UNIVERSITY OF TORINO

Ph.D. in Modeling and Data Science

Final dissertation

Modeling Disease Dynamics in Italian Livestock:
A Multilayer Analysis of Cattle and Buffalo Movement Network

Supervisor: Mario Giacobini

Candidate: Giacomo Zoppi

ACADEMIC YEAR 2023/2024

Abstract

Epidemics are a significant issue that affects various aspects of our lives, both in terms of public health and the global economy. For this reason, several mathematical models have been developed, such as compartmental models, agent-based models, and network-based models, to study the spread of diseases and the role that individual or community behaviors play in them.

This thesis primarily focuses on the Italian cattle and buffalo movement network and the associated risk of disease spread. When applying the model mentioned above to this context, several challenges immediately arise. Indeed, the Italian farms exhibit certain characteristics that are often overlooked and understudied.

Pasture movements are the seasonal displacement of cattle to shared mountain pastures during the summer months. While movements between herds must be recorded, this is not the case for movements between two pastures, complicating both the analysis and the tracking of contacts.

On the other hand, several farms raise both buffalo and cattle in the same spaces, facilitating, through proximity, disease transmission between two different species. The presence of buffalo is often overlooked in studies of the Italian network. Still, they could have a significant role, acting as a reservoir for pathogens that can also infect cattle. Understanding these unique characteristics is crucial to improving disease modeling and implementing effective containment or treatment measures.

In this thesis, we will first aim to observe the impact of cattle movements to pastures on the spread of pathogens and explore ways to improve the available data, highlighting existing issues. Subsequently, we will investigate whether the Italian buffalo network, which has yet to be studied, could sustain the spread of a pathogen. Finally, we will assess whether ignoring the potential cattle-buffalo interaction might lead to an underestimation of epidemic risk.

To achieve this, we will use existing animal movement data to create a contact network between different farms. These farms will be the epidemiological units of interest instead of individual animals. Using a combination of the three models mentioned above, we will simulate the spread of a pathogen to test our hypothesis. Additionally, for the cattle-buffalo interaction, we will adapt the concept of a multilayer network, where each layer will represent the movements of a single species, and the connections between layers will describe the interactions occurring within a single farm.

Our findings emphasize the importance of data quality in modeling disease spread in animal populations. We conclude that both movements to pastures and the dynamics of farms raising both buffalo and cattle play a significant role that should be better considered and studied by improving data collection processes beforehand. Furthermore, a comprehensive approach appears crucial for disease control, as focusing on a single species for certain pathogens could lead to underestimations of the actual risk. Future work could incorporate various environmental factors or wildlife interactions to provide an even more complete representation.

Summary

Chapter 1, *Motivation*, introduces the key topics addressed in this thesis, outlining the primary research questions. This chapter describes how various factors, such as animal movement networks and interactions between different species, contribute to the spread of infectious diseases among livestock. The importance of addressing gaps in existing data and developing models to predict disease spread is emphasized, introducing the analyses that follow.

In Chapter 2, *Compartmental Models*, we then provide some theoretical background about compartmental models. These models categorize populations into compartments based on disease status, such as susceptible, infected, recovered, and deceased. Exploring the dynamics within and between these compartments can help gain insights into how diseases spread and survive in populations.

In Chapter 3, we explore *Agent-Based Models* (ABMs), which simulate the actions and interactions of individual agents within a population. These models are particularly valuable for capturing the heterogeneity of populations and the complex behaviors that can influence disease transmission. By modeling individual behaviors and their interactions, ABMs provide a more granular perspective on disease dynamics, allowing for the exploration of scenarios that may not be captured by simpler models. However, these models are more complex and require careful parameter tuning to function correctly.

Their strength lies in simulating how contacts or interactions between individuals occur, but this may not be necessary if such information is already available.

In Chapter 4, we focus on *Network-based models*, which are central to the analyses conducted in this thesis. These represent the population as a network of nodes connected by edges that describe interactions between them. They are especially useful for understanding how diseases can spread through direct and indirect connections within the network. This chapter discusses the integration of temporal data and real-world movement patterns to enhance the predictive accuracy of these models, providing a critical tool for epidemiological research. Finally, it introduces a combined approach of the three models presented that will be used in the remainder of the thesis.

In Chapter 5, *Diseases*, we examine specific diseases that will be considered in the following chapters. Understanding the unique features of each disease is crucial for developing appropriate model and effective strategies to mitigate their spread within and between livestock populations. In particular, Bubaline herpesvirus 1, a virus that primarily affects buffalo and can cause respiratory illness and reproductive issues, and Q fever, a disease that infects multiple species and can lead to severe consequences in humans, are described.

In Chapter 6, *Cattle*, we turn our attention to the cattle movement network in Italy. Utilizing data from the National Animal Identification and Registration Database (NDB), we analyze the patterns and implications of cattle movements, with a particular focus on the challenges posed by missing or inaccurate data. This Chapter also explores how cattle movements to and from pastures contribute to the spread of diseases and the complexities involved in managing these risks.

In Chapter 7, *Buffalo*, we extend our analysis to the buffalo movement network. We investigate the role of buffalo farms in disease transmission, analyzing if this smaller network can support the spread of a pathogen. This is a novel analysis, as the Italian buffalo movement network has not been deeply studied in the literature. This finding is particularly relevant considering the presence of farms where both buffalo and cattle are raised together, referred to as *mixed* farms from now on. Their interactions could influence overall disease dynamics.

Finally, in Chapter 8, *Interaction on Mixed Farms*, we consider the interactions between cattle and buffalo network through the *mixed* farms. This Chapter explores how these interactions can impact disease transmission, considering factors such as shared environments, cross-species transmission, and the potential for mixed farms to help the diffusion of a pathogen. The findings emphasize the need for an integrated approach to disease management that accounts for the complexities and unique challenges posed by mixed-species farms.

The last Chapter, *Conclusions*, suggests future directions for research, including the development of more sophisticated models to capture the nuances of mixed-species interactions and their implications for disease spread. Additionally, it underscores the critical role of data quality in achieving accurate and reliable results. Good data is essential not only for model accuracy but also for the practical implementation of disease management strategies.

Contents

6 Cattle 49

List of Tables

List of Figures

Chapter 1

Motivation

Epidemics constitute a significant threat to public health and global economy. The spread of diseases is influenced by numerous factors playing different roles, such as individual behavior, interactions, and the characteristics of communities. Understanding these factors and how they interact can help develop effective control or prevention strategies.

A diverse range of mathematical models and approaches have been developed to study these elements. Among these are compartmental models, agent-based models, and network-based models [1].

Compartmental models provide a simplified but effective understanding of the progression of an epidemic through a population. A population is divided into groups, called compartments, representing different stages of an infection.

Typically, compartments include *susceptible*, *infected*, *recovered*, and *deceased*. Individuals move between these compartments based on defined rates of *transmission*, *recovery*, and, eventually, *death*. These models are simple to build, flexible, and provide an intuitive representation of disease spread.

However, they can oversimplify the problem and disregard the complexities of real-world epidemics. Specifically, they focus on trends at the population level without accounting for individual behavior or spatial heterogeneity.

Agent-based models offer a more complex approach at the individual level. Instead of dividing the population into compartments, they represent individuals as *autonomous agents*, each with specific attributes (e.g., age, health conditions) or behaviors (e.g., interaction with others). These *agents* follow rules based on their characteristics. The collective behavior and the epidemic dynamics arise from the interactions of these individuals, each following their own rules.

Unlike compartmental models, this approach allows for the representation of individual differences and some geographical factors. However, they are far more complex and require a large amount of data in order to calibrate and validate the model.

Network-based models are a powerful approach to studying the spread of infectious diseases by considering both the social and physical connections between individuals or groups. In this model, the population is represented as a network where individuals (or groups) are the nodes, and the edges represent the connections through which the disease can be transmitted. Transmission rules are defined based on factors such as the individual properties associated with different nodes or edge strength. In this way, an accurate representation of the *social* structure is available that allows, among other things, to identify key individuals or groups that play a critical role in the diffusion of a pathogen.

It is also possible to evaluate the potential effectiveness of targeted control measures focusing on specific network structures, such as the isolation or vaccination of key nodes or the implementation of restriction of movements.

They require data on connections and some individual properties in order to provide a good model of the real world.

The choice between these models depends on the specific research question and available data. Compartmental models can be good for an initial analysis or as a way of simulating a possible trend of pathogen diffusion. Agent-based models can offer a better representation of the real world but require more data and computational resources. Also, the individual behavior could be too specific and require strong assumptions.

Network-based models provide a valuable middle ground, capturing intrinsic social structure and offering valuable information for targeted interventions while requiring less computational effort. These models become even more attractive when considering the spread of diseases not between humans but between animals. The advantage of compartmental or agent-based models, indeed, is that they can simulate contacts between humans without actual movement data.

Animal movement data for various species have become increasingly available in recent years. The use of this data allows researchers to rely on simulations only for the intrinsic stochastic element of the epidemics. It is then possible to obtain accurate models while also being computationally efficient.

Thus, there is no need to *simulate* movements through compartments or agents since the *actual* movements can be used to construct the relationships and contacts between animals or farms. Therefore, in a network approach, animal movement can be used to create edges between animals or farms. This helps to recreate the real contact and movement structure, allowing us to search for particular elements like *key nodes*. Finding contact patterns can help to reconstruct the pathways that a pathogen took and better understand the potential existing risks in a susceptible population [2, 3].

Furthermore, compartmental models, agent-based models, and network-based models can be combined to overcome their respective limitations, resulting in computationally efficient models that can accurately predict the spread of a pathogen [4, 5].

It is also true that these data are not always easy to obtain for every species or every moment of time, especially for rare or non-domestic species.

Given that the movements of livestock are influential in the diffusion of infectious diseases, the European Union requires these to be registered and entered into a national veterinary database, following regulation (EU) 2016/429, which mandates the identification and registration of animal movements to ensure traceability and facilitate prevention.

In Italy, these data are collected by the Italian National Animal Identification and Registration Database (NDB) of the Ministry of Health. Movements of cattle, buffalo, horses, pigs, and sheep are available, with different granularity levels.

Among these species, cattle represent an important production reality in the peninsula, and their veterinary role and the related movement network have been studied in depth [6, 7].

Nevertheless, Italy presents specific characteristics that could influence the spread of epidemics among cattle and are often overlooked.

One of them is the movements to pastures. They usually take place in two periods. At the beginning of the summer, the herds go up to the mountain pastures, where there is fresher grass, which contributes to the production of higher quality milk. At the end of the season, herds descend to the valley.

The key feature of pastures is that they are shared by different farms and have no structural divisions with the outside environment, thus allowing contact with wild animals.

The proximity between animals from different herds and the use of common pastures can favor the spread of infectious diseases. Also, grazing cattle can be exposed to parasites, such as ticks, that again can work as vectors.

Additionally, infected animals arriving from farms with different pathogens can introduce new diseases to the local population, potentially leading to outbreaks. This thesis will analyze this issue, answering **what is the impact of cattle movements to pastures in the diffusion of pathogens?**

As briefly mentioned before, the animals are moved at the beginning of the summer period to a pasture and then returned to their herd at the end of the season. However, data on pasture movement is often lacking or unreliable, hindering the construction of accurate models and risk assessment of epidemics. In particular, animals may be moved from one pasture to another without passing through any farm. This type of movement is absent from the Italian National Database, where movements between pastures are registered as two fictitious movements. First, from the pasture of origin to the herd where the cattle belong, then from this herd to the arrival pasture, as can be seen in figure 1.1.

This hampers both network analysis, as the actual number of movements is increased, and contact tracing, complicating any control actions. This thesis will analyze this issue to understand the impact of the identified procedural shortcomings and propose potential corrections to the institutions. We will try to answer **how could we improve the available data to study this phenomenon?**

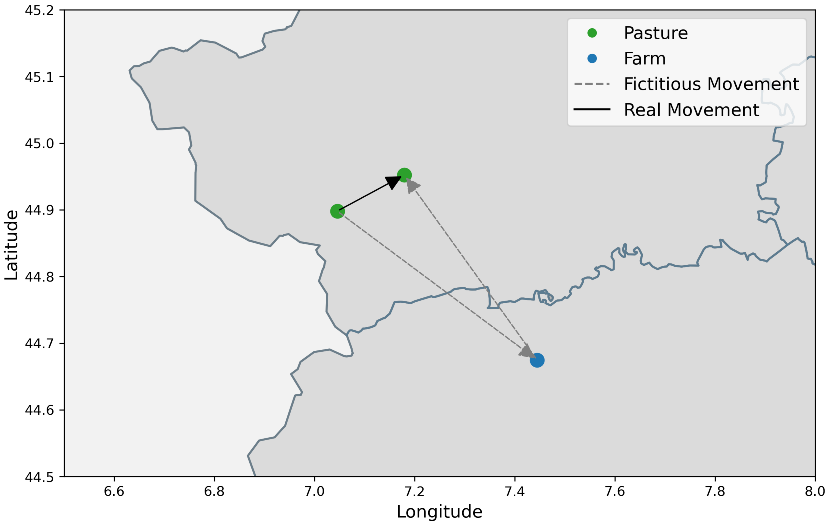


Figure 1.1. **Example of fictitious movement to pastures.**

Another interesting feature of Italian farms is that there are more or less 2500 farms in Italy that raise buffaloes, and more than two-thirds of them also raise bovines.

There is a sharing of space between these two species, and this proximity can once again favor the spread of infectious diseases or exposure to common parasites.

Moreover, the buffalo movement network in Italy is not only overlooked when analyzing the cattle movement network but it has also never been studied in the literature.

Given this lack of research on the role of buffalo in Italy, we will first consider the buffalo movement network to analyze its properties and see **if this network could support the spread of a pathogen.**

Following, we will see **if disregarding this cattle-buffalo interaction network in spreading models could lead to an underestimation of the epidemic risk**. Indeed, buffaloes can serve as a reservoir for bovine pathogens, and their interaction with cattle can facilitate disease transmission.

The use of mathematical and computational models is an approach for studying epidemics in the veterinary field, yielding satisfactory results. In particular, network-based models are an efficient method when data are available.

However, when these data are not considered or are missing, important errors can be made in the study of the spread of a pathogen. In this thesis, we try to understand the extent of this error and what can be done to improve these analyses.

Chapter 2

Compartmental Models

Compartmental models examine population dynamics by splitting the population of interest into categories called compartments. These compartments describe different states or characteristics of the individuals. Each compartment is homogeneous, so everyone within it has similar characteristics. This seemingly simple approach is able to capture even complex behaviors, which is why it is widely used. A more comprehensive treatment, both from a theoretical and applied perspective, can be found in other sources [8, 9, 10, 11].

The fundamental concept relies on the interaction between these compartments. The model defines how individuals transition between compartments based on their interactions with others or the environment.

This structure can be adapted to study disease dynamics. An example of this can be derived from the work presented in 1927 by Kermack and McKendrick [12]. Consider a disease that makes the individual immune after infection and compartmentalize people based on their relation with the disease into susceptible (S), infected (I), and recovered/removed (R) categories. From these labels also comes the name of this model, the SIR model.

In detail:

- Susceptible (S): healthy individuals that can contract the disease.
- Infected (I): individuals who have the disease and can transmit it to others.
- Recovered (R): individuals who have recovered from the disease and are then immune. In some versions of this model, the recovered compartment also considers the individuals killed by the disease.

The model assumes:

- A closed population with no births, deaths from causes different from the disease, or migration.
- Fixed infection rate at which infected individuals come into contact with susceptible individuals and infect them.
- Fixed recovery rate at which individuals recover and become immune.
- No incubation period.
- Individuals are infectious for the whole length of the disease.
- Homogeneous population with spatial or social structure.

Starting from some initial conditions, the transition between the groups, and so the evolution of the disease, are described by a partial differential equation model:

$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

where S is the number of susceptible people, I is the number of infected people, and R is the number of recovered people at a given time. Moreover, t is the time, β represents the infection rate, and γ is the recovery rate.

A typical evolution of the three compartments can be observed in figure 2.1.

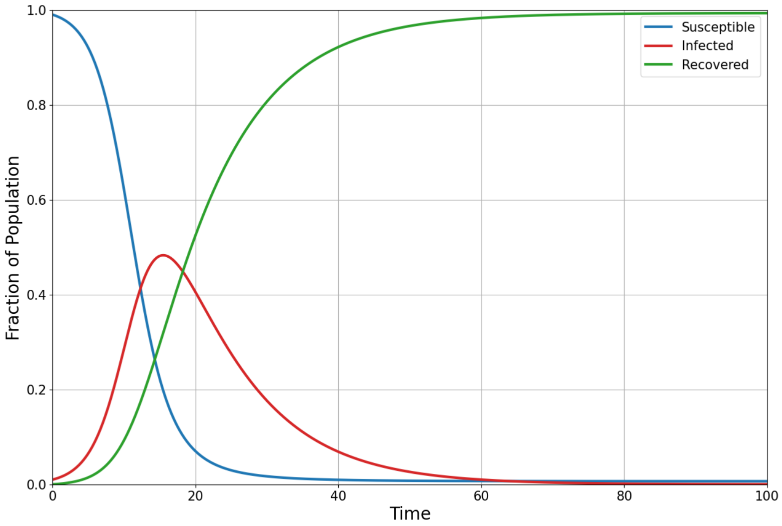


Figure 2.1. **Example Simulation of a SIR model.** 1% of infected population, $\beta = 0.5, \gamma = 0.1$

Some statistics can be extracted from this model. A key quantity is R_0 also called *epidemiological threshold*,

$$R_0 = \frac{\beta S}{\gamma}$$

this quantity expresses the expected number of individuals infected by a single infectious individual if the entire population is susceptible.

This model can be readily adapted to diverse scenarios. By modifying the underlying partial differential equations and introducing new compartments, we can capture the dynamics of a wide range of systems.

Suppose, for example, that it interests us to divide individuals who develop immunity to the disease and those who are deceased into two different categories. We need to add a deceased (D) compartment and change the equation of the partial differential equation model accordingly. Keeping the same notation used for the SIR model and adding N as the total population, we obtain:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} \\
\frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I - \mu I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \mu I
\end{aligned}$$

where t is the time, β represents the infection rate, γ is the recovery rate, and μ is the mortality rate.

Additionally, we could study diseases that do not give long-lasting immunity. Thus, infected individuals may return being susceptible. We will then have only the infected (I) and susceptible (S) compartments, the SIS model.

Keeping the same notation used for the SIRD model and adding N as the total population, the model is defined by the following partial differential equation model:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I$$
$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

where t is the time, and γ is the recovery rate.

Many pathogens have an incubation or latency period. So, there are individuals that have been infected by the disease but are not infectious yet. These individuals will form the exposed (E) compartment. This setting is studied using the SEIR model. A typical evolution of these four compartments can be observed in figure 2.2.

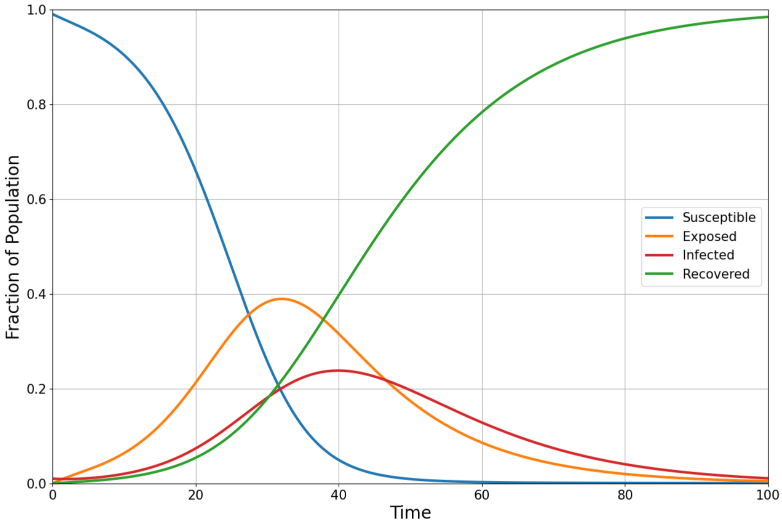


Figure 2.2. **Example Simulation of a SEIR model.** 1% of infected population, $\beta = 0.75, \alpha = 0.075, \gamma = 0.1$

Suppose that the latency period is a random variable with exponential distribution with parameter α, then we can define this model with the following equations:

$$\frac{dS}{dt} = -\frac{\beta S I}{N}$$
$$\frac{dE}{dt} = \frac{\beta S I}{N} - \alpha E$$
$$\frac{dI}{dt} = \alpha E - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

Other models can be implemented in a similar manner, adding or modifying the various compartments according to specific needs.

Chapter 3

Agent Based Models

Agent-based modeling (ABM) is a technique used to formalize models inspired by the real world where multiple *individuals* interact with each other or with the environment. Indeed, ABM focuses on individual behaviors and the resulting collective impact on the system [13]. A more detailed and in-depth discussion of the topic can be found in other resources [14, 15, 16, 17, 18, 19].

The key feature of this approach lies in defining individuals, called agents. Whereas for some models, such as equation-based modeling, the elements of the system are represented through average properties of representative groups or individuals; in this case, for each agent, the relationship it will have with other agents and with the overall system is explicitly defined [13].

This allows for great flexibility of the model [20]. As agent definitions are individual-based, any degree of heterogeneity can be incorporated. Heterogeneity can also be generated through the model's rule-based structure. Indeed, the framework can accommodate diverse rule sets from various disciplines (e.g., integrating medical and economic factors influencing individual agents).

Consequently, this model proves valuable for studying the evolution of diverse phenomena. Its underlying rules and interactions are repeatedly applied over time. It will then converge to one or more equilibrium states, if they exist, without requiring a definition of those.

Agent-based modeling provides the additional advantage of reconstructing the path to equilibrium while quantifying each agent's contribution to the system. This capability extends beyond equilibrium analysis, enabling the exploration of how global phenomena emerge from agent interactions and, conversely, how these phenomena influence individual conditions.

Agent definitions can also incorporate geographical elements and their relation with the environment to model local interactions, adequately representing any possible spatial relationship.

In short, agent-based modeling is particularly useful for studying how a global situation depends on the individual behavior of the agents. ABMs are built from the "bottom-up", defining the agents and their attributes. The agents interact with each other and the environment, causing an impact on the overall behavior of the system. These models are used in a variety of fields, including social sciences, economics, and ecology.

While incorporating complex rules or a large number of agents expands the range of phenomena that can be studied, it also presents significant challenges. Mathematical analysis becomes substantially more difficult, and even simulations demand considerable computational resources.

