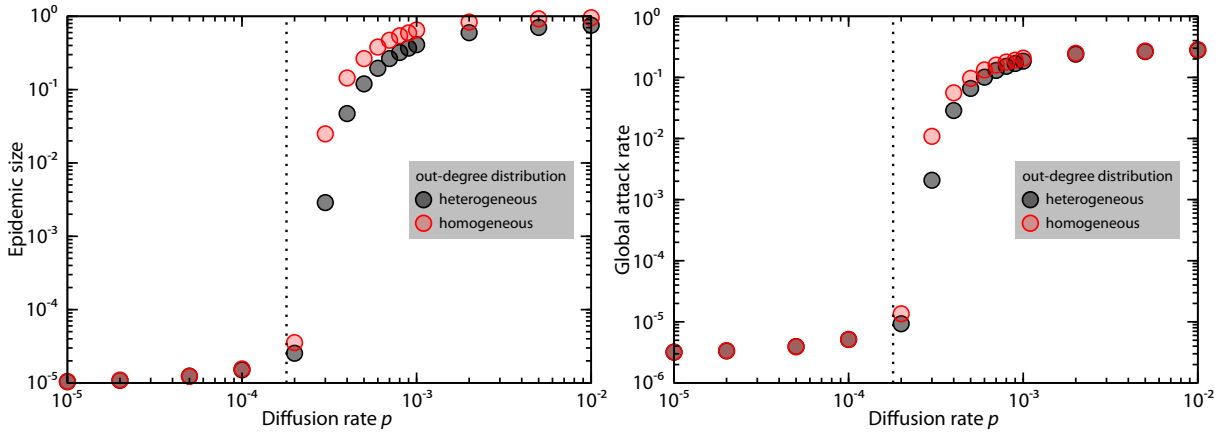

Figure 2.2: *Average epidemic size and global attack rate as a function of p on subpopulation networks with homogeneous and heterogeneous out-degree distributions. On the left panel we display the fraction of subpopulation affected and on the right panel the fraction of livestock who got infected by the end of the epidemic. Both networks show the same critical value of the diffusion rate above which the disease spreads to an appreciable fraction of subpopulations or livestock. The in-degree distribution of both networks is homogeneous.*

displays the average fraction of subpopulations and the proportion of livestock affected by the epidemic as a function of diffusion rate as we switch from a homogeneous to a heterogeneous in-degree distribution while keeping the out-degree distribution fixed. The difference between the critical values of diffusion rate are substantial in that p_c in the case of homogeneous in-degree distribution is 900% higher than that of the heterogeneous case. The two network topologies are not just quantitatively but also qualitatively different such that, in the thermodynamic limit $V \to \infty$, heterogeneous in-degree distribution suppresses the phase transition, i.e., $p_c \to 0$.

2.5 Conclusions

In this chapter, we have proposed a theoretical framework in order to derive the conditions for the global epidemic invasion of subpopulations coupled with exchanges of livestock. Our analyses have extended the mean-field approximation based on degree-block variables[7,56] to directed random graphs. This extension has allowed us to understand the impact of

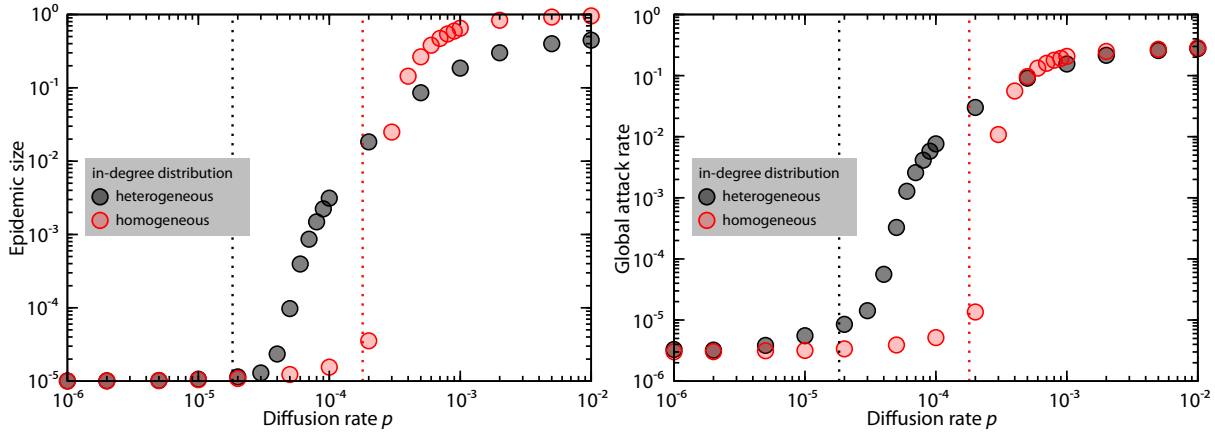

Figure 2.3: *Average epidemic size and global attack rate as a function of p on subpopulation networks with homogeneous and heterogeneous in-degree distributions. On the left panel we display the fraction of subpopulation affected while on the right the fraction of livestock who got infected by the end of the epidemic. The difference between the critical values of the diffusion rate above which the disease spreads to an appreciable fraction of subpopulations or livestock can be clearly marked as the in-degree distribution of the subpopulation networks changes. The out-degree distribution of both networks is homogeneous.*

heterogeneous network topologies on the dissemination of livestock diseases.

We have shown that there is a subpopulation reproduction number R_* that depends not only on the basic reproduction number, but also on the infectious period of the disease, the diffusion rate of livestock, and the network architecture. The basic reproduction number R_0 is responsible for the disease to spread in a single subpopulation, in which if $R_0 > 1$, then the disease can spread to a finite fraction of livestock. However, in the case of subpopulations coupled with livestock trade, this condition is necessary but not sufficient. There is a second threshold parameter R_* that determines the extinction or invasion of the disease at the global scale. In the case of structured livestock populations, the infectious period of the disease is as important as the animal-to-animal transmission rate. This is easy to understand as the disease finds more occasions to export itself to other subpopulations if the infectious period is longer. If the disease parameters and network architecture are fixed, there is a critical value of livestock diffusion rate below which the epidemic is confined to a tiny fraction of subpopulations and dies out in a finite amount of time. We have also shown that the

heterogeneity in the in-degrees favors the disease spread at global scale. In particular, the phase transition is suppressed in the thermodynamic limit if the in-degree distribution is heavy-tailed with an exponent $\gamma < 3$.

All the analyses presented in this chapter are based on several simplifications which need to be addressed. One of the assumptions we have made is that there are no demographical changes in the system. This means that we have ignored natural birth and death processes as well as the importation and exportation of livestock. Even though this is a crude approximation of the system's description, it might be suitable for highly transmissible diseases with short timescales such as Foot and Mouth Disease (FMD). We have restricted our analyses to the giant strongly connected component of the trade network, meaning that we have discarded the in- and out-components (must be dominated by sources and sinks, respectively) as well as other weakly connected components. Given that the Giant Strongly Connected Component (GSCC) covers a very high percentage of premises in livestock trade networks aggregated over relatively large time windows, this assumption is reasonable as a starting point. For its mathematical simplicity, we have assumed that all the livestock move at a constant rate in that the rate of movement to a specific neighboring subpopulation is equally weighted with the inverse of the out-degree of the subpopulation at origin. This assumption can be relaxed, however, calling for empirical observations on the form of the diffusion rates. Moreover, we have considered that the movement rates do not change over time. This means that we restrict our analyses to trade networks aggregated over extended time windows, treating the links as static objects. This assumption can be proper if the time scale of the disease is longer than the time scale of the inter-events (the time between consecutive movements along the very same link). There has been an increasing amount of research on dynamical networks with the availability of data in recent years[20,88]. However, there are many theoretical and technical issues to be addressed. Another assumption on the architecture of the networks is that there are no correlations between the degrees of connected nodes. Literature on the livestock trade network analyses does not provide any

information on the amount and form of the correlations. One exceptional paper[75] states that there are disassortative correlations obtained by assuming no directions on the links. The last restriction we have imposed that there is only a single species of livestock in each subpopulation.

Some of these assumptions will be the subject of future research (including in the next chapter) and may call for more sophisticated theoretical approaches. However, we need the light of empirical observations. It has been demonstrated that animal registration and movement data sets are crucial for modeling of transmissible livestock disease dissemination. These data allow us to deploy policies for prevention and control as well as the assessment of their effectiveness. Despite the vast number of assumptions we have made, our study has attempted for the first time to develop a theoretical framework to address some of the most interesting questions on the dissemination of infectious livestock diseases and suggests a basis for future studies on this direction. We continue this work in chapter 3 where we include source nodes, sink nodes, birth, death, and importation processes. Although the analysis increases in complexity, it reveals a very significant second global invasion threshold which was not previously revealed.

Chapter 3

Global epidemic invasion thresholds in directed cattle subpopulation networks having source, sink, and transit nodes

Through the characterization of a metapopulation cattle disease model on a directed network having source, transit, and sink nodes, we derive two global epidemic invasion thresholds. The first threshold defines the conditions necessary for an epidemic to successfully spread at the global scale. The second threshold defines the criteria that permit an epidemic to move out of the giant strongly connected component and to invade the populations of the sink nodes. As each sink node represents a final waypoint for cattle before slaughter, the existence of an epidemic among the sink nodes is a serious threat to food security. We find that the relationship between these two thresholds depends on the relative proportions of transit and sink nodes in the system and the distributions of the in-degrees of both node types. These analytic results are verified through numerical realizations of the metapopulation cattle model.

3.1 Introduction

Today's computing and technological resources have enabled the compilation of large-scale reliable data sets. Computational and network sciences have provided substantial advances in many fields through these data[12,20,21,22,23]. Not the least among these fields is the study of disease spread through human populations[24,25]. The spatial characterization of a spreading disease can be captured through a division of the host population into relatively distinct and discrete patches of subpopulations, each with a unique geographical location[89,90]. The disease spread then occurs within subpopulations possessing infectious individuals, and the transfers of individuals between pairs of locations allows the process to spread through the system[13,14,15,26]. For spatially structured populations of well-defined social groups interfaced through individuals' movements, epidemics have been successfully characterized using metapopulation dynamics[27,28,29,30,31,32,33,34,35,36,37,38,39]. Mobility dynamics have been used to study systems coupled by both individual movements and aggregations of movements as effective forces of infection between subpopulations[32,34,35,41,42,43,44,45,46,47,48,49,50,91]. Recently, data-driven computational models have employed such metapopulation approaches to describe the large-scale geographical spread of infectious diseases[6,39,51,52,53,54,55].

The study of metapopulation systems has been adapted to complex networks where each node hosts a subpopulation of individuals and these individuals flow between nodes on mobility processes. This has led to the exploration of contagion processes taking place on complex network structures and interconnected by complex mobility models[19,56,57,58]. A global epidemic invasion threshold has been discovered that defines whether a contagion process spreading successfully in a single subpopulation will spread to a notable fraction of the other subpopulations or will die out in a finite amount of time in the thermodynamic limit[7,17,19,57,58,59,60,61]. This threshold parameter can be determined from the characteristics and parameters of the contagion process, the mobility model, and the subpopulation network structure.

A successful metapopulation approach is based on detailed knowledge of the contagion pro-

cess, the mobility model, and the spatial subpopulation network structure[8,62,63,64,65,66,67,68]. In the study of human infectious disease outbreaks, the mobility arcs of the subpopulation network can and have been considered undirected and symmetrical[8,53]. When livestock systems are modeled in subpopulation networks, very different processes drive the movements of individual cattle within the livestock industry. Recent collection, compilation, and analyses of detailed data describing individual livestock movements have enabled the construction of livestock movement networks in various European countries[69,70,71,72,73,74,75]. These studies have found that the arcs of livestock networks are highly unidirectional due to specializations of subpopulations (dairy farms, markets, slaughterhouses, and such) and that an animal rarely returns to its origin[92]. From livestock movement databases, data-driven computational models of geographical spread of livestock disease have increased in popularity in the literature[5,77,78,79]. The conventional models have implemented mass-action laws and distance-based kernels and have not been driven by the movement databases[4,76,93,94]. Apart from some computational models using mechanistic approaches in the metapopulation structure, these models consider livestock premises as their basic units and miss the resolution level of individual animals[80,81].

In this chapter, we design a metapopulation network model and investigate the existence of the novel global epidemic invasion threshold for directed networks. We illustrate these directed system results on a basic livestock industry model. Our model system includes directed cattle movements leading to a natural division of source, transit, and sink subpopulations. Within cattle systems, source nodes may represent cow-calf or dairy farms that rarely purchase cattle, transit nodes may represent grazers or backgrounders, and sink nodes may represent large feed/finishing operations (final waypoints prior to slaughter). In section 3.2, our metapopulation livestock system is defined on directed networks connected through cattle movements and death, birth, and importation processes are described within each subpopulation. With the use of degree-block variables, we find the quasi-equilibrium population levels across the network that result from this combination of population flows[17]. In section

3.3, we add a classic susceptible-infected-recovered model within each subpopulation and determine the directed system version of the undirected global epidemic invasion threshold R_* [17,19]. This threshold defines whether a disease can be expected to spread through the system or to quickly die out. Beyond the global invasion threshold for directed metapopulation systems, we discover a second invasion threshold that defines whether a disease outbreak will likely circulate within transit subpopulations or will likely spill over into the sink subpopulations, consequently threatening food safety. We utilize computational resources to extensively simulate this model, demonstrating the analytical observations through numerical exploration in section 3.4. Section 3.5 summarizes this chapter with conclusions and discussions.

3.2 Metapopulation model of livestock industry

We describe a generic livestock system through a directed subpopulation network where each node i represents a premises or small, well-mixed region with a homogeneous population of cattle $N_i(t)$. Cattle flow between nodes on directed arcs, with each arc described by its origin node i and its destination node j. In directed networks, each node may have a set of arcs coming into the node from an in-neighborhood and a set of arcs leaving the node to its out-neighborhood. The number of arcs arriving to node i is the in-degree of node i and is represented by k_i^{in}. Likewise the number of arcs departing node i is the out-degree k_i^{out}. Each node i then has a pair of node degrees which we represent by a degree vector $\vec{k_i} \equiv (k_i^{in}, k_i^{out})$. We classify each of V nodes based on their respective node degrees. For every node i having $k_i^{in} = 0$, we refer to node i as a "source" node. "Sink" nodes are all nodes such that $k_i^{in} \neq 0$ and $k_i^{out} = 0$. The remaining nodes, each node i having both node degrees non-zero, we describe as "transit" nodes. We denote the fractions of sources and sinks in the network respectively by η_{out} and η_{in}.

The life cycle of our cattle begins as they are either born or imported into the set of source

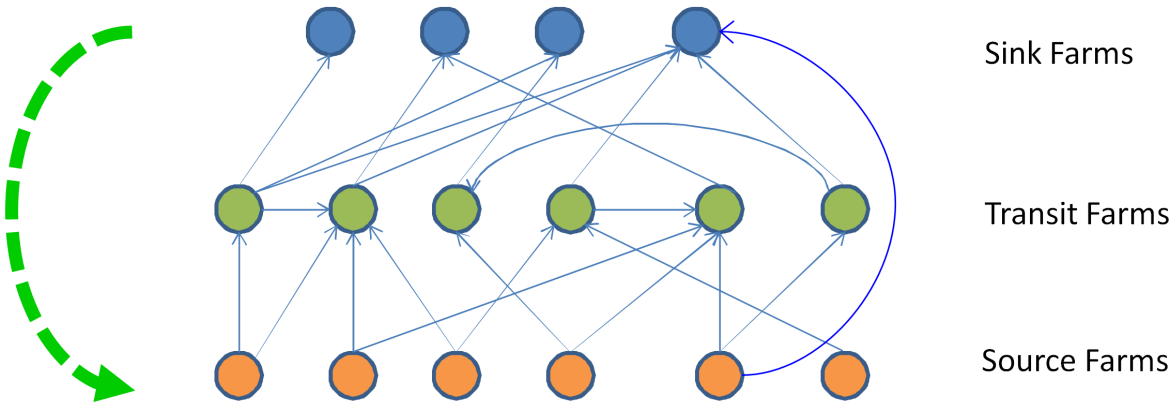

Figure 3.1: *A simple directed network composed of source, transit, and sink nodes which are colored orange, green, and blue, respectively. Individuals can move from source nodes into transit or sink nodes and from transit nodes to other transit nodes or sink nodes. Individuals of any disease state are removed from the sink nodes through a death process and recycled as susceptible individuals into the source nodes through birth and importation processes (shown as a green dashed arrow).*

nodes. After arriving at a sink node, cattle have the possibility to terminate their life cycle when they move off to a slaughter house at a rate of δ per individual. The journey of each individual from its source (or the node at which the individual is initiated) can lead to either a sink node or perpetual wandering among the transit nodes through a movement process. These movements occur inward to each transit or sink node j from its in-neighborhood and outward from each source or transit node i to its out-neighborhood. For each source or transit node i, each individual moves out on one of the k_i^{out} arcs departing the node with a uniform movement rate

$$d_{ij} = \frac{p}{k_i^{out}} \ , \tag{3.1}$$

where p represents the total per capita diffusion rate from all nodes having $k_i^{out} \neq 0$. Let $\overline{N}$ be the system average number of cattle per node. We maintain the total system population of cattle $\overline{N}V$ by recycling the slaughtered cattle back into the system through the birth and importation processes as new susceptible cattle in the source nodes. Figure 3.1 illustrates this recycling process on a simple directed network with the nodes sorted by type. The

fraction of recycled cattle that are selected for birth rather than importation is p_β. Further details are available in appendix B, section B.3.

The first step in characterizing the three nodes classes, source, transit, and sink, is to determine the equilibrium distributions of their populations. We explore a mean-field approximation to the livestock system by assuming the statistical equivalence of each group of nodes (and their subpopulations) that share the same joint-degree vector $\vec{k_i}$. This allows us to introduce "degree-block" variables which are indexed by $\vec{k_i}$ rather than individual node i. Such an approximation is not novel and has been successfully demonstrated in several dynamical systems[7,17,19,56,57]. The degree-block variables are then defined from the node variables and statistical equivalence is assumed for every node within the same degree-block, that is, sharing the same degree $\vec{k_i}$. The average population size of nodes at time t in degree-block $\vec{k_i}$ is then

$$N_{\vec{k}}(t) \equiv \frac{1}{V_{\vec{k}}} \sum_{i|\vec{k_i}=\vec{k}} N_i(t) \ , \tag{3.2}$$

where $V_{\vec{k}}$ is the number of nodes having degree $\vec{k}$. The arc-based variables can also be defined between degree-blocks rather than between individuals. The average diffusion rate on arcs from nodes with degree $\vec{k}$ to nodes with degree $\vec{k'}$ is defined to be

$$d_{\vec{k}\vec{k'}} \equiv \frac{1}{E_{\vec{k}\vec{k'}}} \sum_{i|\vec{k_i}=\vec{k},j|\vec{k_j}=\vec{k'}} d_{ij} = \frac{p}{k_i^{out}} \ \forall k_i^{out} \neq 0 \ , \tag{3.3}$$

where $E_{\vec{k}\vec{k'}}$ is the number of arcs originating from nodes with degree $\vec{k}$ and terminating at nodes with degree $\vec{k'}$. Note that, since the movement as described in equation 3.1 depends only on the degree, the average degree-block expression for movement is equivalent. We present a detailed discussion and further derivations regarding these degree-block variables in appendix section B.1. Now working with degree-block variables, we derive the average population sizes of nodes in each class by degree $\vec{k}$ at the quasi-equilibrium of the collective process of cattle movement, birth, death, and importation. At this point, let us introduce a

notation to distinguish the degrees of each class of nodes. We denote the degrees with the source class, transit class, and sink class as $\vec{k}^{(1)}$ ($k_i^{in} = 0 \; \forall \; i$), $\vec{k}^{(2)}$ ($k_i^{in} \neq 0$ and $k_i^{out} \neq 0 \; \forall \; i$), and $\vec{k}^{(3)}$ ($k_i^{in} \neq 0$ and $k_i^{out} = 0 \; \forall \; i$), respectively. For $0 < p_\beta < 1$, we find the equilibrium configuration of the source class of nodes to be

$$N^*_{\vec{k}^{(1)}}(t) = \frac{k^{out(1)}}{\langle k^{out} \rangle} \frac{\eta_{in}(1-\eta_{in})\,\delta}{\eta_{out}\left[(1-\eta_{out})\,\delta + \eta_{in} p\right]} \overline{N} \; , \tag{3.4}$$

where $\langle k^{out} \rangle$ is the average out-degree of the node in the network. The quasi-equilibrium populations of the source nodes are distributed proportional to their respective out-degrees. This dependence arises from the assumption that the importations to source nodes depend on the node's out-degrees as described in appendix section B.1. The transit and sink nodes likewise depend primarily on their node degrees, but the in-degrees determine their equilibrium population distributions. For transit nodes and sink nodes we have respectively

$$N^*_{\vec{k}^{(2)}}(t) = \frac{k^{in(2)}}{\langle k^{out} \rangle} \frac{(1-\eta_{out})\,\delta}{\left[(1-\eta_{out})\,\delta + \eta_{in} p\right]} \overline{N} \text{ and} \tag{3.5}$$

$$N^*_{\vec{k}^{(3)}}(t) = \frac{k^{in(3)}}{\langle k^{out} \rangle} \frac{(1-\eta_{out})\,p}{\left[(1-\eta_{out})\,\delta + \eta_{in} p\right]} \overline{N} \; . \tag{3.6}$$

This dependence of the equilibrium populations on the in-degrees of each node is reminiscent of a directed random walk process and implies a stronger dependence of the equilibrium configurations on the cattle movement process rather than the cattle recycling processes. The derivations of these results are included in appendix section B.1 as well as their characteristic relaxation times and the boundary situations of $p_\beta = 0$ and $p_\beta = 1$. It is worth remembering that these derivations depend on the assumption of no correlations between the degrees of connected nodes. In the following section, we explore a disease process on top of these mobility dynamics assuming that the system has reached the quasi-equilibrium configurations before the introduction of the disease.

3.3 Global epidemic invasion thresholds of livestock model

Let us consider the addition of disease to our livestock model for the purpose of exploring the global invasion threshold R_* of epidemics in the livestock system[7,17,19]. Using the classic Susceptible-Infected-Recovered (SIR) disease model, we study the global invasion threshold in our directed network system[16]. The consideration of three classes of nodes reveals the potential for the existence of a second global invasion threshold. This new threshold determines whether the disease will remain within one class of nodes or break out into a second class, creating larger epidemics. Before the introduction of the disease, we assume that the livestock populations are fully susceptible and that, within each node, the subpopulations are well-mixed. Inside each node, an infected individual transmits infection to a susceptible individual upon contact at rate β. The model assumes a mass-action law to describe the force of infection acting on each susceptible individual that is proportional to the prevalence of infected individuals in the node. This force of infection is equivalently the rate at which each susceptible individual becomes infected. The recovery rate μ is the rate at which each infected individual recovers from the disease. Once recovered, an individual remains in the recovered state until and unless it happens to be recycled through the slaughter processes. This classic model can be characterized by the basic reproductive number $R_0 = \beta/\mu$ which is the average number of secondary infections generated by a typical infected individual in its first infectious time period[16]. The reproductive number R_0 serves as a threshold parameter at the individual level. Only if $R_0 > 1$, then the disease can spread to a finite proportion of the population. If $R_0 < 1$, the disease will die out in a finite amount of time and only impact a minuscule fraction of the susceptible population (zero in the thermodynamic limit of $N \to \infty$). In our case, this threshold determines the growth of the disease within each the subpopulation of each node. Recently, a new threshold for metapopulation models has been introduced[7,17,19,57,59,82,83]. This threshold considers the generations of infected nodes

through a branching process during the initial stages of a disease outbreak. For the branching process model, the approximated growth rate of the number of infected locations is defined as R_*. More correctly, R_* is a subpopulation reproductive number which is the average number of infected subpopulations generated from a typical infected subpopulation in a fully susceptible, structured metapopulation system[7,17,19,57,59,60,61]. The popular title for R_* is the global epidemic invasion threshold because, if $R_* > 1$, the epidemic will impact a finite fraction of the subpopulations of the system (in the thermodynamic limit of $V \to \infty$). Similarly, if $R_* < 1$, the epidemic will reach only a minuscule fraction of nodes and will die out in finite time. Thus, $R_* = 1$ describes the system-level epidemic invasion threshold for metapopulation models.

Let us consider the characterization of the three classes of nodes in this directed system through the R_* branching model. The branching model assumes that the epidemic is in its early stages with a R_0 value just above 1 and that a majority of the subpopulations are uninfected and susceptible. This model considers the n^{th} generation number of diseased subpopulations with degree $\vec{k}^{(x)}$ from node class x, $D^n_{\vec{k}^{(x)}}$, as a function of the three sets $\{D^{n-1}_{\vec{k}^{(1)}}\}$, $\{D^{n-1}_{\vec{k}^{(2)}}\}$, and $\{D^{n-1}_{\vec{k}^{(3)}}\}$ of the $(n-1)^{\text{th}}$ generation. The "infection" of a node occurs when infected cattle move from one node into another fully susceptible node, wherein they serve as a seed for the SIR process. Considering first the source nodes, we note that they possess no incoming degrees and thus will not receive infections from inward moving, infected cattle. Source nodes may contain disease among their subpopulations, but the number of source nodes having infection is a non-increasing number. Within the branching model, $D^n_{\vec{k}^{(x)}}$ is specifically the number of nodes newly infected having degree $\vec{k}^{(x)}$ and it does not count pre-existing outbreaks. Therefore, $D^n_{\vec{k}^{(1)}}$ is identically equal to 0 for every generation, and the source nodes do not contribute to the global epidemic invasion threshold. Incoming arcs to the transit nodes can arrive from both source nodes and transit nodes. Likewise for the sink nodes, their incoming arcs arrive from both of the other node classes. However, as $D^n_{\vec{k}^{(1)}} = 0 \ \forall n$, only the newly infected nodes of the transit nodes $\{D^n_{\vec{k}^{(2)}}\}$ are considered

in the branching model as sources of infection. Seeing that the diseased subpopulations of the sink nodes depend only on the transit nodes and the transit nodes depend only on themselves, we proceed to derive the global epidemic invasion threshold R_* for the directed metapopulation system from the system-level outbreak among the transit nodes: $D^n_{\vec{k}^{(2)}}$ as a function of the set $\{D^{n-1}_{\vec{k}^{(2)}}\}$. Thus considering the disease model described in section 3.2, we derive the global invasion threshold for our directed metapopulation system as

$$R_* = \frac{2p\delta\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}\ . \tag{3.7}$$

The derivation of equation 3.7 is included in appendix section B.2. The most notable trait of this system's R_* is its dependence on the topology degree distributions, namely the moment $\langle k^{in}\rangle$ of the full network and the moment $\left\langle\left(k^{in(2)}\right)^2\right\rangle$ of the transit nodes' in-degree distribution. Several dynamical processes on undirected networks have demonstrated that heterogeneity in node degrees lowers the threshold value, encouraging the spread of disease[12]. For directed subpopulation networks, this heterogeneity is particularly that of the in-degrees of the transit nodes, which comprise the strongly connected component of the networks. Simply stated, the greater the diversity found among the in-degrees of the transit nodes, the larger the value will be computed for R_*, increasing the probabilities of R_* being greater than 1 and a disease breaking out across the metapopulation system. A common control strategy which is often considered for the control of a disease outbreak is movement restriction. We can explore the restriction of movement by deriving the critical value p_c of the movement rate p such that the global epidemic invasion threshold is equal to 1. Appendix section B.2 also includes the derivation of p_c with the result of

$$p_c = \frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}\ . \tag{3.8}$$

If the movement parameter p of the livestock system is greater (less) than p_c, then the value of R_* will be greater (less) than 1 and the epidemic can be expected to spread (die out). The dependence of p_c on the heterogeneity of the transit nodes' in-degrees is similar to that of R_*. If the heterogeneity is greater, p_c will be lower, thus increasing the opportunities for a disease to spread among the transit nodes. Furthermore, we consider the spread of disease to the sink nodes and the dependence of this process on the successful arrival of infected cattle from the transit nodes. For the sink nodes, let us define according to the same branching model a second invasion threshold which describes whether the disease will spread from the transit nodes into the sink nodes or not. We define a transit-to-sink invasion threshold R_*^{TS} and derive it within appendix section B.2 to be

$$R_*^{TS} = \frac{2p\delta\left(1-\eta_{out}\right)\eta_{in}\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(3)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}\ . \tag{3.9}$$

This R_*^{TS} is an interesting threshold in that it describes the tipping point for a disease to move from the primary class of disease-sustaining nodes into a second class. Primarily, R_*^{TS} and R_* differ by the dependence of R_*^{TS} on the in-degree distribution of the sink nodes rather than the transit nodes. The critical movement rate p_c^{TS} for this threshold is given by

$$p_c^{TS} = \frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}\ . \tag{3.10}$$

The derivation of equation 3.10 is included in appendix section B.2. If the movement parameter p of the livestock system is greater (less) than p_c^{TS}, then the value of R_*^{TS} will be greater (less) than 1 and the epidemic can be expected to (not) spread from the transit nodes to the sink nodes. For higher levels of heterogeneity among the in-degrees of the sink nodes, p_c^{TS} will be lower, thus increasing the opportunities for a disease to spread out of the transit nodes into the sink nodes through the movements of infected cattle.

