Metapopulation and Network Models for Plant and Human Infections

Matthew Baister

Department of Mathematics and Statistics

University of Strathclyde

Glasgow, Scotland, UK

Doctor of Philosophy

2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Chapter 5 describes the author's original research which has been conducted in collaboration with the co-authors: Ewan McTaggart, Paul McMenemy, Itamar Megiddo and Adam Kleczkowski.

Signed: Matthew Baister

Date:

Abstract

Plant pests and diseases infiltrate countries via international trade. Domestic trade moves the pest far beyond the point of entry, making eradication efforts challenging. Understanding the risks and control strategies surrounding trade is vital in mitigating the damage from invasions. While attention has been given to the structure of the international trade network, little is known about the within-UK network structure due to unregistered plant movement and limited data availability.

We address this gap by constructing a directed and weighted network, using sales data from four plant nurseries. We find that nurseries specialise sales towards one of the four customer groups (commercial, consumer, nursery, retailer). This allows us to group nodes into classes differing by sales patterns, with trade volumes and customer numbers highly variable within and between classes.

Using centrality measures, we identify nodes at higher risk of transmitting pests or diseases. We find that edge weights significantly affect node centrality, emphasising the importance of trade volumes in similar network models. Node centrality is robust to small changes in market structure, and customers' contribution to network structure is minimal.

We extend our network model to a compartmental metapopulation Susceptible-Infected framework and investigate the effect of different seedings, compare inspection strategies and conduct a cost-benefit analysis. Our results demonstrate that disease spread faster

when originating in a nursery that primarily sells to other nurseries. We find the utility of inspecting consignments depends on the frequency of inspecting nursery stock.

Finally, we identify inspection efficacy regions when nurseries benefit from more frequent inspections, considering cost. From our analysis of network structure, trade, and disease dynamics, this research provides guidance for targeted surveillance, intervention, and control to mitigate the spread of plant pests and diseases.

We further use network and metapopulation methods to model the spread of COVID-19 in care homes.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"A disease known is half cured."

This old adage demonstrates that we have known for a long time the importance of understanding diseases. The spread of infectious diseases is known to have plagued humans, animals and plants for many thousands of years [3–5]. Infectious diseases are caused by various types of pathogens including bacteria, viruses, fungi and parasites [6]. Fungal diseases were described by Hippocrates in 500 BC, while the Chinese and Egyptians had knowledge of parasites dating back to 3000 BC [5, 7]. However, the scientific understanding of microorganisms was held back for many centuries by superstition and religious beliefs [5, 7]. This understanding remained limited until the discovery of bacteria, made possible by the microscopes developed by Antony Van Leeuwenhoek in the late 17th century [8]. From that period until the 19th century, numerous discoveries contributed to the dominance of the germ theory of disease over spontaneous generation [5]. The first scientific writings on fungal diseases were by David Gruby and Agostino Bassi in the first half of the 19th century, and in the latter half came the first discovered virus, the tobacco mosaic virus [7, 9].

The pathogens discussed so far are just one subset of a larger group called *pests*. In simple terms, "A pest is an organism that man sees as harmful or annoying" [10]. In addition to pathogens, there are four other types of pest: arthropods, vertebrates, weeds and nema-

todes [10]. *Fishel* provides significant dates related to pesticide use, including the use of Sulphur from pre-Roman civilisations as early as 2000 BC, and the use of Hellebore by the Romans in 100 BC to control rat, mice and insect populations [11]. Natural substances such as Rotenone, Nicotine and Paris Green were used in the following years until the first synthetic pesticide, Dichlorodiphenyltrichloroethane (DDT), was created in 1873 [11]. The development and use of pesticides continued unabated until legislation was introduced in 1910 to prevent harm to humans and animals [11]. From this point onward, pesticide use continued as regulators introduced further protective legislation [11]. However, due to health concerns and the rise of pesticide resistance, the fight against pests remains ongoing.

As we have highlighted, the threat from pests and diseases has been ever-present throughout history. Natural selection has, over the millennia, put pressure on all living things to co-evolve with their native pests and diseases [12–14]. All animals have evolved immune systems and many have developed symbiotic relationships with other organisms as defence mechanisms against these threats [14, 15]. These defence mechanisms are by no means perfect, but have been effective enough to bring us to the world we see today.

These defences can fail substantially when faced with a pest or disease that is non-native to the organism's environment [16–18]. In an increasingly globalised world, where animals, plants, insects, food and people are transported across borders, we also carry the burden of introducing these non-native pests and diseases.

The *Columbian Exchange* of 1492 to 1800 should serve as a harsh reminder of the consequences of such introductions. During this period, the movement of Europeans to the Americas introduced many foreign diseases such as smallpox, measles, typhus and cholera which led to the decimation of the native Americans, who had no immunity against these diseases [19]. To avoid repeating similar catastrophes, it is crucial to understand the pathways through which these non-native pests and diseases spread.

It is within this context that this thesis, co-funded by DEFRA (Department for Environment Food and Rural Affairs) and the University of Strathclyde, focuses on the role of domestic plant trade in spreading non-native plant pests and diseases within the United Kingdom (UK).

In this introductory chapter we emphasise the importance of understanding plant pests and diseases by highlighting the damage caused by previous epidemics. We show the role that trade, both domestic and international, plays in the spread of plant pests and diseases. Next, we define the concepts from network theory and mathematical epidemiological modelling that we use throughout subsequent chapters in the thesis. We end the chapter with a review of the relevant literature to display the state of the art in both trade network construction and modelling the spread of disease on networks. This literature review highlights an existing gap, which our research in subsequent chapters aims to fill.