Additionally, rules could be dependent on some parameters (e.g., reproductive rate in a population) that are challenging to estimate and not easily available in the literature.

Inaccurately representing individuals can introduce substantial errors when modeling real-world phenomena that depend on specific individual characteristics or interactions not observed and, therefore, not captured by the model.

3.1 ABMs for disease spread

By simulating the behavior of individual animals as agents, ABM can capture the complexities of animal populations, their interactions, and environmental factors that influence disease transmission [21].

Animal agents can possess a variety of attributes, including species, sex, location, or health status. This model can capture the heterogeneity of animal interactions while simulating individual behaviors. Additional aspects of animal life, such as herd dynamics or environmental factors, can be included. Moreover, both spatial and temporal dynamics of disease spread can be considered, enabling the exploration of diverse disease scenarios and the evaluation of possible control strategies.

Despite its power and flexibility, ABM presents some challenges. Disease transmission is influenced by factors such as contact between susceptible and infected animals, proximity, contact duration, and animal movement patterns like grazing or migration. To obtain useful simulations of real-world phenomena, all the above elements must be accurately integrated.

Precise data on animal populations, behaviors, and disease transmission parameters are essential for building a robust model.

The problem here arises from the fact that those parameters are often difficult to find in the literature or must be estimated. In addition, there may be some elements that are not known or not observed and, consequently, not included in the model, which makes model validation an important issue.

Furthermore, complex interactions or large populations can demand substantial computational resources.

As briefly discussed in the first chapter, obtaining high-quality data is fundamental. Indeed, incomplete or biased data can compromise model outcomes. ABMs often generate complex entities with numerous variables and interactions, making calibration, validation, and interpretation demanding.

Disease spread is inherently influenced by stochastic factors, such as environmental conditions and individual variations. Balancing the level of detail of the individual agent with the more general scale of the disease outbreak is crucial. Excessive detail can result in computational inefficiency, while insufficient detail may limit the model's ability to capture important dynamics.

Validating these models using real-world data is often made difficult by the internal model complexity, the stochastic nature of the studied event, and limited data availability. Despite these limitations, ABM remains a valuable tool for understanding animal disease dynamics and informing disease control strategies, but it may not always be the most suitable approach.

Chapter 4

Network-based models

Graph theory helps identify the structural properties and key nodes of a network (e.g., centrality, connectivity, clustering) while providing a mathematical framework to model and analyze complex networks. In this chapter, we will introduce the notions and concepts used throughout the remainder of the thesis. For a more comprehensive overview, other sources can be consulted [22, 23, 24, 25].

Network-based models offer significant advantages in several fields, including epidemic studies. Animal networks carry valuable information about the structure of a population and the interaction dynamics of individuals. In particular, livestock movement data (e.g., trade between farms, transhumance) enables the construction of realistic networks and the simulation of disease spread among animals.

Network Theory

A graph is a set of objects linked together. The objects are called nodes (or vertices), connected by links called edges. A network is any system that can be represented as a graph, but the two concepts are closely related and often used interchangeably.

More formally, an undirected graph G is defined by a pair of sets (V, E). V is a (countable) set of elements, called *vertices*, and E is a set of *unordered* pairs of vertices, called *edges*. We will refer to an edge that joins the vertices $i \in V, j \in V$ as (i, j). or (j, i). The vertices i and j are said to be *adjacent* or *connected* [22].

A directed graph D is defined as a pair of sets, (V, E). Here, V is a set of vertices (nodes), and E is a set of *ordered* pairs of vertices representing directed edges. The key difference between directed and undirected graphs is that in a directed graph, an edge from vertex i to vertex j does not guarantee the existence of the reverse edge j to i.

Graphs can be defined by the adjacency matrix $A = \{a_{ij}\}$, where

$$a_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{if } (i, j) \notin E \end{cases}$$

The adjacency matrix of an undirected graph is symmetrical $(x_{ij} = x_{ji})$ since if $(i, j) \in E$, then also $(j, i) \in E$. For directed graphs, instead, the matrix A is not necessarily symmetrical.

In real-world networks, connections often vary in strength or frequency. To describe this, we can assign a *weight* to each edge. This weight represents the intensity of the connection between two nodes. The adjacency matrix can then be adapted to this definition:

$$a_{ij} = \begin{cases} w_{ij} & \text{if } (i, j) \in E \\ 0 & \text{if } (i, j) \notin E \end{cases}$$

where w_{ij} is the weight assigned to the edge (i, j).

A *path* between two vertices i and j is a sequence of edges and vertices that begins at vertex i and terminates at vertex j. The path length refers to the number of edges included in this path.

The distance l_{ij} between vertices i and j is the length of the shortest possible path connecting them. There may be more than one shortest path between two nodes in a graph. If no path exists between them, the distance is conventionally set to infinity.

Average path length is a metric in network analysis that represents the average number of links needed to traverse all the possible pairs of nodes, following the shortest possible path.

It is a measure of the so-called *small world* property, where connections between any two nodes are often far shorter than the network's size. A lower average path length indicates a more efficient network with rapid dissemination, while a higher value suggests a slower and less connected structure on average.

Another statistic, called *diameter*, can be computed to express the size of a graph. Let l_{ij} be the distance between node i and node j, then the diameter of a graph G is defined as the maximum value of l_{ij} for any pair of distinct nodes i and j in the graph G.

$$d_G = \max_{i,j} l_{ij}$$

The *degree* k_i of a vertex i is defined as the number of edges that are connected to that vertex in an undirected network. Given the definition of adjacency matrix, we have:

$$k_i = \sum_j x_{ji} = \sum_j x_{ij}$$

In a directed network, we define the *in-degree* ($k_{in,i}$) as the number of edges arriving in a vertex, and the *out-degree* ($k_{out,i}$) as the number of edges coming from a vertex. Usually, the degree of a vertex in a directed network is the sum of the in-degree and of the out-degree, $k_i = k_{in,i} + k_{out,i}$.

$$k_{in,i} = \sum_j x_{ji} \qquad k_{out,i} = \sum_j x_{ij}$$

The *degree distribution* $P(k)$ is the fraction of nodes within a network with degree k. Thus, if we have n nodes in a network and n_k of them have degree k,

$$P(k) = \frac{n_k}{n}$$

Both the in-degree and out-degree distribution can be computed. These distributions can be classified as either homogeneous or heterogeneous. Homogeneous networks have a similar degree value for the majority of nodes, while heterogeneous networks display a wide range of degree values. A particular type of heterogeneous network is the scale-free network, where the degree distribution follows a power law. This means that the quantity $P(k)$ is a power of k, more precisely

$$P(k) \sim k^{-\alpha}$$

with α called scale parameter. Scale-free networks are characterized by the presence of a few highly connected nodes, known as "hubs," which possess significantly more connections compared to the average node in the network.

In graph theory, the importance of a vertex is called centrality. The degree of a node is one possible measure of centrality, along with others like closeness centrality, betweenness centrality, and the clustering coefficient.

Closeness centrality measures how easily a vertex can reach other vertices in the network. It is calculated as the average of the shortest path lengths from the vertex to all other nodes in the graph.

$$g_i = \frac{1}{\sum_{j \neq i} l_{ij}}$$

The betweenness centrality of a vertex is calculated as the number of shortest paths between all pairs of other vertices in the graph that pass through that particular vertex. This centrality measure reflects the importance of a node as a bridge between different parts of the network. If σ_{hj} is the number of shortest paths from h to j and $\sigma_{hj}(i)$ is the number of shortest paths from h to j that pass through i, the betweenness centrality is defined as

$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$$

Clustering refers to the tendency to form interconnected nodes. This means that if vertex i is connected to vertex j and vertex j is connected to vertex h, there's a higher chance vertex i will also be connected to vertex h. This phenomenon is often referred to as *triadic closure* and is a fundamental property of social networks.

The *clustering coefficient* measures the connectedness of a local neighborhood. The clustering coefficient of a vertex i is denoted as $C(i)$ and, if the degree of i is k_i and e_i is the number of links in the neighborhood, is defined as

$$C(i) = \frac{e_i}{k_i(k_i - 1)/2}$$

where e_i can be computed using the adjacency matrix A:

$$e_i = \frac{1}{2} \sum_{jh} a_{ij} a_{jh} a_{hi}$$

An important network statistic related to the weight of edges is the strength s_i of a vertex. It is defined as

$$s_i = \sum_{j \in V(i)} w_{ij}$$

where $V(i)$ is the set of neighbours of i.

The number of edges L in a graph with N vertex is between $N-1$ and $\binom{N}{2}$. The graph can be *sparse* if the number of edges is considerably less than the maximum number of edges, or *dense* if the number of edges is close to the maximum number of edges. Density is defined as

$$D = \frac{L}{\frac{N(N-1)}{2}}$$

If $D \ll 1$, then a graph is sparse.

When studying a network, an important element is the possibility of going from one vertex to another. If every vertex is reachable from any other vertex, then the network is said to be connected. A network can then be divided into different connected components. In particular, the size of the largest possible connected component called the giant connected component (GCC), is relevant.

If the network is directed, different definitions are needed. A directed network can be decomposed into a giant weakly connected component (GWCC), which is the giant component of the corresponding undirected graph, and other smaller components.

The GWCC can be decomposed again into a giant strongly connected component (GSCC) in which for every pair of nodes, there is a *directed* path linking them, the giant in-component (G-IN), grouping the nodes from which is possible to reach the GSCC following a directed path, and the giant out-component (G-OUT), grouping the nodes that can be reached from the GSCC following a directed path.

4.1 Network models for disease spread

Traditional epidemic models often assume random and homogeneous mixing throughout a population. This does not reflect reality, where people have specific contact patterns or different numbers of connections. Network epidemic models address this by incorporating heterogeneous social structures into the equation.

The spread of pathogens is heavily influenced by how individuals interact. The number, duration, and characteristics of these interactions, along with spatial heterogeneity, play a key role in how easily a pathogen can be transmitted. Therefore, network analysis is a valuable tool for understanding disease transmission. It allows us to describe contact networks, analyze their properties, and ultimately simulate how a disease might spread [1].

In network analysis, a vertex represents an individual unit depending on the scale of the study. This could be an individual human, animal, school, farm, or others.

Edges represent connections between these units that can potentially lead to disease transmission. These connections can be direct contact, for example, between humans, or indirect effect, like spatial proximity between farms.

For example, animal movement networks can be depicted as directed networks. In these networks, connections represent the movements of animals between locations (e.g., farms) and so have a specific direction. This means the adjacency matrix for animal movement networks will not be symmetric.

This network can be used to simulate the diffusion of a disease. Links between farms represent the potential for infection due to the movement of infectious animals. By introducing an initial number of infected animals or farms, we can analyze how the disease spreads through the network and evaluate the effectiveness of potential control strategies like vaccination (if applicable) or surveillance of farms.

While the described approach of static network analysis is powerful for simulating disease spread, it has limitations. It might not fully capture the dynamics of situations that change over time, such as variations or seasonality in animal movement patterns. To overcome this issue, dynamic networks are introduced [26].

Dynamic networks are constantly changing. Unlike static networks, the links between nodes are not fixed but can be formed, broken, or modified over time. For example, in a social network, friendships may be established and later dissolved.

Extending the previous notation, a dynamic graph G can be defined by a triplet (V, E, F), where V and E are vertex and edge sets as before. F is a function that goes from E to a set of pairs of natural numbers and indicates the times in which a certain edge is active. For example, $F(a, b) = (1{,}1), (3{,}5)$ indicates that the edge (a, b) exists when $t = 1$ and from $t = 3$ to $t = 5$.

The definition of connectivity and paths needs to be redefined. A *time-respecting* path [27] is a sequence of connections that link a starting point to an ending point.

The key aspect is that each connection in the sequence occurs after the one preceding it in time. A time-respecting path can, for example, spread a disease from the starting to the ending point, while a non time-respecting path cannot.

Some statistics described at the beginning of this chapter, like the *degree* of a node, can be adapted to the dynamic network by considering each temporal *snapshot* of the network as a static one. Others, such as closeness or betweenness, require some additional assumption about what is considered as *shortest* path [28, 29].

The temporal aspect of disease outbreaks, particularly factors like latency periods and infectious windows, is critical for understanding how they spread. Latency represents periods where individuals are infected by a pathogen but not infectious yet. Understanding the duration of this hidden window is essential and can help to implement effective interventions, such as isolation measures.

Furthermore, some diseases have a limited window of infectivity. This time frame, influenced by internal factors like immunity or external ones like treatment, determines how long an infected agent can transmit the disease to others. The length of this window denotes the pace of an outbreak and can help with quarantine periods.

By incorporating these temporal aspects, it is possible to create more accurate models to predict the development of an outbreak, allowing for targeted interventions to reduce disease spread.

As briefly introduced in the first chapter, when using animal movement data, it is not necessary to simulate contacts between different individuals or agents. Since the movement data already capture the spatial and temporal dynamics of animals, it is sufficient to analyze these patterns without explicitly modeling individual interactions.

However, it is true that compartmental models and agent-based models offer some interesting properties for modeling disease spread or population dynamics, which could be explored further.

In the previous chapters, we briefly introduced SIR, Agent-based and network-based models. All of them are powerful tools to simulate complex systems such as disease spread, even if they present some problems. In this work, we will employ a network-based model that seeks to exploit the advantages of both compartmental and agent-based models while mitigating their respective limitations.

The network-based model effectively captures the underlying social structure formed by animal movements. By analyzing the connections between farms, we can gain valuable insights into potential animal contacts and the elements that influence disease transmission patterns.

Compartmental models offer a foundational framework by classifying individuals into susceptible, infected, and recovered (or other) categories, facilitating the tracking of disease progression.

ABMs enhance this approach by introducing individual agents with unique attributes and behaviors, enabling more detailed modeling of interactions. This flexibility allows for the simulation of diverse animal populations or farm types by incorporating specific characteristics into the agents.

By combining these three approaches, we can leverage both the information provided by movement data and the flexibility of agent-based models while preserving geographical and temporal information. This eliminates the need to simulate movement and allows us to characterize different farms and track disease progression.

4.2 Combined Approach

In this thesis, we will use this approach to simulate the spread of pathogens. We will utilize existing animal movement data to create a contact network between different farms. The farms themselves will be the epidemiological units of interest rather than individual animals. We will study the spread of a pathogen over time by categorizing the various farms into compartments at different time points. Farms will be considered susceptible if they contain susceptible animals, while they will be considered infectious if they contain infectious animals. Consequently, any movement from an infectious farm to a susceptible farm could result in a new infection. The edges between the different nodes, therefore, represent a "contact" with respect to the pathogen. Our farms can thus be viewed as agents following specific rules with predetermined contacts. This approach allows us to adjust the infection or recovery parameters for different agents, thereby addressing various modeling and simulation needs without bearing the computational burden of simulating individual contacts.

We can, therefore, trace the evolution of the pathogen using compartmental models, customize the parameters, and simulate infections through agents while simultaneously utilizing the structure provided by the movement network.

Chapter 5

Diseases

A disease is a condition that disrupts the normal functioning of an organism, often causing specific signs and symptoms. Common causes of disease are external infections by pathogens, which are microscopic invaders like bacteria, viruses, fungi, or parasites, or internal dysfunctions. The severity of a disease depends on the specific pathogen, the immune response of the host, and other factors. Understanding both pathogens and the diseases they cause is crucial for developing effective treatments and preventative measures. In this thesis, we will mainly focus on animal diseases.

5.1 Animal diseases

The study of animal health has multiple implications. One key area of impact is animal welfare, where the focus is on keeping animals free from pain, hunger, and distress. This not only improves the ethical treatment of animals but can also contribute to their overall health and productivity. Healthy animals translate to a positive economic impact for the farm. Reduced illness and mortality rates lead to higher yields of milk, eggs, meat, or other products.

Additionally, healthy animals require fewer treatments, lowering overall production costs. Lastly, some diseases, called zoonoses, are transmissible between animals and humans, posing a significant public health risk. Studying animal health allows us to identify and monitor potential zoonotic diseases, preventing outbreaks and protecting human health.

5.1.1 Bubaline herpesvirus 1 (BuHV-1)

Bubaline herpesvirus 1 (BuHV-1) is a serious pathogen specifically affecting water buffalo (Bubalus bubalis) [30]. It falls under the alpha-herpesvirus subfamily, sharing similarities with other herpesviruses, including those that infect humans [31]. This virus poses health and economic concerns for water buffalo populations.

BuHV-1 mainly spreads through respiratory droplets expelled by infected animals, including calves. This transmission can occur through close contact or shared airspace in crowded barns [32].

A concerning aspect of BuHV-1 is its ability to enter a latent state after the initial acute phase of the infection. During this latency period, the virus goes dormant and becomes difficult to detect. However, the virus can be reactivated under stressful conditions for the animal, such as sun exposure or transportation, leading to new outbreaks within the herd [33].

While infected animals typically recover from the initial respiratory illness, BuHV-1 remains within them in a latent state for life. Some animals may not show any further signs of the virus, but others experience periodic reactivations. These reactivations can have significant economic consequences. Reduced milk production, abortion [34], and increased susceptibility to other infections are all potential outcomes of BuHV-1 outbreaks [35].

5.1.2 Q Fever

Q fever is an infectious disease caused by the bacterium *Coxiella burnetii* [36, 37, 38]. This pathogen can infect a wide range of animals, including mammals (also humans), birds, reptiles, and others. In animals, the disease usually does not cause visible symptoms, but when it manifests clinically, it can induce abortions or infertility and so can generate a significant economic impact for affected farms [39, 40]. Infected ruminants excrete *Coxiella burnetii* through different biological fluids, including milk, feces, and urine, as well as birth or abortion products [41, 42, 43], and it can survive several weeks in nature (e.g., in contaminated hay) and be spread by the wind. Animals can be infected through the fluids just mentioned, by inhaling the bacteria, or through tick bites.

It poses a significant threat to both livestock and public health [36]. Its ability to survive for extended periods in the environment facilitates its spread among animals, with inhalation being the primary route of transmission to humans.

In cattle farms, the economic impact of Q fever is substantial. Infected animals experience a decline in milk production and can suffer abortions [44]. These effects translate to significant financial losses for farmers. Additionally, human Q fever outbreaks can strain healthcare resources and lead to lost productivity due to illness.

Cattle, sheep, and goats are considered among the primary reservoirs for human contamination [45, 46, 36]. Aerosol is the primary mode of human contamination; it may occur directly from fluids of infected animals, and human-to-human transmission is extremely rare, although there have been cases of this through blood transfusions [47]. Many infected people have no symptoms, but it can cause fever, headache, diarrhea, and vomiting.

In some severe cases, it can result in pneumonia or hepatitis. Women infected during pregnancy may risk miscarriage or preterm delivery [41]. A small percentage of people ($\leq 5\%$) develop a more serious infection known as chronic Q fever. This can lead to endocarditis and may be fatal if not treated properly. In 2021, there was an outbreak of 14 cases among tourists in Italy caused by an infected calf. This incident highlighted the problem of animal-to-human transmission and the challenges of diagnosing the disease [48]. Due to its zoonotic effects and economic implications for livestock [40], Q fever represents a significant concern that impacts both public and animal health.

5.1.3 Disease selection

These diseases were chosen to address some of the concerns and research questions presented in Chapter 1.

Given that the buffalo movement network has not been studied in the literature, we ask whether this network can support the spread of a pathogen. To answer this question, it is appropriate to use a pathogen that primarily infects only buffaloes, has documented cases of infection, and is sufficiently infectious to observe its spread. The bubaline herpesvirus 1 (BuHV-1) meets these criteria and will therefore be considered in section 7.2 to analyze this research question.

After this, another concern is raised regarding the cattle-buffalo interaction in farms that raise both species. Since we want to determine if omitting this interaction in spreading models could lead to an underestimation of the epidemic risk, we need to consider a pathogen that infects both species and has an impact on the livestock. Given also the presence of documented outbreaks, Q fever meets our requirements and will therefore be modeled in section 8.3 to address our research question.

Chapter 6

Cattle

To study the relationship between cattle farms and the size of this phenomenon in the Italian territory, we extracted from the National Animal Identification and Registration Database (NDB) of the Ministry of Health the movements of cattle from 2017 to 2020. We gathered data regarding the origin and destination of movements, their characteristics, and the number of animals that moved.

6.1 Duplicate Movements

We briefly mentioned in Chapter 1 an issue concerning the collection of pasture movement data within the national database (NDB). In what follows, we will analyze this issue and introduce the solution we proposed for these analyses. We will also conclude with a possible modification regarding data entry that could easily resolve this problem. Considering the importance of the cattle industry in Italy and in order to analyze the potential impact of pastures within the Italian movement network, we decided to cross-reference the collected data with other sources to test their accuracy.

Specifically, we considered the 2020 cattle movement data from an Italian region, Piedmont. This region has a high presence of cattle, a large number of pastures, and precise data collection conducted by the ASL (local health unit who must authorize grazing movements).

Analyzing these two datasets reveals some differences, as shown in Figure 6.1, where the number of animals moved 'From' or 'To' Pasture is represented with weekly aggregation.

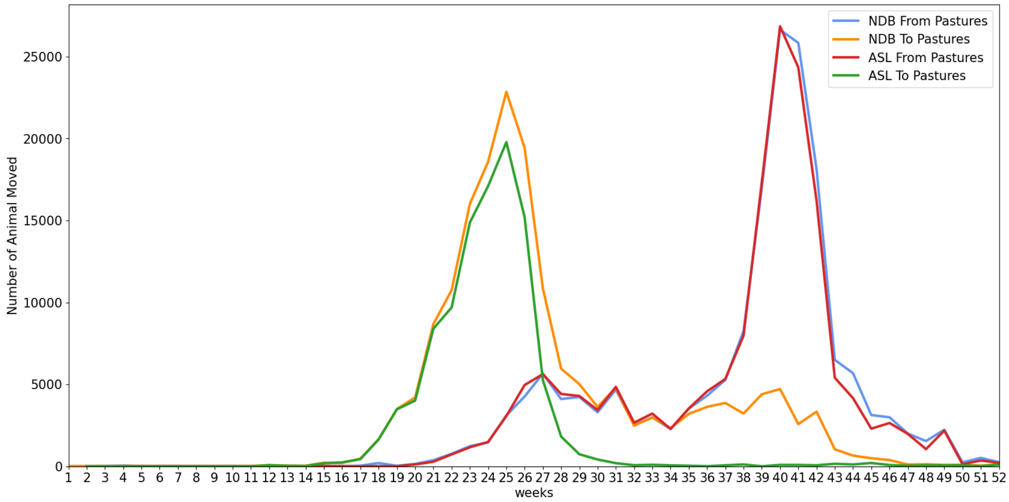


Figure 6.1. **Number of animals moved from or towards pastures.** Weekly Aggregation, Piedmont 2020.