Let us examine the relationship between p_c^{TS} and p_c. If the thresholds described by p_c

and p_c^{TS} are ordered such that $p_c^{TS} < p_c$, the scenarios having movement rates above p_c will initiate disease outbreaks within the sink nodes as well as the transit nodes. However, if a network's parameters are such that $p_c^{TS} > p_c$, then there will be a second transition in the size of the epidemic as the mobility parameter is increased. This two transition situation is depicted in figure 3.2. Through derivations shown in appendix section B.2, we find that $p_c^{TS} < p_c$ when the the ratio of the second moments of the two classes' in-degree distributions $\left\langle \left(k^{in(2)}\right)^2 \right\rangle / \left\langle \left(k^{in(3)}\right)^2 \right\rangle$ is less than the ratio of the sink nodes to the transit nodes $\eta_{in}/\left(1-\eta_{out}-\eta_{in}\right)$. The reverse is also true, that is

$$\frac{\left\langle \left(k^{in(2)}\right)^2 \right\rangle}{\left\langle \left(k^{in(3)}\right)^2 \right\rangle} > \frac{\eta_{in}}{(1-\eta_{out}-\eta_{in})} \Rightarrow p_c^{TS} > p_c \ . \tag{3.11}$$

The reverse case shown in equation 3.11 is the more interesting of the two as it implies the existence of two sequential movement thresholds. The two thresholds describe a two-step process of disease first breaking out in the transit nodes and secondly, with a sufficiently high movement rate, spreading successfully into the sink nodes. The results of this section depend on the assumption of statistical equivalence within each degree block and the assumption that there are no correlations between the degree vectors of connected nodes. We have considered a simplistic set of outgoing movement rates to be uniformly p and time independent. Within each node, only a single species of livestock and a single, well-mixed type of livestock population have been considered.

3.4 Stochastic simulation of livestock model

We conducted numerical simulations of the described livestock model to verify our analytical results. We track individual cattle through time as births, importations, disease spread, and movements take place. In these Monte Carlo simulations, we explore the impacts of the diffusion rate, the movement rate of cattle to slaughter, and the network topology on the

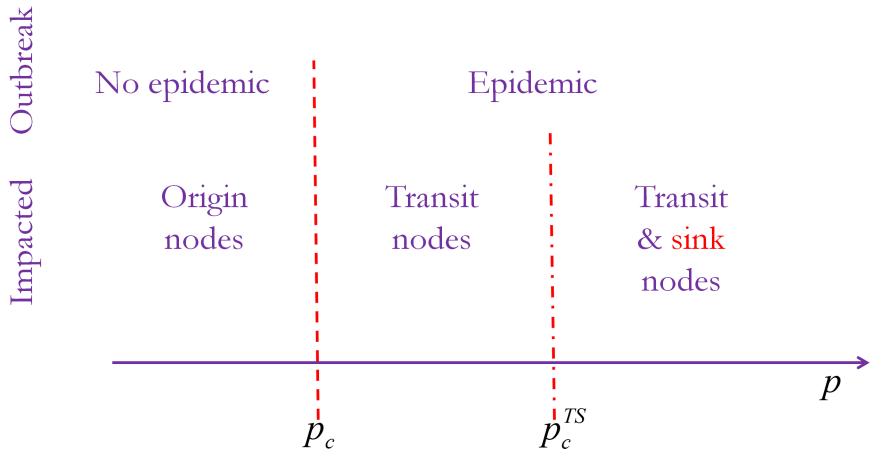

Figure 3.2: *Two transitions are observed as the cattle movement rate p increases past the pair of critical movement rates, p_c and p_c^{TS}. When the movement rate is less than p_c, the disease only impacts the nodes where the disease has originated (or very close by) and it is unable to spread to a significant fraction of the nodes. When the movement rate is greater than p_c and less than p_c^{TS}, the disease will impact a significant fraction of the transit nodes with no significant impact on the sink nodes. When the movement rate is greater than both critical movement rates, the disease will spread to a notable fractions of both transit and sink nodes.*

resulting disease outbreaks.

3.4.1 Dynamical processes

We implemented the model described in section 3.2 and follow individual cattle through their lives in the system. Synthetic livestock networks were generated using a modified uncorrelated configuration model[95]. The populations of each node were initiated through a multinomial distribution, allotting $\overline{N}V$ animals across V nodes, with probabilities proportional to the equilibrium populations described in equations 3.4-3.6. We divide time into discrete intervals of length Δt. Within nodes having a non-zero out-degree, individual animals move to neighboring populations at the rate d_{ij} of equation 3.1. Considering the time interval, $d_{ij}\Delta t$ is the probability that, within the discrete time step, an individual in location i moves to a location j within its out-neighborhood. During each time step, an individual within a source node will replicate itself with probability $\beta_k(t)\,\Delta t$. Also, within the source

nodes, individuals arrive through importation with a probability of $\epsilon_k(t)\,\Delta t$. Each animal in a sink node leaves for a slaughter house with probability $\delta\Delta t$ within a time step.

On top of this system of movement, birth, and death, we simulate the Susceptible-Infected-Recovered (SIR) disease model. We assume independence between the demographic processes and the disease process. Each subpopulation is divided into three states, assigning each individual to one of $S_i(t)$, $I_i(t)$, or $R_i(t)$. Susceptible (S) individuals are infected by infected (I) animals that share their same node at the same time step. Assuming a homogeneous mixing of the subpopulation of node i, we compute the probability of a susceptible (S) individual transitioning to the infected (I) state as

$$\lambda_i(t)\,\Delta t = \beta \frac{I_i(t)}{N_i(t)} \Delta t \ , \tag{3.12}$$

where β is the transmission rate of infection. The prevalence of infectious cattle in the subpopulation is captured as $I_i(t)/N_i(t)$. The infected (I) individual recovers with probability $\mu\Delta t$ and remains in the recovered (R) state until it passes out of a sink node on its way to a slaughter house. The new cattle arriving through the birth and importation processes are added to the susceptible (S) portion of their node's subpopulation. This fresh flow of susceptible individuals permits the existence of endemic outbreaks.

3.4.2 Numerical results

The directed subpopulation network structure has been designed with $V = 10^4$ nodes, having fractions of sources, transits, and sinks of 0.45, 0.45, and 0.10, respectively. This allotment of nodes to classes is similar to the observed distribution in the Italian network of cattle premises[88]. Both the in- and out-degree distributions (denoted by $P_{v-o}(k)$ and $P_{v-i}(k)$ in appendix section B.1) follow a power-law distribution. The exponent of the in-degree distribution is -2.1 and all degrees are between $k_{min} = 1$ and $k_{max} = 100$. As the network structures are generated from a stochastic model, we conduct the following experiments on

10 realizations of this network configuration. We consider an average population size of $\overline{N} = 10^5$ and thus a system population of $V\overline{N} = 10^9$ head. We study the livestock system with a time step of $\Delta t = 1$ day. These simulations are designed to characterize the global epidemic invasion thresholds through their dependence on the mobility rate p of the sources and sinks as well as their relative independence from the slaughter rate δ of the sink nodes. Therefore, for all simulations, we have fixed the birth re-introduction fraction at $p_\beta = 0.8732$ and the disease parameters, assigning the basic reproduction number to $R_0 = 1.2$ and the infectious period to $\mu^{-1} = 7$ days. The infection rate follows as $\beta = \mu R_0$. In each simulation, we initiate the populations through a multinomial distribution with the probabilities given by equations 3.4 - 3.6. We initiate the disease with $I_i(0) = 10$ infected individuals within a subpopulation of a single node i chosen at random from the nodes having minimal out-degree in the giant strongly connected component of the network. This limits the choice of the seed subpopulation to a transit node, and thus there is no initial infection in the source nodes nor the sink nodes. The nature of the population recycling processes allows a potential endemic state to occur in the system. Therefore, we let the disease progress in the metapopulation system until it either dies out or satisfies our criteria as endemic. These endemic criteria are twofold: a total number of new cases over time exceeding the system population of 10^9 and/or an epidemic duration of 100 simulation years. In the results we present here, all realizations producing a successful outbreak in the initially seeded subpopulation have been included in our results. A successful outbreak is considered as an outbreak resulting in at least 1% of the node's population contracting the disease. We conduct simulations of each set of parameters until we collect 5000 successful outbreaks for that set, 500 per each realization of the livestock network.

We vary the mobility rate p and describe the resulting outbreaks through four variables: the duration of the epidemic measured in days, the number of total cases caused by the epidemic over the total system population $V\overline{N}$ (global attack rate), the number of subpopulations having a secondary case occur normalized by the network size V (epidemic size), and the

fraction of outbreaks resulting in endemic situations as defined by our criteria. Figure 3.3 demonstrates the outbreaks of epidemics that take place as the movement rate p is increased past the critical thresholds. Sub-figures 3.3.A-C further demonstrate the initial slow growth that occurs between the first and second threshold and the much more rapid growth as the disease explodes past the second threshold defined by p_c^{TS}. With further simulations, we explored the impact of the death rate δ on the epidemics in figure 3.4. Although there was no significant dependence of either of the critical movement rates on δ, we did observe variations in the epidemic behavior. On the logarithm scale, the epidemics remain relatively small until crossing both invasion thresholds. For intermediate ranges of δ, a smoother transition to intense epidemics is seen, while the higher and lower values produce very sudden transitions in epidemic strengths. Finally, we consider the composition of our criteria to classify an outbreak realization as endemic, as shown in figure 3.5. For the higher movement rates, endemic situations arise primarily from massive outbreaks, but for slower movement rates and very low death rates, the endemic states result from long-lived outbreaks. Although figure 3.5 suggests which criteria is more likely responsible for the endemic situations in each set of parameters, figure 3.5 should not be considered independently of the fraction of endemic cases, as shown in figure 3.4.D.

3.5 Conclusions

In this chapter, we have designed a metapopulation model of a simplified livestock system for the purpose of exploring global invasion thresholds on directed metapopulation networks that contain source and sink nodes. With death, birth, importation, and movement processes, we have derived degree-block, mean-field solutions to the quasi-equilibria of the node populations for each type of node. We added the classic SIR model to the system and have derived directed-network version of the global epidemic invasion threshold R_* that describes the conditions necessary for an epidemic to emerge at the global scale in equation 3.7. Fur-

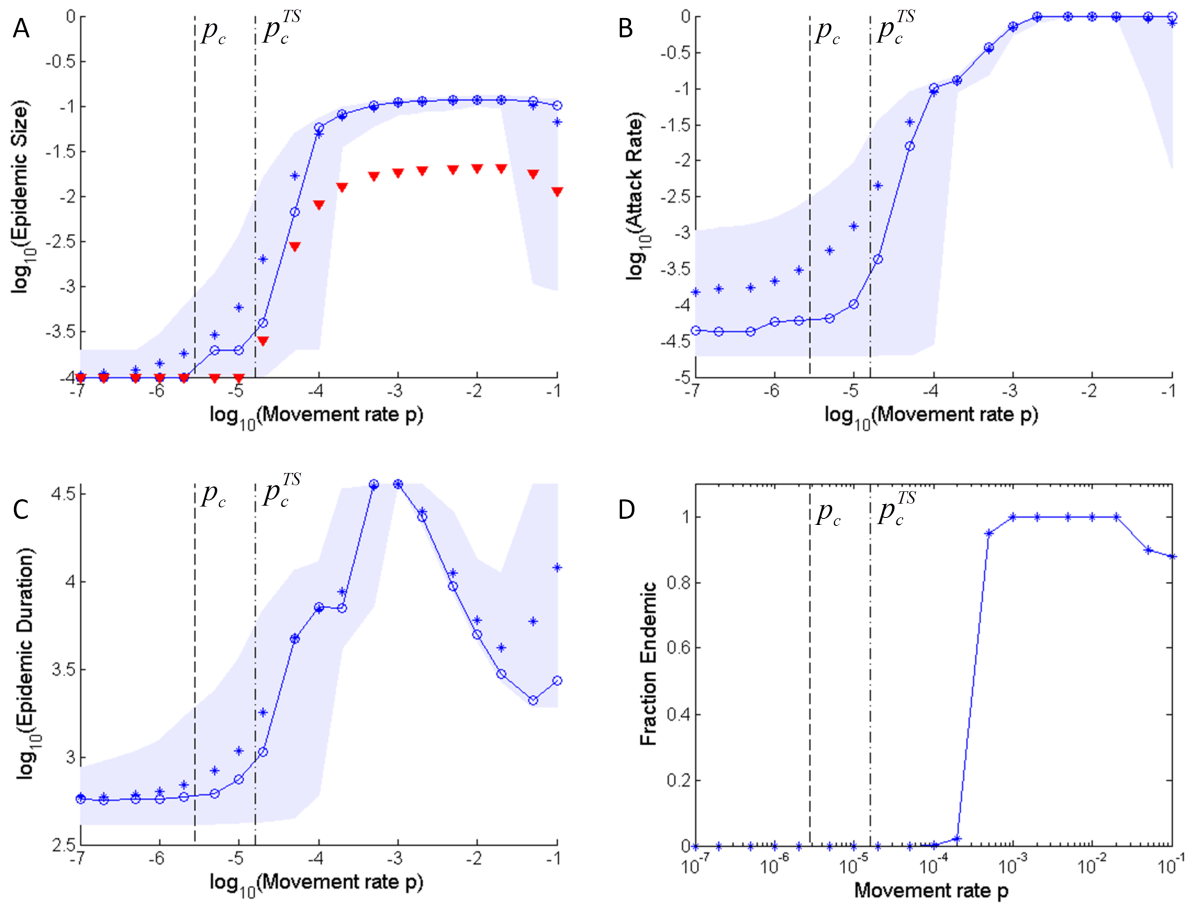

Figure 3.3: *(A) The distribution of the epidemic size as a function of the mobility rate p (with the sample averages denoted by stars, the sample medians denoted by circles connected on the line, and the 95% confidence interval denoted by the shaded region) rises as it passes the two invasion thresholds where p_c for global invasion is marked with a vertical dashed line and p_c^{TS} for invasion of sink nodes is marked with a vertical dash-dotted line. The inverted red triangles represent the average number of infected sink nodes as a fraction of the system size V. When the average number of infected sink nodes was less than one, it was rounded up to one for an artificial floor of $\log_{10}(1/V) = -4$. (B) The distribution of the global attack rate follow the same notations as A. (C) The distribution of the epidemic duration in days follows the same notations as A. (D) The fraction of successful outbreaks that resulted in endemic situations is plotted against the movement rate p. All sub-figures represent realizations that consider a per capita per day death rate of* 0.02.

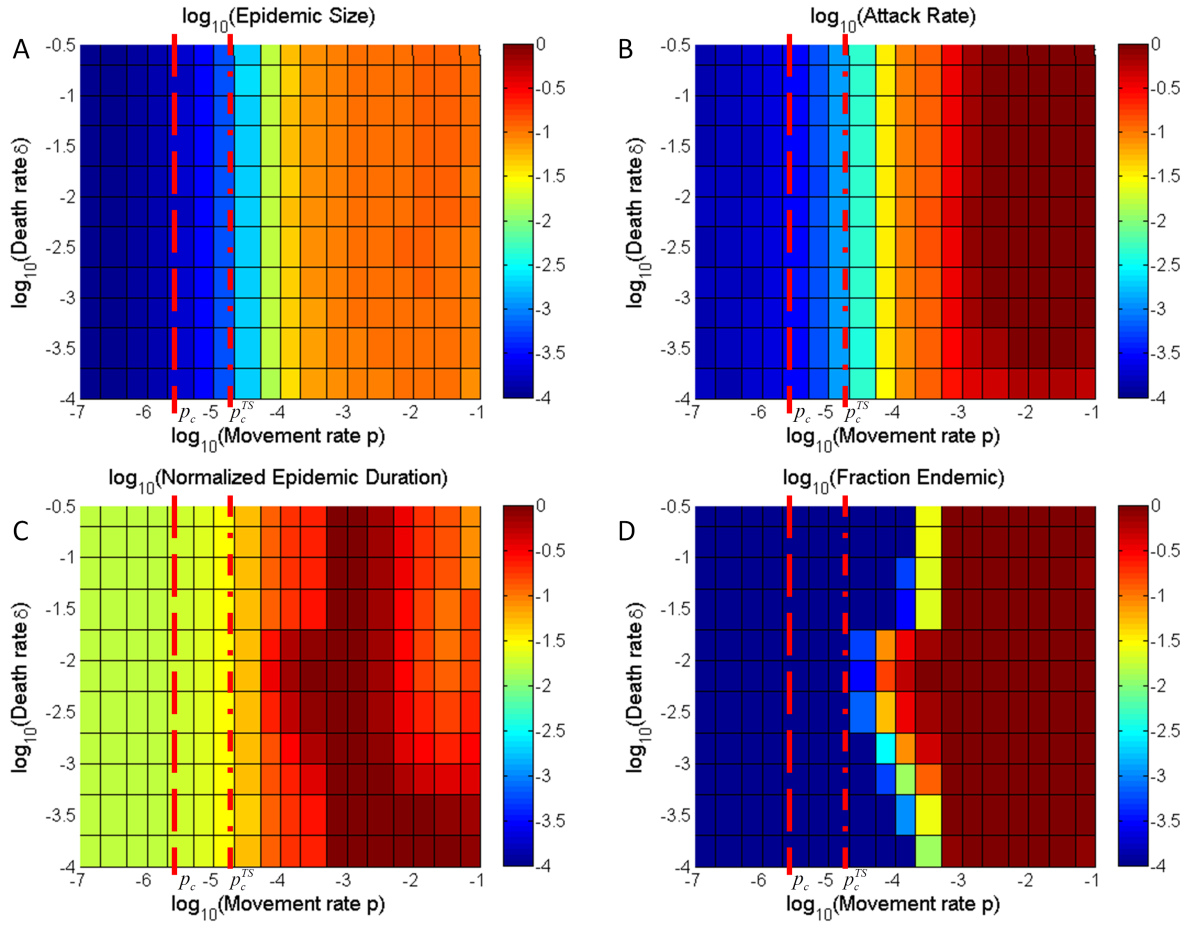

Figure 3.4: *(A) The logarithm of the average of the epidemic size as a function of the mobility rate p and the death rate δ is colored on the third dimension through the color scheme, where dark red represents values approaching 10^0 and dark blue represents values near 10^{-4}. The critical movement rate p_c is marked with a red dashed line and, similarly, p_c^{TS} is marked with a red dash-dotted line. (B) The logarithm of the average of the global attack rate follow the same notations as A. (C) The averages of the logarithm of the epidemic duration are normalized by 100 years and follow the same notations as A. (D) The logarithm of the fraction of successful outbreaks that resulted in endemic situations is plotted against the movement rate p and death rate δ. A small value of 10^{-4} has been added to the fractions of endemic situations to avoid the impossibility of plotting 0 on the logarithmic scale.*

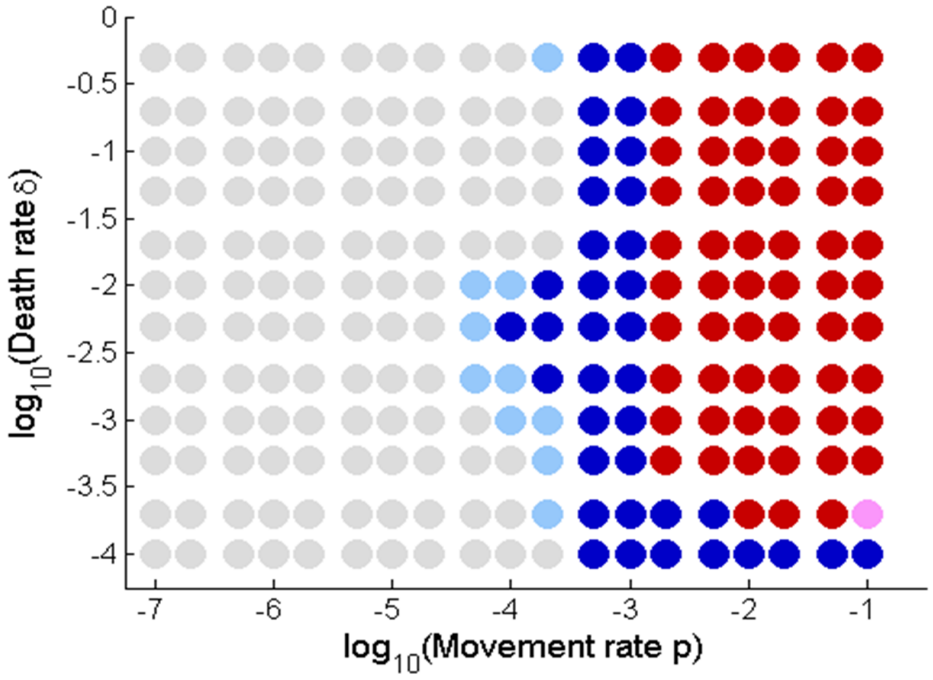

Figure 3.5: *A system of endemic classification is plotted against the movement rate p and death rate δ. Each circle represents a set of parameters and the color of the circle suggests whether the endemic states were more likely due to long-lasting epidemic (*100 *year criteria) or due to significant population turnover (*10^9 *total cases criteria). Grey indicates that neither the attack rate nor the duration distributions had medians nor upper ends on the confidence intervals meeting the endemic criteria. Light blue indicates that the upper end of the confidence interval of the duration met the duration criteria, but neither parameter of the attack rate met the attack rate criteria. Dark blue indicates that the median of the duration met the duration criteria, but, at most, the upper end of the confidence interval of the attack rate met the attack rate criteria. Dark red indicates that the median of the median of the attack rate met the attack rate criteria, but, at most, the upper end of the confidence interval of the duration met the duration criteria. Light purple indicates both distributions had upper ends of their confidence intervals which met the respective criteria, but both medians did not meet the endemic criteria. The respective colorings for light red and dark purple are not contained within the results captured by this figure.*

thermore, our analysis of the sink node populations has produced a second, novel global invasion threshold that describes the conditions necessary for an epidemic to break out of the giant strongly connected component of a directed network and into the populations of the sink nodes in equation 3.9. For these two thresholds, we have derived the critical movement rates in equations 3.8 and 3.10 and extracted a unique dependence in equation 3.11 of the pair's relationship on the ratios of the second moments of the node-type in-degree distributions and the fractions of each type of node, respectively, for the transit and sink nodes. The existence of this second global invasion threshold creates two transitions in the potential significance of emerging epidemics. The first transition enables a potential epidemic to invade the cattle populations. The second transition permits the epidemic to move with cattle to the slaughter facilities and to pose a risk to human populations.

The analyses and results of this chapter have come with a few notable assumptions and room remains for more general results to be derived. The most critical assumptions have been made regarding the movements of livestock. We have considered a constant rate of movement p that occurs uniformly to any node in the originating node's out-neighborhood. This has ignored both the heterogeneity and the strong seasonal components of actual livestock systems. Instead, we have considered static metapopulation networks interconnected through fixed routes of cattle flows. Such assumptions would only be valid in the situation where the disease dynamics occur on a time scale that is significantly shorter than the time scale of the seasonal variations in the movements. Also related to the structures of cattle movements, we have only considered networks which possess no correlations between the degrees of connected nodes. All of the above assumptions could be relaxed if there exists data characterizing the livestock trade network under consideration. Further restrictions we have imposed include the importation of only susceptible individuals (see appendix section B.3) and the study of only a single disease and single livestock species and type within the system.

These assumptions motivate future work, especially along the lines of the construction of

the livestock movement networks and traffic functions. The significance of these inputs for the system model and respective analyses drives the need for real-world data to characterize these movement inputs. A data-driven livestock metapopulation movement system, when combined with the results presented in this chapter, would enable the evaluation of different livestock disease control strategies, especially movement-restriction-based methods. Despite the simplistic livestock movements which we have considered, our work has produced the global epidemic invasion threshold for directed metapopulation networks as well as the second novel threshold. The consideration of epidemic control strategies through these thresholds enables an immediate assessment of the strategy effectiveness. In the next chapter, we develop an optimal estimation method of deducing movement parameters to address the limited availability of data for US cattle systems. The method of chapter 4 is flexible and can be adapted to estimate parameters for other livestock such as swine[96].

Chapter 4

Dance of the Calves: An estimation of cattle movement parameters in the Central States of the US

The characterization of cattle demographics, and especially movements, is an essential component in the modeling of dynamics in cattle systems, yet for cattle systems in the United States (US), this is missing. Through a large-scale maximum entropy optimization formulation, we estimate cattle movement parameters to characterize the movements of cattle across 10 Central States and 1034 counties of the United States. Inputs to the estimation problem are taken from the United States Department of Agriculture National Agricultural Statistics Service database and are pre-processed in a pair of tightly constrained optimization problems to recover non-disclosed elements of data. We compare stochastic subpopulation-based movements generated from the estimated parameters to operation-based movements published by the United States Department of Agriculture. For future Census of Agriculture distributions, we propose a series of questions that enable improvements for our method without compromising the privacy of cattle operations. Our novel method to estimate cattle movements across large US regions characterizes county-level stratified subpopulations

of cattle for data-driven livestock modeling.

4.1 Introduction

Livestock systems serve significant roles for many regions across the world, yet past outbreaks of disease have shown that they can possess a number of vulnerabilities[3,4,76,77,78,79]. The livestock systems of the United States (US), though strictly regulated, may yet be found susceptible to foreign diseases such as Rift Valley Fever[5]. The successful modeling and analysis of livestock epidemics for any region relies heavily on an understanding of the underlying system components. The three most critical elements in a practical epidemic model are the disease progression model, the geo-spatial characterization of the susceptible populations, and the spatial-temporal description of the interactions of individuals within the system[6,7,8]. The models of disease progression are several and often independent of the region studied[12,20,25,28,97]. Data-driven, spatial characterizations of populations are available through regularly conducted censuses (censi)[1,98]. The third element, the interactions of individuals within the system, represents the set of spatial movements of individuals. When considering system-wide outbreaks of disease, the impact of movement parameters has been shown to be as significant as that of epidemic parameters in metapopulation models[17,19]. Domestic livestock systems fit well in such metapopulation models because the movements of livestock are controlled and the individuals are restricted to reside within populations rarely defined by their choice. Within the US, livestock movements are controlled by the cattle industries, primarily beef, dairy, breeding, and showmanship.

Within Europe, motivated by outbreaks of Foot and Mouth Disease, a number of governments have designed and implemented animal tracking systems, even including the resolution of individuals' daily movements. The databases created by these studies have generated very detailed characterizations of livestock movements for a number of European nations[69,70,71,72,73,74,75,77,78,79,80,99,100,101,102,103,104,105,106]. No similar program has yet to be

implemented for the United States, although some have long been in preparation[107]. In the US, a cultural appreciation of personal privacy from the government, strong competition between meat production companies, and a US Federal privacy protection law restrict the ability of the government to collect and release livestock data at a finer spatial resolution than is currently done through the United States Department of Agriculture's (USDA) Census of Agriculture[1]. To address this challenge, a number of survey-based methods have been used to study livestock movements across small regions[108,109,110,111,112,113]. However, the national scale of US cattle trade and the potential for livestock diseases to impact the entire country necessitate movement data, models, or estimates to be determined for larger regions. Recently a study has been published of a nation-wide movement estimation based on a 10% sample of veterinary records from State border-crossing cattle shipments[92]. This impressive study, although the first of its magnitude, only captured shipments of cattle that crossed state borders. Although it offers a picture of state-to-state shipment counts, the method used is unable to capture the livestock movements within each state.

In this chapter, we formulate a large, convex optimization problem to estimate parameters describing the movements of cattle within 10 Central States of the United States. We collect cattle population and aggregated movement data from the USDA's database and optimally estimate anonymous data points to construct a database of inputs for an estimation of cattle movement parameters. We design the estimation method to produce a high resolution of cattle demographic and movement parameters and to include the minimal set of assumptions. Our results produce county-to-county movement probabilities among stratified subpopulations as well as birth, slaughter, and expiration rates of cattle for 1034 US counties. In section 4.2, we describe the USDA data structures and challenges present in the database. We estimate non-disclosed data points and discuss the mapping of USDA data to inputs for our estimation formulation. In section 4.3, we formulate the estimation problem and describe the maximum-entropy objective and the flexible set of linear constraints with parameters sculpted to the USDA data set and as few assumptions as possible. We solve

the optimal estimation problem and display a subset of the results in section 4.4. Section 4.5 summarizes this chapter with a discussion of the results and a series of questions we propose to be added to future agricultural census distributions.

4.2 Data collection and structure

Every five years, the United States Department of Agriculture (USDA) conducts the United States Census of Agriculture[1]. The National Agricultural Statistics Service (NASS) of USDA then summarizes and publishes a large set of data covering livestock, crops, operator demographics, and much more[114]. As the most comprehensive and clean database of US livestock statistics, the Census of Agriculture as presented in the NASS database is used for our estimation of cattle movement parameters. In particular, we use data from the 2007 Agricultural Census as the 2012 data was not published at the time of this study. The data of interest to this work comes from section 13, titled "Cattle and Calves", on page 10 of the 2007 Agricultural Census. Section 13 also has a set of related instructions located on page 2 of the instruction sheet appended to the Agricultural Census[1]. From the US Census Bureau and their 2010 Census (of humans) in the United States, we use the centers of human population for each county[98]. We include these geographical points to consider a basic quantification of distance for the cattle movement estimation. Adding this geography to the data from the NASS database, we estimate sets of parameters to characterize cattle movements in the States of Arkansas, Colorado, Iowa, Kansas, Minnesota, Missouri, Nebraska, Oklahoma, South Dakota, and Texas[9,115]. The US beef production feedlot structure produces more frequent and larger flows of cattle than the typical grazing structure, and notably, these 10 Central States form the core of the US feedlot industry[92,109]. As outlined in figure 4.1, these States contain 1034 US counties with more than 51 million head of cattle out of the 96.3 million head reported in the 2007 Agricultural Census[114].