1.1   Plant pests and diseases and their significance

The global impact of plant pests and diseases is significant, from decimating horticultural and agricultural yields to accelerating climate change and the destruction of ecosystems [20–29]. Plant diseases can spread in a multitude of natural ways; such as soil and roots of the plant, contaminated water sources, dispersal of splashed water, aerial dispersal and insect vectors [30–32]. In addition to their natural movement, plant pests also spread through human trade of plants. An example includes the introduction of *Thaumetopoea processionea*, Oak Processionary Moth (OPM), to the UK in 2005 from the importation of OPM eggs, hidden in the canopy of Oak trees [33]. Since 2012 there has been a UK government programme to manage OPM, and over the last five years, £10 million have been invested into Oak health, managing OPM and outbreak response research [34].

Evidence suggests that OPM spread further in the UK via trade from Europe, with close to 70 recorded interceptions in July 2019 [35]. The moths are not the primary problem, it is the larvae and caterpillars that prey on the Oak trees that inflict the damage [35]. These

pests feed on the leaves of Oak trees and, when they do so in large quantities, they can weaken the trees, increasing susceptibility to other pests and harsh environmental conditions [35]. There is also a threat to humans and other animals from this pest - the older caterpillars shed hairs when disturbed, and contact with these hairs can cause skin rashes, eye irritation, breathing difficulties and sore throat [35]. The UK Plant Health Risk Register, at the time of writing, has given OPM the highest score of 5 for both the likelihood of entry and establishment to the UK and value at risk assuming no mitigation measures [36]. The impact (damage to host and/or industry) has been given a score of 4 and the risk of spread is 3 out of 5. Through mitigations such as regulation, surveillance and research, the value at risk does not change but the likelihood of entry and establishment, impact and further spread are reduced to 3, 3 and 2 respectively. For more information on the relative risk scores, see [37]. DEFRA have identified the most common pathways for the introduction and spread of this pest as plants for planting (except seeds, bulbs and tubers) and Roundwood of Oak when bark is present [36]. It is worth noting, however, that Public Health England concluded in 2015 that this exposure is not likely to cause severe impacts on human health [38].

There have been plant diseases with severe, indirect impacts towards humans such as the 1942 fungal infection *Cochlioblus miyabeanus* (Brown Spot). The disease spread through rice fields in Bengal (India) leading to an estimated 91% loss of yield in rice, contributing to the Bengal famine [21]. In a similar situation, the fungal pathogen *Phytophthora infestans* (Potato Late Blight) caused the Irish potato famine in the 1840s leading to the death of approximately one million people and the emigration of another million [22]. Potato Late Blight also severely affected the Netherlands in 1845, who were at the time the second highest potato-consuming country in the world [39]. In 1845, potato yield in the Netherlands fell by 75%, which led to increased poverty, death rates and social unrest [39]. In 1927, *Tracheomycosis*, Coffee Wilt disease, emerged near Bangui (Central African Republic) and the spread of this disease was so severe that over 50% of the coffee-producing areas were destroyed in Côte d'Ivoire and in the Democratic Republic of Congo [23].

The fungal pathogen *Phytophthora ramorum* (*P. ramorum*) known to cause Sudden Oak Death has been particularly devastating in the United States (US) and Europe [24]. *P. ramorum* is known to have multiple hosts; if infecting foliage plants (like *Rhododendron*), infection will cause necrotic lesions on leaves and shoot dieback (progressive death). When *P. ramorum* causes canker infection to hosts like Oak and Beech, this usually results in bleeding cankers and tree death [24]. Since the 1990s, *P. ramorum* has caused the deaths of millions of trees in California and Oregon in the US [25], with disease spread of *P. ramorum* via domestic plant trade being seen in the US [40–44]. The disease is thought to have been introduced to the UK via infected nursery stock from the European Union [18], and some analyses have also found implications of the domestic spread of *P. ramorum* in the UK [45, 46].

Hemileia vastarix, another fungal pathogen of great significance, is known to cause the plant disease Coffee Rust [28]. Discovered in Ceylon (Sri Lanka) in 1867, Coffee Rust is known to affect the leaves of plants in the genus *Coffea*, causing premature defoliation and lowering yields in subsequent years by as much as 10 times [28]. Interestingly, the devastation caused by this disease in Asia in the 1800s played a significant role in making tea the drink of choice in England [47]. After causing major losses in Ceylon, the disease would then go on to infect nearly all coffee-growing regions across the world [28, 29]. In 1970, Coffee Rust was reported in Brazil [29] and by the 1980s, the fungus spread to South and Central America and Mexico [47]. There has been speculation that Coffee Rust spread from Africa to Brazil by the wind, though it is believed more likely to have been spread by trade [28, 47]. Human-mediated introduction of Coffee Rust has been attributed to Nicaragua, El Salvador, Honduras, Guatemala, Mexico, Ecuador and Columbia, with infected coffee seedlings being the cause of its introduction to Bolivia and Peru [47]. Given that coffee is one of the most important products in international trade [29], the economic losses from Coffee Rust over the past 150 years have been tremendous.