This difference is more evident when looking at the cumulative sum of animals moved, shown in Figure 6.2.

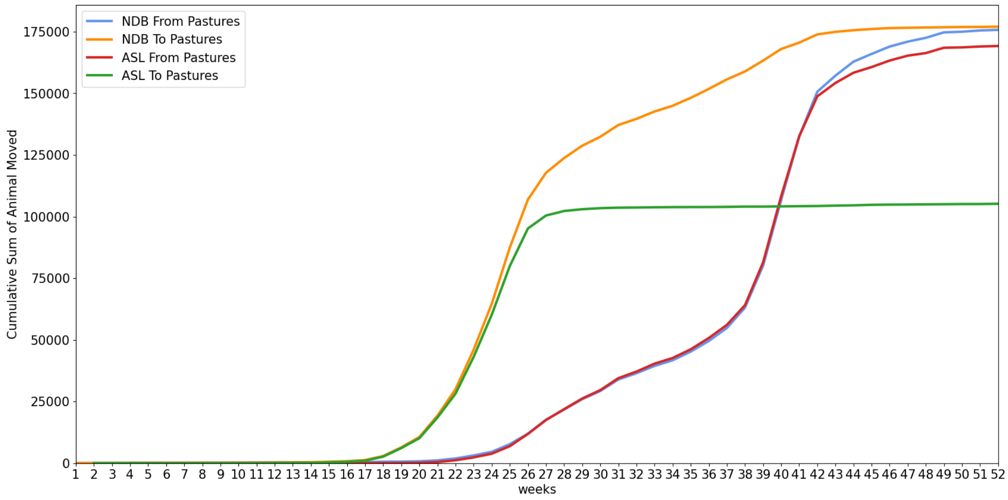


Figure 6.2. **Cumulative sum of number of animals moved.** Comparison of movement data between ASL and NDB. Weekly aggregation.

The movements ’To Pasture’, for example, reach significantly different values by the end of the year. The movements ’From Pasture’ show a similar pattern, but the ASL data include additional movements compared to the NDB data, specifically those ’Between Pastures’.

Regarding this point, we can analyze the ’Between Pastures’ movements, shown in Figure 6.3, along with the ’To Pasture’ and ’From Pasture’ movements from the ASL dataset. From the latter, movements with both the origin and destination as a ’Pasture’ node have been removed, as these movements were already included in the ’Between Pastures’ dataset. These movements are not present in the ’To Pasture’ movements and, therefore, have not been removed.

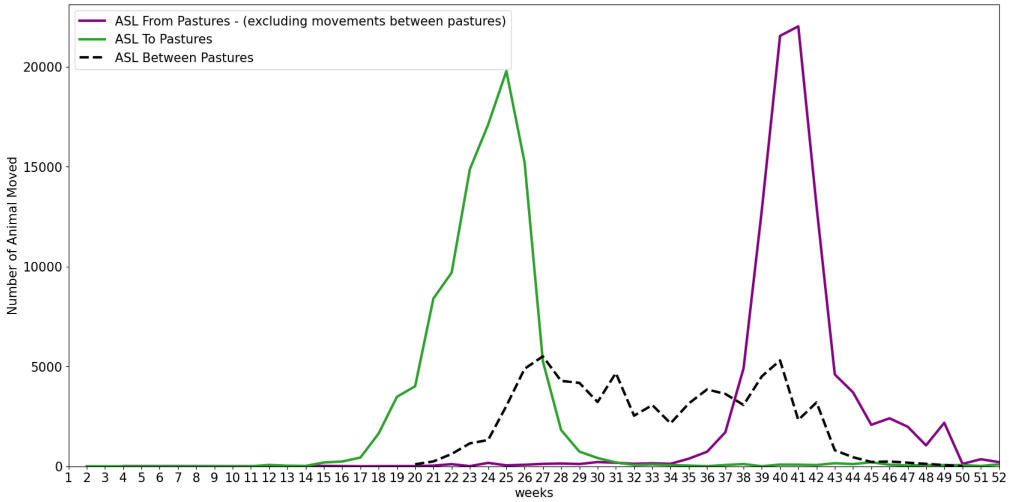


Figure 6.3. **Movements from, towards, and between pastures.** ASL Dataset. Weekly aggregation.

Looking at Figure 6.3, one can notice that the trend of ’Between Pastures’ movements is not dissimilar to that observed for the NDB dataset for ’From Pasture’ and ’To Pasture’ movements shown in Figure 6.1.

We examine this relationship more closely in Figure 6.4, where the number of cattle moved ’From Pasture’ and ’To Pasture’ with NDB data and ’Between Pastures’ with ASL data are represented. The number of cattle moved ’Between Pastures’ appears to coincide, particularly in weeks 30-35, with the other two curves. It is important to emphasize again that there is no movement in the NDB dataset that has both origin and destination as a ’Pasture’ node.

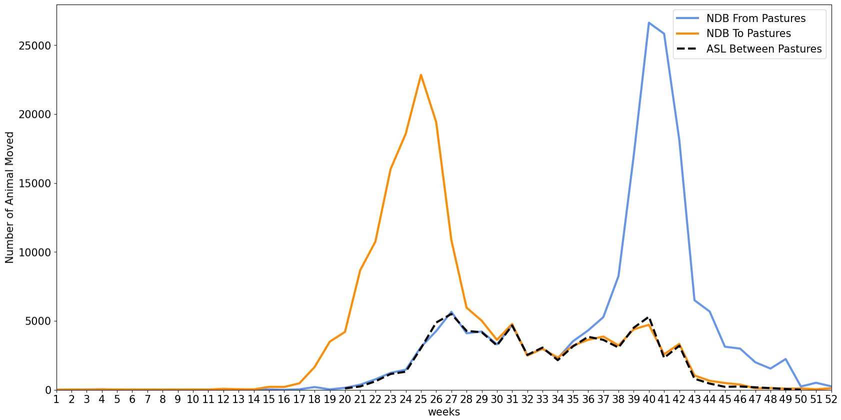


Figure 6.4. **Movements from, towards, and between pastures.** ASL and NDB Dataset. Weekly aggregation.

Moreover, the number of cattle moved ’From Pasture’ or ’To Pasture’ (NDB data) during the summer weeks appears to coincide, as can be seen in Figure 6.5 (showing weeks 28 to 37).

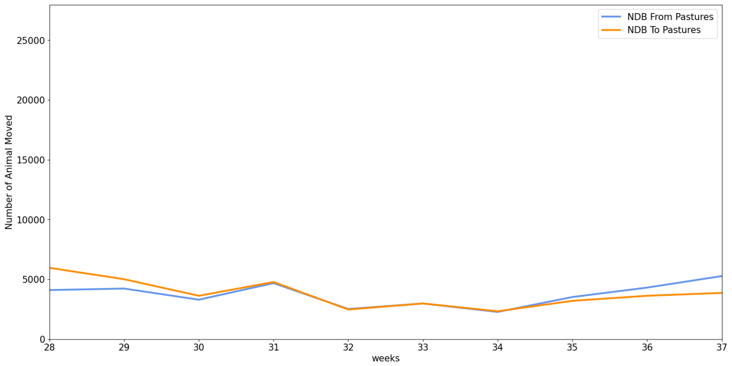


Figure 6.5. **Magnification of movements from, towards, and between pastures.** ASL and NDB Dataset. Weekly aggregation.

From these observations, one might think that what is recorded as a 'Between Pastures' movement by the ASL is entered in the NDB as a 'From Pasture' movement to a non-pasture node (e.g., a farm), and then from this same node, a 'To Pasture' movement is recorded in the same week or day. Some examples in the datasets support this hypothesis, where we have matching 'real' codes of arrival and departure pastures, but in NDB, there is an additional movement towards an intermediate node.

Finally, in Figures 6.6 and 6.7, the number of animals moved, and the corresponding cumulative sum is represented again after subtracting from the NDB movements the number of animals moved 'Between Pastures' by ASL, to avoid double counting. The differences that were previously present have now almost completely disappeared.

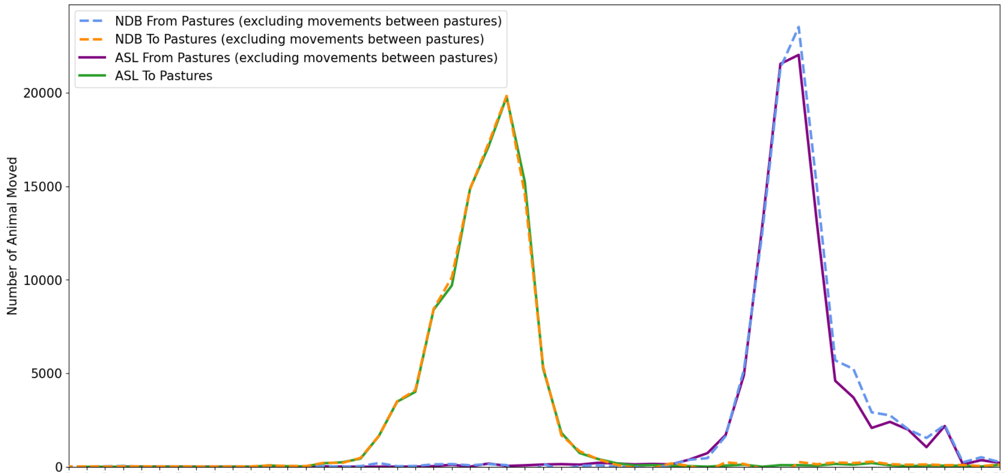


Figure 6.6. **Movements from, towards, and between pastures after correction.** ASL and NDB Dataset. Weekly aggregation.

It is, therefore, very likely that what is actually a movement between two Pastures in the ASL data appears as a 'double' movement in the NDB.

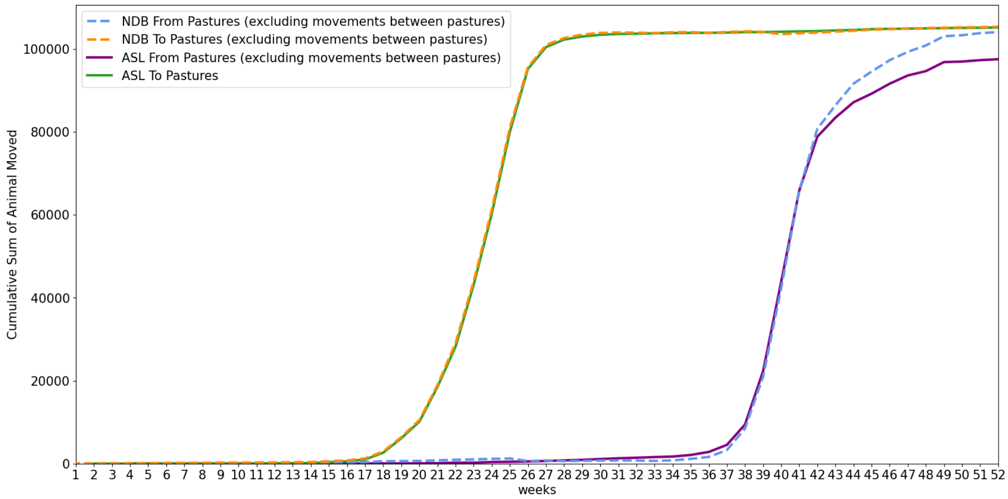
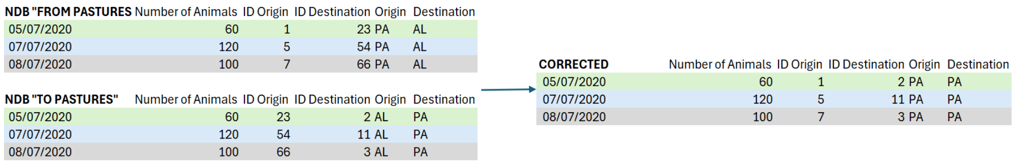


Figure 6.7. **Movements from, towards, and between pastures after correction.** ASL and NDB Dataset. Weekly aggregation.

Therefore, we propose the following correction, an example of which is visible in Figure 6.8:

- All pairs of movements such that the destination of one movement is the origin of another are identified and kept if both movements happen on the same day.
- If such movement follows the pattern $PA \rightarrow AL \rightarrow PA$, the pair of movements is deleted and replaced by a single $PA \rightarrow PA$ movement with the same number of animals moved.



NDB "FROM PASTURES	Number of Animals	ID Origin	ID Destination	Origin	Destination
05/07/2020	60	1	23	PA	AL
07/07/2020	120	5	54	PA	AL
08/07/2020	100	7	66	PA	AL

NDB "TO PASTURES"	Number of Animals	ID Origin	ID Destination	Origin	Destination
05/07/2020	60	23	2	AL	PA
07/07/2020	120	54	11	AL	PA
08/07/2020	100	66	3	AL	PA

CORRECTED	Number of Animals	ID Origin	ID Destination	Origin	Destination
05/07/2020	60	1	2	PA	PA
07/07/2020	120	5	11	PA	PA
08/07/2020	100	7	3	PA	PA

Figure 6.8. **Example of the proposed correction.**

In all the analyses that follow, the corrected movement data have been used as proposed.

This error could be due to an inability on the part of the farmers to correctly enter the $PA \rightarrow PA$ movement. Therefore, a modification of this could lead to a significant and important improvement in the available data.

6.2 Cattle Network Analysis

As previously introduced, cattle represent a significant productive reality both worldwide and in Italy. In Figure 6.9, we can see how there is an almost widespread presence of farms within the Italian territory. Indeed, all the farms that have been either the origin or destination of a movement between 2017 and 2020 are represented.

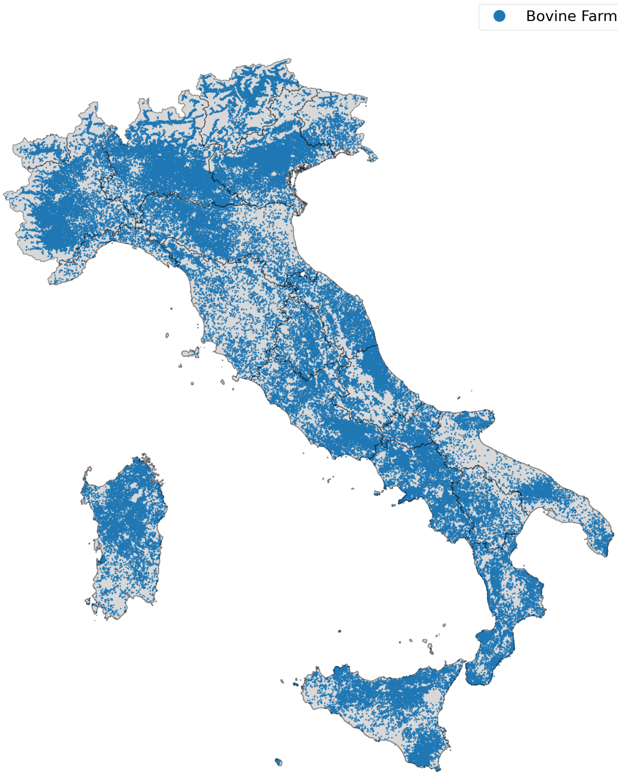


Figure 6.9. **Distribution of cattle farms in Italy.**

However, it is not only farms that give rise to animal movements in the Italian territory. Over the four years we have considered, there are other types of premises, and in Table 6.1, we can observe their distribution.

Table 6.1. **Number of premises from 2017 to 2020.** Cattle network.

Name	Number
Farms	140477
Pastures	11624
Others	1131

Given the importance of cattle in Italy and the magnitude of the phenomenon we are studying, numerous analyses and research have been conducted on this topic.

As justified in the introductory chapters on methodology, a network approach allows us to examine the intrinsic characteristics of this reality, uncovering some important elements that would not otherwise be observable. In what follows, we will, therefore, refer to the origins and destinations of a movement as nodes.

With a similar idea, Dutta *et al*, in [49], considered data from 2005 to 2009 regarding French cattle movement. They created different networks considering different spatial granularities and temporal windows. Additionally, they identified specificities related to the main types of animals and farms.

In this thesis, a weekly aggregation of data will often be used because weekly aggregation offers a good balance between data detail and sparsity. It maintains a level of detail sufficient to detect any variations or movement patterns that might not emerge with monthly aggregation. It is also important to note that movement data are manually uploaded by farmers, which could lead to unintentional human errors on the day of recording. Weekly aggregation also "corrects" this issue.

Moreover, daily aggregation might be too sensitive to rapid daily and random variations, such as weather events or local holidays. Again, weekly aggregation helps to contain these random variations, providing a more stable and representative view of the movements without losing the importance of short-term dynamics, as would happen with monthly aggregation.

In summary, weekly aggregation balances the need for detail and the ability to manage data, reducing the noise of daily fluctuations and capturing periodic patterns that would otherwise be lost.

Natale *et al.*, in [6], show the influence of the network structure on the dynamics and size of a hypothetical epidemic and give useful indications on the effects of targeted removal of nodes based on the centrality of premises within the network of animal movements.

Bajardi *et al.*, in [7], develop the previous idea. Through spatial disease simulations, they detect spreading paths that are stable across different initial conditions. Paths also allow the identification of nodes, called sentinels, which have a large probability of being infected and provide critical information on the outbreak's origin.

Cattle network

Table 6.2 shows the number of active cattle farms (those with at least one movement during the reference period) and movements between them, divided by year.

Table 6.2. **Number of active farms and movements by year.** Cattle network.

year	active cattle farms	movements between cattle farms
2017	99641	667315
2018	95657	654527
2019	93107	650713
2020	92491	642949

Figure 6.10 presents the aggregated indegree and outdegree distributions of the cattle network for each year, represented by orange circles and green diamonds, respectively. The four plots show similar, highly heterogeneous patterns. Indeed, we can see how we have some farms with a high indegree while the same is not true regarding the outdegree.

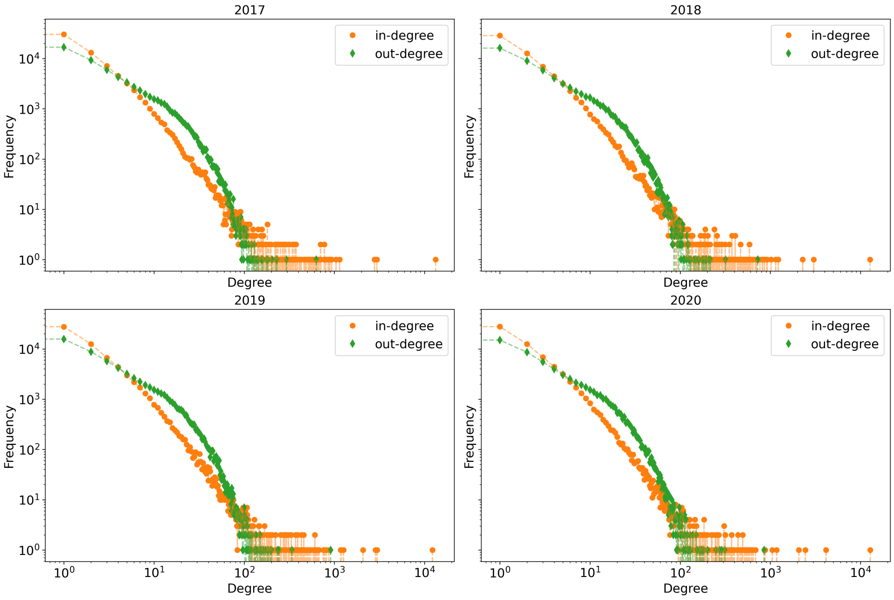


Figure 6.10. **Degree Distribution**. Cattle network, yearly aggregation.

This disparity might indicate a network structure where larger farms buy from smaller ones. This is shown also in Figure 6.11, where the indegree and outdegree for each month, averaged on the four years, are computed. Consistent with the previous observation, the average indegree also remains slightly higher throughout the months.

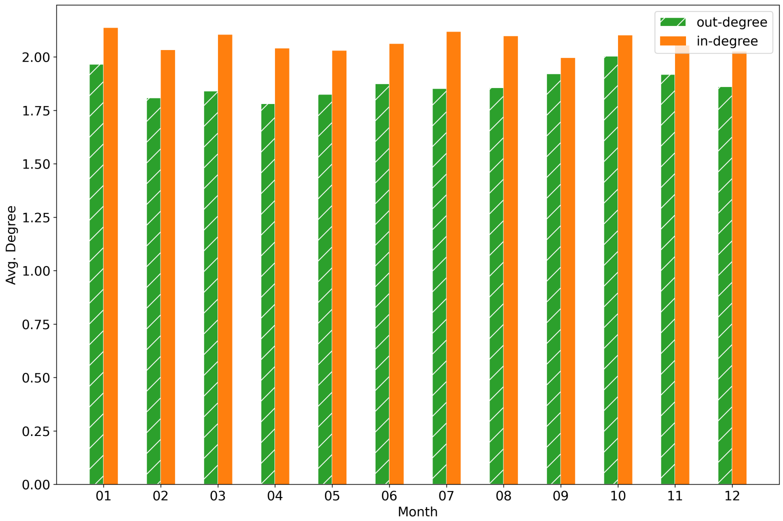


Figure 6.11. **Average degree per month.** Cattle network. Monthly average across all active nodes for each of the four years.

A high in-degree does not always correlate with a high out-degree. While this may be true for some farms primarily engaged in the buying and selling of animals, it does not apply to farms specializing in milk production (which often sell a large number of male cattle, resulting in higher out-degree and lower in-degree) or meat production (which tend to buy male cattle, leading to higher in-degree and lower out-degree) [50].

In Figure 6.12, the number of active edges and active nodes per week are plotted.

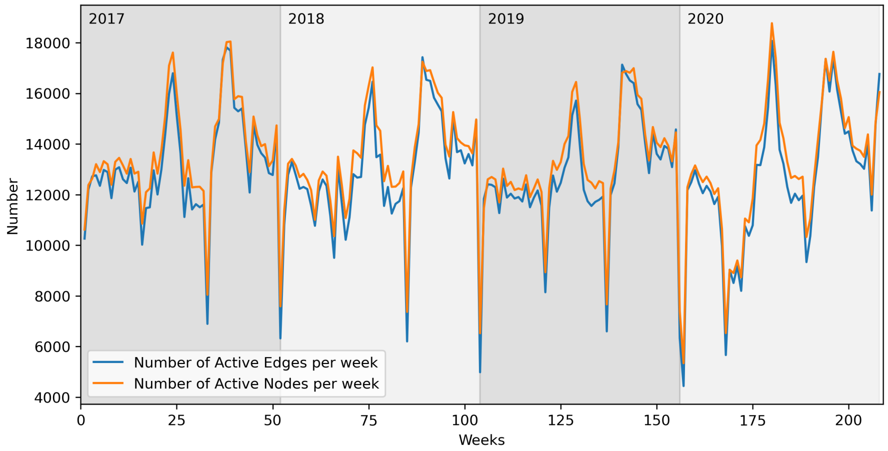


Figure 6.12. **Active nodes and edges.** Cattle network, weekly aggregation.