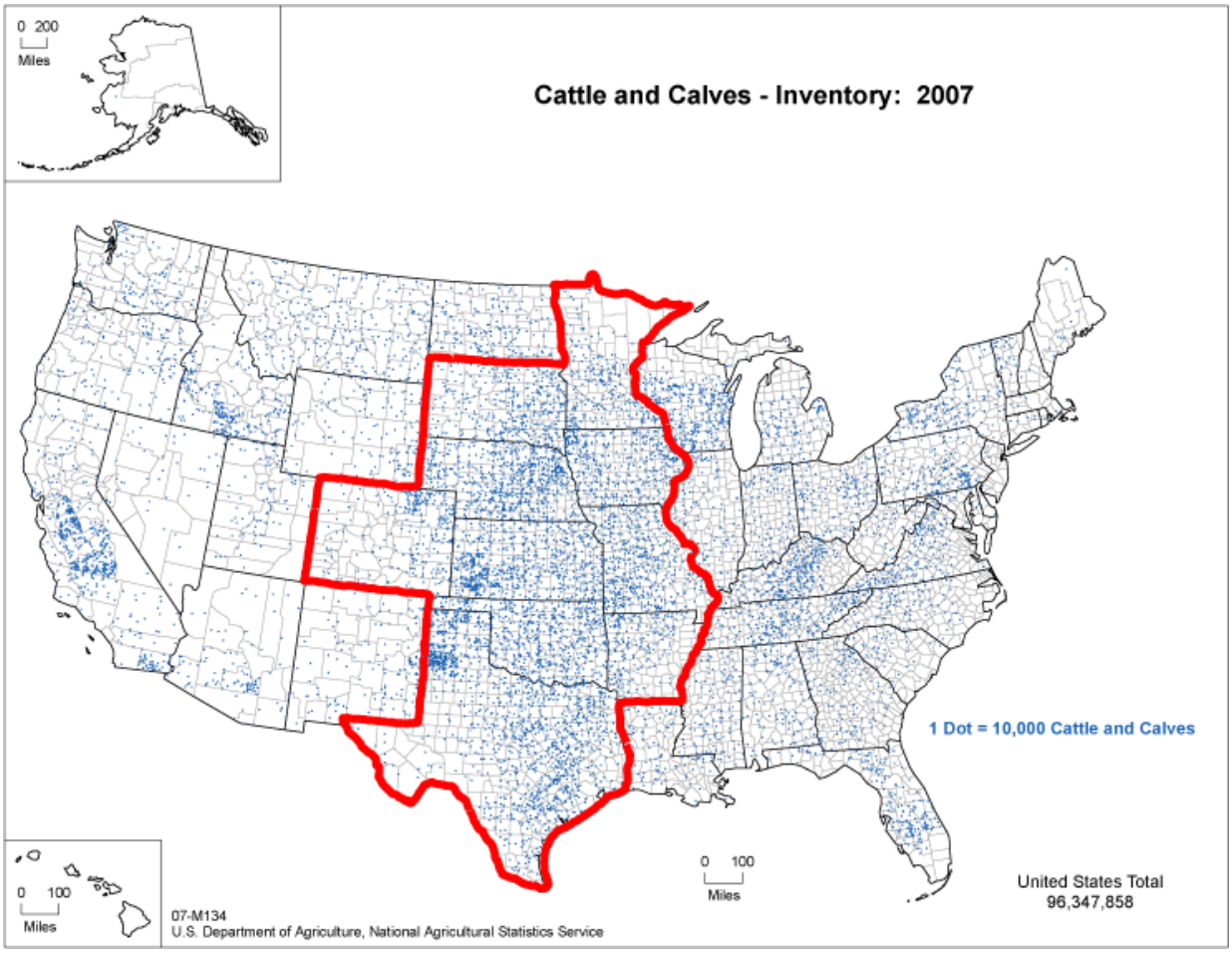

Figure 4.1: *The* 10 *Central States of interest are outlined with a red trace over the population distribution of cattle in the United States, as provided by the United States Department of Agriculture*[114]*. Each blue dot represents* $10,000$ *head of cattle.*

4.2.1 Data structure of USDA NASS

From section 13 of the 2007 Agricultural Census, we are primarily interested in the responses to questions concerning the populations and movements of cattle. For cattle populations, the Agricultural Census identifies the total number of cattle (Question 3), the number of dairy cows kept for production of milk (Question 2.b), and the number of cattle, including calves, who were in a preslaughter feed program (Question 5) on December 31, 2007. For the movements of cattle, composed of all sales and shipments, the Agricultural Census captures the total number of cattle "sold or moved" during 2007 (Question 4) and the total number of cattle shipped directly to a slaughter market from a preslaughter feed program during 2007 (Question 6)[1]. Although these data are collected for each individual operation, the statistics of the cattle populations are reported only through aggregated distributions that are delineated by the operation's county, the type of cattle, and the number cattle of a particular type. The sizes of populations are sorted into seven standard ranges: $1-9$ cattle, $10-19$ cattle, $20-49$ cattle, $50-99$ cattle, $100-199$ cattle, $200-499$ cattle, and 500 or more cattle. The statistics of the cattle movements sort the responses by the total yearly movements (or slaughter) across the same 7 standard ranges, the type of cattle, and the county of the originating operation. It is worth noting that the total dairy cattle population and the total preslaughter population of a given county are subpopulations of the total cattle population for that county. Similarly, the total number of cattle shipped to slaughter from a preslaughter feed program is a fraction of the total cattle movements (sales and shipments) for each county.

According to appendix A of the 2007 Agricultural Census Summary and State Data report, the data presented in the NASS database has undergone some initial processing and systematic error correction[2]. This results in a very consistent database and the potential errors induced by these methods have been quantified in the same appendix. Even with these diligent efforts, there remain two significant challenges in utilizing the data to characterize the movements of cattle. The first concerns the resolution of the timescale of the data.

As a summary of the entire year 2007, these data fail to capture any seasonal fluctuations in the cattle populations and movements[116]. This challenge arises from the administration process of the Agricultural Census and we acknowledge its significance; however, we find no comprehensive, data-driven solution to the seasonality challenge and consider only mean-field probabilities in our estimation. The second challenge is posed by the direct sorting of the census responses into the 7 standard ranges rather than preserving any operation-based connections between data points. Therefore, a population of 50 dairy cattle might belong to any operation having a total number of cattle greater than or equal to 50 (4 possible size ranges) without having any connection to the size of its entire operation. Similarly, the sizes of shipments have no direct connections to the size of the originating operations besides a few loose feasibility restrictions.

The data, as it comes from NASS, has been released in such a way that the information of individual farms and cattle operations is not revealed. This is done intentionally by USDA to comply with Title 7 of the US Code[2]. To maintain this anonymity, critically selected elements of the data have not been disclosed. We will estimate these non-disclosed data points through a pair of tightly-constrained convex problems and then include them in the inputs for our main problem, the estimation of cattle movement parameters across the 10 Central States.

4.2.2 Data structure for estimation problem

To estimate the non-disclosed entries in the original data, we construct a pair of optimization problems, one for the population data and a second for the movement and slaughter data. The objective of both formulations is a maximum entropy function. For the population distributions and by each State, we maximize the entropy of the distributions of each cattle type given by $Type_A = \{Dairy, Preslaughter, All\ Cattle\}$, where the distributions are normalized by their respective State totals. Likewise, the entropy of the normalized distributions describing the totals of cattle slaughter shipments and cattle movements is

maximized across shipment type, $ShipType = \{All\ Shipments, Slaughter\}$, shipment size, $Size_A = \{z1\_9,\ z10\_19,\ z20\_49,\ z50\_99,\ z100\_199,\ z200\_499,\ z500\_up\}$, and county. We present the formulations of these problems in appendix section C.1. These problems are solved for each of the 10 States, filling in all non-disclosed data entries. We quantify the dependence of our inputs on these estimations with both the fraction of cattle and the fraction of populations estimated for each State in appendix section C.2.

We would prefer to represent the system of cattle through three subpopulations rather than the two subpopulations and total population of $Type_A$. The mapping of the first set $Type_A$ to a set of three cattle subpopulation types is not trivial as it requires the expression of a relationship among the 3 cattle types of $Type_A$. For the county totals of these three types, the relationship $Tc^x_{Dairy,c} + Tc^x_{Preslaughter,c} \leq Tc^x_{AllCattle,c}$ holds, where $Tc^x_{\cdot,c}$ represents the respective county total for county c, and x indicates that this is a variable to be estimated. However, the relationship is not guaranteed if we consider the stratification of the populations by size as

$$Pop_{Dairy,c,i} + Pop_{Preslaughter,c,i} \leq Pop_{AllCattle,c,i}\ \forall\,(c \in County, i \in Size_A)\ , \tag{4.1}$$

where $Pop_{\cdot,c,i}$ represents the respective population total for operations with size i and in county c. Rather, Broomfield County in the State of Colorado, as do a number of other counties, reports data in violation of this relationship. Broomfield County is a irregularly cut county on the north side of Denver, Colorado and has 2 dairy farms with total populations in the range $z100\_199$. The county, however, reports 2 populations of Dairy cattle in the size range $z50\_99$ and no total ($All\ Cattle$) populations in the size range $z50\_99$. Thus the left-hand side of inequality 4.1 would be non-zero while the right-hand side is identically 0 for $i = z50\_99$, $c = Broomfield\_Colorado$. This discrepancy arises from other cattle residing at both of these operations that raise the total operation populations into the next size range. We find that through an aggregation of the sizes into 3 ranges, $Size_B = \{z1\_19,\ z20\_199,\ z200\_up\}$, it becomes feasible to assume the relationship of inequality 4.1 for each

size j in $Size_B$.

Let us define a new cattle type $Beef$, representing all cattle that are not serving as dairy cattle nor in a preslaughter feed program, as the difference between the total population $All\ Cattle$ and the two subpopulation types, $Dairy$ and $Preslaughter$. The $Beef$ type represents a diverse set of cattle operations, including grazing, backgrounding, and breeding services. The name $Beef$ is chosen for simplicity with the assumption that this is the majority role served by cattle in this type. At this point, we describe the cattle subpopulations $Pop_{t,c,j}$ by cattle type t in $Type_B = \{Dairy, Preslaughter, Beef\}$, county c, and size j in $Size_B = \{z1\_19,\ z20\_199,\ z200\_up\}$. The set of subpopulations $\{Pop^R_{t,c,j}\}$ results from the solution of the first of the data-patching optimization formulations and represents an aggregation of all cattle type-based subpopulations fitting the subpopulation descriptors, yielding only 9 subpopulations per county. County totals for sales $Tc^{(s),x}_{AllMovement,c}$ and slaughter $Tc^{(s),x}_{Slaughter,c}$, as well as the distributions of yearly totals of sales $Sales^x_{AllMovement,c,i}$ and slaughter $Sales^x_{Slaughter,c,i}$ stratified by county c and size i in $Size_A$, were completed through the solution of the second data-patching optimization problem of appendix section C.1.

4.3 Cattle movement parameter estimation

We formulate a non-linear, yet convex, optimization problem with an objective to maximize the entropy of the out-going distributions of each subpopulation[117]. The formulation is nearly linear with the exception of the objective function, or equivalently stated, this formulation contains only linear constraints. The choice of maximum entropy for the problem objective aims to predict the solution having the minimal assumptions (maximum uncertainty) beyond the information contained within the set of constraints[117]. We chose this form of objective to not force any artificial objective in our estimation. Nevertheless, assumptions have been made in the design of both the variables and the constraints. We

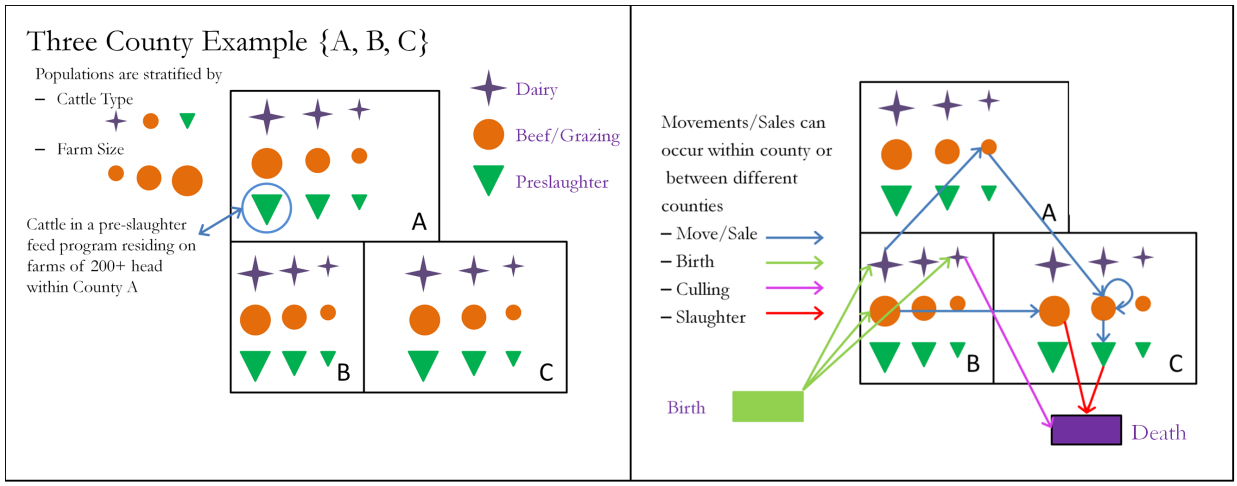

Figure 4.2: *The subpopulation stratification by cattle type and operation size into nine subdivisions per county is depicted in the left pane. Given the subpopulation classifications, cattle movements, birth, culling, and slaughter occur according to parameters estimated in the optimization problem described in section 4.3 in the right pane.*

assume that

- There are no outgoing movements from preslaughter feed programs except for the outgoing movements of cattle for slaughter,
- Cattle classified as dairy cattle do not move into preslaughter feed programs,
- Populations of preslaughter feed cattle having population sizes of 200 head of cattle or more are responsible for all shipments to slaughter that result in yearly totals of 500 or more head shipped from a single premise,
- All sub-populations remain constant on a year-to-year basis, and
- The counties considered form a closed system with no significant movement into or out of the system.

4.3.1 Problem formulation

The central portion of this chapter revolves around the formulation to estimate the cattle movement parameter $p^x_{t_1,j_1,t_2,j_2,dist}$, which is a probability that represents the movement

process from an origin subpopulation of type t_1 and size j_1 to a destination subpopulation of type t_2 and size j_2 with a distance falling in a discrete distance range $dist$ between the origin and destination counties. We mark the decision variables of this formulation with a superscript x to distinguish them from the parameters of the problem. This formulation also estimates the birth $bt^x_{c_1,t_1,j_1}$, expiration of utility (cull) $dt^x_{c_1,t_1,j_1}$, and slaughter $sl^x_{c_1,t_1,j_1}$ probabilities for each sub-population of each county. The three types of cattle are now rearranged into the set $Type_B$ as $\{Dairy, Preslaughter, Beef\}$, with size ranges defined by $Size_B = \{z1\_19,\ z20\_199,\ z200\_up\}$. The set of discrete distance ranges used in this formulation is called $Distance$ and is defined as $\{d0, d100, d200, d500, d1000, d_{toofar}\}$. In three adjacent counties, figure 4.2 summarizes first the subpopulation stratification and secondly a few possible birth-to-death flows for cattle as outlined by the data structure of the following optimization problem. The number noted in each distance range between two county centers should be read as the maximum distance in miles of the range, with the minimum defined by the previous level. For example, $d500$ indicates a distance between the two county centers falling between 200 and 500 miles. The closest range, $d0$, is assigned for any pair of counties with centers less than 10 miles apart, as well as each county with itself. The formulation has an objective to maximize the entropy of the outgoing distributions of all sub-populations as follows:

Maximize J

where

$$J = \sum_{c_1}\sum_{t_1}\sum_{j_1}\Bigg[- st^x_{c_1,t_1,j_1}\log\left(st^x_{c_1,t_1,j_1}\right) - \ sl^x_{c_1,t_1,j_1}\log\left(sl^x_{c_1,t_1,j_1}\right) - \ dt^x_{c_1,t_1,j_1}\log\left(dt^x_{c_1,t_1,j_1}\right)$$
$$+\sum_{c_2}\sum_{t_2}\sum_{j_2} -p^x_{t_1,j_1,t_2,j_2,D(c_1,c_2)}\log\left(p^x_{t_1,j_1,t_2,j_2,D(c_1,c_2)} + 1.0 - f_{t_1,j_1,t_2,j_2,D(c_1,c_2)}\right)\Bigg]$$
$$\text{and } c_1 \in County, t_1 \in Type_B, j_1 \in Size_B,$$
$$c_2 \in \{County | D\left(c_1, c_2\right) \neq d_{toofar}\}, t_2 \in Type_B, j_2 \in Size_B \ . \tag{4.2}$$

The out-going probability distribution for the cattle of each subpopulation is completed through the inclusion of a probability to remain or stay, $st^x_{c_1,t_1,j_1}$, in the origin subpopulation. The sum of the entropy of these distributions composes our objective function. We have implemented industrial constraints with a set of parameters $\{f_{t_1,j_1,t_2,j_2,dist}\}$ described in the following constraints, which forces a subset of the movement probabilities $p^x_{t_1,j_1,t_2,j_2,dist}$ to zero. We account for this by introducing a complimentary $(1.0 - f_{t_1,j_1,t_2,j_2,dist})$ in the logarithm of $p^x_{t_1,j_1,t_2,j_2,dist}$ to avoid the computation of the natural logarithm of zero.

Subject to:

Constraints on Statistical rules

$$p^x_{t_1,j_1,t_2,j_2,dist} \leq f_{t_1,j_1,t_2,j_2,dist} \ \forall\left(t_1, j_1, t_2, j_2, dist\right) \tag{4.3}$$

$$\sum_{c_2 \in County | D(c_1,c_2) \neq d_{toofar}} \ \sum_{t_2 \in Type_B} \ \sum_{j_2 \in Size_B} p^x_{t_1,j_1,t_2,j_2,D(c_1,c_2)} + dt^x_{c_1,t_1,j_1} + sl^x_{c_1,t_1,j_1} + st^x_{c_1,t_1,j_1} = 1.0 \ \forall\left(c_1, t_1, j_1\right) \tag{4.4}$$

Inequality 4.3 serves to restrict the probabilities of movement, $p^x_{t_1,j_1,t_2,j_2,dist}$, to be less or equal to 1 as $f_{t_1,j_1,t_2,j_2,dist}$ takes on a value of 1 in the general case. By taking a value of 0, it further

prevents the movement of cattle from *Dairy* subpopulations to *Preslaughter* subpopulations ($t_1 = Dairy$ and $t_2 = Preslaughter \Rightarrow f_{t_1,j_1,t_2,j_2,dist} = 0$) and the out-going shipments of cattle from preslaughter subpopulations ($t_1 = Preslaughter \Rightarrow f_{t_1,j_1,t_2,j_2,dist} = 0$). Equality constraint 4.4 ensures that the sum of the out-going probability distributions, the same distributions that are considered in the objective function, is equal to 1.

Constraints on Movement data

$$\sum_{t_1 \in Type_B} \sum_{j_1 \in Size_B} \sum_{c_2 \in County | D(c_1,c_2) \neq d_{toofar}} \sum_{t_2 \in Type_B} \sum_{j_2 \in Size_B} Pop^R_{t_1,c_1,j_1} p^x_{t_1,j_1,t_2,j_2,D(c_1,c_2)}$$
$$+ \sum_{t_1 \in Type_B} \sum_{j_1 \in Size_B} Pop^R_{t_1,c_1,j_1} sl^x_{c_1,t_1,j_1} + PN^{mov}(c_1) = \frac{Tc^{(s),x}_{AllMovements,c_1}}{R_C} \; \forall (c_1) \tag{4.5}$$

$$\sum_{j_1 \in Size_B} Pop^R_{Preslaughter,c_1,j_1} sl^x_{c_1,Preslaughter,j_1} + PN^{slt}(c_1) = \frac{Tc^{(s),x}_{Slaughter,c_1}}{R_C} \; \forall (c_1) \tag{4.6}$$

$$Pop^R_{Preslaughter,c_1,z200\_up} sl^x_{c_1,Preslaughter,z200\_up} + PN^{slt500}(c_1) \geq \frac{Sales^x_{Slaughter,c_1,z500\_up}}{R_C} \; \forall (c_1) \tag{4.7}$$

$$D_{mov} = \sum_{c_1 \in County} \left[|PN^{mov}(c_1)| + |PN^{slt}(c_1)| + PN^{slt500}(c_1) \right] \tag{4.8}$$

Equality constraint 4.5 sums the total sales and shipments originating in each county c_1 and tries to equate the total to the total sales and shipments defined by USDA NASS for county c_1, allowing a small amount of discrepancy through an error or roughness term $PN^{mov}(c_1)$. This discrepancy is permitted due to data challenges discussed in section 4.2. The scaling term, $R_C = 52.0$ weeks/year, converts the timescale of the estimation prob-

lem from a yearly to weekly basis for the estimated probabilities. Equality constraint 4.6 equates the total slaughter from preslaughter feed subpopulations in each county c_1 to the respective, scaled data value from USDA NASS, again with a discrepancy term for each county $PN^{slt}(c_1)$. Inequality 4.7 ensures that the largest yearly slaughter counts (500 or more head) are accredited to the largest (200 or more head) preslaughter subpopulations. This inequality requires a discrepancy term $PN^{slt500}(c_1)$ due to seasonality challenges in the USDA NASS data set. The discrepancy terms are collected in equality 4.8. Although we represent equality 4.8 here with absolute value operators, the actual implementation linearizes the terms through a two-variable decomposition of the unrestricted variable that allows us to minimize the resulting value as if it were an absolute value[18]. We retain the absolute value operators for simplicity in the formulation description.

Constraints on Population conservation

$$Leaving(c_1, t_1, j_1,) = Pop^R_{t_1,c_1,j_1} \sum_{c_2 \in County | D(c_1,c_2) \neq d_{toofar}} \sum_{t_2 \in Type_B} \sum_{j_2 \in Size_B} p^x_{t_1,j_1,t_2,j_2,D(c_1,c_2)} \; \forall (c_1, t_1, j_1) \tag{4.9}$$

$$Coming(c_1, t_1, j_1,) = \sum_{c_2 \in County | D(c_2,c_1) \neq d_{toofar}} \sum_{t_2 \in Type_B} \sum_{j_2 \in Size_B} Pop^R_{t_2,c_2,j_2} p^x_{t_2,j_2,t_1,j_1,D(c_2,c_1)} \; \forall (c_1, t_1, j_1) \tag{4.10}$$

$$Leaving(c_1, t_1, j_1,) - Coming(c_1, t_1, j_1,) + \left(dt^x_{c_1,t_1,j_1} + sl^x_{c_1,t_1,j_1} - bt^x_{c_1,t_1,j_1}\right) Pop^R_{t_1,c_1,j_1} + PN^{pop}(c_1, t_1, j_1,) = 0.0 \; \forall (c_1, t_1, j_1) \tag{4.11}$$

$$D_{pop} = \sum_{c_1 \in County} \sum_{t_1 \in Type_B} \sum_{j_1 \in Size_B} |PN^{pop}(c_1, t_1, j_1,)| \tag{4.12}$$

Equality constraints 4.9 and 4.10 sum, respectively, the originating and arriving flows of cattle for each subpopulation of each county. Equality constraint 4.11 then defines the total flux of every subpopulation in the system to be 0 with small exceptions allowed through the discrepancy terms $PN^{pop}\,(c_1, t_1, j_1,)$. Equality constraint 4.12 serves to aggregate the discrepancies. Here again in equality 4.12, we retain the absolute value operator for simplicity in the formulation description[18].

Constraints on Industrial insights and discrepancies

$$D_{mov} + D_{pop} \leq f_{min} P^{\text{tot}}_{AllCattle} \tag{4.13}$$

$$r^{expire-min}_{t_1} \leq dt^{x}_{c_1,t_1,j_1} \leq r^{expire-max}_{t_1} \quad \forall\,(c_1, t_1, j_1) \tag{4.14}$$

$$r^{slaughter-min}_{t_1} \leq sl^{x}_{c_1,t_1,j_1} \leq r^{slaughter-max}_{t_1} \quad \forall\,(c_1, t_1, j_1) \tag{4.15}$$

$$r^{birth-min}_{t_1} \leq bt^{x}_{c_1,t_1,j_1} \leq r^{birth-max}_{t_1} \quad \forall\,(c_1, t_1, j_1) \tag{4.16}$$

In inequality 4.13, we constrain the total discrepancies (counted in head of cattle) of the movements and net population fluxes to be less than a fraction f_{min} of the total cattle in the system. The value of f_{min} is determined by first solving the linear problem composed of the set of constraints of this formulation with an objective to minimize the system discrepancies. The value of f_{min} is then taken as the ratio of the optimal objective value to the total system population and rounded up to the next highest thousandth. The inequality pairs 4.14, 4.15, and 4.16 provide constraints by cattle type on the feasible probabilities used to describe the respective expiration, slaughter, and birth processes for each subpopulation.

4.4 Optimization results

We solved the cattle movement parameter problem of section 4.3 using the AIMMS Modeling System of Paragon Decision Technology[118]. The complete formulation is composed of 81142 constraints with 80107 variables and the objective function. For the error limit, a value of $f_{min} = 0.012$ was obtained through the method described following inequality 4.13, representing a limit of 1.2% of the total number of cattle, $51,252,890$. The limits on the demographic probabilities attempt to capture loose bounds on feasible average rates of birth, culling, and slaughter. We assume that dairy cattle are not sent to slaughter houses through a slaughter rate $r_{Dairy}^{slaughter-max} = r_{Dairy}^{slaughter-min} = 0$, but rather through a culling process $r_{Dairy}^{expire-min} = (312 \text{ weeks})^{-1}, r_{Dairy}^{expire-max} = (104 \text{ weeks})^{-1}$. We bound the expected birthing rate of dairy cattle as $r_{Dairy}^{birth-min} = (62 \text{ weeks})^{-1}$ and $r_{Dairy}^{birth-max} = (36 \text{ weeks})^{-1}$. The mixed collection of cattle, $Beef$, are allowed a reasonable birth rate as well $r_{Beef}^{birth-max} = (52 \text{ weeks})^{-1}, r_{Beef}^{birth-min} = 0$, but the $Preslaughter$ individuals are not $r_{Preslaughter}^{birth-min} = r_{Preslaughter}^{birth-max} = 0$. The $Beef$ cattle have a maximum average useful lifespan defined by $r_{Beef}^{expire-min} = (520 \text{ weeks})^{-1}$ and they join the other two types in minimal useful life as $r_{Beef}^{expire-max} = r_{Preslaughter}^{expire-max} = (104 \text{ weeks})^{-1}$. The $Preslaughter$ population is assumed to not have a minimal expiration rate $r_{Preslaughter}^{expire-min} = 0$, but they have the highest feasible slaughter rate of $r_{Preslaughter}^{slaughter-max} = (2 \text{ weeks})^{-1}$. Lastly the $Beef$ populations have a feasible range for their slaughter rates of $r_{Beef}^{slaughter-min} = 0$ to $r_{Beef}^{slaughter-min} = (13 \text{ weeks})^{-1}$. The upper limits on the slaughter rates are quite high, but we explain the need for this later in this section.

4.4.1 Cattle movement parameters

As the focus of this study, the cattle movement parameters $p^x_{t_1,j_1,t_2,j_2,dist}$ express the probability that, within a week's duration, an individual in a subpopulation of type t_1 and size j_1 will move or be shipped to a subpopulation of type t_2 and size j_2 at a (county-to-county)

Source → Destination	$d0$	$d100$	$d200$	$d500$	$d1000$
$D, z1\_19 \rightarrow B, z1\_19$	0.2687743	0.0390975	0.0171808	0.0807097	0.1742045
$D, z20\_199 \rightarrow B, z1\_19$	0.0	0.0	0.0	0.0	0.0
$D, z200\_up \rightarrow B, z1\_19$	0.0360669	0.0076309	0.0007037	0.0	0.0
$D, z1\_19 \rightarrow B, z20\_199$	0.2088205	0.0370924	0.0201817	0.1386829	0.2043643
$D, z20\_199 \rightarrow B, z20\_199$	0.0	0.0	0.0	0.0004745	0.0
$D, z200\_up \rightarrow B, z20\_199$	0.0204587	0.0056272	0.0005575	0.0008731	0.0
$D, z1\_19 \rightarrow B, z200\_up$	0.1948755	0.0350860	0.0159960	0.1036465	0.3731019
$D, z20\_199 \rightarrow B, z200\_up$	0.0	0.0	0.0	0.0	0.0013401
$D, z200\_up \rightarrow B, z200\_up$	1.7686314	0.0004226	0.0004293	0.0241026	0.0005980

Table 4.1: *Estimated movement parameters* $p^x_{t_1,j_1,t_2,j_2,dist} \cdot 10^3$*, Dairy to Beef*

distance of $dist$. Table 4.1 presents a subset of these probabilities that express the movement of cattle from $Dairy$ subpopulations to $Beef$ subpopulations for 5 ranges of distance. A complete table of the cattle movement probabilities is provided in appendix section C.2 as table C.3. The tables in appendix section C.3 present the birth, expiration, and slaughter probabilities of the 9 subpopulations of each county for a sample of 10 counties from each State. The entire results of the birth, expiration, and slaughter probabilities are too large for this chapter as we are studying 1034 counties. Once having obtained the solution, we revisited the movement data released by the USDA NASS. A significant difference exists between the NASS movement data and these movement parameters we've estimated, namely, that the movements of NASS are summarized from individual premises, but our parameters describe movements to and from collections of premises. We simulated 30 years of virtual cattle movements and shipments for slaughter and summarized the movements into the 3 size ranges of $Size_B$. We compare these results for individual counties, considering that shipments originating from aggregated subpopulations ought to usually be larger than shipments from individual operations. This means that the subpopulation-based results should over represent for larger sizes and perhaps under represent for smaller sizes of shipments. For Ellis County in the State of Kansas, figure 4.3 presents a comparison of our subpopulation based distributions of shipments in blue against the NASS reported yearly totals in red.