The threat from plant pests and diseases in the UK has been growing in recent years. This is illustrated in Figure 1.1, which shows the number of outbreaks affecting trees in Britain [1]. As a consequence, the UK economy has suffered. For example, in 2017 the

Landscape and Horticultural industry contributed £24.2 billion to UK Gross Domestic Product (GDP) alongside £2.9 billion through tourism via parks and gardens [48]. Some estimates argue that *Hymenoscyphus fraxineus* (Ash dieback) cost Britain £15 billion, mainly due to ecosystem service losses [27]. Furthermore, the Dutch Elm disease outbreak caused the loss of almost 30 million Elm trees in the UK between 1970 and 1990 [49]. Therefore, an ill-prepared response to a plant pest or disease can be very damaging to the UK. Plant trade plays a significant role in the spread of plant diseases [50], and so further understanding of how these pests and diseases travel throughout the UK is crucial to mitigating substantial damages.
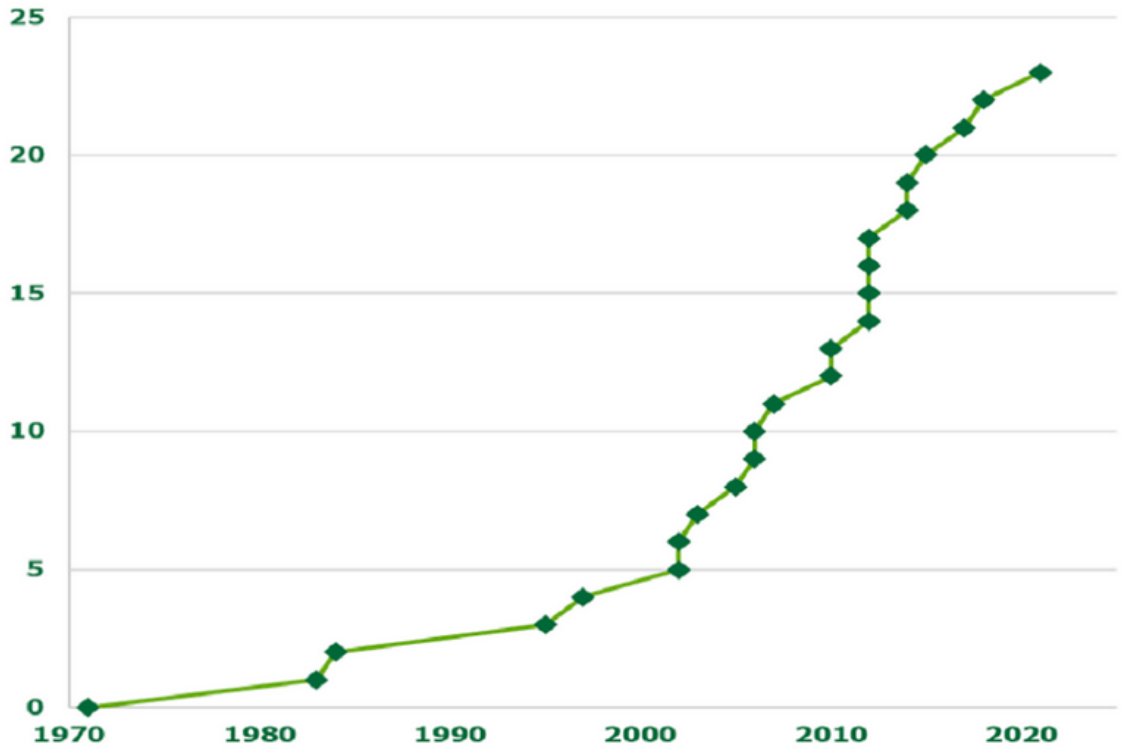
Figure 1.1. Cumulative increase of new pest and disease outbreaks affecting trees in Britain since 1971. The horizontal and vertical axes represent the year and number of outbreaks respectively. Figure from [1].

Xylella is a plant disease caused by the bacterium *Xylella fastidiosa* and is, at time of writing, prevalent in Europe, America and Taiwan [51]. The disease has many hosts and spreads by insect vectors that feed on the xylem fluid of the plant [32], and Xylella is considered to be the most dangerous plant bacteria worldwide [52]. Pierce's Disease, caused

6

by Xylella, destroyed more than 16000 hectares of grapevines in California in the 1880s, with many more outbreaks occurring in the following years [53]. Estimates from 2014 say that Pierce's Disease cost California $104 million per year, including prevention measures and losses to the disease [54]. There are additional losses, such as the available land that could be used for planting if not for the presence of the disease. For example, grapes cannot be grown in the south-eastern states near the Gulf of Mexico because of Xylella [53]. Further examples of the damage caused by the disease include the removal of more than 100 million Citrus trees in Brazil, and the infection of 1-3 million Olive trees in Apulia, Italy [53]. Though not known to be present in the UK yet, the trade of plants for planting is considered Xylella's most likely introductory pathway to the UK [32]. On the UK risk register for unmitigated risks, Xylella has the maximum value at risk of 5, a likelihood of entry and establishment of 4, an impact of 3 and a risk of spread score of 2 [55]. With mitigation consisting of only regulation, the risks can be halved by reducing the likelihood of entry to 2 out of 5. The other risk scores do not change [55]. Thus, if Xylella is introduced to the UK, understanding how to effectively control the further spread of the disease will help the UK avoid similar economic and environmental damage, as seen in Europe and the Americas.

This section has highlighted only some of the substantial damages that plant pests and diseases have caused across the globe. The damage from previous plant pests and diseases serves as motivation for controlling the current threats to the UK (such as Xylella and OPM) and to understand how plant pests and diseases spread, both internationally and domestically. This understanding will help mitigate the damage caused by any subsequent introduced or establishing pest or disease.

1.2 Plant trade as a mode of disease transmission

Increased globalisation has led to the increased trade of plants across the globe [56]. Though plant pests and diseases can travel via other means such as wind, rain and even via soil on hikers' boots [57], plant trade is particularly an issue as it allows the faster, longer range and larger quantity dispersal of pathogens.