Several oscillations occur, caused by seasonal movements typical of the Italian scenario, such as summer grazing in pastures [6, 51, 7]. Despite the fluctuating values, the general behavior seems to be constant. Additionally, we observe that the number of active edges is similar to the number of active nodes, suggesting that while there is a significant number of active farms per week, they do not engage in a large volume of movements.

In Figure 6.13, this characteristic is evident. Here, we have plotted, for each week, the mean indegree and outdegree computed among the active nodes, along with the mean number of animals in each movement (weight). The mean degrees are computed on nodes that have that degree different from 0. On the main plot, we see that the average number of animals moved is relatively low. Additionally, on the zoomed plot, we observe how the indegree and outdegree attain modest values, never exceeding 2. The indegree is higher than the outdegree, suggesting that smaller farms are moving towards the same node.

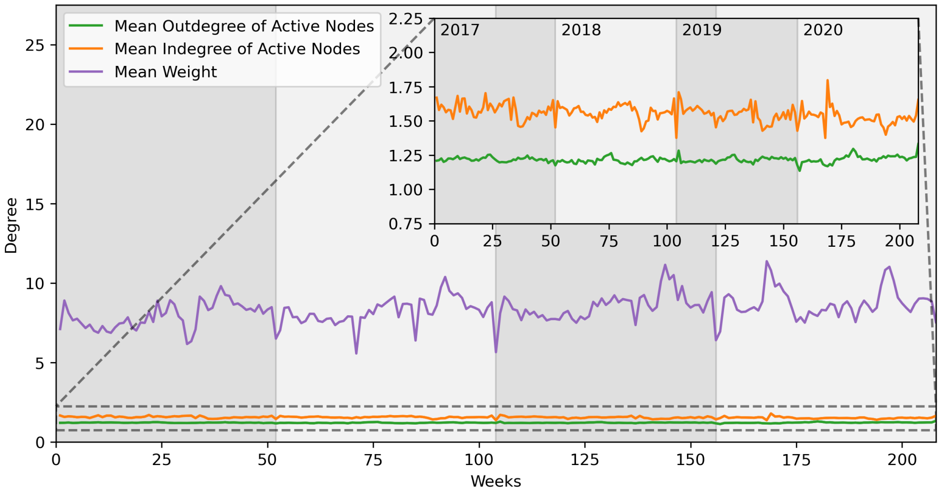


Figure 6.13. **Cattle network statistics.** Main plot: mean outdegree, indegree, and weight. Secondary plot: magnification of outdegree and indegree. Weekly aggregation.

That being said, the number of active farms is significantly large, as evident in Figure 6.9, where we highlighted the geographical location of all the active farms over the four years. These nodes also exhibit strong connectivity. Indeed, the largest weakly connected component, considering only movements from farm to farm each year, encompasses 80% of the farms (and more than 95% if we also consider the movements from and towards the pastures). The largest strongly connected component, instead, is composed of almost 50% of the nodes. This, along with the total number of movements, needs to be taken into account when analyzing the potential implications for the diffusion of a disease.

6.3 Role of Pastures

As briefly introduced in Chapter 1, the Italian cattle industry has unique characteristics that are frequently underestimated.

A significant portion of cattle is periodically moved to shared pastures, where animals from different farms converge. This raises important questions about how these common spaces influence the overall structure and dynamics of the Italian cattle movement network.

To be more precise, movements to pastures occur twice every year: cattle ascend to mountain pastures in early summer for fresher grass and return to the respective farms at the end of the hot season. These shared, open pastures, often without structural barriers, create interaction between animals from different farms and contact with wildlife, increasing the risk of disease transmission. Common grazing areas can also facilitate the spread of infectious diseases and exposure to different parasites, like ticks, which may act as disease vectors. Moreover, animals from farms with different pathogens can introduce new infections, potentially triggering outbreaks in the local population.

This additional risk might not be significant if it corresponded to a less prevalent phenomenon. However, even though we observed in Table 6.1 how the number of 'active' pastures between 2017 and 2020 is a small part of the total (8%), Figure 6.14 shows how the contribution of pastures to the total movements seems to be more relevant in some months of the year.

The solid orange line represents the total number of animals moved weekly. The dashed blue line, on the other hand, represents the total number of animals moved weekly, excluding all movements with the origin or destination being pastures. As can be seen, the blue line maintains a stable trend, while the orange line, seasonally, as one might expect given the nature of the phenomenon, reaches high values

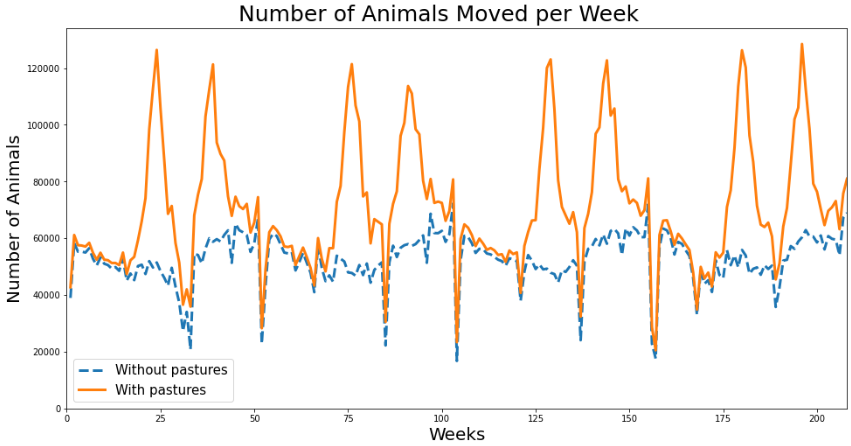


Figure 6.14. **Number of animals moved per week.** Cattle Network.

at the beginning and the end of the summer period. The movements to pastures do not replace the normal movements but increase the total number of animals moved and, therefore, also the risks to the animals, keeping in mind the considerations made earlier.

For this reason, in what follows, we will try to answer **what is the impact of cattle movements to pastures in the diffusion of pathogens?**, the first research question introduced in Chapter 1. Part of this section has been presented in poster form at *EPIDEMICS 8 - 8th International Conference on Infectious Disease Dynamics*[1].

---

[1] G. Zoppi, L. Savini, L. Candeloro, V. Colizza, M. Giacobini, *Analysis of pastures within the Italian cattle movement network and their role in epidemiological phenomena*, Poster Presentation, EPIDEMICS 8, 8th International Conference on Infectious Disease Dynamics, 2021

6.3.1 Impact of Pastures

To analyze the potential impact of pasture movements on the cattle movement network, we decided to use simulations of epidemiological phenomena, following what was described in the introductory chapters on models. We considered a fictitious pathogen and chose farms and pastures as the epidemiological units of interest, grouping these into two compartments: susceptible nodes and infectious nodes.

The idea is indeed to see whether the presence of pastures and their related movements have an impact on the spread of a pathogen and not to model specific characteristics of an existing pathogen, making the SIS model the most suitable for this purpose.

Infection rate	0.9
Healing rate	0.01
Infected nodes as t=0	[10%,1%] selected among all the nodes in the dataset
Dataset	[Complete, no pastures]
Start of simulation	Week 12

Table 6.3. **Parameters of simulation.** *Impact of pastures*

We set the parameters visible in Table 6.3 and used the actual movements that occurred over the four years to simulate the possible evolution of the fictitious pathogen. We started from week 12 to approach the period when pasture movements are present.

The average fraction of infected nodes (calculated on the total existing nodes) obtained in 50 different runs for the respective parameters is presented in Figure 6.15. The standard deviation of the obtained values is also highlighted in shaded tint.

The first row shows the results obtained by infecting 1% of the nodes at the start of the simulations. Similarly, the second row shows the results obtained by infecting 10% of the nodes at the start of the simulations.

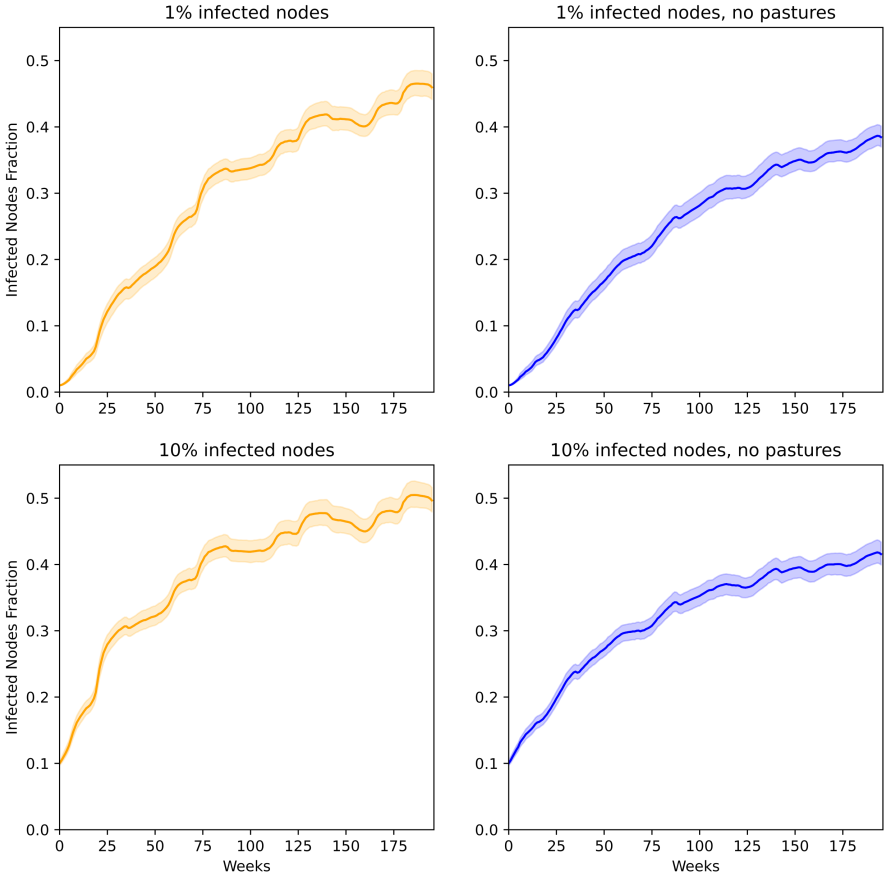


Figure 6.15. **Results of 50 runs.** Shaded tint represents the standard deviation. *Impact of pastures.*

The two columns differ regarding the set of nodes considered in the simulation. The first column (orange graphs) includes farms and pastures in the set, while the second (blue) excludes pastures.

In all four cases presented, we can observe an increase in the fraction of infected nodes over time. This is expected behavior as we have used a high probability of infection and a low probability of recovery. What we are interested in comparing is the value to which the different curves arrive at the end of the simulations (week 196).

If we compare the first column with the second, we can see that, for both rows, in the first case, there is a higher number of infected nodes.

Therefore, the fraction of nodes reached by our fictitious pathogen is greater when pastures are also included in the initial set.

This again indicates the importance of pastures within Italian movements, suggesting that this is a significant and non-negligible characteristic of the Italian context.

However, their impact on the spread of a pathogen could be solely linked to the numerous movements that pass through the pastures themselves.

6.3.2 Pastures as sources of infection

Unlike farms, pastures are almost always located in mountainous territories without fences or other types of boundaries. These lands are home to numerous wild species of various kinds, including mammals, birds, insects, and others. Some of these animals may carry diseases that can infect cattle.

Should this infection occur at the pasture, it could first spread among all the cattle present and then be transmitted to animals from different farms once the pasture period ends, bearing in mind that animals from different farms may share the same pasture or adjacent pastures.

For this reason, we study through the simulation of a fictitious pathogen similar to the one used previously, with parameters visible in Table 6.4. The structure is that of an SIS on a network, with farms and pastures as epidemiological units of study. However, the initially infected nodes are selected only within the pastures. The available nodes include both farms and pastures.

Infection rate	0.9
Healing rate	0.01
Infected nodes as t=0	[10%,1%] selected among the pastures
Dataset	Complete
Start of simulation	Week 12

Table 6.4. **Parameters of simulation.** *Pastures as source of infection*

In Figure 6.16, the results of 50 different simulations (per set of parameters) are presented. In orange, we find the average fraction of infected nodes, while the shaded area indicates the standard deviation of the values achieved.

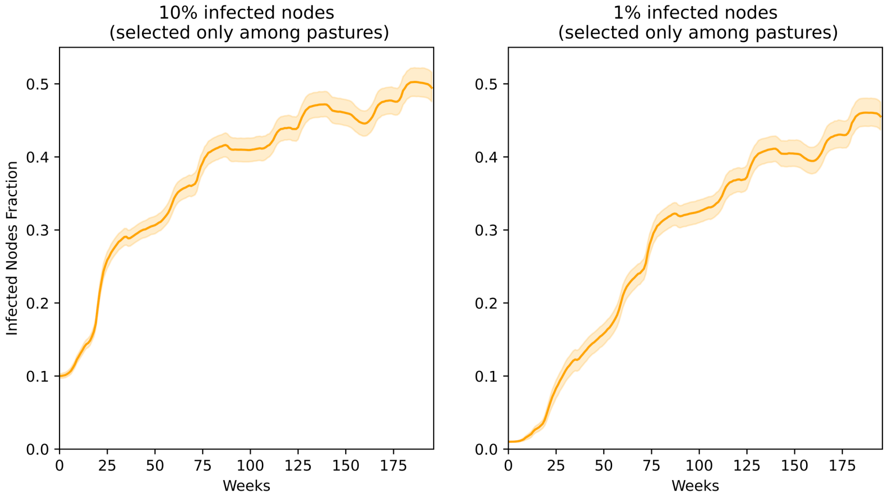


Figure 6.16. **Results of 50 runs.** Shaded tint represents the standard deviation. *Pastures as source of infection.*

For both graphs, which differ by the initial number of infected nodes, an increasing trend of the curve can be observed.

This suggests that a pathogen originating from pastures can survive and spread within farms. It should be noted again how a single pasture can infect multiple farms through the return of animals at the end of the grazing period.

We have thus observed how pastures play a dual fundamental role within the Italian cattle movement network. They are indeed the origin or destination of a large percentage of animal movements at the beginning and end of the summer period. Moreover, the presence of various wild species in pastures could be an important source of diseases for the movement network. For this reason, the role of pastures cannot be disregarded.

These conclusions further emphasize the importance of the discussion regarding duplicate movements. Having proven the central role that movements to pastures have, correcting existing data and making appropriate changes to the data collection system would result in more accurate analyses.

Chapter 7

Buffalo

Farm animals are of huge relevance in Italy, and a considerable number of articles have been published on these subjects. Nonetheless, as we saw in the previous chapter, there could be some peculiarities in the country that have not been thoroughly studied and could lead to inaccuracies in the analysis of phenomena such as disease outbreaks, transmission dynamics, or others.

The topic of pastures is certainly very important and is a distinctive feature of the Italian landscape. However, it is not the only interesting characteristic of livestock farming in this country.

In Italy, there are approximately 6,000 buffalo farms, and this species represents a significant productive reality. Moreover, more than two-thirds of buffalo farms also raise cattle. Henceforth, we will refer to these farms as *mixed* farms.

Therefore, there is a presence of shared spaces between buffalo and cattle, leading to close contact between the two different species, potentially allowing the transmission of certain pathogens among the animals. Additionally, these farms that house both species might need to be considered differently within the analysis of the cattle movement networks.

As briefly discussed in Chapter 1, the buffalo movement network in Italy has been consistently overlooked in analyses of the cattle movement network despite the sharing of spaces described above. Additionally, even the buffalo movement network alone has never been thoroughly studied or documented in the scientific literature.

Given this lack of research, it is essential to examine this reality in detail to assess its role and relevance. Understanding the role of the buffalo network in disease transmission could provide valuable insights for future epidemiological studies and contribute to the development of more effective control measures. Figure 7.1 illustrates the geographical distribution of the farms exclusively housing cattle with blue dots and the distribution of *mixed* farms in Italy with red squares.

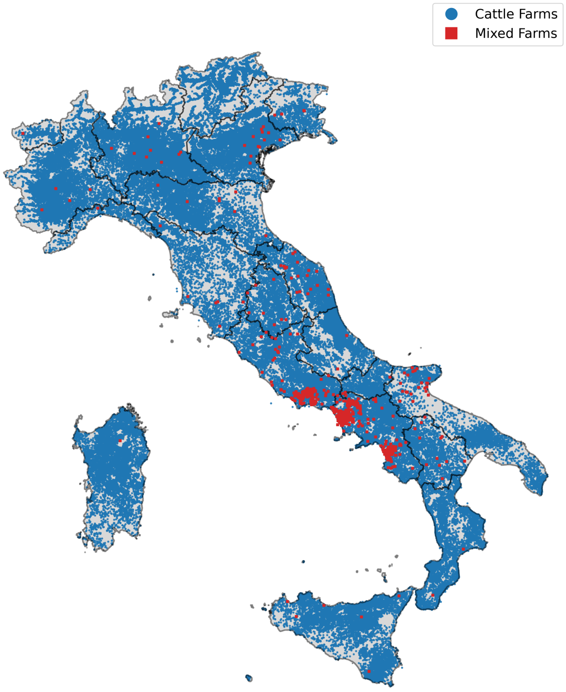


Figure 7.1. **Geographical distribution of cattle farms in Italy.** Blue dots: Farms that raise only cattle. Red squares: *mixed* farms. Shapefile of Italy from GADM data under a CC-BY license https://gadm.org/license.html.

This coexistence of cattle and buffalo within the same environment raises questions about the potential impact it may have on the transmission and survival of different pathogens between these two species. Hence, the often overlooked movements associated with the buffalo industry may significantly contribute to the dissemination of pathogens.

Given these considerations, it is fundamental to understand the impact resulting from the omission of the presence of buffaloes, leading to potential inaccuracies in the evaluation of the overall situation.

To estimate the error made by disregarding the presence of buffalo, we will proceed in two different steps. First, we will consider the buffalo network alone, analyzing it in section 7.1. After this, in section 7.2, we will introduce a pathogen that affects this species and observe whether **the buffalo network supports the spread of a pathogen**, addressing one of the research questions introduced in Chapter 1. Part of this work was presented orally at the *ModAH - Modelling in Animal Health* conference in 2021[1].

Then, in Chapter 8, we will consider the cattle and buffalo networks together, connected by the *mixed* farms. We will then observe the spread of a pathogen on this multilayer network.

---

[1] G. Zoppi, M. Scudeler, L. Savini, L. Candeloro, L. Bertolotti, M. Giacobini, ”Italian Buffalo Movement Network - Qualitative Analysis and Impact on Disease Diffusion”, Oral Presentation, ModAH - Modelling in Animal Health conference, 2021

7.1 Buffalo network analysis

Similar to the previous analyses, we extracted the data from the Italian National Animal Identification and Registration Database (NDB) of the Ministry of Health. From this database, we extracted information regarding the trade of buffaloes in Italy and the related characteristics of the departure and arrival points of these movements. We then further skimmed the data by restricting our analysis to a 4-year time span, from 2017 to 2020.

Therefore, we ended up with two datasets. The first one contains all the movements of buffaloes between 2017 and 2020, with at least one between the origin and the destination in Italy. Actually, the number of movements from or towards a foreign country is negligible. For every movement we have, in addition to the origin and destination already mentioned, the day of the movement, the animal species, and the number of animals moved.

In Table 7.1, we can see the magnitude of the phenomenon that we are treating. Compared to the cattle movements in Table 6.2, we can see that the two phenomena are of very different magnitudes. The buffalo movements are few but increasing. This makes it particularly interesting to observe whether, at this stage, this network can support the spread of a pathogen, both because of its small size and its current growth.

Table 7.1. **Number of active farms and movements by year.** Buffalo network.

year	active buffalo farms	movements between buffalo farms
2017	1243	4791
2018	1242	5327
2019	1237	6372
2020	1305	8075

The other dataset collects all the origins and destinations present in the other dataset. In this, we have a unique (and anonymous) ID for every unit, the facility type (e.g., farm, slaughterhouse, pasture, resting barn, or others), and some geographical identifier (e.g., municipality and coordinates). In Figure 7.2, we can see the productive class of each farm plotted with different colors. We then filled each region with the color of the most present productive class.

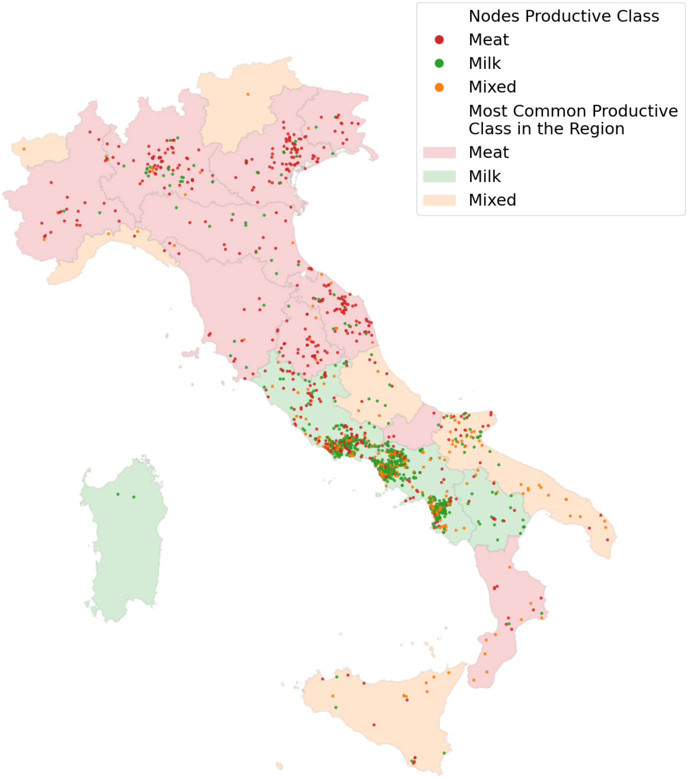


Figure 7.2. **Buffalo Farms.** Geographical distribution and most common productive class in each region.