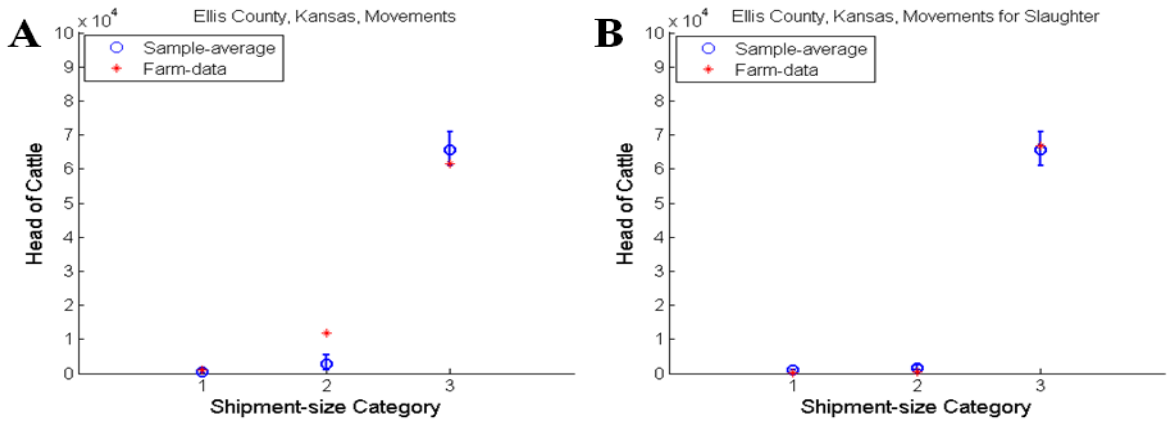

Figure 4.3: *(A) The average yearly totals of All Movements originating from Ellis County, Kansas and generated from the estimated subpopulation cattle movement parameters for the* 3 *size categories of* $Size_B$ *are represented with blue circles and their respective* 99% *confidence intervals are shown with the vertical lines. Shown in red stars are the yearly totals reported by NASS for the same county and aggregated into the ranges of* $Size_B$*. (B) The yearly totals of cattle shipped for slaughter originating in Ellis County are displayed from both the estimated slaughter rates (in blue) and the NASS database (in red) following the same notations as in A.*

The three size categories represent the three ranges of $Size_B$, with the smallest range on the left and the largest on the right side. Trego County, also in the State of Kansas, demonstrates one way in which the year-long resolution of the Agricultural Census is insufficient to express the seasonality of the cattle system. At the time of the 2007 census, Trego County reported no large *Preslaughter* populations of cattle. On the year, however, Trego County was responsible for several large shipments of cattle for slaughter. The census happened to catch the finishing yards at a point in time in which they were empty and thus neither the true capacity nor typical population levels of *Preslaughter* cattle are represented in the NASS database. Figure 4.4.B displays the dramatic mismatch that occurs for the largest slaughter shipment size category.

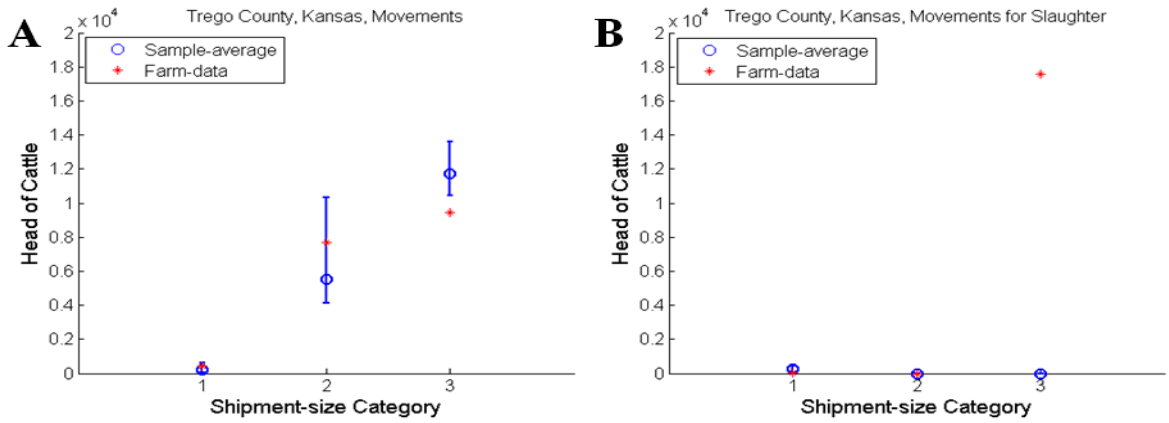

Figure 4.4: *(A) The average yearly totals of All Movements originating from Trego County, Kansas and generated from the estimated subpopulation cattle movement parameters for the* 3 *size categories of* $Size_B$ *are represented with blue circles and their respective* 99% *confidence intervals are shown with the vertical lines. Shown in red stars are the yearly totals reported by NASS for the same county and aggregated into the ranges of* $Size_B$*. (B) The yearly totals of cattle shipped for slaughter originating in Trego County are displayed from both the estimated slaughter rates (in blue) and the NASS database (in red) following the same notations as in A. A large discrepancy occurs in the third size category between the NASS reported slaughter totals and the generated distribution.*

4.5 Discussion and conclusions

We have designed and solved a large-scale optimal estimation problem in an attempt to address the privacy challenges and the need for livestock movement data in the United States. Given the resolution limitations of the data available, we do not try to estimate very detailed parameters, but rather adopt a stratified metapopulation approach where we shape the structure of our variables around the structure of the NASS data. Our approach is limited by the timescale resolution of the census report. This leads to several seasonal challenges that include correctly quantifying populations, identifying periods of higher and lower movement rates, and capturing birthing and slaughtering seasons. The demographic bounds used in our formulation depend on advice from industry experts and are flexibly open to further insights. We simulated and compared subpopulation-based movement distributions with operation-based movements, however this is not a rigorous method of validation. Our problem design includes an assumption on the relationships of cattle types and sizes that

proved to not hold true for all counties, as demonstrated by Broomfield County in the State of Colorado. Lastly, the parameters we estimate fail to consider individual State laws and Veterinary practices as well as the role of wildlife in the interfacing of subpopulations.

A close examination of the results in table C.3 of appendix section C.2 reveals a relatively unrealistic fraction of probabilities that are estimated to take on a value of 0. We believe this to be an artifact that has arisen from the design of the formulation and the nature of the optimization. Without a sufficiently diverse set of constraints, the dimensionality of an optimal solution for the problem will be limited. The objective to maximize entropy would prefer to diversify the results, making as many nonzero as possible, yet it seems too tightly restricted in some way as to allow that to occur in the solution. If we had to select a constraint which would most likely be the cause for these probabilities taking on a value of 0, we would first suspect the tight error limit of inequality 4.13. All other constraints allow a reasonable amount of flexibility in the set of parameters that would satisfy them. As the objective strives to diversify the distributions, we predict that loosening the error limits would result in fewer movement probabilities taking on a value of 0.

From the challenges that this problem held, a few insights were discovered that might improve the success of a future version of these livestock movement estimation methods without posing any threat to the confidentiality of the data. Having examined the design limitations of this study, we would like to propose 3 new questions to be considered for addition to future versions of the US Census of Agriculture. Those being

- "Of the cattle sold/moved in question 4, excluding those sold for slaughter in question 6, how many went to (a) destinations within the county, (b) destinations in a neighboring county, (c) destinations within the same state, (d) destinations in neighboring (bordering) states, (e) destinations further away?"

- "Of cattle that arrive to this operation during 20XX, how many came from (a) a locations within the county, (b) locations in a neighboring county, (c) locations within the same state, (d) locations in neighboring (bordering) states, (e) locations further

away?"

- "How many cattle were born on these premises during the year 20XX?"

Although with many limitations, we have taken a significant first step in tackling the challenges of data in the United States without compromising anonymity through optimization and computation. In chapter 5, we envision the future of this effort in the following chapter and provide a brief example of the incorporation of these results with the results of the previous two chapters.

Chapter 5

Closing Thoughts

The efforts of the previous chapters are certainly not complete. In this closing chapter, we share our ideas concerning the completion of these efforts, demonstrate a rough evaluation of the global epidemic invasion thresholds for the 10 Central States, and summarize the content of this dissertation.

5.1 Future Directions

Much work remains before the two sides of this characterization will fully converge. We propose a series of future objectives that would achieve this convergence and follow it with a brief demonstration of the initial merging of these two parts.

5.1.1 Future work

We envision a data-driven metapopulation cattle demographic and movement model for these 10 Central States that compose the heart of the US cattle feedlot system. This model would be built with nodes representing US counties, would include the culling processes, and would stratify each county's population by three farm size ranges ($Size_B$) and two cattle types (*Preslaughter* and *General*). This would require further expanding the complexity

of the model of chapter 3 as well as revising a few constraints in the data estimation method of chapter 4 and resolving for a new set of parameters. The objectives necessary to reach this data-driven cattle demographic and movement model are

- to modify the single type of cattle into two types, *General* and *Preslaughter*, for the model of chapter 3;
- to introduce 3 size ranges of cattle operations, $1-19$ head, $20-199$ head, and 200 or more head, for the model of chapter 3;
- to replace p and the movement model with a multidimensional movement parameter for the model of chapter 3,
- to expand the single parameter birth and slaughter processes to capture the respective estimated rates for the model of chapter 3,
- to construct the large cattle movement network representing the 10 Central States and determine its degree distributions and network metrics for the model of chapter 3,
- to derive expressions for the global epidemic invasion thresholds and critical movement rates for the expanded model of chapter 3,
- to adjust the daily timescale of the expanded model to the weekly scale of the estimated parameters from the method of chapter 4,
- to add a longer feasible distance range, allowing movements up to 2000 miles as suggested by[92], for the method of chapter 4,
- to revise the $Type_B$ set of the estimation problem to include only *Preslaughter* cattle and a single type, composed of the remaining cattle, *General*, for the method of chapter 4,

- to reduce the dimensionality of the birth, culling, and slaughter parameters from estimating a triple for every subpopulation in every county for the method of chapter 4,
- to quantify the impacts of minor relaxations of the error limits for the updated method of chapter 4,
- to estimate the model parameters through the updated method of chapter 4, and
- to design a software package for the simulation, statistics reporting, and visualization of the data-driven cattle demographic and movement model for further testing and validation.

The further development of this model for US cattle systems would be significantly improved through modifications in the US Census of Agriculture, namely, the questions on distance and direction proposed in section 4.5 of chapter 4 and an increase in the resolution of the timescale from a yearly basis. Apart from these database expansions, another direction of improvement would be the inclusion of additional States in the data estimation to cover the entire United States or further. With these objectives, an initial national modeling system of epidemics in cattle would be ready for testing and validation. This modeling system would serve as a starting point in understanding outbreaks of different diseases in various types of US livestock. Researchers and scientist could then use the system for simulating the impacts of various policies and control strategies. In the next subsection, we consider a brief glimpse at where the proper development of the efforts mentioned in this dissertation may lead.

5.1.2 Are US cattle systems at risk?

We extrapolated the necessary parameters to compute a rough approximation of the critical movement rates p_c and p_c^{TS} from the estimated data of chapter 4 for the 10 Central States.

The computation of the average movement rate $\langle p \rangle$ across the 10 States and the comparison of these values to the critical movement rates reveal the potential for epidemics to break out across these States when considering the current movement rates. We began by converting the timescale of the disease parameters of chapter 3 to a weekly scale with $\mu = 1(week)^{-1}$, $R_0 = 1.2$ (unchanged), and $\beta = \mu R_0$. We averaged the populations across all subpopulations of the results of chapter 4 to yield $\overline{N}^{approx} = 5483.8$ head. To approximate the death rate δ, we averaged the slaughter rates of all $Beef$ and $Preslaughter$ subpopulations to find $\delta^{approx} = 0.01516(week)^{-1}$. For the remaining parameters and the average movement rate $\langle p \rangle$, we simulated 30 years of cattle movements in the same manner as described in subsection 4.4.1 of chapter 4. We captured the average outward movement $\langle p \rangle$ for all $Beef$ and $Dairy$ subpopulations for each week (the time step) and over each year. Concurrently, we measured the resulting network parameters (η_{in}, η_{out}, k_i^{in}, k_i^{out}) for networks representing the weekly and yearly movements. We have considered the dynamic network measurements as produced by the realized movements and not static networks as considered in chapters 2 and 3. Further, we have ignored the potential existence of any degree correlations in the resulting networks from the movement-based network construction. Having collected this approximate set of parameters, we computed the critical movement rates p_c and p_c^{TS} and plot them in comparison to the system average movement rate $\langle p \rangle$ of the estimated cattle movement parameters in figure 5.1.

This first approximation of the incorporation of these two components suggests that the US cattle systems are at a significant level of risk. The average cattle movement rates are 1 to 1.5 magnitudes larger than the thresholds defined by the critical movement rates. This result suggests that an epidemic with parameters similar to the ones implemented would easily invade US cattle populations and reach the cattle's final destinations, possibly compromising the security of the beef supply chain. To impede the progress of this epidemic, a reduction in cattle movement rates of more than 90% would be required. For human influenza-like illnesses, a reduction in mobility rates of one order of magnitude has also been proposed[7].

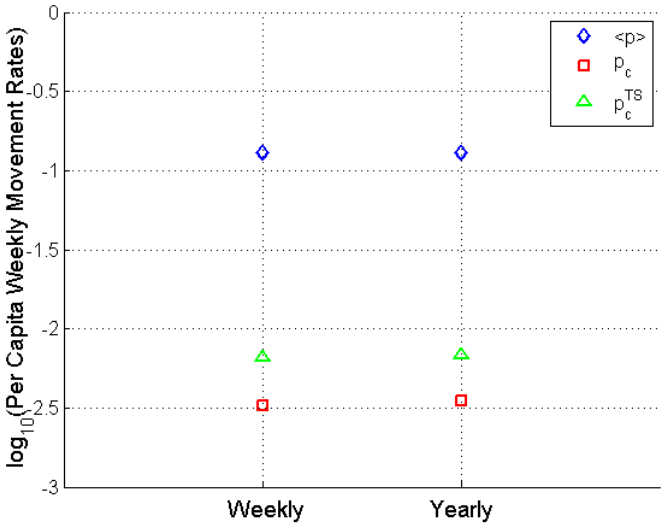

Figure 5.1: *The weekly average and yearly average of weekly individual cattle movement rates* $\langle p \rangle$ *are plotted in blue diamonds on a vertical logarithmic scale in comparison with the approximated critical movement rate* p_c *(shown in red squares) of the global epidemic invasion threshold and the critical movement rate* p_c^{TS} *(shown in green triangles) of the global transit-to-sink invasion threshold over* 30 *years.*

This amount of reduction is based on an approximated comparison, and further work should be completed prior to advising any authorities.

5.2 Recapitulation

This dissertation focuses on improving models and parameters for the characterization of epidemics in metapopulation cattle systems. In chapter 1, we introduce the content and a collection of basic concepts and useful vocabulary for the comprehension of this multi-disciplinary effort. Further work remains to develop this effort into a complete modeling system for each particular livestock species and disease combination. Chapters 2 and 3 design metapopulation cattle demographic and movement models on directed networks. They further determine the quasi-equilibrium distributions of the cattle populations and derive the respective versions of the global epidemic invasion threshold R_*[17,19]. These global epi-

demic invasion thresholds serve to describe the ranges of parameters which either will or will not permit an epidemic to spread through the metapopulation cattle system. In the work of chapter 3 through the study of a more complex directed-movement, metapopulation cattle model, we discovered a second global invasion threshold R_*^{TS}, which defines whether an epidemic will spread beyond the set of transit nodes to the sink nodes of a network system. These sink nodes primarily represent final feedlots that have no outgoing movements except for slaughter. The significance of an epidemic reaching these operations is that the disease has a better chance to reach the slaughter facilities. We have also derived the relationship between the two invasion thresholds and discovered its dependence on the fractional compositions and in-degree distributions of the transit and sink nodes.

The outputs of even the best models are restricted by the quality and resolution of their data inputs. To address this significant challenge in the movements of US cattle, we designed a method to optimally estimate cattle movement parameters from the publicly available data from the USDA Census of Agriculture[1]. We designed a data structure for metapopulation parameters to closely match the USDA data structures. The core of our method is a large, convex optimal parameter estimation problem. This problem constraints are defined to match aggregated movement data, to conserve the year-to-year population levels of cattle, and to follow a general set of industrial constraints. The objective function is based on the maximum entropy method to yield the minimal set of assumptions in our formulation[18]. We use this method to estimate cattle movement and demographic parameters for 1034 Counties of 10 Central States that compose the core of the US cattle industry. Distributions of shipments and slaughter are generated from our estimations, and we compare these to the aggregated farm-level shipment distributions for each County.

The two paths of this characterization effort have yet to converge, thus we outlined a potential list of objectives to achieve this. To conclude this dissertation in chapter 5, we demonstrated a small portion of the potential contributions of the convergence of these efforts. We approximate a comparison of the average movement rates of the metapopulation

cattle system from our estimates of chapter 4 with the critical movement rates of chapter 3. This small, initial example suggests that the current movements and structure of US cattle populations leave them vulnerable to epidemic outbreaks.

Bibliography

[1] USDA - NASS, census of agriculture, October 2013. URL http://www.agcensus.usda.gov.

[2] Tom Vilsack and Cynthia Z.F. Clark. 2007 census of agriculture, united states, summary and state data. Technical Report Volume 1, Part 51, AC-07-A-51, United States Department of Agriculture, December 2009. URL www.agcensus.usda.gov/Publications/2007/Full_Report/usv1.pdf?

[3] Gary W Brester, John M Marsh, and Ronald L Plain. International red meat trade. *Veterinary Clinics of North America: Food Animal Practice*, 19(2):493–518, July 2003. ISSN 0749-0720.

[4] Neil M. Ferguson, Christl A. Donnelly, and Roy M. Anderson. The foot-and-mouth epidemic in great britain: Pattern of spread and impact of interventions. *Science*, 292 (5519):1155–1160, May 2001.

[5] Ling Xue, Lee W. Cohnstaedt, H. Morgan Scott, and Caterina Scoglio. A hierarchical network approach for modeling rift valley fever epidemics with applications in north america. *PLoS ONE*, 8(5):e62049, May 2013.

[6] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Med*, 4(1):e13, January 2007.

[7] Vittoria Colizza and Alessandro Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251(3):450–467, April 2008.

[8] Duygu Balcan, Vittoria Colizza, Bruno Gonalves, Hao Hu, Jos J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106:21484–21489, December 2009.

[9] Dillon M Feuz and Wendy J Umberger. Beef cow-calf production. *Veterinary Clinics of North America: Food Animal Practice*, 19(2):339–363, July 2003.

[10] Derrell S Peel. Beef cattle growing and backgrounding programs. *Veterinary Clinics of North America: Food Animal Practice*, 19(2):365–385, July 2003.

[11] M. E. J. Newman. *Networks: An Introduction*. Oxford University, USA, 1 edition, 2010.

[12] Alain Barrat, Marc Barthlemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, UK, 2008.

[13] Ilkka A. Hanski and Michael Eugene Gilpin. *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press, San Diego, 1997.

[14] David Tilman and Peter M. Kareiva. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, 1997.

[15] Jordi Bascompte and Ricard V. Sol. *Modeling spatiotemporal dynamics in ecology*. Springer, New York, 1998.

[16] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, September 2011.

[17] Duygu Balcan and Alessandro Vespignani. Invasion threshold in structured populations with recurrent mobility patterns. *Journal of theoretical biology*, 293:87–100, January 2012.

[18] Hamdy A. Taha. *Operations Research: An Introduction*. Prentice Hall International, 2007.

[19] Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical Review Letters*, 99(14):148701, October 2007.

[20] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, January 2012.

[21] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, Apr-Jun 2009.

[22] Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R White. Economic networks: the new challenges. *Science*, 325(5939):422–425, July 2009.

[23] Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. DebtRank: too central to fail? financial networks, the FED and systemic risk. *Scientific Reports*, 2:541, August 2012.

[24] Alun L. Lloyd and Robert M. May. How viruses spread among computers and people. *Science*, 292(5520):1316–1317, May 2001.

[25] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, April 2001.

[26] Ilkka Hanski and Oscar E. Gaggiotti. *Ecology, Genetics, and Evolution of Metapopulations*. Elsevier, Academic Press, Amsterdam, 2004.

[27] H W Hethcote. An immunization model for a heterogeneous population. *Theoretical Population Biology*, 14(3):338–349, December 1978.

[28] Robert M. May and Roy M. Anderson. Population biology of infectious diseases: Part II. *Nature*, 280(5722):455–461, August 1979.

[29] R M Anderson and R M May. Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes. *IMA journal of mathematics applied in medicine and biology*, 1(3):233–266, 1984.

[30] Robert M. May and Roy M. Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111, November 1984.

[31] B M Bolker and B T Grenfell. Chaos and biological complexity in measles dynamics. *Proceedings of The Royal Society London Biological Sciences*, 251(1330):75–81, January 1993.

[32] Benjamin Bolker and Bryan Grenfell. Space persistence and dynamics of measles epidemics. *Philosophical Transactions: Biological Sciences*, 348(1325):309–320, May 1995.

[33] L Sattenspiel and K Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical biosciences*, 128(1-2):71–91, August 1995.

[34] Matt J. Keeling and Pejman Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters*, 5(1):2029, 2002.

[35] Alun L. Lloyd and Robert M. May. Spatial heterogeneity in epidemic models. *Journal of Theoretical Biology*, 179(1):1–11, March 1996.

[36] B Grenfell and J Harwood. (meta)population dynamics of infectious diseases. *Trends in ecology & evolution*, 12(10):395–399, October 1997.

[37] Grenfell and Bolker. Cities and villages: infection hierarchies in a measles metapopulation. *Ecology Letters*, 1(1):6370, 1998.

[38] Neil M. Ferguson, Matt J. Keeling, W. John Edmunds, Raymond Gani, Bryan T. Grenfell, Roy M. Anderson, and Steve Leach. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, October 2003.

[39] Steven Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301, June 2007.

[40] O.V. Baroyan, L.A. Genchikov, L.A. Rvachev, and V.A. Shashkov. An attempt at large-scale influenza epidemic modelling by means of a computer. *Bull. Int. Epidemiol. Assoc.*, 18:22–31, 1969.

[41] Leonid A. Rvachev and Ira M. Longini Jr. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3–22, July 1985.

[42] Ira M. Longini Jr. A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, 90(12):367–383, July 1988.

[43] Antoine Flahault and Alain-Jacques Valleron. A method for assessing the global spread of HIV-1 infection based on air travel. *Mathematical Population Studies*, 3(3): 161–171, 1992.

[44] Rebecca F Grais, J Hugh Ellis, and Gregory E Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, 18(11):1065–1072, 2003.

[45] Shigui Ruan, Wendi Wang, and Simon A Levin. The effect of global travel on the spread of sars. *Mathematical Biosciences and Engineering: MBE*, 3(1):205–218, January 2006.

[46] D J Earn, P. Rohani, and BT Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of The Royal Society Biological Sciences*, 265(1390): 7–10, 1998.

[47] Pejman Rohani, David J. D. Earn, and Bryan T. Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*, 286(5441):968–971, October 1999.

[48] Matt J. Keeling. Metapopulation moments: Coupling, stochasticity and persistence. *Journal of Animal Ecology*, 69(5):725–736, September 2000.

[49] Andrew W. Park, Simon Gubbins, and Christopher A. Gilligan. Extinction times for closed epidemics: the effects of host spatial structure. *Ecology Letters*, 5(6):747755, 2002.

[50] Alexei Vazquez. Epidemic outbreaks on structured populations. *Journal of Theoretical Biology*, 245(1):125–129, March 2007.

[51] R. F. Grais, J. H. Ellis, A. Kress, and G. E. Glass. Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel. *Health Care Management Science*, 7(2):127–134, May 2004.

[52] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, October 2004.

[53] Vittoria Colizza, Alain Barrat, Marc Barthlemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, February 2006.

[54] Ben S Cooper, Richard J Pitman, W. John Edmunds, and Nigel J Gay. Delaying the international spread of pandemic influenza. *PLoS Med.*, 3(6):e212, May 2006.

[55] T Dirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine*, 12(5): 497–499, May 2006.

[56] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction-

diffusion processes and metapopulation models in heterogeneousnetworks. *Nature Physics*, 3(4):276–282, April 2007.

[57] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics*, 7(7):581–586, July 2011.

[58] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. Heterogeneous length of stay of hosts movements and spatial epidemic spread. *Scientific Reports*, 2(476), June 2012.

[59] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7(1):46–89, February 1997.

[60] Paul C. Cross, James O. Lloyd-Smith, Philip L. F. Johnson, and Wayne M. Getz. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters*, 8(6):587595, 2005.

[61] Paul C Cross, Philip L.F Johnson, James O Lloyd-Smith, and Wayne M Getz. Utility of r0 as a predictor of disease invasion in structured populations. *Journal of the Royal Society Interface*, 4(13):315–324, April 2007.

[62] G. Chowell, J. M. Hyman, S. Eubank, and C. Castillo-Chavez. Scaling laws for the movement of people between locations in a large city. *Physical Review E*, 68(6):066102, December 2003.

[63] A. Barrat, M. Barthlemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, March 2004.

[64] R. Guimer, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, May 2005.

[65] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, January 2006.

[66] Roberto Patuelli, Aura Reggiani, Sean P. Gorman, Peter Nijkamp, and Franz-Josef Bade. Network analysis of commuting flows: A comparative static approach to german data. *Networks and Spatial Economics*, 7(4):315–331, December 2007.

[67] Marta C. Gonzlez, Csar A. Hidalgo, and Albert-Lszl Barabsi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[68] Alessandro Vespignani. Predicting the behavior of techno-social systems. *Science*, 325 (5939):425–428, July 2009.

[69] R.R Kao, L Danon, D.M Green, and I.Z Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597):1999–2007, August 2006.

[70] M Bigras-Poulin, R A Thompson, M Chriel, S Mortensen, and M Greiner. Network analysis of danish cattle industry trade patterns as an evaluation of risk potential for disease spread. *Preventive Veterinary Medicine*, 76(1-2):11–39, September 2006.

[71] Michel Bigras-Poulin, Kristen Barfod, Sten Mortensen, and Matthias Greiner. Relationship of trade patterns of the danish swine industry animal movements network to potential disease spread. *Preventive Veterinary Medicine*, 80(23):143–165, July 2007.

[72] S E Robinson and R M Christley. Exploring the role of auction markets in cattle movements within great britain. *Preventive Veterinary Medicine*, 81(1-3):21–37, September 2007.

[73] Filipa M. Baptista, Telmo Nunes, Virgilio Almeida, and Armando Louza. Cattle movements in portugal - an insight into the potential use of network analysis. *Revista Portuguesa de Ciencias Veterinarias*, 107(565-566):35–40, 2008.

[74] Fabrizio Natale, Armando Giovannini, Lara Savini, Diana Palma, Luigi Possenti, Gianluca Fiore, and Paolo Calistri. Network analysis of italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine*, 92 (4):341–350, December 2009.

[75] S Rautureau, B Dufour, and B Durand. Vulnerability of animal trade networks to the spread of infectious diseases: A methodological approach applied to evaluation and emergency control strategies in cattle, france, 2005. *Transboundary and Emerging Diseases*, 58(2):110–120, April 2011.

[76] Matt J. Keeling, Mark E. J. Woolhouse, Darren J. Shaw, Louise Matthews, Margo Chase-Topping, Dan T. Haydon, Stephen J. Cornell, Jens Kappey, John Wilesmith, and Bryan T. Grenfell. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817, October 2001.

[77] D. M. Green, I. Z. Kiss, and R. R. Kao. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2729–2735, November 2006.

[78] Rowland R Kao, Darren M Green, Jethro Johnson, and Istvan Z Kiss. Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of the Royal Society Interface*, 4 (16):907–916, October 2007.

[79] Paolo Bajardi, Alain Barrat, Lara Savini, and Vittoria Colizza. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of The Royal Society Interface*, 9(76):2814–2825, November 2012.

[80] Matt J. Keeling, Leon Danon, Matthew C. Vernon, and Thomas A. House. Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences*, 107(19):8866–8870, May 2010.

[81] Hartmut H K Lentz, Thomas Selhorst, and Igor M Sokolov. Spread of infectious diseases in directed and modular metapopulation networks. *Physical review. E*, 85: 066111, June 2012.

[82] Theodore E. Harris. *The Theory of Branching Processes*. Courier Dover Publications, 1989.