Plant pests and diseases can also be indirectly spread in the trade of other products. The Asian long-horned beetle native to East Asia was introduced to other continents in the trade of solid wood packing material [58]. Other examples include the fungal pathogens Canker stain of Plane (*Ceratocystis platani*) introduced from the US to Europe on wood packing material, and Dutch Elm disease (*Ophiostoma ulmi*), introduced to the UK on Elm logs [16]. The trading data we have access to for this thesis is for live plants and so we focus on this pathway.

The importation of live plants is seen to be the major driver of unintentional introductions of plant pests and diseases [56, 59–62]. When non-native pests and diseases are introduced to environments without an evolutionary resistance, it can be especially damaging, putting whole ecosystems at risk [16–18]. For instance, the importation of the fungus *Cryphonectria parasitica* (Chestnut blight) from Asia to North America in the early 1900s would go on to kill between 3 to 4 billion mature Chestnut trees in the first half of the 20th century [63]. Plant trade on both the international and domestic scale have been separately attributed to the spread of plant pests and diseases.

1.2.1 International trade

The processes underlying non-native species introduction by international trade are complex and temporally dynamic, with many factors affecting the risk of pest/disease introduction. These factors include changes in "...global species distributions, volumes and types of traded commodities, trade partners and technologies, and investments in biosecurity" [17]. The number of new plant disease outbreaks is increasing, which is linked to the increased

global movement of plants and plant products in recent decades [18,64]. This increased international movement of plants is seen to pose the greatest risk in terms of spreading plant pests and diseases due to the large volumes of plants involved in these movements [16,18].

Chapman et al. showed that plant pest invasions are strongly determined by global trade networks using Generalised Linear Mixed-effects Models (GLMMs) and world trade data [50]. The authors tested the relationship between non-native species invasions and international trade networks and analysed the invasions of 422 plant pests in member countries of the European and Mediterranean Plant Protection Organisation (EPPO). GLMMs were used based on the connection strengths of 10 different connectivity indexes, representing the roles of different trade networks, to determine the causality of non-native species invasions. Features such as geographical distance, species prevalence and climate similarity were considered in conjunction with the volume of trade between countries. They distinguished between total and recent invasions of non-native invasive species in their paper. Five trade networks were created on total imports of agricultural commodities, live plants, forest products, fruits & vegetables and seeds. The authors analysed GLMMs with individual and multiple trade networks and found that multiple trade networks yielded a better fit to the data. It was found that the model that best explained all invasions was using the live plants, forest products and seeds networks. In addition, the best-fitting index for recent invasions was the trade network for live plants and forest products.

The accidental transportation of plant pests via trade is not an unusual occurrence. This is illustrated in the statistic that plant pests or diseases were found in 2.6% of all shipments of plants entering the United States between the years 2003 and 2010 [65]. *MacLachlan et al.* noticed that in the US, the rate of discovery for non-native pests did not rise in accordance with the increase in the volume of plant imports in the late 20th century [65]. The authors developed two models: one to estimate past establishments of Hemiptera (an order of insect), in the US, Alaska and Canada using plant imports, and the other modelled the time delay between pest establishment and discovery. The authors found that, although import volumes have increased from all geographic regions, the number of establishments has changed, depending on the region. They imply that that there is a utility to a biosecu-

rity approach that focuses preventative measures on higher-risk regions. Additionally, the authors also estimate that many established pests into the US have not been discovered yet, with a median delay to pest discovery of 80 years and 25% more undiscovered pests than observed discoveries. A surprising result was that the authors found that the average risk from plant imports has fallen over time, perhaps explained by past trade experience and improvements in both technology and policy [65]. This result highlights that the biosecurity efforts made across the globe have had a positive effect on the spread of plant pests and diseases over time.

Governments try to minimise the spread of plant pests and diseases through boarder inspections and control measures such as trade restrictions, quarantine measures and required treatments (e.g., fumigation) [66]. Yet, as we have seen, plant pest and diseases still infiltrate across country borders. Once introduced, the transport of infected material can result in establishment far from the initial point of entry, making eradication efforts challenging [67].

1.2.2 Domestic trade

Significant attention has been put on the role of international trade as a means for spreading plant pests and diseases globally. However, there has been comparatively little research on the role that trade on the national level plays in spreading plant pests and diseases.

Data on trade volumes of plants can be difficult or impossible to obtain, and so researchers turn to using proxies with more available data. In their study of the ornamental horticultural trade in Britain, *Dehnen-Schmutz et al.* used the marketing of plants as a proxy for propagule pressure (a measure of the number of individuals introduced into a nonnative environment [68]), in a model to examine the link between plant market presence and their capability to escape and establish into the wild [59]. The authors sampled plant species for their analysis which were actively advertised in nursery catalogues in the mid 1800s. Market presence in the 19th century for a plant was determined if it was included in one of the eight catalogues chosen from that period. A plant was considered present in the

modern-day market if listed in the *Plant Finder*, an annual publication by the Royal Horticultural Society. The proxy for trade volume used was the number of nurseries selling a species included in both the 19th century list and the latest *Plant Finder* publication. Plant traits were included in their regression models such as: life form (perennial, annual, etc.), plant height, minimum tolerated temperature, recommended propagation method, the native range of species, the native range of genus and garden origin. The authors used non-parametric tests and two logit regression tests in their analysis, which found that market presence strongly explained plant escape but weakly explained establishment. Establishment probability was found to be greater if the plant genus was native to Britain, and this probability increased with the number of continents in a plant's native range. The authors' investigation revealed a correlation between the adoption of invasive ornamental plants and changes in the relative price of those plants, indicating the potential to use price data as an early warning system against possible establishing species.