The majority of the farms are located in the central-southern part of the Italian peninsula, and among those, dairy activities are predominant. The rest of Italy has a smaller number of farms, mostly related to meat production activities.

As briefly mentioned before, a widespread phenomenon among Italian companies is the sharing of available space between animals of different species. In Figure 7.3, we tried to visualize the impact of the above-mentioned situation to understand how widespread it actually is. To be more specific, we plotted all the farms present in our dataset. We then cross-referenced this data with information on cattle farms and highlighted in red those companies that breed both of these species. Numerically speaking, around two-thirds of companies handle both of them.

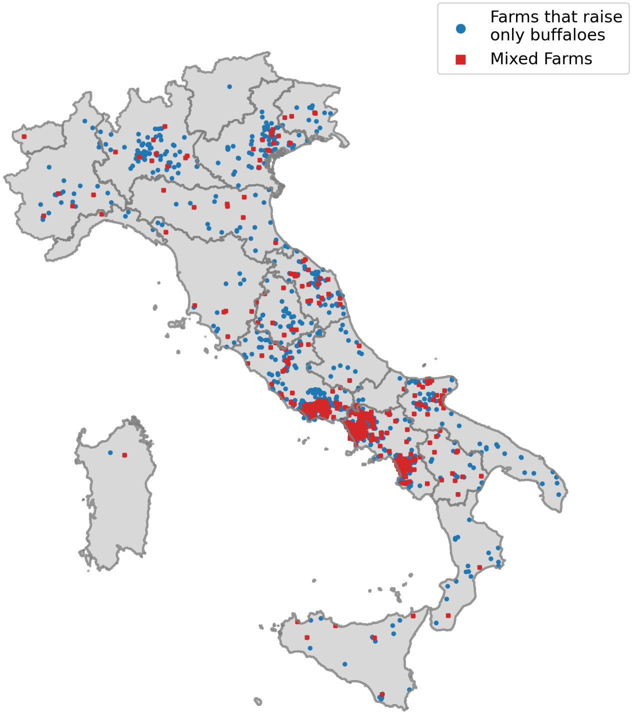


Figure 7.3. **Geographical distribution of buffalo farms in Italy**. Blue dots: Farms that raise only buffaloes. Red squares: mixed farms. Shapefile of Italy from GADM data under a CC-BY license https://gadm.org/license.html.

It is possible to see how the disposition of the farms is similar, with a higher density of those in the southern region of the Italian peninsula, therefore excluding a possible regional-restricted presence of *mixed* farms.

In order to maintain consistency in the notation with the remainder of the section, where the movements will be modeled within a network, we will refer to the origins and destinations as nodes.

As briefly hinted in the previous section, we wanted to model those movements as a network and perform a statistical analysis to look for the characteristics and peculiarities of this reality. To perform a first exploratory analysis, we considered the entirety of the data at our disposal, aggregating them into a single network, not considering, in other words, the temporal dimension of our data but keeping track of possible repeated movements between nodes that is we allowed multiple edges to have the same end nodes.

We discussed how the nodes were geographically distributed in a heterogeneous way in the Italian peninsula, with the majority of them located in the southern part. Accordingly, we sought to test whether this geographic separation was also reflected in a possible separation of movements. So, we divided our 4-year aggregated network into relevant components, analyzing the sizes of the strongly and weakly connected components.

In Table 7.2, we can see how, in both cases, there is one big component, and the remaining nodes are organized in a high number of really small components. The biggest strongly connected component includes more than 500 nodes, while the second largest strongly connected component is just a set of four nodes.

	SCC Size	WCC Size
1.	568	2853
2.	4	4
3.	3	3
4.	3	3

Table 7.2. **Number of nodes in the four biggest connected components.** 4-year aggregated buffalo network.

The same reasoning applies also to the sizes of the weakly connected components. From this, we can conclude that the network has a big core of strongly connected nodes across the peninsula, which allows us to rule out an impact of geographic separation on the movements and a relevant number of small components weakly connected to the above-cited core.

The indegree and outdegree distributions for the buffalo network are also heterogeneous, as is visible in Figure 7.4, where we can see plotted on a log-log scale frequency distribution of the indegree and of the outdegree of the nodes, represented by orange circles and green diamonds, respectively. Differently from the bovine case, we can find some farms that have a high outdegree. This is opposite to what we can see in Figure 6.10, but it can be pointed out that the values here are much smaller than the ones related to the bovine network.

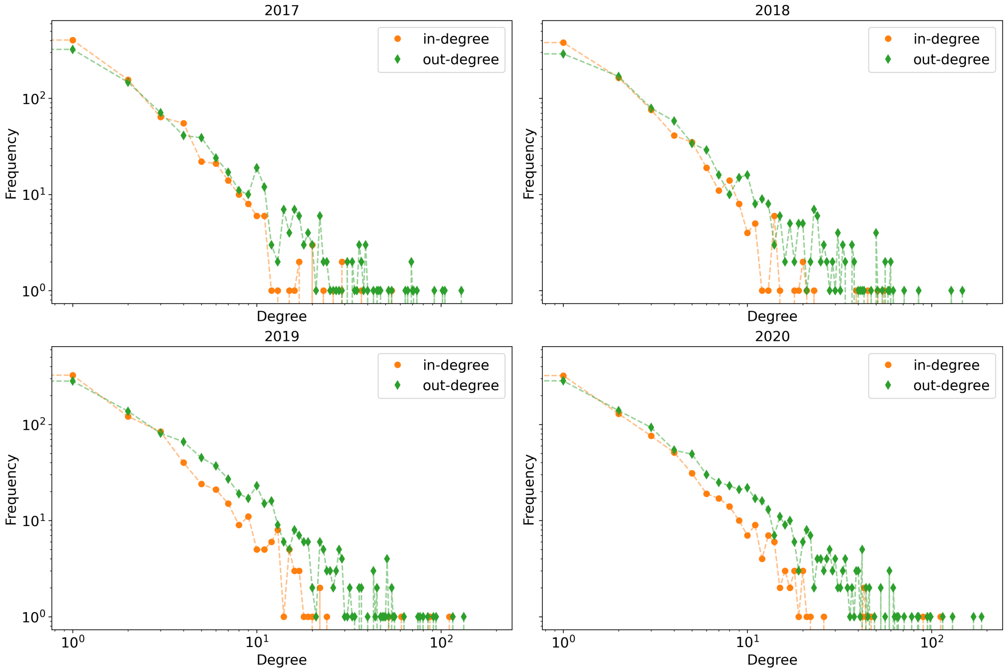


Figure 7.4. **Degree Distribution.** Buffalo network, Yearly aggregation.

Figure 7.5, where the indegree and outdegree for each month, averaged on the four years, are computed, is consistent with the previous figure. Indeed, also the average outdegree is slightly bigger than the average indegree for each month. Even for the buffalo, a high indegree does not always correlate with a high out-degree for reasons similar to those observed in the bovine case.

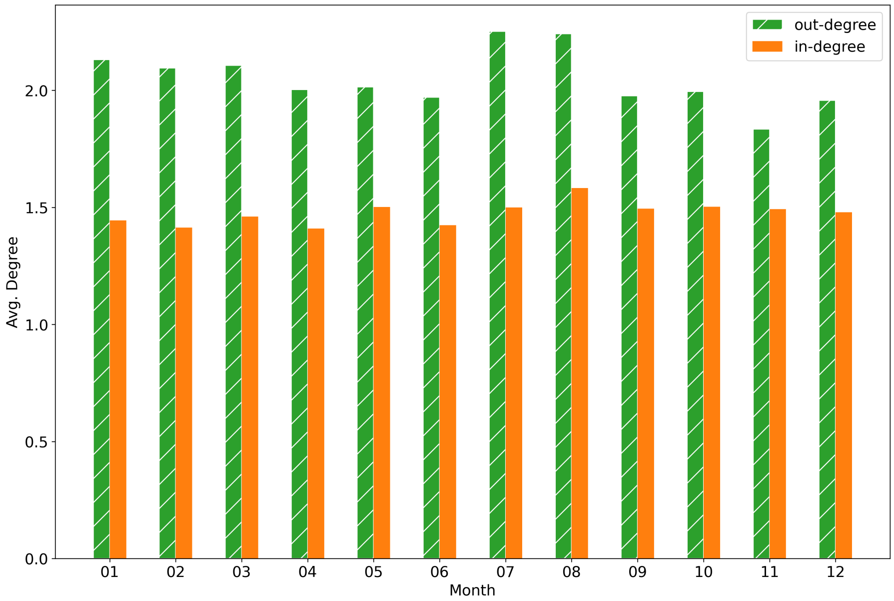


Figure 7.5. **Average degree per month.** Buffalo network. Monthly average across all active nodes for each of the four years.

After studying the aggregate network, we decided to consider the temporal dimension of the movements. So, we created different temporal networks according to different time aggregations and compared their statistical results. To be more specific, we aggregated the movements by day, week, and month. Similarly to the cattle network, we tried to see which choice creates temporal networks that are neither too sparse, so where the temporal snapshots comprehend a number of movements that is too small nor too dense, that is, by aggregating, we lose some potentially useful insights. Again, the weekly aggregation was the right balance between the two.

Thus, in Figure 7.6, we can see the number of active edges and nodes per week. An upward trend is visible, confirming again the increasing activity of these farms.

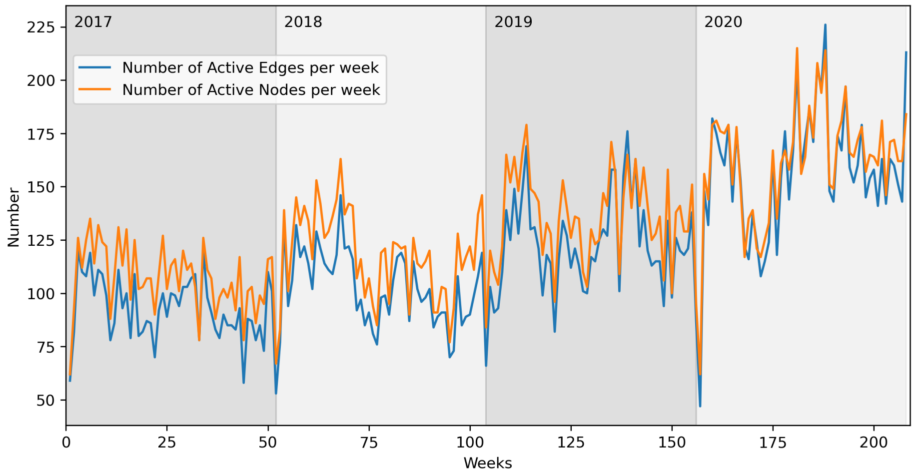


Figure 7.6. **Active nodes and edges.** Buffalo network, Weekly aggregation.

In Figure 7.7, we observe the average outdegree, indegree, and number of animals moved (weight).

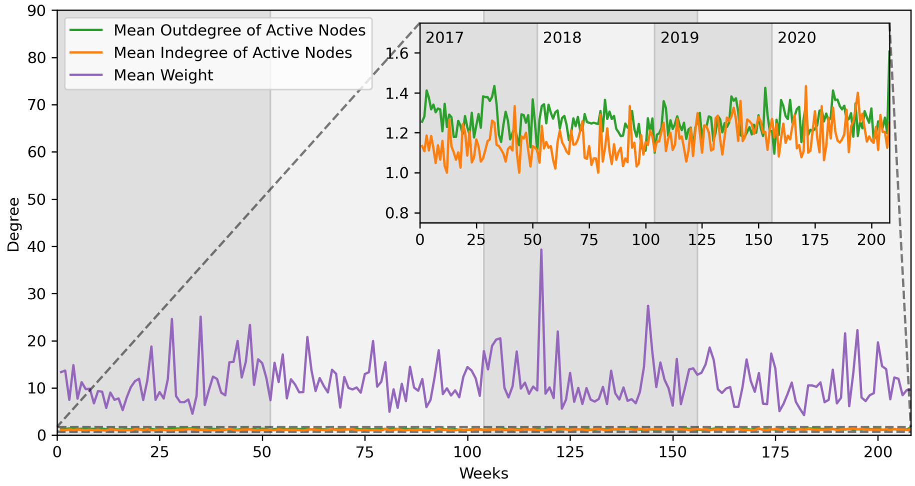


Figure 7.7. **Buffalo network statistics - Weekly aggregation.** Main plot: mean outdegree, indegree, and weight. Secondary plot: magnification of outdegree and indegree.

In the background plot, we can see how the number of animals moved is bigger on average than the cattle network. In the foreground, we observe in a zoomed plot how the degrees have low and similar values, once again an indication of a network with low activity.

These nodes are less connected compared to those in the bovine network. Indeed, the largest weakly connected component encompasses more than 95% of the nodes, but the largest strongly connected component is composed of a bit more than 25% of all the nodes.

7.2 Simulation of BuHV-1 diffusion

After analyzing the buffalo movement network in Italy, we will now consider the research question presented in Chapter 1, namely, **can the buffalo network support the spread of a pathogen?**

To address this, we introduce a pathogen that affects buffalo farms and can lead to various symptoms, including fever, conjunctivitis, rhinitis, respiratory difficulties, and abortions in pregnant females, resulting in harm to the animals' health and economic losses for the farms. This pathogen is the Bubaline Herpesvirus type 1 (BuhV-1), a virus belonging to the Herpesviridae family, as seen in section 5.1.1.

It is a highly infective pathogen transmitted by close contact or aerosol. After the initial infection, it becomes latent after two weeks on average. The infected animal can return infective if it is exposed to stress (transportation) or exposed to the sun.

7.2.1 SILS Model

Again, our epidemiological units will be the farms. However, since the pathogen is highly infective, we can safely suppose that if one animal is infected, then all the animals inside the farm will be infected in a short period of time (less than a week).

We consider the farms as units of our model, and we divide them into different compartments:

- Susceptible farms are farms that contain susceptible animals.
- Infected farms are farms that contain infected and infectious animals.
- Latent farms are farms that contain animals with latent viruses.

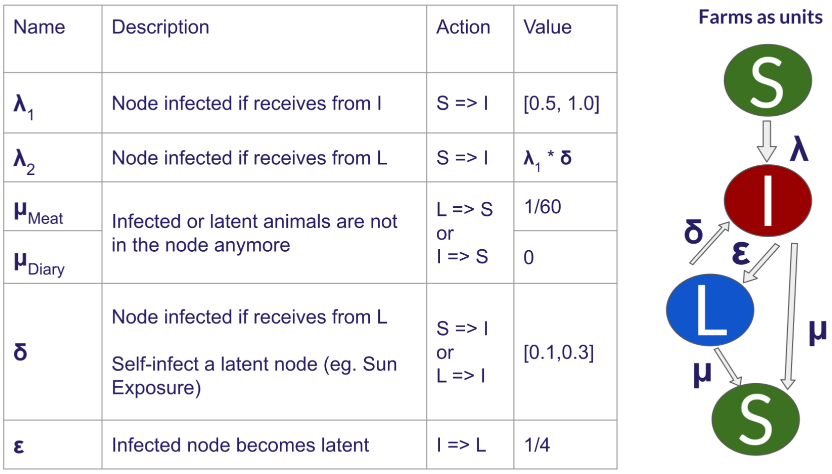
A susceptible farm can be infected if:

- receives animals from an infected farm (it receives infectious animals),
- receives animals from a latent farm (it receives animals with a latent virus, but since the transportation is stressful for the animal, symptoms may reappear).

Given that the animals can return to be infectious if exposed to the sun, a latent node can become an infected node during the periods when the animals are moved to pastures or open spaces (from May to August).

In detail, a farm (latent or infected) can return to be susceptible if all the animals are sold, and so the farm is "cleared". The average life of a male buffalo (meat production and not breeding) is one year. The average life of a female buffalo (milk production) is 20 years. Given that we are aggregating our movement data by week, we set probabilities of returning susceptible to $\frac{1}{60}$ for meat farms and 0 for milk farms.

A graphical explanation of the above is available in Figure 7.8, where we also describe the different parameters used in the simulations.



Name	Description	Action	Value
λ_1	Node infected if receives from I	S => I	[0.5, 1.0]
λ_2	Node infected if receives from L	S => I	$\lambda_1 * \delta$
μ_{Meat}	Infected or latent animals are not in the node anymore	L => S or I => S	1/60
μ_{Diary}			0
δ	Node infected if receives from L Self-infect a latent node (eg. Sun Exposure)	S => I or L => I	[0.1,0.3]
ε	Infected node becomes latent	I => L	1/4

Figure 7.8. **SILS Parameters and Compartments.**

7.2.2 Simulation Results

In Figure 7.9, we can see the results of 50 different runs for each set of parameters and 10% of farms infected at week 0. The red line shows the number of infected farms each week, while the red shading indicates the minimum and maximum values obtained in the runs. The same applies to the blue line and shading for the latent farms. The purple line represents the sum of infected and latent farms, which is the total number of farms where the pathogen is present.

In all four graphs, it is evident that the number of farms where the pathogen is present increases as the weeks progress. This indicates that the herpesvirus is capable of surviving within this network.

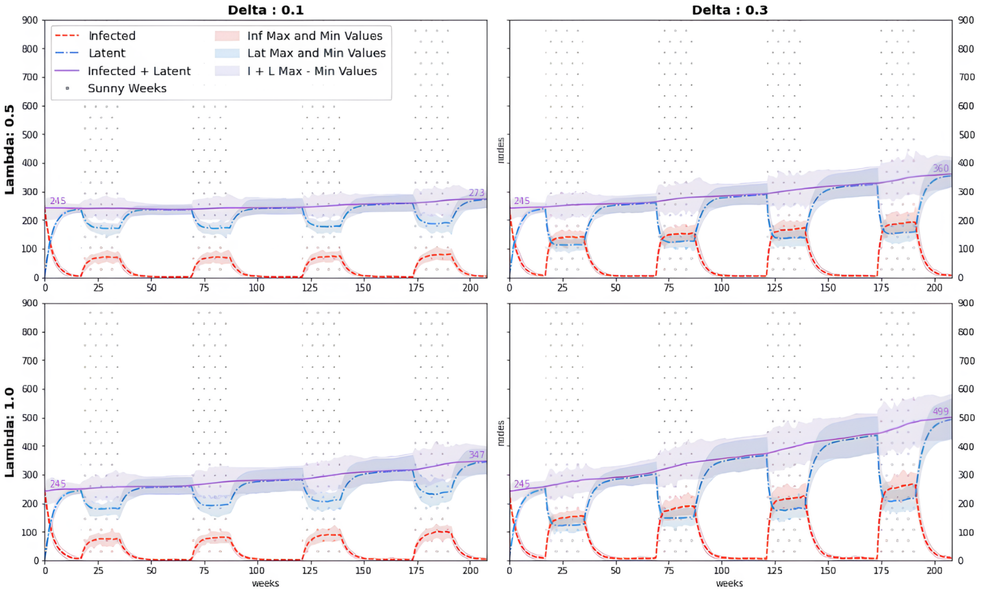


Figure 7.9. **Simulation Results.** Buffalo Farms. 10% of farms infected at week 0

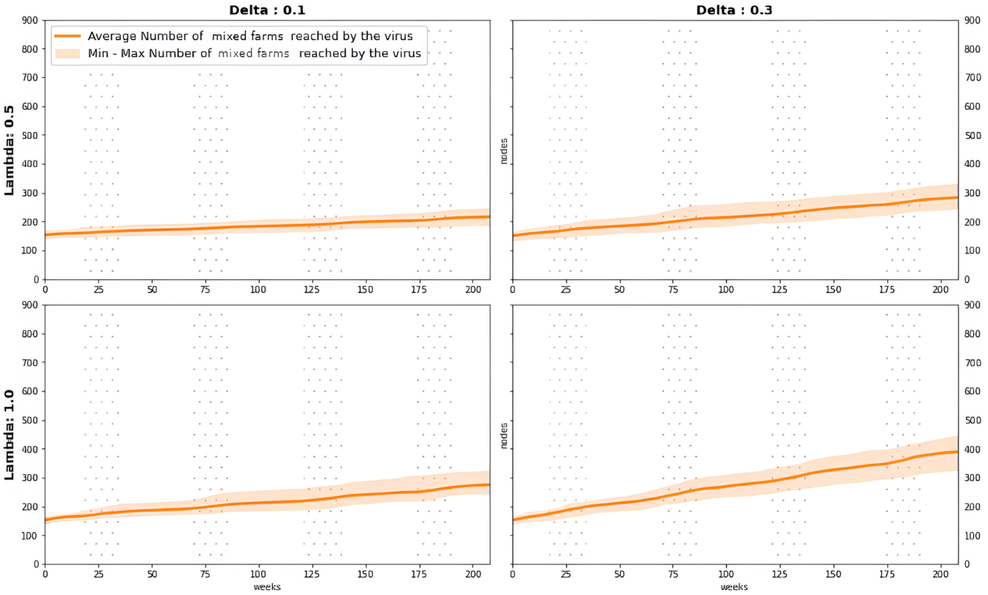


Figure 7.10. **Simulation Results.** Mixed Farms. 10% of farms infected at week 0

In Figure 7.10, we can observe how the pathogen spreads within the *mixed* farms. We used the same parameters as in the previous figure. Here, too, we can see that the number of farms where the pathogen is present increases over time. The behavior we observed across all farms can, therefore, also be seen within the *mixed* farms.

In Figures 7.11 and 7.12, we can see the results of 50 different runs for each set of parameters and 1% of farms infected at week 0. The comments are similar to those for the figures just seen but with a different order of magnitude. The y-axis takes on different values compared to the previous figures. Thus, even by initially infecting a smaller number of farms, the pathogen is free to spread and is capable of surviving within the buffalo movement network.