[83] Alexei Vazquez. Polynomial growth in branching processes with diverging reproductive number. *Physical Review Letters*, 96(3):038702, January 2006.

[84] Norman T. J. Bailey. *The Mathematical Theory of Infectious Diseases and Its Applications*. Charles Griffin, 1975.

[85] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Comb. Probab. Comput.*, 7(3):295305, September 1998.

[86] Michele Catanzaro, Marin Bogu, and Romualdo Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical review. E*, 71:027103, February 2005.

[87] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Correlations in interacting systems with a network topology. *Physical Review E*, 72(6):066130, December 2005.

[88] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical patterns of cattle trade movements. *PLoS ONE*, 6(5):e19869, May 2011.

[89] Claudia Taylor, Achla Marathe, and Richard Beckman. Same influenza vaccination strategies but different outcome across US cities? *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 14(9):e792–e795, September 2010.

[90] Nidhi Parikh, Mina Youssef, Samarth Swarup, and Stephen Eubank. Modeling the effect of transient populations on epidemics in washington DC. *Scientific Reports*, 3 (3152), November 2013.

[91] Leon Danon, Ashley P. Ford, Thomas House, Chris P. Jewell, Matt J. Keeling, Gareth O. Roberts, Joshua V. Ross, and Matthew C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011(284909), March 2011.

[92] Michael G. Buhnerkempe, Daniel A. Grear, Katie Portacci, Ryan S. Miller, Jason E. Lombard, and Colleen T. Webb. A national-scale picture of U.S. cattle movements obtained from interstate certificate of veterinary inspection data. *Preventive Veterinary Medicine*, In Press, 2013.

[93] William H. Hsu, Sohini Roy Chowdhury, and Caterina Scoglio. Mitigation strategies for foot and mouth disease: A learning-based approach. *Int. J. Artif. Life Res.*, 2(2): 4276, April 2011.

[94] Ling Xue, H Morgan Scott, Lee W Cohnstaedt, and Caterina Scoglio. A network-based meta-population approach to model rift valley fever epidemics. *Journal of Theoretical Biology*, 306:129–144, August 2012.

[95] Jie Zhou, Gaoxi Xiao, Limsoon Wong, Xiuju Fu, Stefan Ma, and Tee Hiang Cheng. Generation of arbitrary two-point correlated directed networks with given modularity. *Physics Letters A*, 374(3132):3129–3135, July 2010.

[96] Ranjni J Chand, Benjamin R Trible, and Raymond RR Rowland. Pathogenesis of porcine reproductive and respiratory syndrome virus. *Current Opinion in Virology*, 2 (3):256–263, June 2012.

[97] F. Darabi Sahneh, C. Scoglio, and P. Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, PP(99):1, 2013. ISSN 1063-6692.

[98] US census bureau, October 2013. URL http://www.census.gov/.

[99] European Council. European parliament and european council (2000) regulation (EC) no. 1760/ 2000 of 17 july 2000 establishing a system for the identification and registration of bovine animals and regarding labeling of beef and beef products and repealing council regulation (EC) no. 820/97. Technical Report L 204, Off. J. Eur. Communities, August 2000.

[100] Cerian R Webb. Farm animal networks: unraveling the contact structure of the british sheep population. *Preventive veterinary medicine*, 68(1):3–17, April 2005. ISSN 0167-5877.

[101] Istvan Z. Kiss, Darren M. Green, and Rowland R. Kao. The network of sheep movements within great britain: network properties and their implications for infectious disease spread. *Journal of The Royal Society Interface*, 3(10):669–677, October 2006.

[102] S.E Robinson, M.G Everett, and R.M Christley. Recent network evolution increases the potential for large epidemics in the british cattle population. *Journal of the Royal Society Interface*, 4(15):669–674, August 2007.

[103] M.L. Brennan, R. Kemp, and R.M. Christley. Direct and indirect contacts between cattle farms in north-west england. *Preventive Veterinary Medicine*, 84(34):242–260, May 2008. ISSN 0167-5877.

[104] Matthew C. Vernon and Matt J. Keeling. Representing the UK's cattle herd as static and dynamic networks. *Proceedings of the Royal Society B: Biological Sciences*, 276 (1656):469–476, February 2009.

[105] MAF Biosecurity New Zealand. Review of selected cattle identification and tracing systems worldwide, 2009. URL http://www.biosecurity.govt.nz/files/publications/ris/nait-review-cattleident-systems-worldwide.pdf. 2009/03.

[106] Victoriya V. Volkova, Richard Howey, Nicholas J. Savill, and Mark E. J. Woolhouse.

Sheep movement networks and the transmission of infectious diseases. *PLoS ONE*, 5 (6):e11185, June 2010.

[107] National animal identification system - US department of agriculture, October 2013. URL http://www.usda.gov/nais.

[108] DeeVon Bailey, B. Wade Brorsen, and Michael R. Thomsen. Identifying buyer market areas and the impact of buyer concentration in feeder cattle markets using mapping and spatial statistics. *American Journal of Agricultural Economics*, 77(2):309–318, May 1995.

[109] Forde K., Hillberg-Seitzinger A., Dargatz D., and Wineland N. The availability of state-level data on interstate cattle movements in the united states. *Preventive Veterinary Medicine*, 37(1):209–217, 1998.

[110] T W Bates, M C Thurmond, and T E Carpenter. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three california counties, with reference to control of potential foot-and-mouth disease transmission. *American journal of veterinary research*, 62(7):1121–1129, July 2001.

[111] D. Shields and K. Mathews. 2003 interstate livestock movements USDA ERA outlook report LDP-M-108-01. Technical report, USDA National Agricultural Statistics Service, USDA, 2007. URL http//www.nass.usda.gov/census.

[112] B. J. Dominguez. *Characterization of livestock herds in extensive agricultural settings in southwest Texas*. MS, Texas A&M University, USA, 2007.

[113] Michael J. Tildesley, Thomas A. House, Mark C. Bruhn, Ross J. Curry, Maggie ONeil, Justine L. E. Allpress, Gary Smith, and Matt J. Keeling. Impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences*, 107(3):1041–1046, December 2009.

[114] National agricultural statistics service: NASS, October 2013. URL http://www.nass.usda.gov.

[115] Willard C Losinger, Lindsey P Garber, George W Hill, Stephen E Dornseif, Judith M Rodriguez, and William B Frye. Design and implementation of the united states national animal health monitoring system 19941995 cattle on feed evaluation, and an evaluation of the impact of response biases. *Preventive Veterinary Medicine*, 31(12): 1–14, July 1997.

[116] David A Dargatz, Grant A Dewell, and Robert G Mortimer. Calving and calving management of beef cows and heifers on cowcalf operations in the united states. *Theriogenology*, 61(6):997–1007, April 2004.

[117] Raymond W. Yeung. *Information Theory and Network Coding*. Springer, August 2008.

[118] AIMMS: the modeling system, October 2013. URL http://business.aimms.com/.

[119] J. Gani and D. Jerwood. Markov chain methods in chain binomial epidemic models. *Biometrics*, 27(3):591–603, September 1971.

[120] Jr Longini, I M, M G Hudgens, M E Halloran, and K Sagatelian. A markov model for measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV vaccines. *Statistics in medicine*, 18(1):53–68, January 1999.

[121] Dennis L. Chao, M. Elizabeth Halloran, Valerie J. Obenchain, and Ira M. Longini. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol*, 6(1):e1000656, January 2010.

Appendix A

Global Invasion Threshold I: Derivations and Simulations

A.1 Basics of subpopulation networks

The degree distribution of the subpopulation network will be denoted by $P_v(\vec{k})$ which is the probability that a node chosen at random has joint-degree $\vec{k}$. Given the joint-degree distribution $P_v(\vec{k})$, the in- and out-degree distributions can easily be computed by

$$P_v^{in}(k^{in}) = \sum_{k^{out}} P_v(k^{in}, k^{out}) \text{ and } P_v^{out}(k^{out}) = \sum_{k^{in}} P_v(k^{in}, k^{out}) \ , \tag{A.1}$$

respectively. If there are no single-node degree correlations, i.e., correlations between in- and out- degrees of single nodes, then the joint-degree distribution is simply $P_v(k^{in}, k^{out}) = P_v^{in}(k^{in})P_v^{out}(k^{out})$. We will denote the two-node degree distribution by $P_a(\vec{k}, \vec{k'})$ which is the probability that an arc chosen at random originates from a node with degree $\vec{k}$ and terminates at a node with degree $\vec{k'}$. The degree distribution of nodes at the origins of arcs $P_a^{out}(\vec{k})$ is the probability that there is a node with joint-degree $\vec{k}$ at the origin of an arc

chosen at random and is related to the two- and single-node degree distributions by

$$P_a^{out}(\vec{k}) = \sum_{\vec{k'}} P_a(\vec{k}, \vec{k'}) = \frac{k^{out} P_v(\vec{k})}{\langle k^{out} \rangle} \ . \tag{A.2}$$

The complementary distribution, the degree distribution of nodes at the terminal ends of arcs, $P_a^{in}(\vec{k})$ corresponds to the probability that the node at the terminal end of a randomly chosen arc has joint-degree $\vec{k}$ and can be computed by

$$P_a^{in}(\vec{k}) = \sum_{\vec{k'}} P_a(\vec{k'}, \vec{k}) = \frac{k^{in} P_v(\vec{k})}{\langle k^{in} \rangle} \ . \tag{A.3}$$

If there are no two-node degree correlations, i.e., correlations between joint-degrees of connected nodes, then the two-node degree distribution is simply $P_a(\vec{k}, \vec{k'}) = P_a^{out}(\vec{k}) P_a^{in}(\vec{k'})$. The two-node degree correlations will play a crucial role on the dynamical processes which will be discussed below.

A.2 Infection and mobility dynamics

Since all the individuals in the same compartment X and the same node j are identical in terms of the dynamical processes, we are going refer to the number of such individuals at time t by $X_j(t)$. Then, by definition, the total number of individuals in node j at time t is given by $N_j(t) = \sum_X X_j(t)$. The number of individuals in compartment X and in node j is subject to discrete and stochastic dynamical processes defined by disease and transport operators. The disease operator $\mathcal{D}_j$ represents the change due to the compartment transition induced by the infection dynamics, and the transport operator Ω_X represents the variation due to mobility.

The term $\mathcal{D}_j$ can be written as a combination of a set of transitions $\{\mathcal{D}_j(X, Y)\}$, where $\mathcal{D}_j(X, Y)$ represents the number of transitions from compartment X to Y and is simulated as an integer random number extracted from a multinomial distribution. Then the change

due to infection dynamics reads as

$$\mathcal{D}_j(X) = \sum_Y \left[\mathcal{D}_j(Y,X) - \mathcal{D}_j(X,Y)\right] \quad . \tag{A.4}$$

As a concrete example, let us consider the temporal change in the infectious compartment. There is only one possible transition from the compartment, which is to the recovered compartment. The number of transitions is extracted from the binomial distribution

$$\Pr^{\text{Binom}}(I_j(t), p_{I_j \to R_j}) \quad , \tag{A.5}$$

which is determined by the transition probability

$$p_{I_j \to R_j} \quad = \quad \mu \Delta t \quad , \tag{A.6}$$

and the number of individuals in the compartment $I_j(t)$ (its size). This transition causes a reduction in the size of the compartment. The increase in the compartment size is due to the transitions from the susceptible to infectious compartment. This is also a random number extracted from the binomial distribution

$$\Pr^{\text{Binom}}(S_j(t), p_{S_j \to I_j}) \quad , \tag{A.7}$$

given by the chance of contagion

$$p_{S_j \to I_j} \quad = \quad \lambda_j(t) \Delta t \quad , \tag{A.8}$$

and the number of attempts equal to the number of susceptibles $S_j(t)$. After extracting these numbers from the appropriate distributions, we can calculate the total change $\mathcal{D}_j(I)$

in the infectious compartment as

$$\mathcal{D}_j(I) = \mathcal{D}_j(S, I) - \mathcal{D}_j(I, R) \quad . \tag{A.9}$$

Transport operator Ω_X expresses the total change in compartment sizes due to movements. The variation in X_j can be expressed as a combination of a set of variables $\{\Omega_X(i, j)\}$, where $\Omega_X(i, j)$ corresponds to the number of individuals moving from node i to j. Then the change due to diffusion dynamics is given by

$$\Omega_X(j) = \sum_i \left[\Omega_X(i, j) - \Omega_X(j, i)\right] \quad . \tag{A.10}$$

The $\Omega_X(i, j)$ is an integer random number extracted from the multinomial distribution

$$\Pr^{\text{Multinom}}(X_i(t), \{p_{X_i \to X_\ell} | \ell \in \upsilon_i^{out}\}) \quad , \tag{A.11}$$

determined by the probability of diffusion to subpopulation j

$$p_{X_i \to X_j} = d_{ij} \Delta t \quad , \tag{A.12}$$

and the number of such trails $X_i(t)$. We have assumed that the infection does not alter people's behavior, i.e., all the compartments are identical in their mobility. After each operator is applied to all the compartments, the population sizes are updated. The total outcome of the infection and mobility dynamics during one time step Δt can be summarized as

$$X_j(t + \Delta t) = \tilde{X}_j + \Omega_{\tilde{X}}(j) \quad , \quad \tilde{X}_j = X_j(t) + \mathcal{D}_j(X) \quad . \tag{A.13}$$

Recognize that the stochastic state variables $\{S_j(t),\ I_j(t),\ R_j(t)\}$ define a multivariate Markov chain[119,120,121] in which the present state of the system is determined only by the state of the system in the previous time step.

Appendix B

Global Invasion Thresholds II: Derivations and Simulations

The contents of this appendix serve to supplement the work of chapter 3.

B.1 Solutions of livestock model demographics

In the following we consider an ensemble of directed subpopulation networks that contain "source" and "sink" nodes. We propose population dynamics on the network ensemble, which considers birth, death, and migration processes while preserving the total population of the system.

B.1.1 Subpopulation networks

Imagine to have a directed subpopulation network of made of V nodes. Each node can assume one of three types or classes: source, sink, or transit. While source and sink nodes are connected to the rest of the system via out-going and in-coming links only, respectively, transit nodes have both out- and in-neighborhoods. The probability of finding a node of type source (sink) when a node is chosen at random is η_{out} (η_{in}), whereas $1 - \eta_{out} - \eta_{in}$ is

the probability that we will pick a transit node if a node is selected at random.

For the sake of simplicity, we assume that all the nodes but sources obey the same in-degree distribution $P_{v-o}^{in}\left(k^{in}\right)$, while all the nodes except sinks follow the same out-degree distribution $P_{v-i}^{out}\left(k^{out}\right)$. Moreover, we consider the case that there are no single-node degree correlations between the in- and out-degrees of transit nodes. However, the formulation below can be generalized to cover all the cases in which different types of nodes assume different degree distributions and also that there are single-node degree correlations. The joint-degree distribution $P_v\left(\vec{k}\right)$ of the subpopulation network under the simplifying assumptions above is thus given by

$$P_v\left(\vec{k}\right) = \eta_{out}\delta_{k^{in},0}P_{v-i}^{out}\left(k^{out}\right) + \eta_{in}\delta_{k^{out},0}P_{v-o}^{in}\left(k^{in}\right) + \left(1-\eta_{out}-\eta_{in}\right)P_{v-o}^{in}\left(k^{in}\right)P_{v-i}^{out}\left(k^{out}\right) \ , \tag{B.1}$$

where

$$P_{v-o}^{in}\left(0\right) = 0 \text{ and } P_{v-i}^{out}\left(0\right) = 0 \ . \tag{B.2}$$

Note that $\delta_{x,y}$ here represents the Kronecker delta. We focus on the degree distribution in order to obtain the relationships which are to be used in the following derivations. The in-degree distribution $P_v^{in}\left(k^{in}\right)$ of the subpopulation network is

$$P_v^{in}\left(k^{in}\right) = \sum_{k^{out}} P_v\left(k^{in},k^{out}\right) = \eta_{out}\delta_{k^{in},0} + \left(1-\eta_{out}\right)P_{v-o}^{in}\left(k^{in}\right) \ . \tag{B.3}$$

We can obtain the relationship between the average in-degree of all the nodes excluding sources and the average in-degree $\langle k^{in}\rangle$ of the whole system by

$$\begin{aligned}\left\langle k^{in}\right\rangle = \sum_{k^{in}} k^{in}P_v^{in}\left(k^{in}\right) = \left(1-\eta_{out}\right)\sum_{k^{in}\geq 1} k^{in}P_{v-o}^{in}\left(k^{in}\right) \Rightarrow \\ \sum_{k^{in}\geq 1} k^{in}P_{v-o}^{in}\left(k^{in}\right) = \frac{\left\langle k^{in}\right\rangle}{\left(1-\eta_{out}\right)} \ .\end{aligned} \tag{B.4}$$

The probability that if we pick a node at random we will hit a sink with in-degree k^{in} is $P_v\left(k^{in},0\right)=\eta_{in}P^{in}_{v-o}\left(k^{in}\right)$. Then it follows that

$$\sum_{k^{in}}k^{in}P^{in}_{v}\left(k^{in},0\right)=\eta_{in}\sum_{k^{in}}k^{in}P^{in}_{v-o}\left(k^{in}\right)=\eta_{in}\frac{\langle k^{in}\rangle}{(1-\eta_{out})}\ , \tag{B.5}$$

where we have plugged in equation B.4. These last two equations are going to prove their worth when we consider population dynamics. The out-degree distribution of the subpopulation network is given by

$$P^{out}_{v}\left(k^{out}\right)=\sum_{k^{in}}P_v\left(k^{in},k^{out}\right)=\eta_{in}\delta_{k^{out},0}+\left(1-\eta_{in}\right)P^{out}_{v-i}\left(k^{out}\right)\ . \tag{B.6}$$

We obtain the relationship between the average out-degree $\langle k^{out}\rangle$ of the whole network and that of all the nodes excluding sinks by

$$\begin{aligned}\left\langle k^{out}\right\rangle=\sum_{k^{out}}k^{out}P^{out}_{v}\left(k^{out}\right)=(1-\eta_{in})\sum_{k^{out}\geq 1}k^{out}P^{out}_{v-i}\left(k^{out}\right)\Rightarrow\\ \sum_{k^{out}\geq 1}k^{out}P^{out}_{v-i}\left(k^{out}\right)=\frac{\langle k^{out}\rangle}{(1-\eta_{in})}\ .\end{aligned} \tag{B.7}$$

The probability of encountering a source with out-degree k^{out} when a node is chosen at random is $P_v\left(0,k^{out}\right)=\eta_{out}P^{out}_{v-i}\left(k^{out}\right)$. It then follows that

$$\sum_{k^{out}}k^{out}P^{out}_{v}\left(0,k^{out}\right)=\eta_{out}\sum_{k^{out}}k^{out}P^{out}_{v-i}\left(k^{out}\right)=\eta_{out}\frac{\langle k^{out}\rangle}{(1-\eta_{in})}\ , \tag{B.8}$$

where we have plugged in equation B.7. Similarly, the last two equations are going to be useful in the following where we consider population dynamics on the subpopulation networks described here. Let us denote the pairwise (or joint) node degree distribution by $P_a\left(\vec{k},\vec{k}'\right)$ which is the probability that an arc selected at random has its origin from a node with degree $\vec{k}$ and its termination at a node with degree $\vec{k}'$. The probability to find a node

with degree $\vec{k}$ at the origin of a randomly chosen arc is

$$P_a^{out}\left(\vec{k}\right) = \sum_{\vec{k}'} P_a\left(\vec{k}, \vec{k}'\right) = \frac{k^{out} P_v\left(\vec{k}\right)}{\langle k^{out} \rangle} \ . \tag{B.9}$$

Likewise, the probability to find a node with degree $\vec{k}$ at the destination of a randomly chosen arc is

$$P_a^{in}\left(\vec{k}\right) = \sum_{\vec{k}'} P_a\left(\vec{k}', \vec{k}\right) = \frac{k^{in} P_v\left(\vec{k}\right)}{\langle k^{in} \rangle} \ . \tag{B.10}$$

With an assumption of no correlations between the degrees of connected nodes, the joint node degree distribution becomes $P_a\left(\vec{k}, \vec{k}'\right) = P_a^{out}\left(\vec{k}\right) P_a^{in}\left(\vec{k}'\right)$. The locations of arcs in the subpopulation network as described by $P_a\left(\vec{k}, \vec{k}'\right)$ plays a significant role in the dynamic processes of the system.

B.1.2 Demography in subpopulation networks

Consider now that each subpopulation of joint-degree $\vec{k}$ is occupied by $N_{\vec{k}}(t)$ individuals at time t. Each individual in a subpopulation of joint-degree $\vec{k}$ moves to a subpopulation with joint-degree $\vec{k}'$ in its out-neighborhood at rate $d_{\vec{k}\vec{k}'}$,

$$d_{\vec{k}\vec{k}'} = \frac{p}{k^{out}} \ , \ \forall k^{out} \neq 0 \ , \tag{B.11}$$

yielding a total per capita diffusion rate $d_{\vec{k}}$ of

$$d_{\vec{k}} = p \ , \ \forall k^{out} \neq 0 \ . \tag{B.12}$$

The above diffusion process is obviously defined for all the subpopulations except sinks. Each individual in a sink subpopulation with joint-degree $\vec{k}$, on the other hand, leaves the

subpopulation network or dies at rate $\delta_{\vec{k}}$,

$$\delta_{\vec{k}} = \delta \ , \ \forall k^{out} = 0 \ . \tag{B.13}$$

Each individual in a source subpopulation $\vec{k}$ replicates herself or gives birth to an offspring at rate $\beta_{\vec{k}}(t)$,

$$\beta_{\vec{k}}(t) = p_\beta \delta \frac{\sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)}{\sum_{\vec{l}|l^{in}=0} P_v\left(\vec{l}\right) N_{\vec{l}}(t)} \ , \ \forall k^{in} = 0 \ , \tag{B.14}$$

where p_β, $0 \leq p_\beta \leq 1$, represents the fraction of individuals "recycled" from the death process into the system through the birth process. The remaining fraction, $(1 - p_\beta)$ re-enters the network through an importation process to the source nodes. Notice that the fraction term of equation B.14 is the ratio of the total population in the sink nodes to the total population in the source nodes at time t. Each source subpopulation $\vec{k}$ imports individuals from an external source at rate $\epsilon_{\vec{k}}(t)$,

$$\epsilon_{\vec{k}}(t) = (1 - p_\beta)\, \delta \frac{k^{out}}{\sum_{\vec{l}|l^{in}=0} P_v\left(\vec{l}\right) l^{out}} \sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t) \ , \ \forall k^{in} = 0 \ , \tag{B.15}$$

that is proportional to the out-degree of the source node. Remembering equation B.8 simplifies the importation rate to

$$\epsilon_{\vec{k}}(t) = (1 - p_\beta)\, \delta \frac{(1 - \eta_{in})\, k^{out}}{\eta_{out} \left\langle k^{out} \right\rangle} \sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t) \ , \ \forall k^{in} = 0 \ , \tag{B.16}$$

The rate equation of average population size $N_{\vec{k}}(t)$ in a subpopulation with joint-degree $\vec{k}$ is then

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta_{\vec{k}} N_{\vec{k}}(t)\,\delta_{k^{out},0} - d_{\vec{k}} N_{\vec{k}}(t)\,(1-\delta_{k^{out},0}) \\ &+ \left[\beta_{\vec{k}}(t)\,N_{\vec{k}}(t) + \epsilon_{\vec{k}}(t)\right]\delta_{k^{in},0} \\ &+ k^{in}\sum_{\vec{j}} P_a\left(\vec{j},\vec{k}|\vec{k}\right) d_{\vec{j}\vec{k}} N_{\vec{j}}(t)\,(1-\delta_{j^{out},0})\ . \end{aligned} \tag{B.17}$$

If we insert all the rates in equation B.17 and also assume that there are no correlations between the degrees of connected nodes, we get

$$\begin{aligned}\partial_t N_{\vec{k}}(t) \;=\; &-\delta N_{\vec{k}}(t)\,\delta_{k^{out},0} - p N_{\vec{k}}(t)\,(1-\delta_{k^{out},0}) \\ &+\left[p_\beta\delta\frac{N_{\vec{k}}(t)}{\sum_{\vec{l}|l^{in}=0} P_v\left(\vec{l}\right) N_{\vec{l}}(t)} + (1-p_\beta)\,\delta\frac{(1-\eta_{in})\,k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right]\sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\delta_{k^{in},0} \\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}\sum_{\vec{j}} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\,(1-\delta_{j^{out},0})\ . \end{aligned} \tag{B.18}$$

The first thing to consider is that the average population per node $\overline{N} = \sum_{\vec{k}} P_v\left(\vec{k}\right) N_{\vec{k}}(t)$ is kept invariant over time by the above dynamical process, i.e., $\sum_{\vec{k}} P_v\left(\vec{k}\right)\partial_t N_{\vec{k}}(t) = 0$, which leads to

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta N_{\vec{k}}(t)\,\delta_{k^{out},0} - p N_{\vec{k}}(t)\,(1-\delta_{k^{out},0}) \\ &+\left[p_\beta\delta\frac{N_{\vec{k}}(t)}{\sum_{\vec{l}|l^{in}=0} P_v\left(\vec{l}\right) N_{\vec{l}}(t)} + (1-p_\beta)\,\delta\frac{(1-\eta_{in})\,k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right]\sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\delta_{k^{in},0} \\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}\left[\overline{N} - \sum_{\vec{j}|j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\right]\ . \end{aligned} \tag{B.19}$$

Now let us simplify the equations at the expense of more definitions, which are going to be helpful in the rest of the derivations. Let us denote the total population in all source nodes per all the nodes at time t by $\psi_{out}(t)$,

$$\psi_{out}(t) \equiv \sum_{\vec{k}|k^{in}=0} P_v\left(\vec{k}\right) N_{\vec{k}}(t) \ . \tag{B.20}$$

Similarly, we denote the total population in all sink nodes per all the nodes at time t by $\psi_{in}(t)$,

$$\psi_{in}(t) \equiv \sum_{\vec{k}|k^{out}=0} P_v\left(\vec{k}\right) N_{\vec{k}}(t) \ . \tag{B.21}$$

Substituting equations B.20 and B.21 into the rate equation of $N_{\vec{k}}$, we obtain

$$\begin{aligned} \partial_t N_{\vec{k}}(t) = &-\delta N_{\vec{k}}(t)\,\delta_{k^{out},0} - pN_{\vec{k}}(t)\left(1-\delta_{k^{out},0}\right) \\ &+\left(p_\beta\delta\frac{N_{\vec{k}}(t)}{\psi_{out}(t)} + (1-p_\beta)\,\delta\frac{(1-\eta_{in})\,k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right)\psi_{in}(t)\,\delta_{k^{in},0} \\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}\left(\overline{N}-\psi_{in}(t)\right) \ . \end{aligned} \tag{B.22}$$

It is hard to analyze the above closed system all at once. We thus consider the rate equation for each type of node separately. To clearly distinguish the results from this point onward, we adopt the following notation for the respective node degrees of each type of node: $\vec{k}^{(1)}$ for all source nodes ($k^{in}=0$), $\vec{k}^{(2)}$ for all transit nodes ($k^{in}\neq 0$ and $k^{out}\neq 0$), and $\vec{k}^{(3)}$ for all sink nodes ($k^{in}\neq 0$ and $k^{out}=0$).

Sink nodes

The rate equation of the average population size in a sink node with joint-degree $\vec{k}^{(3)} = \left(k^{in(3)},0\right)$ is

$$\partial_t N_{\vec{k}^{(3)}}(t) = -\delta N_{\vec{k}^{(3)}}(t) + p\frac{k^{in(3)}}{\left\langle k^{out}\right\rangle}\left(\overline{N}-\psi_{in}(t)\right) \ . \tag{B.23}$$

If we multiply both sides of the above equation by $P_v\left(\vec{k}^{(3)}\right)$ and then sum over all the joint-degrees $\vec{k}^{(1)} = \left(k^{in(3)}, 0\right)$, we obtain

$$\partial_t \psi_{in}(t) = -\left(\delta + \frac{\eta_{in} p}{1-\eta_{out}}\right)\psi_{in}(t) + \frac{\eta_{in} p}{1-\eta_{out}}\overline{N} \ , \tag{B.24}$$

where we have substituted in equation B.5. This first order differential equation can be solved as reported in B.1.3, yielding the equilibrium configuration ψ^*_{in},

$$\psi^*_{in} = \frac{\eta_{in} p}{(1-\eta_{out})\,\delta + \eta_{in} p}\overline{N} \ , \tag{B.25}$$

and a characteristic relaxation time of $\left(\delta + \frac{\eta_{in} p}{1-\eta_{out}}\right)^{-1}$. Using the solution for $\psi_{in}(t)$, we similarly solve the differential equation for $N_{\vec{k}^{(3)}}(t)$. The solution, in particular, leads to the equilibrium configuration $N^*_{\vec{k}^{(3)}}$,

$$N^*_{\vec{k}^{(3)}} = \frac{k^{in(3)}}{\langle k^{out}\rangle}\frac{(1-\eta_{out})\,p}{(1-\eta_{out})\,\delta + \eta_{in} p}\overline{N} \ , \tag{B.26}$$

and the characteristic relaxation time δ^{-1} (see section B.1.3).