The authors followed on from this work with a multispecies temporal approach and a Generalised Estimation Equation model that compared invading ornamental non-native species introduced into Britain with introduced species not known to have invaded [69]. In this paper, the authors used the same method of proxy for propagule pressure as in [59]. The authors also used data on the relative price of exotic species in the national horticultural trade, species characteristics and socioeconomic variables [69]. *Dehnen-Schmutz et al.* sampled 506 ornamental plant species and used longitudinal data on the invasion process, market availability and seed prices. Data was used from historical nursery catalogues on the availability and prices of seeds of ornamental species in the horticultural market in Britain from 1885 to 1985. The same plant traits were used in this regression model as in their previous work [59]. The authors found that on average, invading species were sold by more nurseries than non-invading species. The model indicated the existence of a causal relationship between frequent low prices for a species and future escape [69]. The authors stated that their results "provide quantitative evidence for the role of the ornamental plant trade as a vector activity in plant invasions" and highlighted that the "attention to the

threats posed by international trade has missed the critical role of local market development" [69].

The link between domestic trade and disease spread has been shown directly. For example, the trade of host plants for *P. ramorum* in the US have been recorded (Figure 1.2a), in addition to the confirmed *P. ramorum* infected shipments (Figure 1.2b) [2]. From these figures we can see that the structure of both the shipping patterns (a) and infected shipments (b) are similar. These figures show a clear link between domestic plant trade and the spread of plant diseases, and that spread via domestic trade can travel long distances from the west-coast to the east-coast of the US.

(a)   **(b)**
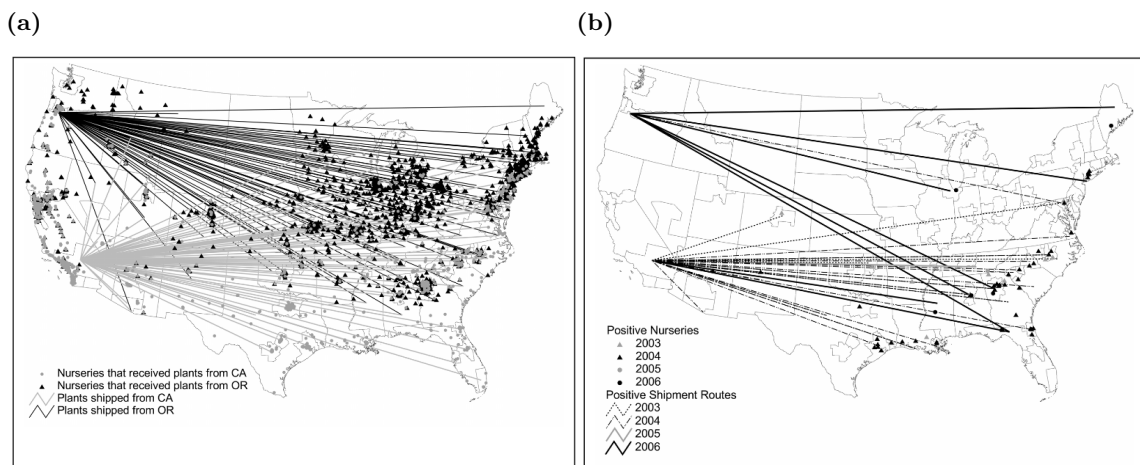
Figure 1.2. (a) The shipping pattern for *P. ramorum* host plants. **(b)** The confirmed *P. ramorum* infected host plant shipments between the years 2003 and 2006. Figures from [2].

So far, we have discussed the damage that plant pests and diseases have had across the globe and, in particular, the role that trade plays in spreading disease. It has been noted that the structure of plant trade can be different, depending on the country of interest [70]. Thus, if we want to understand the structure of UK plant trade, we require a UK specific investigation.

In the next section, we provide some mathematical background for the concepts from network theory and mathematical epidemiology that we use throughout the thesis.

1.3 Background and definitions

In this section we introduce the concepts in both network theory and mathematical epidemiology that we use throughout the thesis. We refer the interested reader to [71–73] for further details and for a general introduction to network theory.

1.3.1 Network theory

A *network*, $G = (V, E)$, comprises of a set V of *nodes*, that usually represent individuals, and a set $E \subset V \times V$ known as *edges* or *links* that connect the nodes together representing an interaction. In our trade network, nodes represent individual traders, while edges represent a sale between buyers and sellers. A list of edges in the network, which is unique up to the ordering in which the edges are presented, is called an *edge-list*.

In the case of a *directed* network, the set of edges E is asymmetric, meaning that a connection from node i to node j does not imply a connection from j to i, *undirected* otherwise. A directed trade network encodes in the edge from i to j the information that i sold goods to j. We use the terms *in-going/out-going* edges of a node to refer to edges ending/starting at said node. For instance, if we consider an edge from i to j we say that this edge is out-going from node i and in-going to node j.

When a network is said to be *weighted*, each edge has an associated real number, representing the strength of the connection. Weights are usually selected to be positive. In a trade network, weights can be used to represent the quantity of goods sold. As such, a larger edge weight can be interpreted as a stronger connection between nodes. In unweighted networks, only the presence of connections is considered.

A network can be represented as a matrix $\boldsymbol{A}$, called an *adjacency matrix*, where the elements $a_{i,j} \in \mathbb{R} \setminus \{0\}$ represent the weight for the edge from node i to node j. If node i and j are not connected by an edge, $a_{i,j} = 0$. For unweighted networks, the values of $a_{i,j}$ are restricted to 1 or 0 to represent connections.