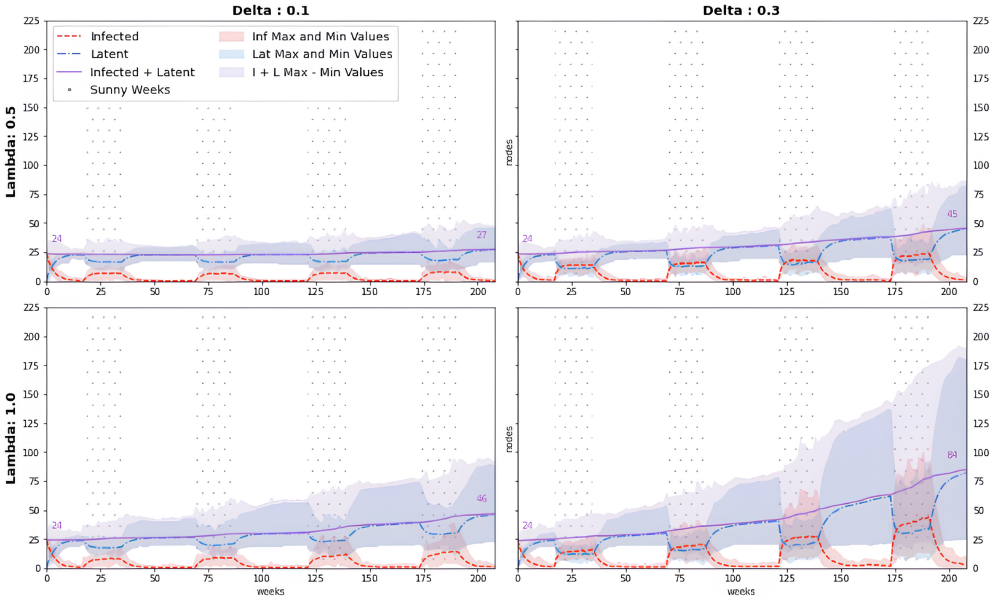


Figure 7.11. **Simulation Results.** Buffalo Farms. 1% of farms infected at week 0

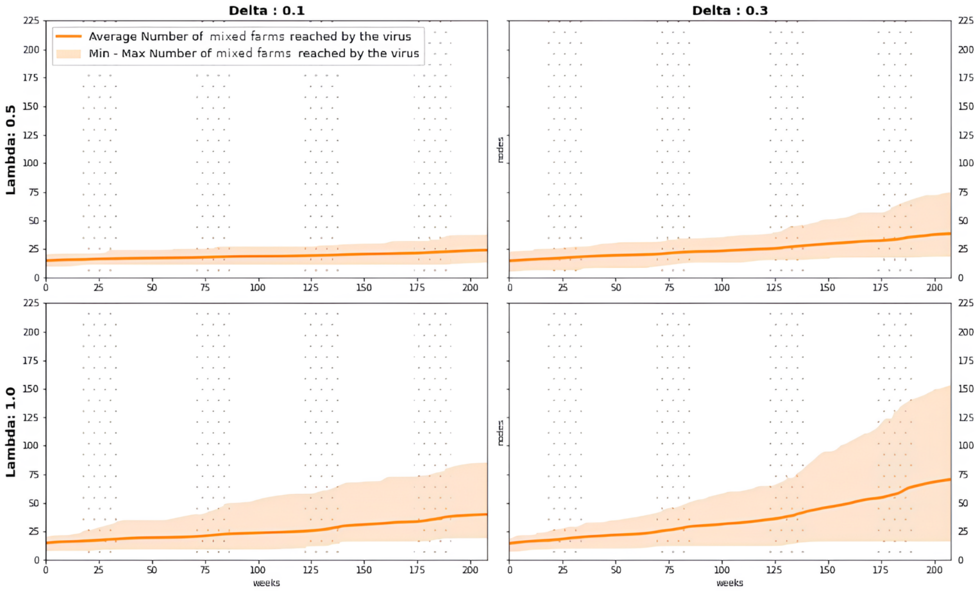
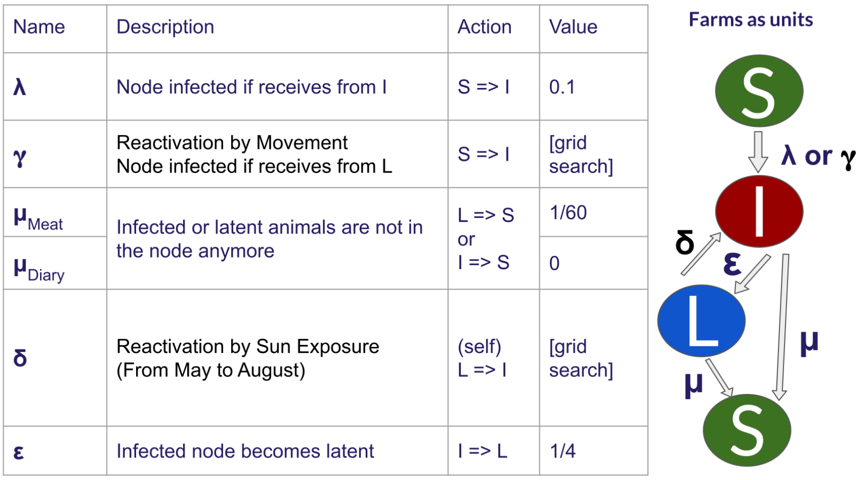


Figure 7.12. **Simulation Results.** Mixed Farms. 1% of farms infected at week 0

7.2.3 Grid Search

From an analysis of the different simulation results, an important element that could impact the outcomes is the parameters used, especially those related to the reactivation of the pathogen. Therefore, we repeated the analyses using the same model but with a lower probability of infection and conducted research regarding the reinfection parameters. We then replaced the λ_2 parameter shown in Figure 7.8 with a γ parameter to study the types of reinfection separately. All the parameters can be found in Figure 7.13.



Name	Description	Action	Value
λ	Node infected if receives from I	S => I	0.1
γ	Reactivation by Movement Node infected if receives from L	S => I	[grid search]
μ_{Meat}	Infected or latent animals are not in the node anymore	L => S or I => S	1/60
μ_{Diary}			0
δ	Reactivation by Sun Exposure (From May to August)	(self) L => I	[grid search]
ε	Infected node becomes latent	I => L	1/4

Figure 7.13. **SILS Parameters and Compartments.**

In Figure 7.14, we can see plotted in a heatmap the fraction of runs that had at least one infected farm in the last three weeks. The vertical axis represents the variation of the γ parameter, which corresponds to reactivation due to stress from movement. The horizontal axis represents the variation of the δ parameter, which corresponds to reactivation due to sun exposure. The latter appears to have a greater impact on the two reasons for reactivation.

In fact, a shift in colors can be observed as one moves to the right. This shift is less noticeable when moving from top to bottom.

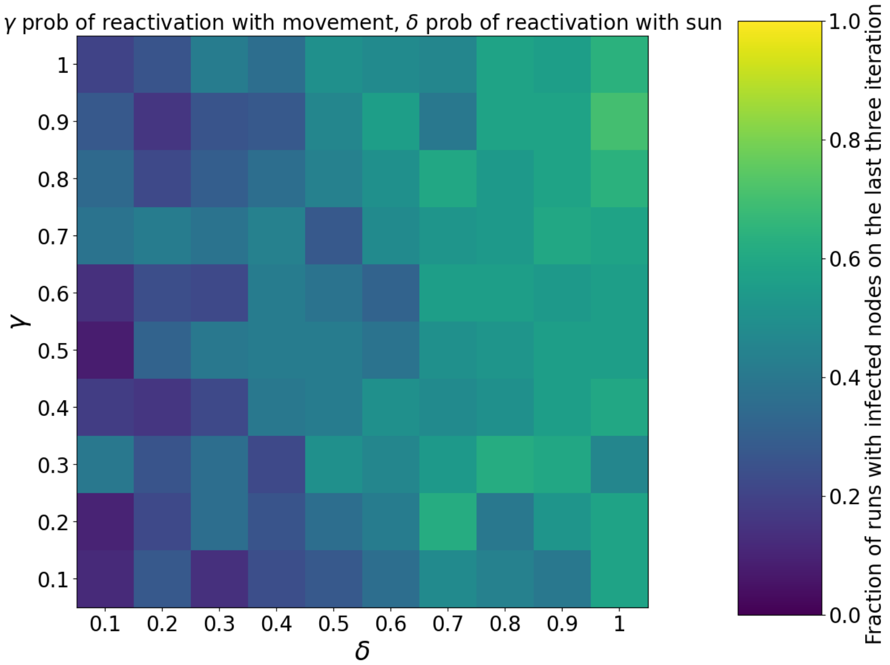


Figure 7.14. **Heatmap. Survival of the epidemic for different values of γ and δ parameters.**

In Figure 7.15, we have another heatmap with the same axis as before. In this case, the values plotted are relative to the attack rate, a variable defined as

$$\text{attack rate} = \frac{\text{number of new cases in the population at risk}}{\text{size of the population at risk}}$$

In this case, as well, a greater impact seems to be due to sun exposure compared to the stress from movement.

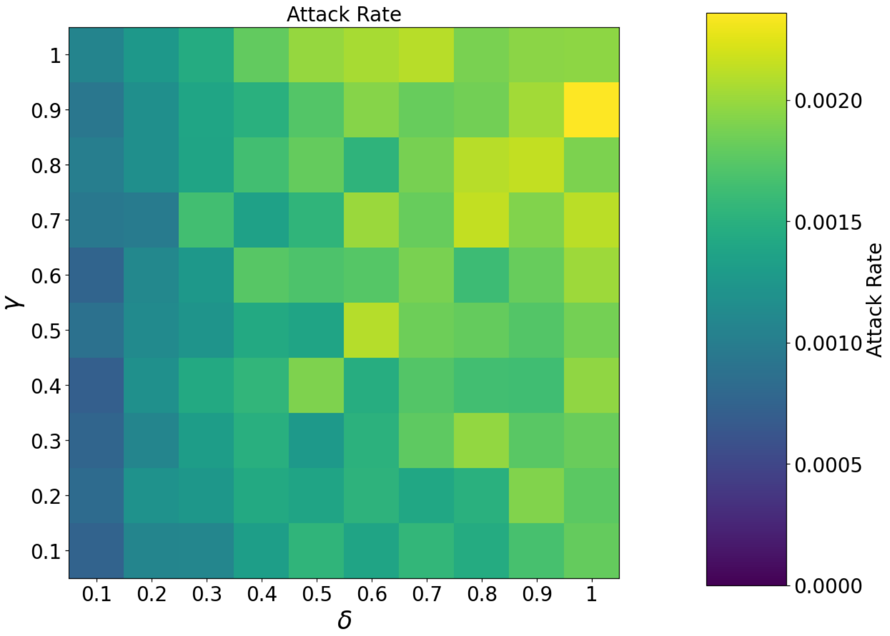


Figure 7.15. **Heatmap. Attack rate for different values of γ and δ parameters.** All the farms considered.

Regarding the attack rate, it is interesting to observe the difference between buffalo-only farms and mixed farms. In Figure 7.16, you can see the different values that this statistic takes. The values are significantly higher for mixed farms.

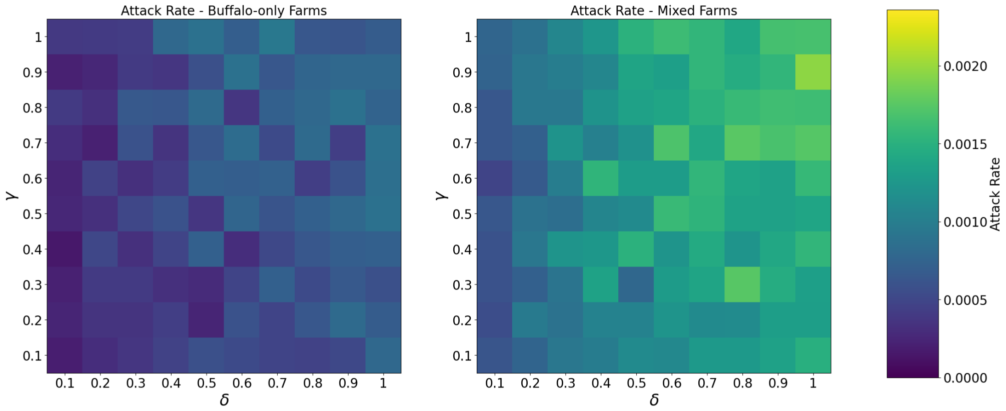


Figure 7.16. **Heatmap. Attack rate for different values of γ and δ parameters.** Left: Farms that raise only buffaloes. Right: Mixed farms.

This difference is more clearly observable in Figure 7.17. The fact that the attack rate is higher in this type of farm underscores the critical role they play. This increased rate could be attributed to greater activity or higher exposure to pathogens, which might not only be significant for the buffalo network but also for the cattle network. The presence of shared spaces between the two species could facilitate the spread of pathogens from one species to another, amplifying the impact on both networks. This highlights the importance of monitoring and managing mixed farms closely to prevent potential cross-species transmission and to safeguard the health of both buffalo and cattle populations. This will be analyzed more thoroughly in chapter 8.

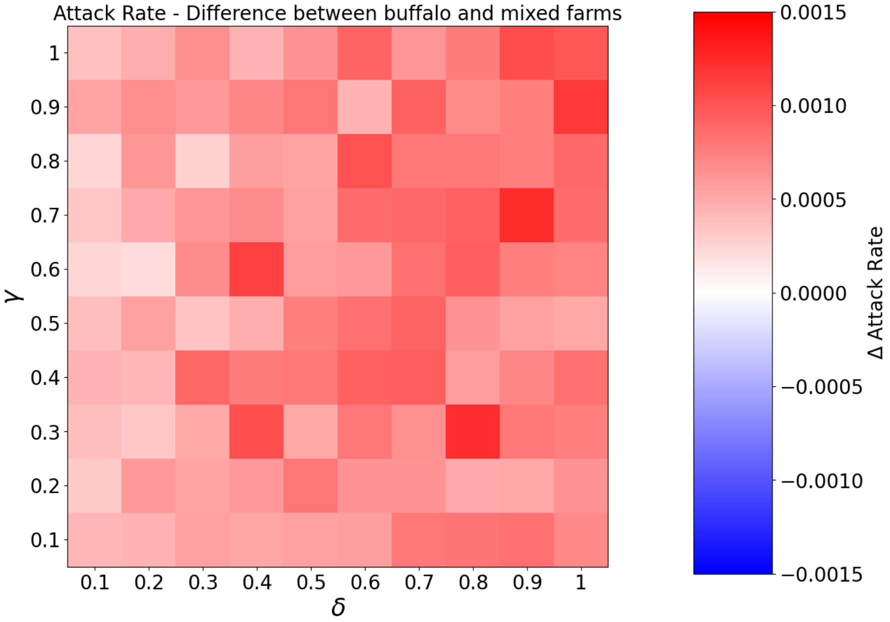


Figure 7.17. **Heatmap. Attack rate for different values of γ and δ parameters.** Difference between Buffalo-only farms and Mixed farms.

Therefore, The Italian Buffalo Movement Network constitutes a non-negligible entity. We have shown that the network is capable of supporting the diffusion of pathogens like BuHV-1, underlying its importance in the context of animal health.

This network should also be carefully considered when examining the spread of diseases across other networks, such as the bovine one. The interconnected nature of these networks should be evaluated with attention. Looking forward, it could be interesting to study the role of sun exposure in the reactivation of latent pathogens.

In the following, we will analyze the buffalo and cattle networks together using a multilayer approach. This integrated analysis would provide a more accurate and comprehensive understanding of how diseases move across species and networks, leading to better prevention and control measures.

Finally, future research should consider other bovine pathogens that might find a reservoir in the buffalo network. Understanding the potential for these pathogens to persist and spread within the buffalo population could be key to preventing their transmission to cattle and managing broader epidemic risks.

In summary, the Italian Buffalo Movement Network could be a critical factor in the epidemiology of livestock diseases. Its influence extends beyond buffalo populations, impacting the broader cattle network and necessitating an integrated approach.

Chapter 8

Interaction on Mixed Farms

Disregarding the buffalo network in disease analysis could lead to significant errors in understanding and managing the spread of pathogens. The buffalo network is linked with the cattle network, and by overlooking the role of buffalo in the transmission dynamics, we risk missing critical pathways through which diseases can spread. Without considering the buffalo network, prevention and control measures might be less effective, potentially leaving gaps that could allow for the resurgence or spread of infections. In this chapter, we will try to see **if disregarding this cattle-buffalo interaction network in spreading models could lead to an underestimation of the epidemic risk**.

The content of this chapter is based on an article titled *The Hidden Role of Buffalo Trade Network in Bovine Epidemic Spreading*, G. Zoppi, L. Savini, L. Candeloro, V. Colizza, M. Giacobini, currently submitted and under review at *PLOS One* journal and part of what follows was presented at *GeoVet 2023, International Conference of Spatial Epidemiology, Geostatistics, and GIS applied to animal health, public health, and food safety*[1].

---

[1] On the interaction between bovine and bubaline trade network in epidemic spreading G. Zoppi, L. Savini, L. Candeloro, V. Colizza, M. Giacobini, "On the interaction between bovine and bubaline trade network in epidemic spreading", Oral Presentation.

As previously mentioned, the coexistence of cattle and buffaloes in *mixed* farms establishes a potential link between the buffalo and cattle networks, providing a pathway for the transmission of diseases between the two. Nodes, where both species are present, can help the spread of diseases facilitated by airborne transmission or through vectors (e.g., ticks).

Multilayer networks are a collection of multiple layers, where each layer represents a different type of relationship or interaction among the same set of nodes that are connected through links (called inter-layer links) that allow for the modeling of complex systems with multiple dimensions of interaction. In the following, we adapted this concept to represent the unique characteristics of livestock movement networks. Specifically, each layer considers trades of a single species (cattle or buffalo), while the connections between layers reflect possible interactions happening on the same farm between those species. In Figure 8.1, we depict this concept.

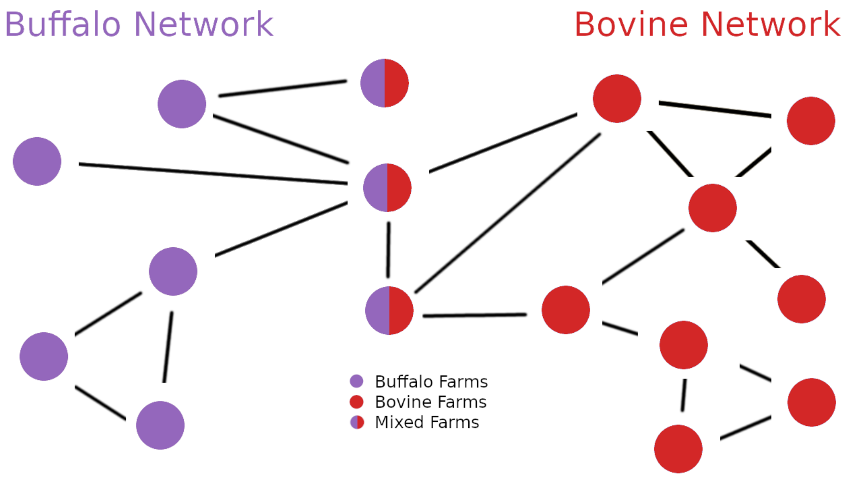


Figure 8.1. **Cattle and buffalo networks interaction.**

To model this aspect in our disease model and enable the potential transmission of infection between the buffalo and cattle networks, we created links between those two through the *mixed* farms [52].

This coexistence of cattle and buffalo within the same environment raises questions about the potential impact it may have on the transmission and survival of different pathogens between these two species. Hence, the often overlooked movements associated with the buffalo industry may significantly contribute to the dissemination of pathogens.

Given these considerations, the primary objective of this study is to understand the impact resulting from the omission of the presence of buffaloes, leading to potential inaccuracies in the evaluation of the overall situation.

In recent times, studies on the prevalence of *Coxiella burnetii* among buffaloes in Italy [53, 39, 54] have found a wide distribution of this pathogen. Thus, considering the particular situation of *mixed* farms and the presence of this pathogen in both cattle and buffalo, a relevant role of the latter can be assumed. Therefore, not considering the totality of movements could lead to underestimating the prevalence of this pathogen or to making incorrect or inefficient decisions.

As described in 5.1.2, Q fever is a disease that can infect both species and is present in Italian territory, for which the impact of the buffalo population has not been thoroughly studied. If their role is confirmed through this analysis, one should contemplate incorporating the buffalo network into network models, particularly those focused on bovine diseases in Italy. An analysis and simulation of the spread of this pathogen could, therefore, be used to discover the actual role of buffalo farms.

8.1 Dataset description

To analyze the role that the buffalo population could have in the spreading of this disease, we considered the movements related to the trades of both cattle and buffaloes in Italy. We extracted data from the Italian National Animal Identification and Registration Database (NDB) of the Ministry of Health, obtaining the starting point, ending point, and date of each movement. We restricted our analysis to national movements for four years, from 2017 to 2020, the period in which the data are more stable and solid. We also removed movements to slaughterhouses because of no epidemiological interest.

For each movement, along with the date, we have information on the type of departure and arrival points, as well as their geographic locations. Our analysis will focus on the farms, categorized by the species they contain. We will examine the number of movements related to each species, taking a network approach to analyze the number of movements arriving at (indegree) or starting from (outdegree) a farm.

The following definitions will be used:

- Cattle farms: all the farms that house cattle, visible in Figure 6.9.
- Cattle network: a network that considers all the movements between different entities housing cattle (i.e., cattle farms and mixed farms).
- Buffalo farms: all the farms that house buffaloes.
- Buffalo network: a network that considers all the movements between different entities housing buffaloes (i.e., buffalo farms and mixed farms).
- *Mixed* farms: farms that house both buffaloes and cattle.

8.2 Mixed farms in the buffalo network

Table 8.1 displays the same data as Table 7.1 but considers the *mixed* farms. Here, "active *mixed* farms" are those *mixed* farms that are either the origin or destination for at least one movement of **buffaloes** in a given period.

It can be seen how the Italian reality of this species has a different order of magnitude than that of cattle. This adds significance to our research question. While the size of this network might suggest it is negligible, the seroprevalence of Q fever, as discussed in the introduction, appears to conflict with this idea.

It is also interesting to note how the number of movements is significantly increasing despite a small rise in the number of active farms, suggesting a growing average number of movements per farm. In 2017, movements to or from buffalo *mixed* farms accounted for more than 90% of the total buffalo movements. By 2020, this percentage had decreased to 80%.

One possible explanation for this shift is that in the past, companies with existing cattle facilities could easily expand to include buffaloes without a significant initial investment. As the market for buffalo milk products, such as buffalo mozzarella, grew in Italy [55], there was a trend toward establishing farms exclusively dedicated to buffaloes.

In Table 8.2, we have a brief statistical analysis of the yearly aggregated, directed, and with possible multiple edges between the same nodes network [3]. The weight of the edges is the number of animals moved in a single movement. In this Table, *mixed* farms (buffalo network) denotes the subset of *mixed* farms that are either the origin or destination for at least one movement of **buffaloes** in a given period. The mean degrees are computed on nodes that have that degree different from 0.