Transit nodes

The rate equation of the average population size in a transit node subpopulation with joint-degree $\vec{k}^{(2)}$ follows very closely the rate equation for sink nodes,

$$\partial_t N_{\vec{k}^{(2)}}(t) = -p N_{\vec{k}^{(2)}}(t) + p\frac{k^{in(2)}}{\langle k^{out}\rangle}\left(\overline{N} - \psi_{in}(t)\right) \ , \tag{B.27}$$

whose solution is reported in section B.1.3. The solution of $N_{\vec{k}^{(2)}}(t)$ assumes the characteristic relaxation time $\max\left(p^{-1}, \left(\delta + \frac{\eta_{in} p}{1-\eta_{out}}\right)^{-1}\right)$ to the equilibrium configuration of the

populations of the transit nodes

$$N^*_{\vec{k}^{(2)}} = \frac{k^{in(2)}}{\langle k^{out}\rangle}\frac{(1-\eta_{out})\,\delta}{(1-\eta_{out})\,\delta+\eta_{in}p}\overline{N}\ . \tag{B.28}$$

Source nodes

The rate equation of the average population in a source node with joint-degree $\vec{k}^{(1)} = \left(0, k^{out(1)}\right)$ is

$$\partial_t N_{\vec{k}^{(1)}}(t) = -pN_{\vec{k}^{(1)}}(t) + \left(p_\beta\delta\frac{N_{\vec{k}^{(1)}}(t)}{\psi_{out}(t)} + (1-p_\beta)\,\delta\frac{(1-\eta_{in})\,k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\right)\psi_{in}(t)\ . \tag{B.29}$$

If we multiply both sides of the above equation by $P_v\left(\vec{k}^{(1)}\right)$ and then sum over all the joint-degrees $\vec{k}^{(1)} = \left(0, k^{out(1)}\right)$, we have

$$\partial_t\psi_{out}(t) = -p\psi_{out}(t) + \delta\psi_{in}(t)\ , \tag{B.30}$$

where we have substituted in equation B.8. As derived in subsection B.1.3, the solution $\psi_{out}(t)$ admits the equilibrium configuration

$$\psi^*_{out} = \frac{\delta}{p}\psi^*_{in} = \frac{\eta_{in}\delta}{(1-\eta_{out})\,\delta+\eta_{in}p}\overline{N}\ , \tag{B.31}$$

with a characteristic time scale of $\max\left(p^{-1}, \left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)^{-1}\right)$. Particular attention is devoted to the rate equation of the degree-block variable $N_{\vec{k}^{(1)}}(t)$ in that we consider three cases of p_β separately.

- If $p_\beta = 1$, meaning that we only consider birth, then the rate equation is given by

$$\partial_t N_{\vec{k}^{(1)}}(t) = \left(-p + \delta\frac{\psi_{in}(t)}{\psi_{out}(t)}\right)N_{\vec{k}^{(1)}}(t)\ . \tag{B.32}$$

The expression in brackets is equal to $\psi_{out}^{-1}\partial\psi_{out}$ (see equation B.30), yielding

$$\frac{\partial_t N_{\vec{k}^{(1)}}(t)}{N_{\vec{k}^{(1)}}(t)} = \frac{\partial\psi_{out}(t)}{\psi_{out}(t)} \ . \tag{B.33}$$

Using the definition of $\psi_{in}(t)$ in equation B.21, we obtain the solution

$$N_{\vec{k}^{(1)}}(t) = \frac{1}{\eta_{out}}\psi_{out}(t) \ . \tag{B.34}$$

This means that the characteristic relaxation time is $\max\left(p^{-1}, \left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)^{-1}\right)$ and the equilibrium configuration is

$$N^*_{\vec{k}^{(1)}} = \frac{1}{\eta_{out}}\psi^*_{out} = \frac{\eta_{in}\delta}{\eta_{out}\left[(1-\eta_{out})\delta + \eta_{in}p\right]}\overline{N} \ . \tag{B.35}$$

- Now let us look at the other extreme of $p_\beta = 0$, meaning that we only consider the importation process. In this case, the rate equation is

$$\partial_t N_{\vec{k}^{(1)}}(t) = -pN_{\vec{k}^{(1)}}(t) + \delta\frac{(1-\eta_{in})\,k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\psi_{in}(t) \ . \tag{B.36}$$

Using the solution for $\psi_{in}(t)$, we can solve the differential equation (see section B.1.3) which yields the equilibrium configuration

$$N^*_{\vec{k}^{(1)}} = \frac{(1-\eta_{in})\,\delta\psi^*_{in}}{\eta_{out}p}\frac{k^{out(1)}}{\langle k^{out}\rangle} = \frac{k^{out(1)}}{\langle k^{out}\rangle}\frac{\eta_{in}(1-\eta_{in})\,\delta}{\eta_{out}\left[(1-\eta_{out})\delta + \eta_{in}p\right]}\overline{N} \tag{B.37}$$

and the characteristic relaxation time $\max\left(p^{-1}, \left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)^{-1}\right)$.

- On the other hand, if $0 < p_\beta < 1$, then we have all the terms of the rate equation, that is, equation B.29. In this case, the differential equation does not seem to be easy

to solve. However, the equilibrium configuration can easily be evaluated, yielding

$$N^*_{\vec{k}^{(1)}} = \frac{(1-\eta_{in})\,\delta\psi^*_{in}}{\eta_{out}p}\frac{k^{out(1)}}{\langle k^{out}\rangle} = \frac{k^{out(1)}}{\langle k^{out}\rangle}\frac{\eta_{in}\,(1-\eta_{in})\,\delta}{\eta_{out}\left[(1-\eta_{out})\,\delta+\eta_{in}p\right]}\overline{N}\ , \tag{B.38}$$

that is exactly the same as the solution for the case of $p_\beta = 0$ (i.e., no birth).

B.1.3 Solutions of rate equations

In the following, we solve the differential equations for the average population sizes in subpopulations of different types of nodes: sink, transit, and source.

Sink nodes

Remember the rate equation $\partial_t\psi_{in}(t)$ in equation B.24 which expresses the change of the total number of individuals in sink nodes. The solution of this first order differential equation is

$$\psi_{in}(t) = e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t}\left(C_{in}+\overline{N}\frac{\eta_{in}p}{1-\eta_{out}}\int e^{\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t}dt\right)\ , \tag{B.39}$$

$$\psi_{in}(t) = C_{in}e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t} + \overline{N}\frac{\eta_{in}p}{(1-\eta_{out})\,\delta+\eta_{in}p}\ , \tag{B.40}$$

where C_{in} is a time-independent variable determined by initial conditions. If we assume that all the subpopulations have the same number of individuals initially, i.e., $N_{\vec{k}^{(1)}}(0) = N_{\vec{k}^{(2)}}(0) = N_{\vec{k}^{(3)}}(0) = \overline{N}$, then we have that $\psi_{in}(0) = \eta_{in}\overline{N}$, leading to

$$\psi_{in}(t) = \eta_{in}\overline{N}\frac{(1-\eta_{out})\,\delta-(1-\eta_{in})\,p}{(1-\eta_{out})\,\delta+\eta_{in}p}e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t} + \eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\,\delta+\eta_{in}p}\ . \tag{B.41}$$

Now we turn our attention to the rate equation of the joint-degree block variable $N_{\vec{k}^{(3)}}(t)$ in equation B.23. The solution is given by

$$
\begin{aligned}
N_{\vec{k}^{(3)}}(t) &= e^{-\delta t}\left(C_{\vec{k}^{(3)}} + p\frac{k^{in(3)}}{\langle k^{out}\rangle}\int e^{\delta t}\left(\overline{N} - \psi_{in}(t)\right)dt\right), \\
N_{\vec{k}^{(3)}}(t) &= C_{\vec{k}^{(3)}}e^{-\delta t} + (1-\eta_{out})\,\overline{N}\frac{p}{(1-\eta_{out})\,\delta + \eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle} \\
&+ (1-\eta_{out})\,\overline{N}\frac{(1-\eta_{out})\,\delta - (1-\eta_{in})\,p}{(1-\eta_{out})\,\delta + \eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t},
\end{aligned}
\tag{B.42}
$$

where $C_{\vec{k}^{(3)}}$ is a constant fixed by the initial condition $N_{\vec{k}^{(3)}}(0) = \overline{N}$. Inserting the initial condition yields

$$
\begin{aligned}
N_{\vec{k}^{(3)}}(t) &= \overline{N}\left(1 - (1-\eta_{out})\frac{k^{in(3)}}{\langle k^{out}\rangle}\right)e^{-\delta t} \\
&+ (1-\eta_{out})\,\overline{N}\frac{(1-\eta_{out})\,\delta - (1-\eta_{in})\,p}{(1-\eta_{out})\,\delta + \eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\
&+ (1-\eta_{out})\,\overline{N}\frac{p}{(1-\eta_{out})\,\delta + \eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle},
\end{aligned}
\tag{B.43}
$$

Transit nodes

The solution of the rate equation of the average population size in a transit node with joint-degree $\vec{k}^{(2)}$ in equation B.27 can be similarly solved by

$$
\begin{aligned}
N_{\vec{k}^{(2)}}(t) &= e^{-pt}\left(C_{\vec{k}^{(2)}} + p\frac{k^{in(2)}}{\langle k^{out}\rangle}\int e^{pt}\left(\overline{N} - \psi_{in}(t)\right)dt\right), \\
N_{\vec{k}^{(2)}}(t) &= C_{\vec{k}^{(2)}}e^{-pt} + (1-\eta_{out})\,\overline{N}\frac{\delta}{(1-\eta_{out})\,\delta + \eta_{in}p}\frac{k^{in(2)}}{\langle k^{out}\rangle} \\
&+\eta_{in}\overline{N}\frac{(1-\eta_{out})\,p\left[(1-\eta_{out})\,\delta - (1-\eta_{in})\,p\right]}{\left[(1-\eta_{out})\,\delta + \eta_{in}p\right]\left[(1-\eta_{out})\,\delta - (1-\eta_{out}-\eta_{in})\,p\right]}\frac{k^{in(2)}}{\langle k^{out}\rangle}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t},
\end{aligned}
\tag{B.44}
$$

where $C_{\vec{k}^{(2)}}$ is a constant fixed by the initial condition $N_{\vec{k}^{(2)}}(0) = \overline{N}$. Inserting the initial condition leads to

$$\begin{aligned} N_{\vec{k}^{(2)}}(t) = \overline{N}\left[1 - \frac{(1-\eta_{out})}{(1-\eta_{out})\delta + \eta_{in}p}\left(\delta + \eta_{in}p\frac{(1-\eta_{out})\delta - (1-\eta_{in})p}{(1-\eta_{out})\delta - (1-\eta_{out}-\eta_{in})p}\right)\frac{k^{in(2)}}{\langle k^{out}\rangle}\right]e^{-pt} \\ +\eta_{in}\overline{N}\frac{(1-\eta_{out})p\left[(1-\eta_{out})\delta - (1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta + \eta_{in}p\right]\left[(1-\eta_{out})\delta - (1-\eta_{out}-\eta_{in})p\right]}\frac{k^{in(2)}}{\langle k^{out}\rangle}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\ +(1-\eta_{out})\overline{N}\frac{\delta}{(1-\eta_{out})\delta + \eta_{in}p}\frac{k^{in(2)}}{\langle k^{out}\rangle}\ , \end{aligned} \tag{B.45}$$

Source nodes

Recall the rate equation in equation B.30 that determines the change of the total population size in the source nodes. This first order differential equation can easily be solved by

$$\begin{aligned} \psi_{out}(t) &= e^{-pt}\left(C_{out} + \delta\int e^{pt}\psi_{in}(t)\,dt\right)\ , \\ \psi_{out}(t) &= C_{out}e^{-pt} + \delta e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta - (1-\eta_{in})p}{(1-\eta_{out})\delta + \eta_{in}p}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t}dt \\ &\quad +\delta e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\delta + \eta_{in}p}dt \\ \psi_{out}(t) &= C_{out}e^{-pt} - \eta_{in}\overline{N}\frac{(1-\eta_{out})\delta\left[(1-\eta_{out})\delta - (1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta - (1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta + \eta_{in}p\right]}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\ &\quad +\eta_{in}\overline{N}\frac{\delta}{(1-\eta_{out})\delta + \eta_{in}p}\ , \end{aligned} \tag{B.46}$$

where the constant C_{out} is given by the initial condition $\psi_{out}(0) = \eta_{out}\overline{N}$, yielding

$$\begin{aligned} \psi_{out}(t) = \eta_{out}\overline{N}\frac{(1-\eta_{out}-\eta_{in})(\delta - p)}{(1-\eta_{out})\delta - (1-\eta_{out}-\eta_{in})p}e^{-pt} \\ -\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta\left[(1-\eta_{out})\delta - (1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta - (1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta + \eta_{in}p\right]}e^{-\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\ +\eta_{in}\overline{N}\frac{\delta}{(1-\eta_{out})\delta + \eta_{in}p}\ . \end{aligned} \tag{B.47}$$

Now we focus on the degree-block variable $N_{\vec{k}^{(1)}}(t)$. In the following, we consider two cases of p_β:

- $p_\beta = 1$. In this case, the rate equation leads to the solution $N_{\vec{k}^{(1)}}(t) = \eta_{out}^{-1}\psi_{out}(t)$.
- $p_\beta = 0$. In this case, the rate equation is given by equation B.36 and the solution follows as

$$N_{\vec{k}^{(1)}}(t) = e^{-pt}\left(C_{\vec{k}^{(1)}} + \delta\frac{(1-\eta_{in})\,k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\int e^{pt}\psi_{in}(t)\,dt\right), \tag{B.48}$$

$$\begin{aligned} N_{\vec{k}^{(1)}}(t) = {} & C_{\vec{k}^{(1)}}e^{-pt} + \delta\frac{(1-\eta_{in})\,k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{(1-\eta_{out})\,\delta-(1-\eta_{in})\,p}{(1-\eta_{out})\,\delta+\eta_{in}p}e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t}dt \\ & +\delta\frac{(1-\eta_{in})\,k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\,\delta+\eta_{in}p}dt \\ N_{\vec{k}^{(1)}}(t) = {} & C_{\vec{k}^{(1)}}e^{-pt} \\ & -\eta_{in}\overline{N}\frac{(1-\eta_{in})(1-\eta_{out})\,\delta\left[(1-\eta_{out})\,\delta-(1-\eta_{in})\,p\right]}{\eta_{out}\left[(1-\eta_{out})\,\delta-(1-\eta_{out}-\eta_{in})\,p\right]\left[(1-\eta_{out})\,\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\ & +\eta_{in}\overline{N}\frac{(1-\eta_{in})\,\delta}{\eta_{out}\left[(1-\eta_{out})\,\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}, \end{aligned} \tag{B.49}$$

where $C_{\vec{k}^{(1)}}$ is a constant set by the initial condition $N_{\vec{k}^{(1)}}(0) = \overline{N}$. Using the initial value, we obtain

$$\begin{aligned} N_{\vec{k}^{(1)}}(t) = {} & \left(1-\overline{N}\frac{\eta_{in}(1-\eta_{in})\,\delta}{(1-\eta_{out})\,\delta-(1-\eta_{out}-\eta_{in})\,p}\frac{k^{out(1)}}{\langle k^{out}\rangle}\right)e^{-pt} \\ & -\eta_{in}\overline{N}\frac{(1-\eta_{in})(1-\eta_{out})\,\delta\left[(1-\eta_{out})\,\delta-(1-\eta_{in})\,p\right]}{\eta_{out}\left[(1-\eta_{out})\,\delta-(1-\eta_{out}-\eta_{in})\,p\right]\left[(1-\eta_{out})\,\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}e^{-\left(\delta+\frac{\eta_{in}p}{1-\eta_{out}}\right)t} \\ & +\eta_{in}\overline{N}\frac{(1-\eta_{in})\,\delta}{\eta_{out}\left[(1-\eta_{out})\,\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}. \end{aligned} \tag{B.50}$$

B.2 Derivation of global invasion thresholds

In this section, we present the derivations of the global epidemic invasion thresholds, their respective critical movement rates, and the relationship between the two thresholds as seen

through a comparison of the critical movement rates.

B.2.1 Global invasion thresholds

We consider the n^{th} generation number of diseased subpopulations with degree $\vec{k}^{(x)}$ from node class x, $D^n_{\vec{k}^{(x)}}$, as a function of the three sets $\{D^{n-1}_{\vec{k}^{(1)}}\}$, $\{D^{n-1}_{\vec{k}^{(2)}}\}$, and $\{D^{n-1}_{\vec{k}^{(3)}}\}$ of the $(n-1)^{\text{th}}$ generation. The "infection" of a node occurs when infected cattle move from one node into another node containing only susceptible individuals. The expression for the branching process as expressed by[7,17] follows as

$$D^n_{\vec{k}} = \sum_{\vec{j}} D^{n-1}_{\vec{j}} P_a\left(\vec{j},\vec{k} \mid \vec{j}\right) j^{\text{out}} p\left(\vec{j},\vec{k}\right) \prod_{m=0}^{n-1}\left(1 - \frac{D^m_{\vec{k}}}{V_{\vec{k}}}\right) \tag{B.51}$$

This branching equation models the n^{th} generation of newly infected nodes by considering that each subpopulation in a node with a (out-) degree of j has the potential to infect j other nodes by the intersection of three events. These events are that a node with degree $\vec{k}$ exists in the out neighborhood of a node with degree $\vec{j}$, $P_a\left(\vec{j},\vec{k} \mid \vec{j}\right)$, that the neighbor with degree k has not been infected in previous generations, $\prod_{m=0}^{n-1}\left(1 - \frac{D^m_{\vec{k}}}{V_{\vec{k}}}\right)$, and that the disease will spread from the node with degree j to the node with degree k, $p\left(\vec{j},\vec{k}\right)$. For the classic SIR model[84], $p\left(\vec{j},\vec{k}\right)$ is given by

$$p\left(\vec{j},\vec{k}\right) = 1 - R_0^{-\lambda_{\vec{j}\vec{k}}}\ , \tag{B.52}$$

where $\lambda_{\vec{j}\vec{k}}$ is the number of infected individuals moving from the infected subpopulation to the fully susceptible subpopulation. The classic SIR model solves for

$$\lambda_{\vec{j}\vec{k}} = \frac{\alpha}{\mu} N_{\vec{j}} d_{\vec{j}\vec{k}}\ , \tag{B.53}$$

where α is the final size of an SIR epidemic. Furthermore, the probability $p\left(\vec{j},\vec{k}\right)$ can be approximated as $p\left(\vec{j},\vec{k}\right) \simeq \alpha N_{\vec{j}} d_{\vec{j}\vec{k}} \left(R_0 - 1\right)/\mu \simeq 2\left(1 - R_0^{-1}\right)^2 N_{\vec{j}} d_{\vec{j}\vec{k}}/\mu$ if we assume $R_0 \simeq 1$. The assumption that we are considering an epidemic in its early stages allows the approximation $\prod_{m=0}^{n-1}\left(1 - \frac{D_{\vec{k}}^m}{V_{\vec{k}}}\right) \simeq 1$ and yields the approximated branching model

$$D^n = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}} D_{\vec{j}}^{n-1} P_a\left(\vec{j},\vec{k} \mid \vec{j}\right) j^{\text{out}} N_{\vec{j}} d_{\vec{j}\vec{k}} \ . \tag{B.54}$$

Note that this approximation of the process ignores bidirectional arcs and also has an assumption that the next generation is only infected by the immediately previous generation and none prior. This second assumption will remove the impact of the initial condition of the sink and sources nodes. Let us expand this expression to correctly consider the 3 classes of nodes. The lack of in-degrees for the source nodes and out-degrees for the sink nodes yield three expressions for the source and sink branching model on directed networks,

$$D_{\vec{k}^{(1)}}^n = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \left[\sum_{\vec{j}^{(1)}} D_{\vec{j}^{(1)}}^{n-1} 0 j^{\text{out}(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(1)}} + \sum_{\vec{j}^{(2)}} D_{\vec{j}^{(2)}}^{n-1} 0 j^{\text{out}(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(1)}}\right] = 0, \ \ n > 0 \ , \tag{B.55}$$

$$\begin{aligned} D_{\vec{k}^{(2)}}^n = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \Bigg[&\sum_{\vec{j}^{(1)}} D_{\vec{j}^{(1)}}^{n-1} P_a\left(\vec{j}^{(1)},\vec{k}^{(2)} \mid \vec{j}^{(1)}\right) j^{\text{out}(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(2)}} \\ &+ \sum_{\vec{j}^{(2)}} D_{\vec{j}^{(2)}}^{n-1} P_a\left(\vec{j}^{(2)},\vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{\text{out}(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(2)}}\Bigg], \ \ n > 0 \ , \end{aligned} \tag{B.56}$$

$$\begin{aligned} D_{\vec{k}^{(3)}}^n = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \Bigg[&\sum_{\vec{j}^{(1)}} D_{\vec{j}^{(1)}}^{n-1} P_a\left(\vec{j}^{(1)},\vec{k}^{(3)} \mid \vec{j}^{(1)}\right) j^{\text{out}(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(3)}} \\ &+ \sum_{\vec{j}^{(2)}} D_{\vec{j}^{(2)}}^{n-1} P_a\left(\vec{j}^{(2)},\vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{\text{out}(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(3)}}\Bigg], \ \ n > 0 \ . \end{aligned} \tag{B.57}$$

From equation B.55, when $n > 0$ we have that $D^n_{k_1} = 0$, and equations B.56 and B.57 simplify to expressions that only consider infections arriving from the transit class.

$$D^n_{\vec{k}^{(2)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{\text{out}(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(2)}}, \quad n > 1 \tag{B.58}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{\text{out}(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(3)}}, \quad n > 1 \tag{B.59}$$

With the traffic functions as mentioned above, these simplify to

$$D^n_{\vec{k}^{(2)}} = \frac{2p}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) N_{\vec{j}^{(2)}}, \quad n > 1\ , \tag{B.60}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2p}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) N_{\vec{j}^{(2)}}, \quad n > 1\ . \tag{B.61}$$

Inserting the quasi-equilibrium populations, $N_{\vec{j}^{(2)}}$, as derived in appendix B.1 yields

$$D^n_{\vec{k}^{(2)}} = \frac{2p\left(1 - \eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1 - \eta_{out}\right)\delta + \eta_{in}p\right]\langle k^{in}\rangle}\left(1 - \frac{1}{R_0}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{in(2)}, \quad n > 1\ , \tag{B.62}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2p\left(1 - \eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1 - \eta_{out}\right)\delta + \eta_{in}p\right]\langle k^{in}\rangle}\left(1 - \frac{1}{R_0}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{in(2)}, \quad n > 1\ . \tag{B.63}$$

As we are considering uncorrelated networks here,

$$P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) = P^{in}_a\left(\vec{k}^{(2)}\right) = \frac{k^{in(2)} P_v\left(\vec{k}^{(2)}\right)}{\langle k^{in}\rangle}\ , \tag{B.64}$$

$$P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) = P^{in}_a\left(\vec{k}^{(3)}\right) = \frac{k^{in(3)} P_v\left(\vec{k}^{(3)}\right)}{\langle k^{in}\rangle}\ . \tag{B.65}$$

Inserting equations B.64 and B.65 simplifies the branching processes to

$$D^n_{\vec{k}^{(2)}} = \frac{2p\left(1-\eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 P_v\left(\vec{k}^{(2)}\right)k^{in(2)}\sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} j^{in(2)}, \quad n>1\ , \tag{B.66}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2p\left(1-\eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 P_v\left(\vec{k}^{(3)}\right)k^{in(3)}\sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} j^{in(2)}, \quad n>1\ . \tag{B.67}$$

Let us define, similar to the undirected analysis of[17],

$$\Theta^n_2 = \sum_{\vec{j}^{(2)}} D^n_{\vec{j}^{(2)}} j^{in(2)} \text{ and } \Theta^n_3 = \sum_{\vec{j}^{(3)}} D^n_{\vec{j}^{(3)}} j^{in(3)}\ . \tag{B.68}$$

Combining these definition with equations B.66 and B.67, we have

$$\Theta^n_2 = \sum_{\vec{j}^{(2)}} D^n_{\vec{j}^{(2)}} j^{in(2)} = \frac{2p\left(1-\eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 \sum_{\vec{k}^{(2)}} P_v\left(\vec{k}^{(2)}\right)\left(k^{in(2)}\right)^2\Theta^{n-1}_2, \quad n>1 \tag{B.69}$$

$$\Theta^n_3 = \sum_{\vec{j}^{(3)}} D^n_{\vec{j}^{(3)}} j^{in(3)} = \frac{2p\left(1-\eta_{out}\right)\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 \sum_{\vec{k}^{(3)}} P_v\left(\vec{k}^{(3)}\right)\left(k^{in(3)}\right)^2\Theta^{n-1}_2, \quad n>1\ . \tag{B.70}$$

Noting that the node-type degree distributions can be expressed as $P_v\left(\vec{k}^{(2)}\right) = \left(1-\eta_{out}-\eta_{in}\right)P_{v^{(2)}}\left(\vec{k}^{(2)}\right)$ and $P_v\left(\vec{k}^{(3)}\right) = \eta_{in}P_{v^{(3)}}\left(\vec{k}^{(3)}\right)$; we derive, from the above equations, the global epidemic invasion threshold, R_*, and the transit-to-sink invasion threshold, R^{TS}_*, respectively as

$$R_* = \frac{\Theta^n_2}{\Theta^{n-1}_2} = \frac{2p\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}\ , \tag{B.71}$$

$$R^{TS}_* = \frac{\Theta^n_3}{\Theta^{n-1}_2} = \frac{2p\left(1-\eta_{out}\right)\eta_{in}\delta\overline{N}}{\mu\left[\left(1-\eta_{out}\right)\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(3)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}\ . \tag{B.72}$$

As seen in equation B.72, the disease process in the sink nodes is driven by the disease process in the transit nodes. This is only the uncorrelated case, with uniform movement

rates, and further work could extend this analysis to other types of networks and mobility patterns. We proceed next to consider the critical movement rates p_c, p_c^{TS} that define the tipping points for these invasion thresholds.

B.2.2 Critical movement rates

Here we extract the critical movement rates that are defined by $R_*(p_c) = 1$ and $R_*^{TS}\left(p_c^{TS}\right) = 1$. These critical rates express the smallest movement rates necessary for the disease to spread among the transit nodes and from the transit nodes to the sink nodes, respectively, as

$$p_c = \frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle - \eta_{in}\mu\left\langle k^{in}\right\rangle^2}\ , \tag{B.73}$$

$$p_c^{TS} = \frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle - \eta_{in}\mu\left\langle k^{in}\right\rangle^2}\ . \tag{B.74}$$

Of significant interest is the relationship between these two critical movement probabilities. If the system and disease outbreak are such that $p_c > p_c^{TS}$, then the disease process will move through both transit and sink nodes easily when the individuals' movement rate exceeds p_c. However, if the reverse is true, three regions of movement rates will be defined wherein a second outbreak situation arises where the disease may persist within the transit nodes, but

not reach the sink nodes. The relationship is then derived as follows.

$$
\begin{gathered}
p_c^{TS} > p_c \\
\Downarrow \\
\frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2} \\
> \frac{\mu\delta\left(1-\eta_{out}\right)\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2} \\
\Downarrow \\
2\delta\overline{N}\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2 \\
> 2\delta\overline{N}\left(1-\eta_{out}\right)\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2 \\
\Downarrow \\
2\delta\overline{N}\left(1-\eta_{out}\right)\left(1-\eta_{out}-\eta_{in}\right)\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle \\
> 2\delta\overline{N}\left(1-\eta_{out}\right)\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle \\
\Downarrow \\
\left(1-\eta_{out}-\eta_{in}\right)\left\langle\left(k^{in(2)}\right)^2\right\rangle > \eta_{in}\left\langle\left(k^{in(3)}\right)^2\right\rangle \\
\Downarrow \\
\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle\left(k^{in(3)}\right)^2\right\rangle} > \frac{\eta_{in}}{\left(1-\eta_{out}-\eta_{in}\right)} \iff p_c^{TS} > p_c \ . \qquad \text{(B.75)}
\end{gathered}
$$

Note that, in this derivation, from the second line to the third, it is assumed that the denominators are positive. If the denominators were not strictly positive, the case would be that the critical rates would be either undefined or negative values, and the respective thresholds would always be greater than unity. We limited our derivation to consider only

the case of positive denominators in this step. The comparison concludes that $p_c^{TS} > p_c$ if and only if the ratio of the second moments of the node-type in-degree distributions (transit nodes over sink nodes) is greater than the ratio of the sink nodes to the transit nodes.

B.3 Stochastic simulation processes

Here we present the details of the numerical implementation of the dynamical processes described in the main text. We denote the number of Susceptible, Infected, and Recovered individuals in node i at time t respectively by $S_i(t)$, $I_i(t)$, and $R_i(t)$. The numbers of individuals in each of these populations is varied by discrete and stochastic dynamics describing the demographic, diffusion, and disease processes. By definition, the number of individuals in the population of node i at time t is given by $N_i(t) = S_i(t) + I_i(t) + R_i(t)$.