Paths in a network are defined as a sequence of nodes, where each consecutive node is connected by an edge and no nodes are revisited. An undirected network is said to be *connected* if there is a path between each pair of nodes in G. Directed networks are described as *weakly-connected* if they are connected when considered as an undirected network, i.e., when ignoring the directions of the edges.

Network centrality measures

A centrality measure quantifies the importance of the nodes in a network, in some specific context. Some of these measures are particularly informative when assessing how the structure of a network affects the spread of infectious diseases.

The *degree* of a node, for an undirected network, is the total number of edges connected to the node. In a directed network, a node has two different degrees. An *out-degree*, the total number of out-going edges from the respective node, and an *in-degree*, the total number of in-going edges to that node. If the network is weighted, there is another degree which incorporates the edge weights, called the *node strength*. The node strength is the sum total of the weights of edges connected to a given node. Similar to the out-degree and in-degree, we can apply the concept of strength in directed networks and define the *out-strength* and *in-strength* of a node by accounting for the sum total of the weights of their out-going and in-going edges.

The out-degree distribution of a network shows the proportion of nodes in a network with a given out-degree. The shape of the distribution can give an indication of the effects from an initial seeding of a disease and the rate of transmission across the network. If most nodes in the network have a large out-degree, then a random initially infected node is most

likely going to be connected to a large number of nodes, which they can now infect. Alternatively, a random seeding in a network where most nodes have a small out-degree would result in fewer secondary infections. Networks with out-degree distributions with heavy right tails are more susceptible to *superspreader events*. A superspreader event is when an infected individual infects far more than the average infected individual [74]. Thus, the nodes with the higher out-degrees are a good candidate for targeted prevention measures.

The in-degree distribution of a network shows the distribution of the risk of infection for nodes, i.e., nodes with larger in-degrees will have a higher chance of receiving infection. Thus, nodes with higher in-degrees are good candidates for surveillance/inspection for the presence of disease.

The in- and out-strength measures are interpreted in the same way as their unweighted counterparts. However, these measures incorporate the inequalities amongst edges. Consequently, nodes which have fewer in-going or out-going links but with larger edge weights can have the potential for a more significant contribution to the spread of disease.

The betweenness centrality measure calculates the extent to which a node appears in the paths between other nodes in the network. An intuitive interpretation of this centrality measure is that betweenness measures how often a node acts as a bridge between others. In the context of disease spread, the higher a node's betweenness centrality, the more likely said node is to receive an infection if an outbreak occurs.

The *hubs* and *authorities* in a network can be identified by Kleinberg's hubs and authority scores [71]. The hub and authority scores are the corresponding eigenvectors for the leading eigenvalue of $A^T A$ and $A A^T$. These scores are defined in reference to one another, where a hub is a node that has outward links to many authorities, and an authority is a node which has many in-going links from hubs. To interpret this measure, a disease starting at a hub has greater potential to start an epidemic, and authorities are of higher risk of

receiving infection. Thus hubs and authorities are potentially good candidates for targeted prevention measures and inspections respectively.

The comparison of weighted and unweighted measures is of particular epidemiological relevance. When a pest/disease has a low transmission rate, the probability of infecting another node will depend more on the volume of plants traded amongst nodes as opposed to the number of out-going links a node has. Whereas, when a pest/disease has a very high transmission rate, the probability of infecting another node will be less dependent/independent on the volume of trade. Thus edge weights can be an important feature to include in network models and affect the interpretation of network structure in the context of the spread of disease.

1.3.2 Compartmental modelling

When we refer to diseases, we are specifically referring to infectious diseases where infected individuals drive further infections. A simple method of modelling disease spread, stemming from [75], is the deterministic SI model. A population is split into a non-infected "Susceptible" compartment and the infectious individuals into an "Infected" compartment. The number of susceptible and infected individuals $(S(t), I(t))$ is calculated over time points t, in specified increments of time Δt. A parameter $\beta \in \mathbb{R}^+$ is used to describe the *transmission rate* of the disease which consists of the population contact rate per unit of time multiplied by the probability of transmission in a contact. For simplicity we consider β to be constant, though this does not need to be the case. It is assumed that the population is well mixed, meaning that the probability an infected individual interacts with a susceptible individual is the same for all susceptible and infected. With an assumed *frequency dependent* force of infection, the total number of interactions between the susceptible and infected populations at time t is $\frac{S(t)I(t)}{N(t)}$ and the total number of interactions resulting in infection is $\beta \frac{S(t)I(t)}{N(t)}$. Here, $N(t)$ represents the total population where $S(t) + I(t) = N(t)$. This simple model can then be represented as difference equations:

$$S(t + \Delta t) = S(t) - \Delta t \; \beta \frac{S(t)I(t)}{N(t)} \qquad (1.1)$$

$$I(t + \Delta t) = I(t) + \Delta t \; \beta \frac{S(t)I(t)}{N(t)} \qquad (1.2)$$

This simple model can be modified in many ways to represent a system more accurately by including, for example, features such as death and birth/growth rates. Note that this model can be transformed to a set of differential equations in continuous time by rearranging Equation 1.1 and Equation 1.2 and taking the limit as $\Delta t \to 0$.

Compartments can be added to describe different stages of infection. For diseases with a latency period, an "Exposed" compartment is added, resulting in a SEI (Susceptible - Exposed - Infected) system. The SIS (Susceptible - Infected - Susceptible) model describes when infected individuals can recover and instantly can become susceptible again. The SIRD (Susceptible - Infected - Recovered - Dead) model includes a "Recovered" compartment for those that gain immunity after infection, and a "Dead" compartment to describe individuals that die during the simulation timeframe.