In Figures 8.2 and 8.3, we compare the buffalo network properties, respectively the mean outdegree and the mean indegree, of *mixed* farms to farms that raise only buffaloes.

The average indegree of the two networks is similar, while the *mixed* farms have a higher average outdegree, an aspect that could, once again, impact the diffusion of a pathogen capable of reaching these farms.

Therefore, *mixed* farms are significant not only for providing a link to the cattle network but also for potentially accelerating the spread of a pathogen, given the considerations mentioned above.

Table 8.1. **Number of active farms and movements by year.** Buffalo network.

year	active buffalo farms	movements between buffalo farms	active *mixed* farms	movements between *mixed farms*
2017	1243	4791	853	4390
2018	1242	5327	850	4705
2019	1237	6372	839	5352
2020	1305	8075	870	6609

Table 8.2. **Statistics of networks by species.** Yearly aggregation.

year	Nodes	Mean indegree	Mean outdegree	Mean weight	Density
2017	cattle farms	8.32	8.83	54.73	5.64e-05
	buffalo farms	5.86	5.78	38.48	0.0029
	mixed farms (buffalo network)	6.11	6.01	37.78	0.0032
2018	cattle farms	8.37	8.82	56.43	5.90e-05
	buffalo farms	6.43	6.14	41.09	0.0031
	mixed farms (buffalo network)	6.66	6.27	40.39	0.0033
2019	cattle farms	8.53	8.91	58.05	6.16e-05
	buffalo farms	8.65	6.97	45.99	0.0038
	mixed farms (buffalo network)	9.03	7.06	44.59	0.0040
2020	cattle farms	8.35	9.04	59.81	6.17e-05
	buffalo farms	10.61	8.39	49.09	0.0044
	mixed farms (buffalo network)	11.04	8.62	48.21	0.0048

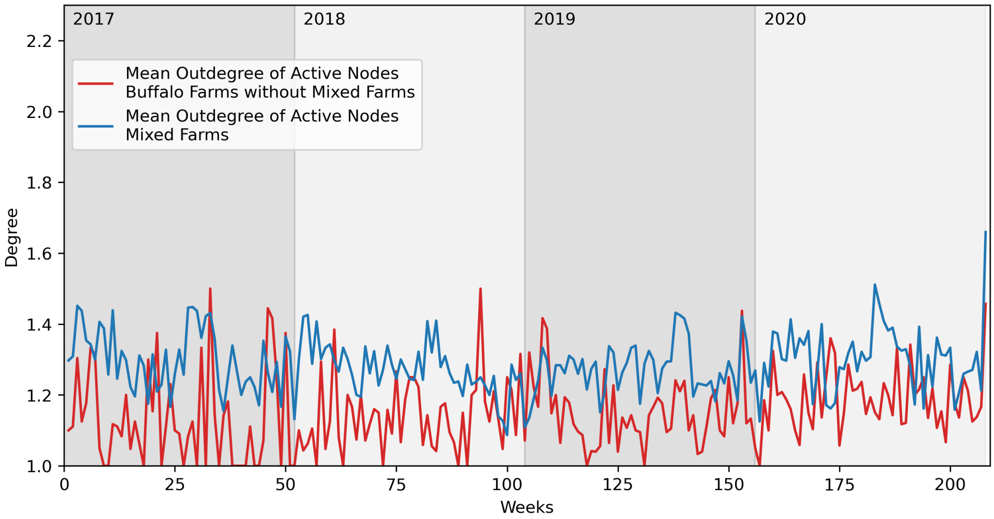


Figure 8.2. **Mean outdegree of farms.** Comparison between only buffaloes and mixed farms.

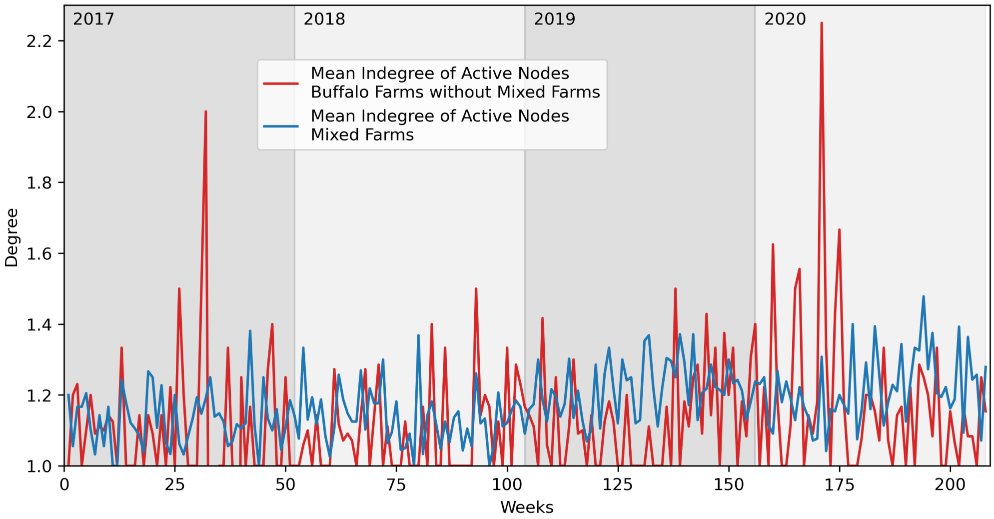


Figure 8.3. **Mean indegree of farms.** Comparison between only buffaloes and mixed farms.

8.3 Simulations and results

We simulated the diffusion of Q fever in our aggregated network, as seen in Figure 8.1. As depicted in Table 8.3, we considered various probabilities of infection and recovery. Starting the simulations, we infected a different number of farms selected from different sets. For the latter, the percentage is relative to the total number of nodes in the aggregated 4-year network.

Table 8.3. **Parameters of the simulations and starting infected set.**

Probability of infection	$\lambda \in \{0.075, 0.1, 0.25, 0.5\}$
Probability of healing	$\mu \in \{0.01, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15\}$
Infected farms	{Buffalo farms, Buffalo farms and *mixed* farms}
Initial Percentage of infected farms (selected among **all** farms)	$\{0.05\%$ (78 nodes)$, 0.5\%$ (780 nodes)$\}$

In light of the considerations in the previous sections, to assess whether the buffalo network alone could indeed facilitate the diffusion of a pathogen, we initiated the simulation at week 0 by infecting a randomly selected subset of nodes, choosing from either buffalo farms or buffalo farms and *mixed* farms.

The first approach avoids prematurely establishing a link with the cattle network and prevents overestimation of the diffusion due to the higher outdegree associated with the *mixed* buffalo nodes.

Therefore, we anticipate the second method to result in a higher diffusion of Q fever in the network. Subsequently, we proceeded to simulate the diffusion of Q fever using real movement data aggregated by week from 2017 to 2020.

Figure 8.4 and Figure 8.5 present 112 histograms representing different scenarios of random selection, infection, and healing probabilities. The values described by the histograms indicate the number of infected farms at the end of each simulation among 100 runs and are shown using a *symlog* scale on the *x-axis*.

They show the results of the simulations with the parameters in Table 8.3 and the initial percentage of infected farms equal to 0.05% and 0.5%, respectively.

The top grids (blue and orange) show the results of the simulations where the initial set of infected farms is composed of the buffalo farms without the *mixed* farms. The bottom grids (green and red), instead, show the results of the simulations where the initial set of infected farms is composed of the buffalo farms with the *mixed* farms.

In the top grid (blue) in Figure 8.4, we can see how only for a low probability of infection combined with a high probability of healing, we do not have infected farms at the end of the simulation. In the other cases, we see how the histograms are located on values bigger than zero, denoting the presence, at the end of the four years, of farms that contain the pathogen.

As expected, the number of infected farms increases when we increase the probability of infection. For example, we can see how, in the three plots in the first column of the blue grid, the histograms shift to the right as we move down the column. The opposite is true as we move right on a line.

In the bottom grid (green) in Figure 8.4, it is interesting to see how starting from a set that also contains the *mixed* farms, those farms that raise both buffaloes and cattle, allows the pathogen to spread from the buffalo network to the cattle network, therefore contributing to the spread to and among cattle farms.

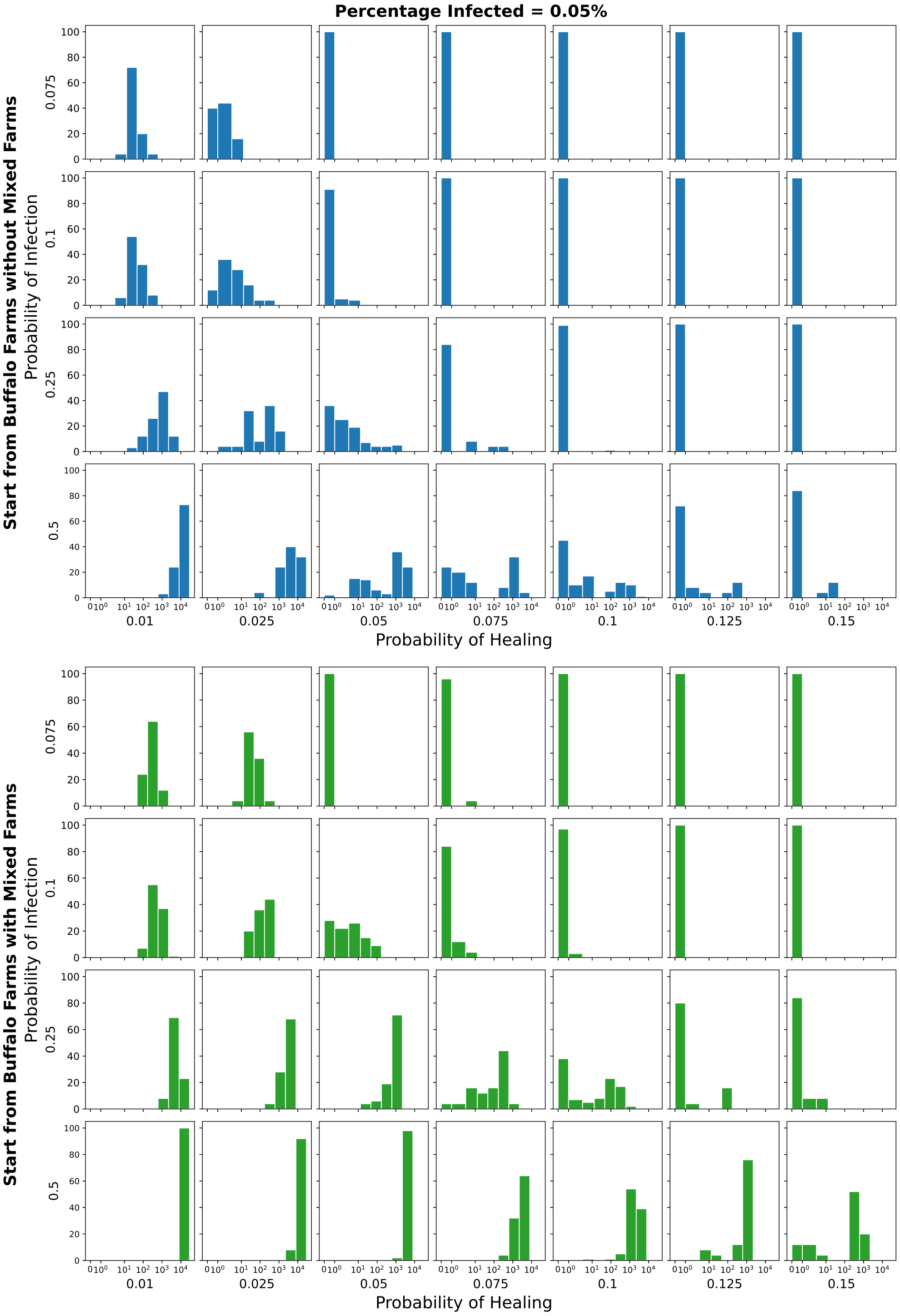


Figure 8.4. **Number of infected farms, among 100 runs, at the end of each simulation.** *Symlog* scale on the x-axis. Initial percentage of infected farms equal to 0.05%.

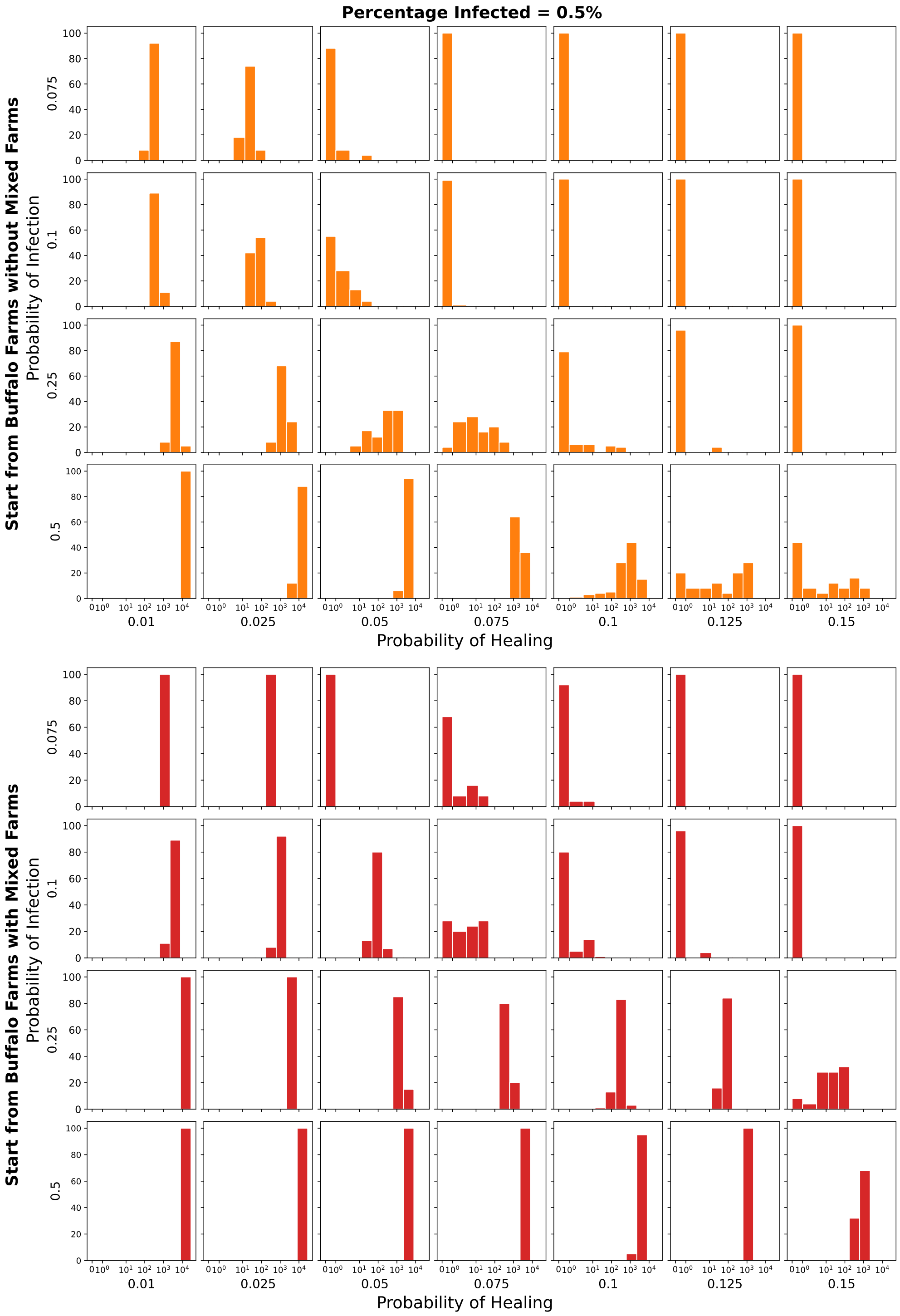


Figure 8.5. **Number of infected farms, among 100 runs, at the end of each simulation.** *Symlog* scale on the x-axis. Initial percentage of infected farms equal to 0.5%.

The general considerations are similar to the ones related to the top grid, but here, even when starting with a small percentage of infected farms with a low probability of infection and a high probability of healing, there are some runs where the pathogen is able to spread and survive into the cattle network.

In Figure 8.5, we can observe a similar behavior to the one described just above. A higher initial number of infected farms results in fewer simulations with no infected farms at the end and a larger number of infected farms overall.

Figures 8.6 and 8.7 represent the same runs, but the histograms describe the number of infected buffalo farms. The color scheme and the considerations are similar to the ones related to Figures 8.4 and 8.5, respectively, obviously with a different order of magnitude. But what is important to point out here is how the pathogen is also able to survive in a smaller network like the buffalo one.

As anticipated, when the infections start from both the buffalo farms and the *mixed* farms, the infection exhibits an easier diffusion into the cattle network. Nevertheless, the disease can still spread into the cattle network even with a small number of initially infected farms selected from the buffalo farms without the mixed farms. Furthermore, even with low infection probabilities, the pathogen, albeit less frequently, spreads into the cattle network, highlighting again a potential risk for this species.

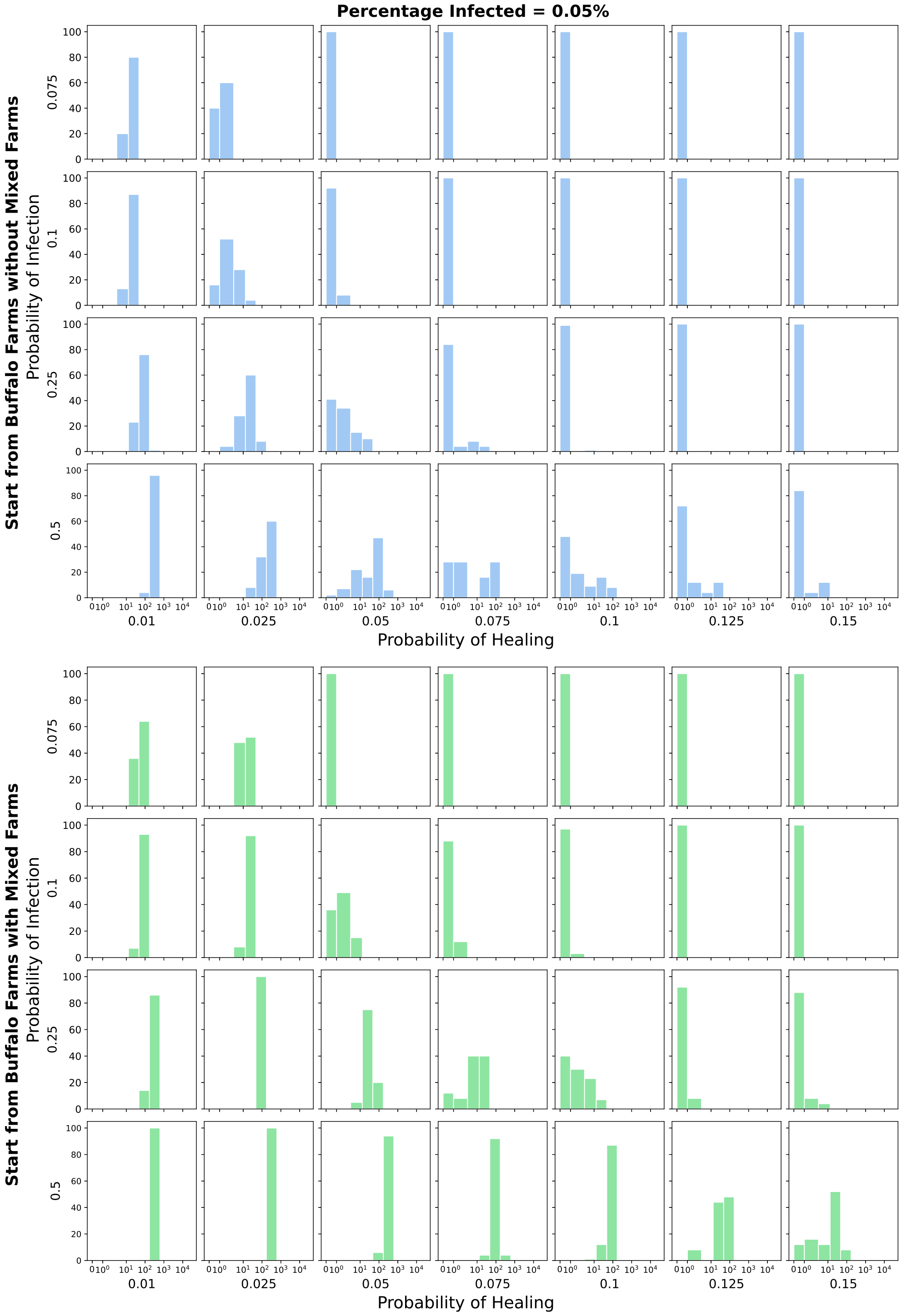


Figure 8.6. **Number of infected *buffalo* farms, among 100 runs, at the end of each simulation.** *Symlog* scale on the x-axis. Initial percentage of infected farms equal to 0.05%.