B.3.1 Disease dynamics

Susceptible individuals in the population of node i at time t receive infections from any infected individuals also present in node i during the time interval Δt with a probability $p_{S_i \to I_i}$ defined as

$$p_{S_i \to I_i} = \lambda_i(t) \Delta t \ , \tag{B.76}$$

where $\lambda_i(t)$ is the per capita force of infection $\lambda_i(t) = \beta I_i(t) / N_i(t)$. The number of individuals $\mathcal{Q}_i(S, I)$ transferring from the susceptible state $S_i(t)$ to the infected state $I_i(t)$ at time t in node i is then extracted from the binomial distribution as

$$\mathcal{Q}_i(S, I) = \Pr^{\text{Binom}}(S_i(t), p_{S_i \to I_i} \Delta t) \ . \tag{B.77}$$

The number of individuals $\mathcal{Q}_i(I,R)$ transferring from the infected state $I_i(t)$ to the recovered state $R_i(t)$ at time t in node i is similarly extracted from the binomial distribution

$$\mathcal{Q}_i(I,R) = \Pr^{\text{Binom}}\left(I_i(t), p_{I_i \to R_i}\Delta t\right) \ , \tag{B.78}$$

where $p_{I_i \to R_i}$ is the per capita recovery probability, $p_{I_i \to R_i} = \mu \Delta t$, within the time interval Δt. The classical SIR model contains only these two transitions. After extracting $\mathcal{Q}_i(S,R)$ and $\mathcal{Q}_i(I,R)$ for all nodes, we update the subpopulations of each node i as

$$\begin{aligned} S_i &\to S_i - \mathcal{Q}_i(S,R) \ , \\ I_i &\to I_i + \mathcal{Q}_i(S,R) - \mathcal{Q}_i(I,R) \ , \\ R_i &\to R_i + \mathcal{Q}_i(I,R) \ . \end{aligned} \tag{B.79}$$

After each iteration of the disease process, we similarly update the population dynamics.

B.3.2 Population dynamics

The number of individuals $N_i(t)$ in subpopulation i at time t is subject to changes in the following time interval Δt due to death, birth, importation, and migration. These events are assumed to occur in the following order:

1. *Death.* Each individual in node i dies at rate δ_i. The number of individuals who die $\mathcal{D}_i = \mathcal{D}_i^S + \mathcal{D}_i^I + \mathcal{D}_i^R$ in subpopulation i within this time interval is a random integer number extracted from three binomial distributions

$$\mathcal{D}_i = \Pr^{\text{Binom}}\left(S_i(t), \delta_i \Delta t\right) + \Pr^{\text{Binom}}\left(I_i(t), \delta_i \Delta t\right) + \Pr^{\text{Binom}}\left(R_i(t), \delta_i \Delta t\right) \ . \tag{B.80}$$

After the extraction of the random numbers $\{\mathcal{D}_i\}$, we update the population sizes by

$$S_i \to S_i - \mathcal{D}_i^S;\; I_i \to I_i - \mathcal{D}_i^I;\; R_i \to R_i - \mathcal{D}_i^R \;. \tag{B.81}$$

The total number of individuals who are dropped out of the system due to death is then given by

$$\mathcal{D} = \sum_i \mathcal{D}_i \;. \tag{B.82}$$

Note that the death rate δ_i is subpopulation dependent and will be assumed to be equal to δ for all subpopulations of sink nodes and 0 otherwise.

2. *Birth and Importation.* We reintroduce the total number of lost individuals $\mathcal{D}$ into the system in order to keep the total population invariant over time. We consider that a fraction p_β of these deaths are introduced back into the system as births and the rest as importations. The total number of births $\mathcal{B}$ is an integer random number extracted from the binomial distribution

$$\mathcal{B} = \Pr^{\text{Binom}}(\mathcal{D}, p_\beta) \;. \tag{B.83}$$

After extracting $\mathcal{B}$, we calculate the number of imported individuals $\mathcal{I}$ by $\mathcal{I} = \mathcal{D} - \mathcal{B}$. Once we compute $\mathcal{B}$ and $\mathcal{I}$, we distribute these individuals in the system as follows:

- *Births.* We assume that node i gets a newly born individual with probability p_{β_i} proportional to its current population:

$$p_{\beta_i} = \frac{r_{\beta_i} N_i(t)}{\sum_h r_{\beta_h} N_h(t)} \;, \tag{B.84}$$

where r_{β_i} is going to take values of 0 or 1 and is introduced for normalization purposes. For now, we will assume that $r_{\beta_i} = 1$ for all the source nodes and 0 for the rest of the nodes. The number of newly born individuals $\mathcal{B}_i$ in node i is

extracted from the multinomial distribution

$$\Pr^{\text{Multinom}}\left(\mathcal{B}, \{p_{\beta_i}\}\right) \ . \tag{B.85}$$

- *Importation.* Each node i receives an imported case with probability $p_{\mathcal{I}_i}$ proportional to its out-degree:

$$p_{\mathcal{I}_i} = \frac{r_{\mathcal{I}_i} k_i^{out}}{\sum_h r_{\mathcal{I}_h} l_h^{out}} \ , \tag{B.86}$$

  where $r_{\mathcal{I}_i}$ is introduced for the purpose of normalization, similar to r_{β_i}, and will take values of 1 for all the source nodes and 0 otherwise. Then, the total number of importations for each (source) subpopulation is given by an integer random number extracted from the multinomial distribution

$$\Pr^{\text{Multinom}}\left(\mathcal{I}, \{p_{\mathcal{I}_i}\}\right) \ . \tag{B.87}$$

After the extraction of the sets $\{\mathcal{B}_i\}$ and $\{\mathcal{I}_i\}$, the source node populations are updated with the addition of susceptible cattle by

$$S_i \to S_i + \mathcal{B}_i + \mathcal{I}_i \ . \tag{B.88}$$

In this implementation, we have assumed that imported individuals are not entering with the system with any infection. This permits the study of a single source of the outbreak, but in general this may not be an accurate representation of a livestock system.

3. *Migration.* Outward migration occurs from both transit and source node subpopulations as they have non-zero out-degrees. Individuals in node i migrate to subpopulation

j in the out-neighborhood v_i^{out} at per capita rate d_{ij}:

$$d_{ij} = \frac{p}{k_i^{out}} \ , \ k^{out} \neq 0 \tag{B.89}$$

yielding the probability of migration pij within the time interval Δt, $p_{ij} = d_{ij}\Delta t$. The number of individuals $\mathcal{M}_{ij}^X$ leaving node i and arriving to node j among each disease state X is an integer random number extracted from the multinomial distributions

$$\Pr^{\text{Multinom}}\left(X_i, \{p_{ij} | j \in v_i^{out}\}\right) \ . \tag{B.90}$$

After the extraction of all the integer numbers $\{\mathcal{M}_{ij}^X\}$, the population sizes of each subpopulation is updated according to

$$\begin{aligned} S_i &\to S_i + \sum_j \left[\mathcal{M}_{ij}^S - \mathcal{M}_{ji}^S\right] \ , \\ I_i &\to I_i + \sum_j \left[\mathcal{M}_{ij}^I - \mathcal{M}_{ji}^I\right] \ , \\ R_i &\to R_i + \sum_j \left[\mathcal{M}_{ij}^R - \mathcal{M}_{ji}^R\right] \ . \end{aligned} \tag{B.91}$$

Appendix C

Dance of the Calves: Formulations and Results

The contents of this appendix serve to supplement the work of chapter 4.

C.1 Data Estimation

Following the discussion of section 4.2, we formulated a pair of maximum entropy estimation problems that are constrained by straightforward rules defined by the US Department of Agriculture database. The maximum entropy method does not recover a significant amount of diversity with the estimated values, but selects the sets of values that are as homogeneous as possible while obeying all constraints. This method seeks such values because the homogeneity of maximum entropy represents the minimal assumptions possible in the problem design.

C.1.1 Population data estimation

With the following formulations, we estimate any undisclosed data elements, namely subpopulations $Pop^x_{t,c,i}$, County totals $Tc^x_{t,c}$, and State totals for each size category $Tz^x_{t,i}$. We

defined the sets $County$ as the Counties of the considered State, $t \in Type_A = \{Dairy,$ $Preslaughter$, $All\ Cattle\}$, $i \in Size_A = \{z1\_9,\ z10\_19,\ z20\_49,\ z50\_99,\ z100\_199,\ z200\_499,$ $z500\_up\}$, and $j \in Size_B = \{z1\_19,\ z20\_199,\ z200\_up\}$. From the USDA NASS database, for each State, we collected the numbers of operations $n_{t,c,i}$ of cattle type t having subpopulations with their sizes falling within size range i in county c, the published populations counts $Pop^r_{t,c,i}$ representing all cattle of type t with subpopulations within size range i in county c, the total numbers of cattle $Tc^r_{t,c}$ of type t in county c, the total numbers of cattle $Tz^r_{t,i}$ of type t in size range i in the State, and the total counts of cattle in each type t for the entire State P^{tot}_t. To these parameters, we have added upper limits on the subpopulations $u_{t,i}$ for each cattle type t and size range i, lower limits on the subpopulations $l_{t,i}$ for each cattle type t and size range i, and data indicator parameters, $datPop_{t,c,i}$, $datTc_{t,c}$, and $datTz_{t,i}$, that express the existence of data elements for their respective parameters (The data indicators for non-disclosed data elements are assigned a value of 0 and the remainder are set to 1). The lower and upper limits are set by the limits of the size ranges ($l_{t,z1\_9} = 1$, $u_{t,z1\_9} = 9$, ...), with the exception of the upper limits on the largest size range, $z500\_up$. The largest upper limit $u_{t,z500\_up}$ is set to $Tz^r_{t,z500\_up}$ if $datTz_{t,z500\_up} = 1$, else we set $u_{t,z500\_up} = P^{\text{tot}}_t$. For the population data of each State, we solve

$$\text{Maximize} \sum_{t \in Type_A} \sum_{c \in County} \sum_{i \in Size_A} \left(-\frac{Pop^x_{t,c,i}}{P^{\text{tot}}_t} \log \left[\frac{Pop^x_{t,c,i}}{P^{\text{tot}}_t} + 1.0 - datPop_{t,c,i} \text{sign}\left(Pop^r_{t,c,i}\right) \right] \right) . \tag{C.1}$$

The objective function maximizes the entropy of the three (by cattle type) population distributions and includes additional terms in the logarithm argument similar to the objective function of the formulation of section 4.3. We include the $sign$ function to consider the case when the published data $Pop^r_{t,c,i} = 0$, which would otherwise result in $log\,(0)$. Thus, the 1.0 is removed from the logarithm argument if and only if $datPop_{t,c,i} = 1$ and the corresponding data value $Pop^r_{t,c,i}$ is strictly positive.

Subject to

Constraints on Population values

$$Pop^{x}_{t,c,i} \leq datPop_{t,c,i} Pop^{r}_{t,c,i} + (1.0 - datPop_{t,c,i})\, n_{t,c,i} u_{t,i} \;\forall\,(t,c,i) \tag{C.2}$$

$$Pop^{x}_{t,c,i} \geq datPop_{t,c,i} Pop^{r}_{t,c,i} + (1.0 - datPop_{t,c,i})\, n_{t,c,i} l_{t,i} \;\forall\,(t,c,i) \tag{C.3}$$

The constraints form bounds for the complete set $\{Pop^{x}_{t,c,i}\}$ even if the data element $Pop^{r}_{t,c,i}$ is known from the NASS database. We use $datPop_{t,c,i}$ to represent the existence of the data element. Observe that, when $datPop_{t,c,i} = 1$, inequalities C.2 and C.3 converge to act as an equality constraint $Pop^{x}_{t,c,i} = Pop^{r}_{t,c,i}$. When $datPop_{t,c,i} = 0$, the upper (lower) bound is defined by the number of subpopulations in $Pop^{x}_{t,c,i}$ and the upper (lower) limit of the size range. Although these bounds have the potential to be very loose, they help shape the feasible set of values for $Pop^{x}_{t,c,i}$.

Constraints on County totals

$$Tc^{x}_{t,c} \leq datTc_{t,c} Tc^{r}_{t,c} + (1.0 - datTc_{t,c})\, uTc_{t,c} \;\forall\,(t,c) \tag{C.4}$$

$$Tc^{x}_{t,c} \geq datTc_{t,c} Tc^{r}_{t,c} + (1.0 - datTc_{t,c})\, lTc_{t,c} \;\forall\,(t,c) \tag{C.5}$$

$$uTc_{t,c} = \sum_{i \in Size_A} \left[datPop_{t,c,i} Pop^{r}_{t,c,i} + (1.0 - datPop_{t,c,i})\, n_{t,c,i} u_{t,i} \right] \;\forall\,(t,c) \tag{C.6}$$

$$lTc_{t,c} = \sum_{i \in Size_A} \left[datPop_{t,c,i} Pop^{r}_{t,c,i} + (1.0 - datPop_{t,c,i})\, n_{t,c,i} l_{t,i} \right] \;\forall\,(t,c) \tag{C.7}$$

$$\sum_{i \in Size_A} Pop^x_{t,c,i} = Tc^x_{t,c} \; \forall \, (t, c) \tag{C.8}$$

Inequalities C.4 and C.5 follow the same data indicator controlled constraint structure as inequalities C.2 and C.3. The upper and lower limits for the county totals $Tc^x_{t,c}$ are constructed in equality constraints C.6 and C.7. Given constraints C.2, C.3, and C.8, the bounds computed in C.6 and C.7 are redundant; however, we included them to help describe the solution space.

Constraints on Size totals

$$Tz^x_{t,i} \leq datTz_{t,i} Tz^r_{t,i} + (1.0 - datTz_{t,i}) \, uTz_{t,i} \; \forall \, (t, i) \tag{C.9}$$

$$Tz^x_{t,i} \geq datTz_{t,i} Tz^r_{t,i} + (1.0 - datTz_{t,i}) \, lTz_{t,i} \; \forall \, (t, i) \tag{C.10}$$

$$uTz_{t,i} = \sum_{c \in County} \left[datPop_{t,c,i} Pop^r_{t,c,i} + (1.0 - datPop_{t,c,i}) \, n_{t,c,i} u_{t,i} \right] \; \forall \, (t, i) \tag{C.11}$$

$$lTz_{t,i} = \sum_{c \in County} \left[datPop_{t,c,i} Pop^r_{t,c,i} + (1.0 - datPop_{t,c,i}) \, n_{t,c,i} l_{t,i} \right] \; \forall \, (t, i) \tag{C.12}$$

$$\sum_{c \in County} Pop^x_{t,c,i} = Tz^x_{t,i} \; \forall \, (t, i) \tag{C.13}$$

Constraints C.9 through C.13 repeat the structure of constraints C.4 through C.8 for the totals $Tz^x_{t,i}$ of each size category.

Constraints on State totals

$$\sum_{c \in County} Tc^{x}_{t,c} = P^{\text{tot}}_{t} \; \forall\,(t) \tag{C.14}$$

$$\sum_{i \in Size_A} Tz^{x}_{t,i} = P^{\text{tot}}_{t} \; \forall\,(t) \tag{C.15}$$

Equality constraints C.14 and C.15 state that the totals of the sub-totals must equal the State total P^{tot}_{t} of the respective population type.

Constraints on Population relations

$$Pop^{R}_{t,c,z1\_19} = Pop^{x}_{t,c,z1\_9} + Pop^{x}_{t,c,z10\_19} \; \forall\,(t) \tag{C.16}$$

$$Pop^{R}_{t,c,z20\_199} = Pop^{x}_{t,c,z20\_49} + Pop^{x}_{t,c,z50\_99} + Pop^{x}_{t,c,z100\_199} \; \forall\,(t) \tag{C.17}$$

$$Pop^{R}_{t,c,z200\_\text{up}} = Pop^{x}_{t,c,z200\_499} + Pop^{x}_{t,c,z500\_up} \; \forall\,(t) \tag{C.18}$$

$$Pop^{R}_{Dairy,c,j} + Pop^{R}_{Preslaughter,c,j} \leq Pop^{R}_{AllCattle,c,j} \; \forall\,(c \in County, j \in Size_B) \tag{C.19}$$

Constraints C.16-C.18 aggregate the populations from the size ranges of $Size_A$ to those of $Size_B$. Inequality C.19 defines the relationship discussed in section 4.2 and enables us to assemble the third cattle type of $Type_B$, $Beef$, as $Pop^{R}_{Beef,c,j} = Pop^{R}_{AllCattle,c,j} - Pop^{R}_{Dairy,c,j} - Pop^{R}_{Preslaughter,c,j}$ for every county c and $j \in Type_B$.

C.1.2 Sales and shipments data estimation

For Sales, Movements, and Slaughter data, we consider another a set of shipment types $ShipType = \{All\ Shipments, Slaughter\}$. We formulated the parallel problem for the non-disclosed data elements which describe the movements and slaughter of cattle. To estimate shipments $Sales^{x}_{q,c,i}$, County totals $Tc^{(s),x}_{q,c}$, and State totals for each size category $Tz^{(s),x}_{q,i}$ we solve

$$\text{Maximize} \sum_{q \in ShipType} \sum_{c \in County} \sum_{i \in Size_A} \left(-\frac{Sales^{x}_{q,c,i}}{S^{\text{tot}}_{q}} \log \left[\frac{Sales^{x}_{q,c,i}}{S^{\text{tot}}_{q}} + 1.0 - datSales_{q,c,i} \text{sign}\left(Sales^{r}_{q,c,i}\right) \right] \right) . \tag{C.20}$$

Subject to

Constraints on Population values

$$Sales^{x}_{q,c,i} \leq datSales_{q,c,i} Sales^{r}_{q,c,i} + (1.0 - datSales_{q,c,i})\, n^{(s)}_{q,c,i} u^{(s)}_{q,i} \ \forall\, (q \in ShipType, c \in County, i \in Size_A) \tag{C.21}$$

$$Sales^{x}_{q,c,i} \geq datSales_{q,c,i} Sales^{r}_{q,c,i} + (1.0 - datSales_{q,c,i})\, n^{(s)}_{q,c,i} l^{(s)}_{q,i} \ \forall\, (q, c, i) \tag{C.22}$$

Constraints on County totals

$$Tc^{(s),x}_{q,c} \leq datTc^{(s)}_{q,c} Tc^{(s),r}_{q,c} + \left(1.0 - datTc^{(s)}_{q,c}\right) uTc^{(s)}_{q,c} \ \forall\, (q, c) \tag{C.23}$$

$$Tc^{(s),x}_{q,c} \geq datTc^{(s)}_{q,c} Tc^{(s),r}_{q,c} + \left(1.0 - datTc^{(s)}_{q,c}\right) lTc^{(s)}_{q,c} \ \forall\, (q, c) \tag{C.24}$$

$$uTc_{q,c}^{(s)} = \sum_{i \in Size_A} \left[datSales_{q,c,i} Sales_{q,c,i}^{r} + (1.0 - datSales_{q,c,i})\, n_{q,c,i}^{(s)} u_{q,i}^{(s)} \right] \forall\, (q,c) \qquad \text{(C.25)}$$

$$lTc_{q,c}^{(s)} = \sum_{i \in Size_A} \left[datSales_{q,c,i} Sales_{q,c,i}^{r} + (1.0 - datSales_{q,c,i})\, n_{q,c,i}^{(s)} l_{q,i}^{(s)} \right] \forall\, (q,c) \qquad \text{(C.26)}$$

$$\sum_{i \in Size_A} Sales_{q,c,i}^{x} = Tc_{q,c}^{(s),x} \;\forall\, (q,c) \qquad \text{(C.27)}$$

Constraints on Size totals

$$Tz_{q,i}^{(s),x} \leq datTz_{q,i}^{(s)} Tz_{q,i}^{(s),r} + \left(1.0 - datTz_{q,i}^{(s)}\right) uTz_{q,i}^{(s)} \;\forall\, (q,i) \qquad \text{(C.28)}$$

$$Tz_{q,i}^{(s),x} \geq datTz_{q,i}^{(s)} Tz_{q,i}^{(s),r} + \left(1.0 - datTz_{q,i}^{(s)}\right) lTz_{q,i}^{(s)} \;\forall\, (q,i) \qquad \text{(C.29)}$$

$$uTz_{q,i}^{(s)} = \sum_{c \in County} \left[datSales_{q,c,i} Sales_{q,c,i}^{r} + (1.0 - datSales_{q,c,i})\, n_{q,c,i}^{(s)} u_{q,i}^{(s)} \right] \forall\, (q,i) \qquad \text{(C.30)}$$

$$lTz_{q,i}^{(s)} = \sum_{c \in County} \left[datSales_{q,c,i} Sales_{q,c,i}^{r} + (1.0 - datSales_{q,c,i})\, n_{q,c,i}^{(s)} l_{q,i}^{(s)} \right] \forall\, (q,i) \qquad \text{(C.31)}$$

$$\sum_{c \in County} Sales_{q,c,i}^{x} = Tz_{q,i}^{(s),x} \;\forall\, (q,i) \qquad \text{(C.32)}$$

Constraints on State totals

$$\sum_{c \in County} Tc_{q,c}^{(s),x} = S_q^{\text{tot}} \; \forall \, (q) \tag{C.33}$$

$$\sum_{i \in Size_A} Tz_{q,i}^{(s),x} = S_q^{\text{tot}} \; \forall \, (q) \tag{C.34}$$

Constraint on Shipment relations

$$Tc_{Slaughter,c}^{(s),x} \leq Tc_{AllMovement,c}^{(s),x} \; \forall \, (c,i) \tag{C.35}$$

The primary structural difference in this pair of optimization problems is the aggregations and relations of the last constraints of each. The shipment formulation only relates the two types through county totals $Tc_{q,c}^{(s),x}$ and not through the sub-elements as in the population formulation. For the objectives of this chapter, there is not a need to add any assumptions to the relationship between the slaughter shipments and the total shipments beyond the assumption-less relationship of inequality C.35. The aggregated populations $Pop_{t,c,j}^{R}$, the county shipment totals $Tc_{q,c}^{(s),x}$, and the largest category of slaughter shipments $Sales_{Slaughter,c,z500\_up}^{x}$ compose the set of inputs for the estimation of cattle movement parameters described in section 4.3.

C.2 Results of Optimization

In the appendix 4.A, the formulations of two optimization problems that estimate all non-disclosed elements of the particular USDA NASS data sets which are used in this chapter. We quantify the amount of estimated data for each State in tables C.1 and C.2. Table C.1

presents the numbers of populations estimated for each State by count and percentage as well as the number (head) of cattle assigned across these populations. The populations are counted by summing over all three types of $Type_A = \{Dairy,\ Preslaughter,\ All\ Cattle\}$. This method of counting induces double counting since cattle in the first two types also belong to the third type, but offers a systematic quantification of the amount of data estimated. The percentage of cattle assigned through the estimation demonstrates that in situations where many populations are estimated the significance of these populations is less $< 7\%$. The States of Kansas and Texas are exceptions to the trend of small fractions of the total cattle being assigned through estimation. These two States appear to have relatively similar percentages of estimated populations when compared to the other States, yet they assign larger percentages of the cattle totals. These larger numbers of cattle suggest that a higher number of counties in these States possess only a few large cattle operations, where the sparsity and the size of the operations necessitate the non-disclosure of their data elements. Table C.2 presents a parallel quantification for the estimations of the shipment distributions with counts similarly aggregated over $ShipType = \{All\ Shipments,\ Slaughter\}$. The results of table C.2 are comparable to those of table C.1.

State	**Count**	**Count %**	**Head**	**Head %**
Arkansas	170	10.79	55,748	3.06
Colorado	221	16.44	178,472	4.46
Iowa	391	18.81	149,802	2.52
Kansas	524	23.76	1,153,577	12.20
Minnesota	387	21.18	132,701	3.83
Missouri	500	31.33	112,753	2.51
Nebraska	399	20.43	552,159	5.89
Oklahoma	330	20.41	381,148	6.55
South Dakota	282	20.35	112,648	2.62

Continued on next page

Table C.1 – *Continued from previous page*

State	**Count**	**Count** %	**Head**	**Head** %
Texas	697	13.07	1,513,115	8.81

Table C.1: Totals and percentages of estimated cattle populations

State	**Count**	**Count** %	**Head**	**Head** %
Arkansas	139	13.24	75,326	7.56
Colorado	160	17.86	165,694	3.20
Iowa	165	11.90	163,735	2.75
Kansas	327	22.24	1,958,683	13.71
Minnesota	182	14.94	108,631	4.74
Missouri	334	20.93	213,464	8.18
Nebraska	242	15.16	1,144,166	8.98
Oklahoma	200	18.55	407,340	8.74
South Dakota	156	16.88	155,108	4.37
Texas	720	20.25	3,055,251	17.38

Table C.2: Totals and percentages of estimated cattle shipments

Table C.3 displays the solution for $p^x_{t_1,j_1,t_2,j_2,dist}$ sorted by the origin and destination pairs in the left column and the distances between the county centers in the rows. The cattle types of $Dairy$, $Beef$, and $Preslaughter$ are denoted respectively by D, B, and P to be brief.

Source → Destination	$d0$	$d100$	$d200$	$d500$	$d1000$
$D, z1\_19 \rightarrow D, z1\_19$	0.3748829	0.2107100	0.0732444	0.1286620	0.2043053
$D, z20\_199 \rightarrow D, z1\_19$	2.6880019	0.0551637	0.0144204	0.0026045	0.0
$D, z200\_up \rightarrow D, z1\_19$	0.3308997	0.0030519	0.0024367	0.0067531	0.0050407
$D, z1\_19 \rightarrow D, z20\_199$	0.0142987	0.0078986	0.0004281	0.0012359	0.0051593
$D, z20\_199 \rightarrow D, z20\_199$	0.0	0.0	0.0	0.0	0.0
$D, z200\_up \rightarrow D, z20\_199$	0.0	0.0	0.0	0.0	0.0
$D, z1\_19 \rightarrow D, z200\_up$	0.1359276	0.0187450	0.0155348	0.0996421	0.1292015
$D, z20\_199 \rightarrow D, z200\_up$	0.0	0.0	0.0	0.0	0.0
$D, z200\_up \rightarrow D, z200\_up$	0.0	0.0	0.0	0.0	0.0
$D, z1\_19 \rightarrow B, z1\_19$	0.2687743	0.0390975	0.0171808	0.0807097	0.1742045
$D, z20\_199 \rightarrow B, z1\_19$	0.0	0.0	0.0	0.0	0.0
$D, z200\_up \rightarrow B, z1\_19$	0.0360669	0.0076309	0.0007037	0.0	0.0
$D, z1\_19 \rightarrow B, z20\_199$	0.2088205	0.0370924	0.0201817	0.1386829	0.2043643
$D, z20\_199 \rightarrow B, z20\_199$	0.0	0.0	0.0	0.0004745	0.0
$D, z200\_up \rightarrow B, z20\_199$	0.0204587	0.0056272	0.0005575	0.0008731	0.0
$D, z1\_19 \rightarrow B, z200\_up$	0.1948755	0.0350860	0.0159960	0.1036465	0.3731019
$D, z20\_199 \rightarrow B, z200\_up$	0.0	0.0	0.0	0.0	0.0013401
$D, z200\_up \rightarrow B, z200\_up$	1.7686314	0.0004226	0.0004293	0.0241026	0.0005980
$B, z1\_19 \rightarrow D, z1\_19$	0.1549233	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow D, z1\_19$	0.3219837	0.0046190	0.0	0.0	0.0
$B, z200\_up \rightarrow D, z1\_19$	0.0093844	0.0	0.0	0.0	0.0
$B, z1\_19 \rightarrow D, z20\_199$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow D, z20\_199$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow D, z20\_199$	0.0	0.0	0.0	0.0	0.0

Continued on next page

Table C.3 – *Continued from previous page*

Source → Destination	$d0$	$d100$	$d200$	$d500$	$d1000$
$B, z1\_19 \rightarrow D, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow D, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow D, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z1\_19 \rightarrow B, z1\_19$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow B, z1\_19$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow B, z1\_19$	0.0116617	0.0	0.0	0.0	0.0
$B, z1\_19 \rightarrow B, z20\_199$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow B, z20\_199$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow B, z20\_199$	1.4740526	0.0	0.0	0.0	0.0
$B, z1\_19 \rightarrow B, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow B, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow B, z200\_up$	0.8815763	0.0	0.0	0.0	0.0
$B, z1\_19 \rightarrow P, z1\_19$	0.0934555	0.0015552	0.0005454	0.0	0.0044008
$B, z20\_199 \rightarrow P, z1\_19$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow P, z1\_19$	0.0018910	0.0	0.0	0.0002278	0.0
$B, z1\_19 \rightarrow P, z20\_199$	0.2044223	0.0084513	0.0004080	0.0001565	0.0005884
$B, z20\_199 \rightarrow P, z20\_199$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow P, z20\_199$	0.1459641	0.0	0.0007071	0.0002750	0.0
$B, z1\_19 \rightarrow P, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z20\_199 \rightarrow P, z200\_up$	0.0	0.0	0.0	0.0	0.0
$B, z200\_up \rightarrow P, z200\_up$	0.6325839	0.0	0.0	0.0008162	0.0

Table C.3: Estimated movement parameters $p^x_{t_1,j_1,t_2,j_2,dist}$.

10^3

C.3 Further Results

Tables C.4, C.5, and C.6 display the solutions of the slaughter, expiration, and birth probabilities for the first 10 alphabetical counties in each of the 10 States. To receive an electronic copy of the full set of these results, or of the full solutions of the two formulations described in appendix A, please contact the author to place your request.