The spread of plant pests and diseases can be modelled using these techniques by compartmentalising a population of plants into different stages of infection [76–78]. Additionally, plant pests can be modelled compartmentally in conjunction with their hosts in vector-host systems [79].

Spatial structure can be added to these compartmental models by segmenting the population into a "population of populations" [80]. A disease spreads within and across subpopulations by the movements of individuals [80]. Modelling these systems then brings forward the question of how subpopulation contacts are structured. Network theory can help answer this question if we represent each subpopulation as a node in a network and travel/interaction between subpopulations as edges. This idea is referred to as the *net-*

worked metapopulation and is used for modelling the large-scale spatial transmission of diseases [81].

1.3.3   Network construction methods

To construct a general model network first, we require a definition of the nodes in the network and of what the edges between nodes represent. Defining nodes can be a challenge because there may be uncertainty regarding how many qualitatively different types of nodes the system being described has. Defining edges will be driven by the research question in mind, although deciding what the edge weights represent may be difficult, or whether to include edge weights at all. These challenges can arise due to limited available data.

In the context of constructing a plant trade network, there is commonly insufficient data available. This is likely due to the fact that acquiring large amounts of data to inform a trade network is difficult and expensive. Data from a significant number of sellers and traders can be difficult to obtain for several reasons. Sellers and traders can be reluctant to give access to commercially sensitive sales data due to worries of potential reputation damage. For example, being associated with diseased stock. Usually, this obstacle is overcome if the seller/trader has a sense of trust in the researcher, established through previous working relationships. Understandably, this is difficult to achieve with a substantial number of sellers and traders in a particular industry. Given that this obstacle has been overcome, it can be of high cost in time and money to collate and clean sales data from many different sellers and traders due to the distinct and potentially incompatible ways businesses can store and record this data. Given these limitations, a more achievable goal is to obtain large data sets from a smaller number of sellers and traders.

When data is unavailable to directly analyse a network representation of a system, analysis can be conducted on many idealised networks with different properties as an alternative. These network qualities include for example the degree distribution and the level of *clustering* (the extent to which neighbours of nodes are also directly connected) in the network.

Examples include algorithms to produce networks which are *random*, *scale-free*, *local* and *small-world*.

A network is *random* when edges between nodes are assigned randomly, with a given probability [82]. The Erdős-Rényi model produces a random graph using two parameters N (number of nodes) and L (number of links) [82]. An alternate formulation used a parameter p which represents the probability of assigning a link to any two nodes instead of L.

Scale-free networks are largely studied, due to their frequent observation in nature [83]. A network is *scale-free* if the degree distribution follows a power law, meaning the fraction of nodes in a network that have a degree of at least k follows the distribution $k^{-\gamma}$. The parameter $\gamma > 0$ controls the spread of this distribution, where a larger value represents a network with fewer nodes that have a high degree. It is worth noting that in most scale-free networks that have been studied, the value of γ lies between 2 and 3 [84]. It is within this parameter region where the degree distribution has finite mean but infinite variance [71].

There are many existing algorithms to construct a scale-free network, one of these is the *Barabási-Albert* model [85]. The algorithm starts with a small number of nodes and edges, and in each subsequent step, a node is added with a number of edges which is at most equal to the current size of the network [85]. Nodes preferentially attach to nodes with larger degrees, meaning the probability that a node acquires a new link is proportional to its degree [85].

Local networks have high clustering and high path lengths, such as regular lattices [86]. A network is *small-world* if it is local and shortcuts are added to lower the path length in the network [86]. Small-world networks can be constructed with the *Watts-Strogatz* algorithm [87]. This algorithm starts with a regular ring lattice and for each edge in the network, it is rewired with probability p [87].

When constructing a trade network, the algorithms described above can be useful if there is sufficient data or existing research to make assumptions about the network structure. When investigating a novel trade network, however, these resources may not exist. When there has been no data on plant trade networks, research has been conducted on comparing many different assumed network structures, with a focus on its implication on disease spread [70, 88, 89].

Other than retroactive node classification [70], it is unclear how to implement these methods when a trade network involves nodes which are split into categories. An example of node categorisation would be to label nodes as either "buyers" or "sellers" in a simple *bipartite* trade network. A network is *bipartite* if nodes can be split into two sets, where edges only join nodes between sets [71]. The categorisation of nodes in real-world networks may be of relevance to the objectives of the researcher(s) and can help apply the results from network analysis to the real world. Hence, for these systems, node categories should be incorporated into the network construction approach.

1.3.4   Edge weights and network models

Edge weights (or flows) in a network can represent, for example, distances, costs and volumes of trade. When a network is unweighted, the strengths of the connections between nodes are not considered. For plant trade networks and the spread of disease, the volume of trade between nodes would affect the probability of transmission from node to node. To illustrate this, a consignment of 10 plants from a nursery would have a lower probability to include an infected plant than a consignment of 1 000 plants, assuming that the nursery is not fully infected. Hence, when the strengths of connections between nodes differ and are an important factor, trade flows should be included in the network construction. There are many existing methods to model trade flows, some of which we discuss in this section.

The simplest way of modelling edge weights in a network is to assign weights based on data. An example which we have already discussed includes the work of *Chapman et al.*, who used geographical distance, species prevalence, climate similarity and world trade data to

inform the edge weights on their model of global trade networks [50]. Another example includes *McKelvey et al.* who estimated what proportion of total commodities consisted of nursery stock to approximate trade volumes [2]. In absence of similar data to inform edge weights, there are other methods that can be used.