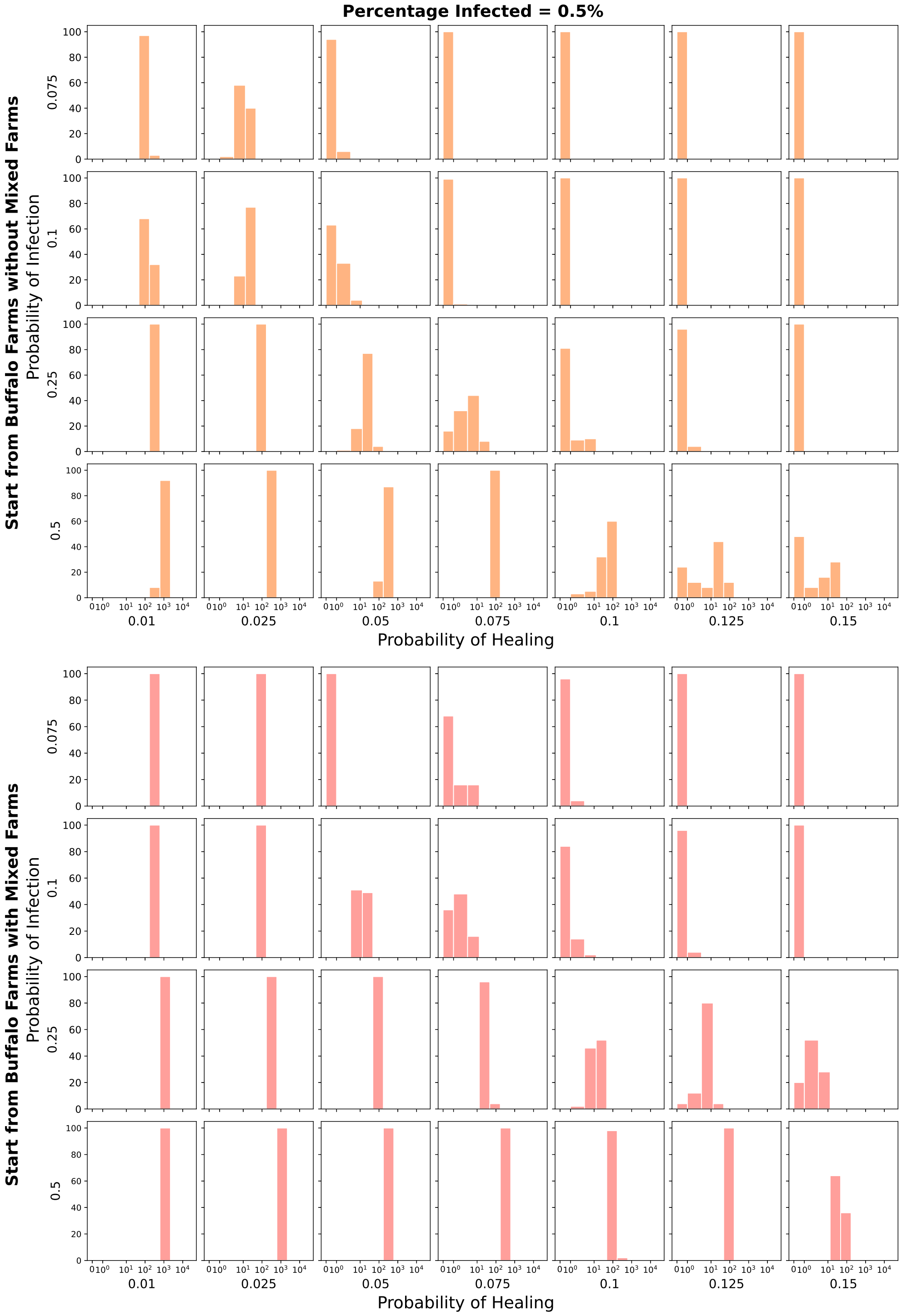


Figure 8.7. **Number of infected *buffalo* farms, among 100 runs, at the end of each simulation.** *Symlog* scale on the x-axis. Initial percentage of infected farms equal to 0.5%.

Figure 8.8 presents the average number of infected farms over 100 runs with parameters λ=0.5, μ=0.1 and starting by infecting 0.5% of all the nodes selected among the buffalo farms without the mixed farms (last row and fifth column in the orange histogram panel). Indeed, the blue line represents the number of infected farms. So, at time 0, it will be composed of only buffalo farms. The total number of infected farms decreases at the beginning and then, around time 30, increases again. This is because as time goes on, while the number of infected buffalo farms decreases, an increasing number of cattle farms are infected, as we can see from the increasing behavior of the red line. Thus, we can observe that the disease enters the cattle network and, from that point onward, it is able to spread.

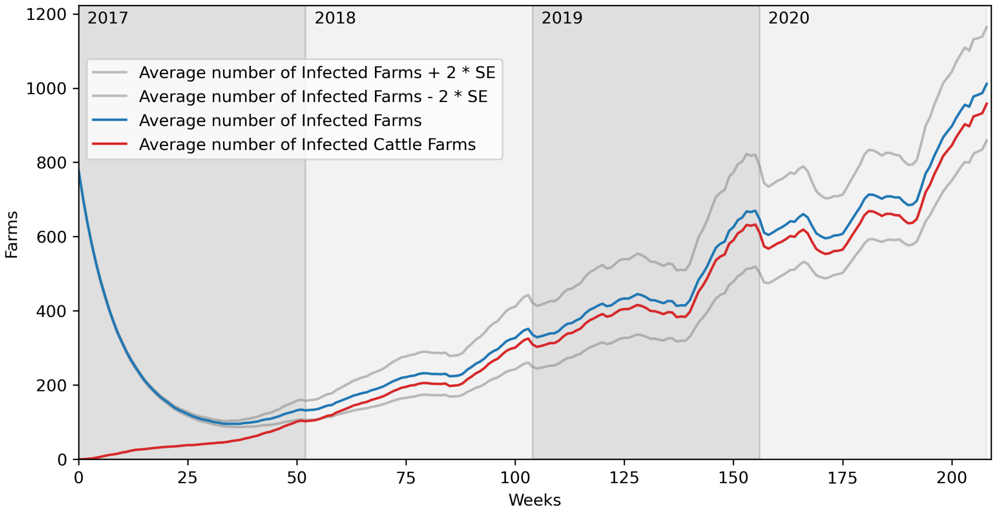


Figure 8.8. **Average number of infected farms over 100 runs.** λ=0.5, μ=0.1. Initial percentage of infected nodes equal to 0.5%. Nodes selected among the buffalo farms without the mixed farms. SE = Standard Error.

Figure 8.9 presents the average number of infected farms over 100 runs with the same parameters, but the nodes are selected among the buffalo farms with the mixed farms (last row and fifth column in the red histogram panel). Therefore, at time 0, the blue line represents both buffalo and mixed farms. Unlike Figure 8.8, the blue line does not initially decline. Instead, there is an immediate increase in the number of infected cattle farms due to the presence of the pathogen in mixed farms, facilitating its spread within the cattle network.

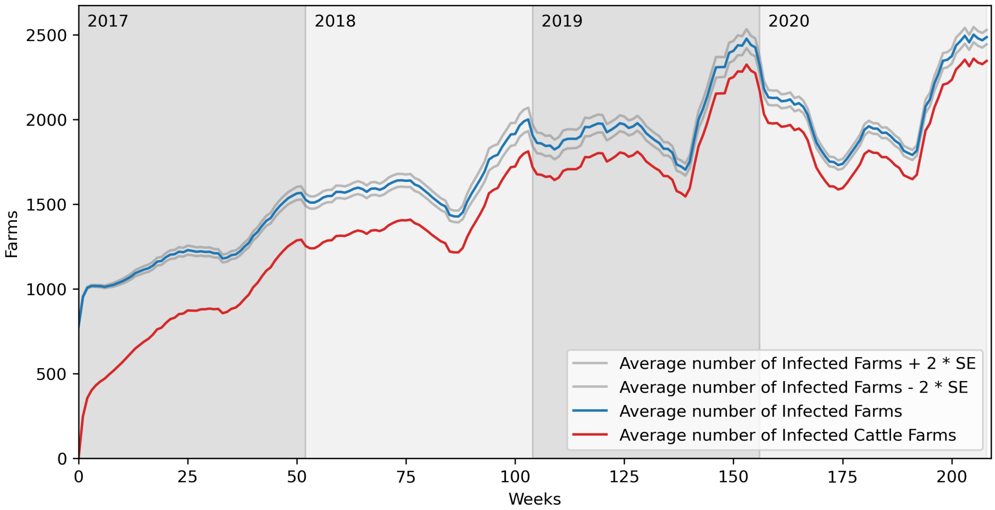


Figure 8.9. **Average number of infected farms over 100 runs.** λ=0.5, μ=0.1, Initial percentage of infected nodes equal to 0.5%. Nodes selected among the buffalo farms with the mixed farms. SE = Standard Error.

To summarize, we have observed the capability of Q fever to spread within the cattle network, even when initiated from a small and poorly connected subset of farms that raise only buffaloes, so without infecting any cattle or *mixed* farms at the beginning. This holds even under a low probability of infection and a high probability of healing. It is also interesting to add how the pathogen survives in the buffalo network.

Despite the substantial difference in the network sizes, the peculiarity of buffalo farms in Italy renders the buffalo network non-negligible when analyzing the reality of Italian bovines. Indeed, we have seen how a pathogen can spread from the Buffalo network to the bovine one. Excluding the buffalo network would, therefore, lead to a major loss of information and a consequent error in the analysis of the cattle network alone.

Not only this, but also the fact that the pathogen can survive, as seen in most simulations in the buffalo network, suggests that it is fundamental to consider both networks when trying to deal with Q fever.

While the buffalo network in Italy is considerably smaller than the bovine network, the close proximity of buffalo and cattle increases the risk of pathogen transmission. Farms that contain both species effectively connect the two networks, facilitating the spread of the disease.

Conclusions

This thesis explored the complexities surrounding disease transmission in livestock populations, particularly emphasizing how crucial data quality is for accurate modeling. By applying advanced network-based modeling techniques, the research has underscored the importance of having reliable data to understand and predict how diseases spread among animals.

A significant contribution of this work lies in the novel analysis of the Italian buffalo movement network, which, to date, has not been studied in the literature. Indeed, buffaloes represent an important reality in Italy, yet their movement patterns have largely been overlooked in epidemiological studies. This thesis has revealed that this network can support the spread of a pathogen and, for this reason, should not be neglected. Furthermore, another significant insight from this study is the role that pastures and *mixed* farms play in the spread of diseases within the Italian livestock network. These elements have a substantial impact on disease transmission and should be included in epidemiological models. By tackling the issues caused by incomplete, inaccurate, or disregarded movement data, this research has provided a clearer picture of the epidemiological risks linked to livestock movements.

The study has demonstrated that enhancing the quality of input data leads to more dependable predictions about disease transmission.

This is particularly important for developing effective disease control strategies, suggesting that interventions based on high-quality data are more likely to succeed in preventing and managing the spread of diseases among livestock.

Moreover, this study has emphasized the need for better data collection practices, particularly in under-researched areas like buffalo farming, and the integration of multiple data sources in veterinary epidemiology. The research suggests that a comprehensive approach, one that considers both biological and social factors influencing disease transmission, is essential for creating effective control measures.

Looking forward, there are several promising areas for further research. Expanding the models to account for additional factors, like environmental conditions or wildlife interactions, could provide deeper insights into the complex forces driving disease dynamics. Additionally, applying these models to different regions or species could test their broader applicability and help refine the approach.

In summary, this thesis adds to the understanding of veterinary epidemiology by highlighting the critical role of data quality in disease modeling and introducing novel insights into the role of the buffalo network in disease transmission. The methods and findings presented here not only improve our grasp of how diseases spread among livestock but also how better disease control and prevention strategies could be developed. As the challenges of emerging diseases and global livestock movements continue to grow, the insights from this research will be relevant in protecting both animal and public health.

Bibliography

[1] C. Lanzas and S. Chen, “Complex system modelling for veterinary epidemiology,” *Preventive Veterinary Medicine*, vol. 118, pp. 207–214, Feb. 2015.

[2] C. Dubé, C. Ribble, D. Kelton, and B. McNab, “A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development,” *Transboundary and Emerging Diseases*, vol. 56, pp. 73–85, Apr. 2009.

[3] B. Martínez-López, A. M. Perez, and J. M. Sánchez-Vizcaíno, “Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine,” *Transboundary and Emerging Diseases*, vol. 56, pp. 109–120, May 2009.

[4] C. Gopalappa, H. Balasubramanian, and P. J. Haas, “A new mixed agent-based network and compartmental simulation framework for joint modeling of related infectious diseases- application to sexually transmitted infections,” *Infectious Disease Modelling*, vol. 8, pp. 84–100, Mar. 2023.

[5] M. Eden, R. Castonguay, B. Munkhbat, H. Balasubramanian, and C. Gopalappa, “Agent-based evolving network modeling: A new simulation method for modeling low prevalence infectious diseases,” *Health Care Management Science*, vol. 24, pp. 623–639, Sept. 2021.

[6] F. Natale, A. Giovannini, L. Savini, D. Palma, L. Possenti, G. Fiore, and P. Calistri, “Network analysis of Italian cattle trade

patterns and evaluation of risks for potential disease spread," *Preventive Veterinary Medicine*, vol. 92, pp. 341–350, Dec. 2009.

[7] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, "Dynamical Patterns of Cattle Trade Movements," *PLoS ONE*, vol. 6, p. e19869, May 2011.

[8] D. J. Daley and J. M. Gani, *Epidemic Modelling: An Introduction.* No. 15 in Cambridge Studies in Mathematical Biology, 1st paperback edition, reprinted 2005: Cambridge University Press, 2005.

[9] M. J. Keeling, M. J. Keeling, and P. Rohani, *Modeling Infectious Diseases in Humans and Animals.* Princeton, NJ: Princeton Univ. Press, 2008.

[10] F. Brauer, P. Van Den Driessche, J. Wu, J. M. Morel, F. Takens, and B. Teissier, eds., *Mathematical Epidemiology*, vol. 1945 of *Lecture Notes in Mathematics.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[11] F. Brauer and C. Castillo-Chávez, *Mathematical Models in Population Biology and Epidemiology.* No. 40 in Texts in Applied Mathematics, New York, NY: Springer, 2. ed ed., 2012.

[12] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, pp. 700–721, Jan. 1997.

[13] J. M. Galán, L. R. Izquierdo, S. S. Izquierdo, J. I. Santos, R. Del Olmo, and A. López-Paredes, "Checking Simulations: Detecting and Avoiding Errors and Artefacts," in *Simulating Social Complexity* (B. Edmonds and R. Meyer, eds.), pp. 119–140, Cham: Springer International Publishing, 2017.

[14] U. Wilensky and W. Rand, *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo.* Cambridge, Massachusetts London, England: The MIT Press, 2015.

[15] D. Kim, *New Horizons in Modeling and Simulation for Social Epidemiology and Public Health.* Wiley, 1 ed., Feb. 2021.

[16] G. N. Gilbert, *Agent-Based Models.* No. 153 in Quantitative Applications in the Social Sciences, Los Angeles: Sage Publ, 3. pr ed., 2010.

[17] M. Resnick, *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds.* Complex Adaptive Systems, Cambridge, Mass.: MIT Press, 8. printing ed., 2002.

[18] S. F. Railsback and V. Grimm, *Agent-Based and Individual-Based Modeling: A Practical Introduction.* Princeton: Princeton University Press, 2012.

[19] G. N. Gilbert and K. G. Troitzsch, *Simulation for the Social Scientist.* Maidenhead: Open Univ. Press, 2. ed., repr ed., 2011.

[20] J. M. Epstein and R. L. Axtell, *Growing Artificial Societies: Social Science from the Bottom Up.* The MIT Press, Oct. 1996.

[21] L. Perez and S. Dragicevic, “An agent-based approach for modeling dynamics of contagious disease spread,” *International Journal of Health Geographics*, vol. 8, no. 1, p. 50, 2009.

[22] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks.* Cambridge University Press, 1 ed., Oct. 2008.

[23] A.-L. Barabási, *Network Science.* Cambridge: Cambridge University Press, 2016.

[24] M. van Steen, *Graph Theory and Complex Networks: An Introduction.* S.l.: Maarten van Steen, 2010.

[25] M. E. J. Newman, *Networks: An Introduction.* Oxford ; New York: Oxford University Press, 2010.

[26] J. Enright and R. R. Kao, "Epidemics on dynamic networks," *Epidemics*, vol. 24, pp. 88–97, Sept. 2018.

[27] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, pp. 97–125, Oct. 2012.

[28] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, "Analysing information flows and key mediators through temporal centrality metrics," in *Proceedings of the 3rd Workshop on Social Network Systems*, (Paris France), pp. 1–6, ACM, Apr. 2010.

[29] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Physical Review E*, vol. 85, p. 026107, Feb. 2012.

[30] C. Caruso, R. Prato, F. Ingravalle, D. Vecchio, A. Sciarra, M. Ternavasio, L. Ceccarelli, A. Martucciello, G. Galiero, E. De Carlo, and L. Masoero, "Prevalence of antibodies against Bubaline herpesvirus (BuHV-1) among Mediterranean water buffalo (*Bubalus bubalis*) with implications in buffalo trade," *Veterinary Quarterly*, vol. 36, pp. 184–188, Oct. 2016.

[31] F. A. Murphy, *Veterinary Virology.* San Diego: Academic Press, 3rd ed ed., 1999.

[32] A. Martucciello, A. Balestrieri, C. Righi, G. Cappelli, E. Scoccia, C. Grassi, S. Brandi, E. Rossi, G. Galiero, D. Gioia, G. Fusco, F. Feliziani, E. De Carlo, and S. Petrini, "Evaluation of an Immunization Protocol Using Bovine Alphaherpesvirus 1 gE-Deleted Marker Vaccines against Bubaline Alphaherpesvirus 1 in Water Buffaloes," *Vaccines*, vol. 11, p. 891, Apr. 2023.

[33] F. Fenner, N. J. MacLachlan, and E. J. Dubovi, *Fenner's Veterinary Virology.* Amsterdam: Elsevier, 5th edition ed., 2017.

[34] C. Esposito, F. Fiorito, G. Miletti, F. Serra, A. Balestrieri, B. Cioffi, C. Cerracchio, G. Galiero, E. De Carlo, M. G. Amoroso, and G. Fusco, "Involvement of herpesviruses in cases of abortion among water buffaloes in southern Italy," *Veterinary research communications*, vol. 46, pp. 719–729, Sept. 2022.

[35] M. Amoroso, F. Corrado, E. De Carlo, M. Lucibelli, A. Martucciello, A. Guarino, and G. Galiero, “Bubaline herpesvirus 1 associated with abortion in a Mediterranean water buffalo,” *Research in Veterinary Science*, vol. 94, pp. 813–816, June 2013.

[36] M. Maurin and D. Raoult, “Q Fever,” *Clinical Microbiology Reviews*, vol. 12, pp. 518–553, Oct. 1999.

[37] E. Angelakis and D. Raoult, “Q fever,” *Veterinary Microbiology*, vol. 140, pp. 297–309, Jan. 2010.

[38] F. Winter, C. Schoneberg, A. Wolf, B. U. Bauer, T. L. Prüfer, S. F. Fischer, U. Gerdes, M. Runge, M. Ganter, and A. Campe, “Concept of an Active Surveillance System for Q Fever in German Small Ruminants—Conflicts Between Best Practices and Feasibility,” *Frontiers in Veterinary Science*, vol. 8, p. 623786, Feb. 2021.

[39] G. Ferrara, B. Colitti, U. Pagnini, D. D’Angelo, G. Iovane, S. Rosati, and S. Montagnaro, “Serological Evidence of Q Fever among Dairy Cattle and Buffalo Populations in the Campania Region, Italy,” *Pathogens*, vol. 11, p. 901, Aug. 2022.

[40] M. Van Asseldonk, J. Prins, and R. Bergevoet, “Economic assessment of Q fever in the Netherlands,” *Preventive Veterinary Medicine*, vol. 112, pp. 27–34, Oct. 2013.

[41] G. Borriello, G. Iovane, and G. Galiero, “Q fever in domestic ruminants,” *Large Animal Review*, vol. 16, pp. 273–283, Dec. 2010.

[42] C. Eldin, C. Mélenotte, O. Mediannikov, E. Ghigo, M. Million, S. Edouard, J.-L. Mege, M. Maurin, and D. Raoult, “From Q Fever to Coxiella burnetii Infection: A Paradigm Change,” *Clinical Microbiology Reviews*, vol. 30, pp. 115–190, Jan. 2017.

[43] “Signs and symptoms of Q fever | CDC.” https://www.cdc.gov/qfever/symptoms/index.html, Jan. 2019.

[44] C. Saegerman, F. Grégoire, and L. Delooz, “Diagnosis of Coxiella

burnetii Cattle Abortion: A One-Year Observational Study," *Pathogens*, vol. 11, p. 429, Apr. 2022.

[45] A. Tavella, K. Trevisiol, G. Morosetti, E. Robatscher, C. Piffer, and A. Fugatti, "Q fever: The importance of surveillance in the Autonomous Province of Bolzano (Italy)," *International Journal of Infectious Diseases*, vol. 21, p. 189, Apr. 2014.

[46] M. Georgiev, A. Afonso, H. Neubauer, H. Needham, R. Thiery, A. Rodolakis, H. Roest, K. Stark, J. Stegeman, P. Vellema, W. van der Hoek, and S. More, "Q fever in humans and farm animals in four European countries, 1982 to 2010," *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, vol. 18, p. 20407, Feb. 2013.

[47] W. van der Hoek, F. Dijkstra, B. Schimmer, P. M. Schneeberger, P. Vellema, C. Wijkmans, R. ter Schegget, V. Hackert, and Y. van Duynhoven, "Q fever in the Netherlands: An update on the epidemiology and control measures," *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, vol. 15, p. 19520, Mar. 2010.

[48] A. Chiesa, C. Onza, N. Sulcaj, A. Torggler, G. Morosetti, F. Conforti, E. Kofler, G. Moretto, P. Sinigaglia, L. Signorini, C. Piffer, M. G. Zuccaro, A. Tavella, and A. Matteelli, "Coxiella burnetii outbreak, Northern Italy 2021," *One Health*, vol. 17, p. 100604, Dec. 2023.

[49] B. L. Dutta, P. Ezanno, and E. Vergu, "Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period," *Preventive Veterinary Medicine*, vol. 117, pp. 79–94, Nov. 2014.

[50] B. Iotti, E. Valdano, L. Savini, L. Candeloro, A. Giovannini,

S. Rosati, V. Colizza, and M. Giacobini, “Farm productive contexts and the dynamics of bovine viral diarrhea (BVD) transmission,” *Preventive Veterinary Medicine*, vol. 165, pp. 23–33, Apr. 2019.

[51] A. Darbon, E. Valdano, C. Poletto, A. Giovannini, L. Savini, L. Candeloro, and V. Colizza, “Network-based assessment of the vulnerability of Italian regions to bovine brucellosis,” *Preventive Veterinary Medicine*, vol. 158, pp. 25–34, Oct. 2018.

[52] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, “Mathematical Formulation of Multilayer Networks,” *Physical Review X*, vol. 3, p. 041022, Dec. 2013.

[53] A. Fanelli, A. Trotta, F. Bono, M. Corrente, and D. Buonavoglia, “Seroprevalence of Coxiella burnetii in dairy cattle and buffalo from Southern Italy: C.burnetii in Southern Italy,” *Veterinaria Italiana*, vol. 56, pp. 193–197, Dec. 2020.

[54] M. Riccò, A. Baldassarre, S. Corrado, and F. Marchesi, “Seroprevalence of Coxiella burnetii in Occupational Settings: A Meta-Analysis of Italian Studies,” *Zoonotic Diseases*, vol. 3, pp. 38–51, Feb. 2023.

[55] G. dell’Orefice, “La Mozzarella di Bufala è cresciuta il doppio degli altri formaggi Dop.” https://www.ilsole24ore.com/art/la-mozzarella-bufala-e-cresciuta-doppio-altri-formaggi-dop-AErNteeC, Jan. 2023.