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Allen_Kansas	P	0.5	0.036512368	0.033738401
	D	0	0	0
	B	0.009230631	0.010457562	0.007486202
Anderson_Kansas	P	0.110336388	0.093087101	0.023255814
	D	0	0	0
	B	0.009626369	0.011963276	0.007258185
Atchison_Kansas	P	0.329837167	0.1932723	0.023255814
	D	0	0	0
	B	0.008852841	0.006444239	0.00806116
Barber_Kansas	P	0.495192308	0.172171946	0.035871536
	D	0	0	0
	B	0.015204733	0.016840585	0.013123124
Barton_Kansas	P	0.495192308	0.185803167	0.043736865
	D	0	0	0
	B	0.011760191	0.013612247	0.004658081
Bourbon_Kansas	P	0.119362722	0.039255861	0.048816568
	D	0	0	0
	B	0.008461298	0.015055827	0.007137421
Brown_Kansas	P	0.5	0.030375102	0.025395369

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0	0	0
	B	0.012261904	0.01085365	0.008656168
Butler_Kansas	P	0.083982065	0.040468081	0.041906406
	D	0	0	0
	B	0.016454385	0.022280143	0.014439649
Chase_Kansas	P	0.5	0.06873808	0.040341958
	D	0	0	0
	B	0.019976392	0.019796885	0.013368885
Chautauqua_Kansas	P	0.5	0.031011781	0.495192308
	D	0	0	0
	B	0.01133813	0.015032186	0.008972601
Arkansas_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.017942335	0	0.019670195
Ashley_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.009037572	0.006588376	0.076923077
Baxter_Arkansas	P	0.5	0.495192308	0.495192308
	D	0	0	0
	B	0.015777869	0.01046699	0.013148033
Benton_Arkansas	P	0.5	0.495192308	0.495192308
	D	0	0	0
	B	0.007363092	0.006957121	0.00704525

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Boone_Arkansas	P	0.172865974	0.040994943	0.495192308
	D	0	0	0
	B	0.007423113	0.010207347	0.007849095
Bradley_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.010429088	0.001657369	0.017421946
Calhoun_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.013839735	0.009905294	0.076923077
Carroll_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.007834269	0.010446174	0.008222271
Chicot_Arkansas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.027957615	0.000620787	0.015409635
Clark_Arkansas	P	0.023255814	0.495192308	0.495192308
	D	0	0	0
	B	0.009266604	0.007784026	0.009533164
Adams_Colorado	P	0.051547394	0.049326205	0.495192308
	D	0	0	0
	B	0.008948174	0.013501349	0.003481196
Alamosa_Colorado	P	0.397373165	0.495192308	0.495192308
	D	0	0	0

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.032068358	0.012498428	0.003606102
Arapahoe_Colorado	P	0.086108753	0.495192308	0.495192308
	D	0	0	0
	B	0.008733645	0.020601471	0.002244155
Archuleta_Colorado	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.076923077	0.017150424	0.021813076
Baca_Colorado	P	0.495192308	0.028683181	0.038064109
	D	0	0	0
	B	0.017125729	0.012592866	0.006278118
Bent_Colorado	P	0.235839423	0.201277225	0.031941835
	D	0	0	0
	B	0.019648351	0.016180213	0.001168923
Boulder_Colorado	P	0.12099356	0.495192308	0.023255814
	D	0	0	0
	B	0.009161603	0.005428329	0.004044157
Broomfield_Colorado	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.076923077	0.076923077	0.076923077
Chaffee_Colorado	P	0.5	0.495192308	0.495192308
	D	0	0	0
	B	0.013937867	0.014442408	0.006391623
Cheyenne_Colorado	P	0.385232552	0.041487515	0.023255814

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0	0	0
	B	0.032230188	0	0
Adair_Iowa	P	0.08044667	0.037194898	0.023255814
	D	0	0	0
	B	0.010715823	0.008675187	0.007509519
Adams_Iowa	P	0.118137402	0.023255814	0.023255814
	D	0	0	0
	B	0.012379519	0.006601231	0.007812842
Allamakee_Iowa	P	0.031380403	0.02430315	0.023255814
	D	0	0	0
	B	0.011822923	0.008790575	0
Appanoose_Iowa	P	0.212919265	0.023255814	0.023255814
	D	0	0	0
	B	0.008550749	0.008732926	0.007911649
Audubon_Iowa	P	0.023255814	0.023255814	0.023255814
	D	0	0	0
	B	0.010317745	0.011199666	0.008207629
Benton_Iowa	P	0.202356581	0.023255814	0.036682697
	D	0	0	0
	B	0.011218631	0.002839699	0.008215449
Black_Hawk_Iowa	P	0.024470922	0.023255814	0.023255814
	D	0	0	0
	B	0.012815887	0.008196206	0.010250613

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Boone_Iowa	P	0.076327051	0.048974686	0.023255814
	D	0	0	0
	B	0.009852973	0.006226539	0.003807872
Bremer_Iowa	P	0.023255814	0.023255814	0.023255814
	D	0	0	0
	B	0.018062367	0.001609874	0.008413108
Buchanan_Iowa	P	0.095232698	0.023255814	0.023255814
	D	0	0	0
	B	0.01658426	0	0.010008496
Aitkin_Minnesota	P	0.094095988	0.495192308	0.495192308
	D	0	0	0
	B	0.009414384	0.003805876	0.008915079
Anoka_Minnesota	P	0.026166117	0.044030492	0.495192308
	D	0	0	0
	B	0.017678372	0	0.076923077
Becker_Minnesota	P	0.040570189	0.023255814	0.023255814
	D	0	0	0
	B	0.01027713	0.000147299	0.006313001
Beltrami_Minnesota	P	0.192810937	0.033838393	0.495192308
	D	0	0	0
	B	0.010125093	0.011269042	0.004419177
Benton_Minnesota	P	0.256082008	0.023255814	0.023255814
	D	0	0	0

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.009303556	0	0
Big_Stone_Minnesota	P	0.14106771	0.023255814	0.023255814
	D	0	0	0
	B	0.026281989	0.011251358	0.020637848
Blue_Earth_Minnesota	P	0.023255814	0.025909508	0.025031289
	D	0	0	0
	B	0.012861263	0.006442737	0.000144225
Brown_Minnesota	P	0.23404349	0.023255814	0.024324221
	D	0	0	0
	B	0.010990401	0.014762487	0.002716561
Carlton_Minnesota	P	0.095296062	0.107508242	0.495192308
	D	0	0	0
	B	0.008897847	0	0.001764042
Carver_Minnesota	P	0.023255814	0.023255814	0.023255814
	D	0	0	0
	B	0.012989327	0.014215573	0
Adair_Missouri	P	0.113873759	0.023255814	0.495192308
	D	0	0	0
	B	0.015949232	0.010638529	0.007637658
Andrew_Missouri	P	0.208097364	0.183503826	0.023255814
	D	0	0	0
	B	0.008233105	0.003147495	0.00953856
Atchison_Missouri	P	0.5	0.029339206	0.028649477

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0	0	0
	B	0.012527719	0.009438482	0.013123369
Audrain_Missouri	P	0.242757101	0.023255814	0.028213303
	D	0	0	0
	B	0.009932618	0	0
Barry_Missouri	P	0.180302112	0.029109465	0.034977089
	D	0	0	0
	B	0.007603623	0.005135756	0.00702282
Barton_Missouri	P	0.031466518	0.036964084	0.050047837
	D	0	0	0
	B	0.00858394	0.00491464	0.007555362
Bates_Missouri	P	0.281939475	0.023255814	0.050304968
	D	0	0	0
	B	0.00832711	0.010737224	0.001841125
Benton_Missouri	P	0.142049026	0.023255814	0.023255814
	D	0	0	0
	B	0.008294364	0.005774341	0.007584192
Bollinger_Missouri	P	0.059999129	0.042487346	0.032044741
	D	0	0	0
	B	0.007385009	0.007971676	0.008206904
Boone_Missouri	P	0.218554343	0.103200221	0.495192308
	D	0	0	0
	B	0.00792694	0.008447651	0.007660508

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Adams_Nebraska	P	0.201144758	0.041179405	0.042548066
	D	0	0	0
	B	0.020900713	0.024872508	0.014937561
Antelope_Nebraska	P	0.074839164	0.056532324	0.033339779
	D	0	0	0
	B	0.016399335	0.017640911	0.006322647
Arthur_Nebraska	P	0.495192308	0.091503327	0.341841758
	D	0	0	0
	B	0.076923077	0.046621832	0
Banner_Nebraska	P	0.5	0.5	0.026193458
	D	0	0	0
	B	0.043763141	0.015056123	0.001422003
Blaine_Nebraska	P	0.495192308	0.495192308	0.023255814
	D	0	0	0
	B	0.063652728	0.028987279	0.006018489
Boone_Nebraska	P	0.081795671	0.04860457	0.032027837
	D	0	0	0
	B	0.020898201	0.012793252	0.006304617
Box_Butte_Nebraska	P	0.5	0.060151633	0.039079407
	D	0	0	0
	B	0.019690691	0.016087895	0.011842795
Boyd_Nebraska	P	0.5	0.057585913	0.033045794
	D	0	0	0

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.025353629	0.014851542	0.00373573
Brown_Nebraska	P	0.5	0.034195102	0.04378072
	D	0	0	0
	B	0.0242513	0.02096682	0.008992986
Buffalo_Nebraska	P	0.220161259	0.023255814	0.038124145
	D	0	0	0
	B	0.011184891	0.012502991	0.007065253
Adair_Oklahoma	P	0.0859375	0.495192308	0.495192308
	D	0	0	0
	B	0.007818441	0.014269724	0.007138971
Alfalfa_Oklahoma	P	0.495192308	0.037011834	0.078730238
	D	0	0	0
	B	0.014818868	0.015349528	0.003717114
Atoka_Oklahoma	P	0.484475001	0.023255814	0.495192308
	D	0	0	0
	B	0.007946291	0.011040068	0.001014804
Beaver_Oklahoma	P	0.140429312	0.039618098	0.028946517
	D	0	0	0
	B	0.01953867	0.021694779	0.016281579
Beckham_Oklahoma	P	0.463181287	0.096744207	0.495192308
	D	0	0	0
	B	0.009249394	0.011239221	0.00716536
Blaine_Oklahoma	P	0.482273065	0.028375258	0.028619376

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0	0	0
	B	0.017861551	0.012601011	0.01085919
Bryan_Oklahoma	P	0.073464254	0.023255814	0.025800989
	D	0	0	0
	B	0.007666567	0.011135779	0.002225489
Caddo_Oklahoma	P	0.241503031	0.038457463	0.495192308
	D	0	0	0
	B	0.008644533	0.011702546	0.005810811
Canadian_Oklahoma	P	0.241394806	0.403693352	0.049321851
	D	0	0	0
	B	0.008612362	0.012283982	0.010959348
Carter_Oklahoma	P	0.093753108	0.104225157	0.495192308
	D	0	0	0
	B	0.007525577	0.014768178	0.007132077
Aurora_SouthDakota	P	0.1553729	0.025934539	0.023255814
	D	0	0	0
	B	0.029418367	0.014039367	0.004550125
Beadle_SouthDakota	P	0.290465625	0.023255814	0.023255814
	D	0	0	0
	B	0.022776183	0.018050471	0.006241368
Bennett_SouthDakota	P	0.221462035	0.023255814	0.495192308
	D	0	0	0
	B	0.02942767	0.018426908	0.006482326

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Bon_Homme_SouthDakota	P	0.089939776	0.036915117	0.023255814
	D	0	0	0
	B	0.016629124	0.012038847	0.004723059
Brookings_SouthDakota	P	0.311031245	0.023255814	0.023255814
	D	0	0	0
	B	0.016648927	0.013362236	0.008401178
Brown_SouthDakota	P	0.5	0.080357143	0.025677008
	D	0	0	0
	B	0.013674176	0.014563087	0.005185603
Brule_SouthDakota	P	0.141412544	0.039371962	0.023255814
	D	0	0	0
	B	0.028170523	0.024201846	0.004269409
Buffalo_SouthDakota	P	0.495192308	0.023255814	0.074527789
	D	0	0	0
	B	0.047489304	0.032072039	0
Butte_SouthDakota	P	0.5	0.042644015	0.495192308
	D	0	0	0
	B	0.011151988	0.01633013	0.009699981
Campbell_SouthDakota	P	0.190711884	0.023255814	0.032997901
	D	0	0	0
	B	0.076923077	0.020677038	0.005531828
Anderson_Texas	P	0.205991536	0.032203105	0.036288044
	D	0	0	0

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.006823256	0.009094842	0.007365105
Andrews_Texas	P	0.495192308	0.095961538	0.495192308
	D	0	0	0
	B	0.023982103	0.040550594	0.004026146
Angelina_Texas	P	0.064148406	0.02804128	0.495192308
	D	0	0	0
	B	0.007051296	0.008145837	0.008589059
Aransas_Texas	P	0.5	0.495192308	0.495192308
	D	0	0	0
	B	0.012332315	0.020462644	0.004441283
Archer_Texas	P	0.463581851	0.023255814	0.076579896
	D	0	0	0
	B	0.011354851	0.014409959	0.006038685
Armstrong_Texas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.030039875	0.015952475	0.01240551
Atascosa_Texas	P	0.078324742	0.040192143	0.023255814
	D	0	0	0
	B	0.006404975	0.002946461	0.001532337
Austin_Texas	P	0.023255814	0.118222788	0.495192308
	D	0	0	0
	B	0.006394747	0.008172323	0.007622772
Bailey_Texas	P	0.5	0.024905863	0.054255332

Continued on next page

Table C.4 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0	0	0
	B	0.020569594	0.027219799	0.013538062
Bandera_Texas	P	0.495192308	0.495192308	0.495192308
	D	0	0	0
	B	0.006923185	0.004045219	0.016307025

Table C.4: Estimated slaughter probabilities $sl^{x}_{c_1,t_1,j_1}$

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Allen_Kansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009066039
	B	0.009615385	0.009615	0.009615385
Anderson_Kansas	P	0.009615385	0.009615	0.005621609
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Atchison_Kansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Barber_Kansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.003282366
Barton_Kansas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.007749956

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.009615385	0.009615	0.009615385
Bourbon_Kansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.005741	0.009615385
Brown_Kansas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009166129
Butler_Kansas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.001923077	0.001923	0.001923077
Chase_Kansas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.003246176
Chautauqua_Kansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.007775092
Arkansas_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Ashley_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Baxter_Arkansas	P	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.009615385	0.009615	0.009615385
	B	0.001923077	0.009615	0.004650145
Benton_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.007070624
	B	0.009615385	0.009615	0.009615385
Boone_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.008950444
Bradley_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Calhoun_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Carroll_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.008470125
Chicot_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Clark_Arkansas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Adams_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Alamosa_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Arapahoe_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Archuleta_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.00434	0.001923077
Baca_Colorado	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bent_Colorado	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Boulder_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Broomfield_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.009615385	0.009615	0.009615385
Chaffee_Colorado	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Cheyenne_Colorado	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Adair_Iowa	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Adams_Iowa	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Allamakee_Iowa	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Appanoose_Iowa	P	0.009615385	0.003894	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Audubon_Iowa	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Benton_Iowa	P	0.009615385	0.009615	0

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.009615385	0.009615	0.008770647
	B	0.009615385	0.009615	0.009615385
Black_Hawk_Iowa	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Boone_Iowa	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bremer_Iowa	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Buchanan_Iowa	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Aitkin_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Anoka_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Becker_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Beltrami_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Benton_Minnesota	P	0.009615385	0.000765	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Big_Stone_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Blue_Earth_Minnesota	P	0.003578613	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Brown_Minnesota	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Carlton_Minnesota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Carver_Minnesota	P	0.009615385	0.009615	0.006773745
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Adair_Missouri	P	0.009615385	0.005991	0.009615385
	D	0.003205128	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.001923077	0.009615	0.009311731
Andrew_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Atchison_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Audrain_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Barry_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.008907378
	B	0.009615385	0.009615	0.009615385
Barton_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bates_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Benton_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.003205128	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bollinger_Missouri	P	0.007325991	0	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Boone_Missouri	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Adams_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.001923077	0.001923	0.001923077
Antelope_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.007830446
	B	0.009615385	0.009615	0.009615385
Arthur_Nebraska	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Banner_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Blaine_Nebraska	P	0.009615385	0.009615	0.001017718
	D	0.009615385	0.009615	0.009615385
	B	0.001923077	0.009615	0.009615385
Boone_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Box_Butte_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.005118296
Boyd_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Brown_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.007625446
Buffalo_Nebraska	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.005460091
	B	0.009615385	0.009615	0.009578172
Adair_Oklahoma	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.007656286
	B	0.009615385	0.006286	0.009615385
Alfalfa_Oklahoma	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Atoka_Oklahoma	P	0.009615385	0.003566	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Beaver_Oklahoma	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.001923077	0.001923	0.001923077
Beckham_Oklahoma	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009484965
Blaine_Oklahoma	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.006758381
	B	0.001923077	0.009615	0.005528052
Bryan_Oklahoma	P	0.009615385	0.002592	0.009615385
	D	0.009615385	0.009615	0.007019634
	B	0.009615385	0.009615	0.009615385
Caddo_Oklahoma	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Canadian_Oklahoma	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.005396559
Carter_Oklahoma	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.006145	0.009615385
Aurora_SouthDakota	P	0.009615385	0.009615	0.000425186
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Beadle_SouthDakota	P	0.009615385	0.009615	0

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.009615385	0.009615	0.009615385
	B	0.001923077	0.009615	0.009615385
Bennett_SouthDakota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bon_Homme_SouthDakota	P	0.009615385	0	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Brookings_SouthDakota	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.008689305
Brown_SouthDakota	P	0.009615385	0.009615	0
	D	0.003205128	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Brule_SouthDakota	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Buffalo_SouthDakota	P	0.009615385	0.00528	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Butte_SouthDakota	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.007202598

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Campbell_SouthDakota	P	0.009615385	0.009615	0.007468162
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Anderson_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Andrews_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Angelina_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Aransas_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Archer_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.007896358
	B	0.009615385	0.009615	0.009615385
Armstrong_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.00447016
Atascosa_Texas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009615385

Continued on next page

Table C.5 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.009615385	0.009615	0.009615385
Austin_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385
Bailey_Texas	P	0.009615385	0.009615	0
	D	0.009615385	0.009615	0.009547419
	B	0.009615385	0.009615	0.003742052
Bandera_Texas	P	0.009615385	0.009615	0.009615385
	D	0.009615385	0.009615	0.009615385
	B	0.009615385	0.009615	0.009615385

Table C.5: Estimated culling probabilities $dt^x_{c_1,t_1,j_1}$

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Allen_Kansas	P	0	0	0
	D	0.016129032	0.018662672	0.027777778
	B	0.019230769	0.019230769	0.018348469
Anderson_Kansas	P	0	0	0
	D	0.016129032	0.018910942	0.025254026
	B	0.019230769	0.019230769	0.019230769
Atchison_Kansas	P	0	0	0
	D	0.016129032	0.019542332	0.016129032
	B	0.019230769	0.015502172	0.019230769

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Barber_Kansas	P	0	0	0
	D	0.016129032	0.017449682	0.025009062
	B	0.019230769	0.019230769	0.019230769
Barton_Kansas	P	0	0	0
	D	0.016129032	0.018100417	0.027777778
	B	0.019230769	0.019230769	0.016904635
Bourbon_Kansas	P	0	0	0
	D	0.016129032	0.019026778	0.016129032
	B	0.019230769	0.019230769	0.019230769
Brown_Kansas	P	0	0	0
	D	0.016129032	0.01954821	0.016129032
	B	0.019230769	0.019230769	0.019230769
Butler_Kansas	P	0	0	0
	D	0.016129032	0.016129032	0.025777242
	B	0.019230769	0.019230769	0.019230769
Chase_Kansas	P	0	0	0
	D	0.016129032	0.018304191	0.025574486
	B	0.019230769	0.019230769	0.019230769
Chautauqua_Kansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Arkansas_Arkansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.018594586	0.006592623	0
Ashley_Arkansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.015163695	0
Baxter_Arkansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Benton_Arkansas	P	0	0	0
	D	0.016129032	0.018729768	0.027777778
	B	0.019230769	0.015884904	0.019230769
Boone_Arkansas	P	0	0	0
	D	0.016129032	0.018321971	0.016129032
	B	0.019230769	0.019230769	0.019230769
Bradley_Arkansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.009386676	0.019230769
Calhoun_Arkansas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.016786466	0
Carroll_Arkansas	P	0	0	0
	D	0.016129032	0.018713055	0.023531374
	B	0.019230769	0.019230769	0.019230769
Chicot_Arkansas	P	0	0	0

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.007529768	0.019230769
Clark_Arkansas	P	0	0	0
	D	0.016129032	0.017251212	0.016129032
	B	0.019230769	0.016591987	0.019230769
Adams_Colorado	P	0	0	0
	D	0.027777778	0.016129032	0.026169606
	B	0.019230769	0.019230769	0.016593653
Alamosa_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.013140765
Arapahoe_Colorado	P	0	0	0
	D	0.016129032	0.017750992	0.016129032
	B	0.019230769	0.01467602	0.007398824
Archuleta_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Baca_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.018235219
Bent_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.013021909

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Boulder_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.012726451	0.012445706
Broomfield_Colorado	P	0	0	0
	D	0.016129032	0.0178693	0.016129032
	B	0	0	0
Chaffee_Colorado	P	0	0	0
	D	0.027777778	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.015094119
Cheyenne_Colorado	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.00643659	0.011153653
Adair_Iowa	P	0	0	0
	D	0.016129032	0.019286284	0.016129032
	B	0.019230769	0.017229809	0.019230769
Adams_Iowa	P	0	0	0
	D	0.016129032	0.019452005	0.016129032
	B	0.019230769	0.014771435	0.019230769
Allamakee_Iowa	P	0	0	0
	D	0.016129032	0.016845997	0.027021995
	B	0.019230769	0.014632652	0.011804873
Appanoose_Iowa	P	0	0	0
	D	0.016129032	0.018996018	0.016129032

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.019230769	0.017831937	0.019230769
Audubon_Iowa	P	0	0	0
	D	0.016129032	0.019231784	0.016129032
	B	0.019230769	0.019230769	0.018719666
Benton_Iowa	P	0	0	0
	D	0.016129032	0.018137318	0.027777778
	B	0.019230769	0.011288747	0.019230769
Black_Hawk_Iowa	P	0	0	0
	D	0.016129032	0.018219969	0.023410277
	B	0.019230769	0.01664206	0.019230769
Boone_Iowa	P	0	0	0
	D	0.016129032	0.019055274	0.016129032
	B	0.019230769	0.014082176	0.014294893
Bremer_Iowa	P	0	0	0
	D	0.016129032	0.018072216	0.024570503
	B	0.019230769	0.009798775	0.019230769
Buchanan_Iowa	P	0	0	0
	D	0.027777778	0.017809639	0.016129032
	B	0.019230769	0.009075366	0.019230769
Aitkin_Minnesota	P	0	0	0
	D	0.027777778	0.016129032	0.016129032
	B	0.019230769	0.011744888	0.019230769
Anoka_Minnesota	P	0	0	0

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.016129032	0.016679414	0.016129032
	B	0.019230769	0.0072468	0
Becker_Minnesota	P	0	0	0
	D	0.027777778	0.016129032	0.023296426
	B	0.019230769	0.00892994	0.015886639
Beltrami_Minnesota	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.015370458
Benton_Minnesota	P	0	0	0
	D	0.027777778	0.017039124	0.024490055
	B	0.019230769	0.007218229	0.011040238
Big_Stone_Minnesota	P	0	0	0
	D	0.016129032	0.016883226	0.019252458
	B	0.019230769	0.019230769	0.019230769
Blue_Earth_Minnesota	P	0	0	0
	D	0.016129032	0.020952834	0.016129032
	B	0.019230769	0.005815579	0.010069174
Brown_Minnesota	P	0	0	0
	D	0.016129032	0.018419416	0.024182449
	B	0.019230769	0.019230769	0.014454431
Carlton_Minnesota	P	0	0	0
	D	0.027777778	0.016129032	0.016432883
	B	0.019230769	0.009034055	0

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Carver_Minnesota	P	0	0	0
	D	0.016129032	0.017495302	0.024388714
	B	0.019230769	0.019230769	0.01146673
Adair_Missouri	P	0	0	0
	D	0.027777778	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Andrew_Missouri	P	0	0	0
	D	0.016129032	0.019765496	0.016129032
	B	0.019230769	0.012484502	0.019230769
Atchison_Missouri	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.018624682	0.019230769
Audrain_Missouri	P	0	0	0
	D	0.027777778	0.01817127	0.016129032
	B	0.019230769	0.008879543	0.011629212
Barry_Missouri	P	0	0	0
	D	0.027777778	0.018639334	0.027777778
	B	0.019230769	0.013886066	0.019230769
Barton_Missouri	P	0	0	0
	D	0.027777778	0.018786645	0.016129032
	B	0.019230769	0.013918873	0.019230769
Bates_Missouri	P	0	0	0
	D	0.027777778	0.019215486	0.027777778

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.019230769	0.019230769	0.014056776
Benton_Missouri	P	0	0	0
	D	0.027777778	0.018928524	0.016129032
	B	0.019230769	0.014885147	0.019230769
Bollinger_Missouri	P	0	0	0
	D	0.027777778	0.016129032	0.016129032
	B	0.019230769	0.017367646	0.019230769
Boone_Missouri	P	0	0	0
	D	0.016129032	0.017179028	0.02158095
	B	0.019230769	0.0171166	0.019230769
Adams_Nebraska	P	0	0	0
	D	0.016129032	0.016129032	0.02516782
	B	0.019230769	0.019230769	0.019230769
Antelope_Nebraska	P	0	0	0
	D	0.016129032	0.018539226	0.027777778
	B	0.019230769	0.019230769	0.018592555
Arthur_Nebraska	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0	0.019230769	0.011942628
Banner_Nebraska	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.012373642
Blaine_Nebraska	P	0	0	0

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.018236777
Boone_Nebraska	P	0	0	0
	D	0.016129032	0.018639569	0.016129032
	B	0.019230769	0.019230769	0.018473283
Box_Butte_Nebraska	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Boyd_Nebraska	P	0	0	0
	D	0.016129032	0.017388702	0.026580313
	B	0.019230769	0.019230769	0.015781254
Brown_Nebraska	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Buffalo_Nebraska	P	0	0	0
	D	0.027777778	0.01828023	0.027777778
	B	0.019230769	0.019230769	0.019230769
Adair_Oklahoma	P	0	0	0
	D	0.027777778	0.018647217	0.027777778
	B	0.019230769	0.019230769	0.019230769
Alfalfa_Oklahoma	P	0	0	0
	D	0.016129032	0.017098853	0.016129032
	B	0.019230769	0.019230769	0.016251749

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Atoka_Oklahoma	P	0	0	0
	D	0.027777778	0.01731787	0.016129032
	B	0.019230769	0.019230769	0.013325167
Beaver_Oklahoma	P	0	0	0
	D	0.027777778	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Beckham_Oklahoma	P	0	0	0
	D	0.016129032	0.017058426	0.016129032
	B	0.019230769	0.019230769	0.019230769
Blaine_Oklahoma	P	0	0	0
	D	0.016129032	0.017714165	0.027777778
	B	0.019230769	0.019230769	0.019230769
Bryan_Oklahoma	P	0	0	0
	D	0.027777778	0.017917387	0.027777778
	B	0.019230769	0.019230769	0.014477112
Caddo_Oklahoma	P	0	0	0
	D	0.016129032	0.017698719	0.016129032
	B	0.019230769	0.019230769	0.018342818
Canadian_Oklahoma	P	0	0	0
	D	0.016129032	0.017855236	0.016129032
	B	0.019230769	0.019230769	0.019230769
Carter_Oklahoma	P	0	0	0
	D	0.027777778	0.016129032	0.016129032

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	B	0.019230769	0.019230769	0.019230769
Aurora_SouthDakota	P	0	0	0
	D	0.027777778	0.017128887	0.02732635
	B	0.019230769	0.019230769	0.016435762
Beadle_SouthDakota	P	0	0	0
	D	0.016129032	0.01734951	0.024956718
	B	0.019230769	0.019230769	0.01854649
Bennett_SouthDakota	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.018476057
Bon_Homme_SouthDakota	P	0	0	0
	D	0.016129032	0.018333642	0.016129032
	B	0.019230769	0.019230769	0.016575628
Brookings_SouthDakota	P	0	0	0
	D	0.016129032	0.01801976	0.026643809
	B	0.019230769	0.019230769	0.019230769
Brown_SouthDakota	P	0	0	0
	D	0.027777778	0.016129032	0.024472893
	B	0.019230769	0.019230769	0.017270986
Brule_SouthDakota	P	0	0	0
	D	0.016129032	0.016584687	0.023963025
	B	0.019230769	0.019230769	0.016535861
Buffalo_SouthDakota	P	0	0	0

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.011762894
Butte_SouthDakota	P	0	0	0
	D	0.027777778	0.016129032	0.022173289
	B	0.019230769	0.019230769	0.019230769
Campbell_SouthDakota	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0	0.019230769	0.017426546
Anderson_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.018076724	0.019230769
Andrews_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.014787344
Angelina_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.017048314	0.019230769
Aransas_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.012774542
Archer_Texas	P	0	0	0
	D	0.016129032	0.017661326	0.027777778
	B	0.019230769	0.019230769	0.018030699

Continued on next page

Table C.6 – *Continued from previous page*

County_State	**Type**	$z1\_19$	$z20\_199$	$z200\_up$
Armstrong_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.019230769	0.019230769
Atascosa_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016282208
	B	0.019230769	0.011278516	0.013382048
Austin_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.017532657	0.019230769
Bailey_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.027777778
	B	0.019230769	0.019230769	0.019230769
Bandera_Texas	P	0	0	0
	D	0.016129032	0.016129032	0.016129032
	B	0.019230769	0.013295198	0.019230769

Table C.6: Estimated birth probabilities $bt^x_{c_1,t_1,j_1}$