In econometrics, the gravity model, and variations of it have been used to model trade flows in networks in absence of trading data. The gravity model stems from the idea that trade flows will be inversely proportional to the distance between nodes (*resistance factor*), and proportional to some measure of mass of the interacting nodes multiplied by a constant, representing gravity. This method of modelling trade flows has been primarily used to model trade and migration across and between countries on a large scale [90, 91]. The gravity model has also been used to model the movement of people between cities in the context of spreading disease [92]. One immediate limitation of this approach is that in an increasingly globalised world, the relationship between the flow of goods and distance does not strictly hold for all systems. Another potential limitation of this method is that it requires estimations of the total amount of flow at a particular node, which a researcher may not have adequate data to estimate. *Stefanouli and Polyzoz* highlight further limitations; calibrating the parameters to data can be difficult and the fitting parameters can vary significantly even within a data set, fitted parameters can have values which cannot be seen in reality, some significant variables are omitted in gravity models, gravity models in heterogeneous population groups can be subject to misspecification and the geographical subdivision into "origin" and "destination" omits other catchment centres which would make the models more realistic [93]. Note that there are other models stemming from physics that model trade flows such as the radiation model, which comes from diffusion dynamics and the intervening opportunities theory [93].

Other approaches have focused on using machine learning methods to predict edge weights based on similarities between nodes [94, 95]. However, these approaches require a large quantity of data for training and testing, which is not always available. As noted by *Fu et*

al. "Link weight prediction is a relatively new topic" [95], and so this area can certainly benefit from further developments.

1.4 Modelling disease spread on networks

Modelling the spread of disease on a network involves classifying nodes in a state of infection over time, where connections in the network drive disease spread from node to node. In this section we discuss the relevant literature for modelling disease spread on networks, with particular emphasis on Susceptible-Infected compartmental models.

Disease spread can be described as a Markov chain model, where the state space of the network consists of all possible combinations of node infection states across the entire network [96]. For the SI model, and a network of N nodes, this results in 2^N states. A system of equations, the *forward Kolmogorov equations*, can be defined to describe how the network can change from one state to another [96]. Having larger networks or increasing the number of compartments can give rise to difficulties in dealing with a system of very high dimension. More techniques such as mean field approximations, individual-based models, dynamic networks and non-Markovian epidemics are discussed in [96].

To simulate the spread of disease on a network in, for example, an SI compartmental model, we have two options: a binary system where nodes are classed as either susceptible or infected (similar to the models discussed in [96]), or a metapopulation model where nodes act as subpopulations of individuals and we model the total number of susceptible and infected within each node.

In a binary system, nodes are classed as either susceptible or infected and the edge weights represent the probability of transmission. All nodes have self-edges whose weights control the internal spread of disease within the node. The question is then how to model the probability of transmission. For the network we will be investigating, the volume of trade across the network is heterogeneous. Therefore, unless the disease is highly infectious, the

volume of trade will affect the probability of transmission from one node to another and thus should be included in an epidemiological model.

In metapopulation models, each node acts as a subpopulation of individuals and the edges represent the movement of susceptible and infected individuals between them. We can either model the proportion or the total number of susceptible and infected individuals in each node. In most real-world systems, subpopulation sizes and the number of individuals moving from one subpopulation to another will be heterogeneous. Modelling this heterogeneous movement of susceptible and infected in terms of proportions of subpopulations is therefore not appropriate.

Alternatively, we can model in the units of individuals as opposed to proportions. This avoids problems with modelling the heterogeneous movement of individuals amongst subpopulations of different sizes. Thus, edge weights in the network are required to represent the movement of a number of susceptible and infected individuals (similar to the work from *McKelvey et al.* [2]). The next question becomes how to model the volume of movement of individuals on each edge in the network.

For our system of UK plant trade, the volume of trade between nodes is heterogeneous. Therefore, we require a way of capturing this heterogeneity. Our data gave information on trade volumes, consisting of several years of sales from thousands of customers. We can use data for each group of customers to model the distribution for the number of plants a customer from that group purchases. Edge weights can then be generated by sampling from these distributions. This approach would result in a weighted, directed network with connectance modelled directly from sales data.

Colizza and Vespignani studied disease spread dynamics in metapopulation networks with heterogeneous connectivity [97]. Their analysis was conducted on a network of arbitrary degree distribution, where disease spread was modelled with a simple branching process. The number of infected subpopulations in the nth generation with degree k could be expressed as a function of the same quantity at the previous generation. Movement across the net-

work was modelled with a parameter p, and edge weights were assigned such that weights from all edges of a given node of degree k were p/k. In their analysis they found a threshold for global disease spread, dependent on mobility across subpopulations and network topology. The results imply that network heterogeneity lowers the threshold for global spread and suggest an inefficacy of using travel restrictions to contain global epidemics. The authors then used parameters from real world transportation networks, performing Monte Carlo simulations on the metapopulation system. The networks analysed agreed with their analytical results. The authors found that a reduction of one order of magnitude of the mobility, p, was insufficient to bring the system below the threshold for global spread. An important note to make about this model is that all edge weights stemming from one node were the same. This is unrealistic for trade networks, where a business will not necessarily sell the same amount of stock to all customers. The investigation in [97] takes into account heterogeneity in how links are distributed and how edge weights differ amongst nodes of different degree. However, heterogeneity in edge weights stemming from one node have not been considered which would more accurately represent certain real-world systems.

1.4.1 Modelling disease spread on trade networks

In this section we highlight previous approaches to modelling the spread of disease, specifically on trade networks. We initially focus on previous network investigations into plant trade, and then present similar work on the trade of animals.

Plant trade

In the review entitled "Network epidemiology and plant trade networks", *Pautasso and Jeger* highlight key developments in plant network epidemiology [98]. Research on disease spread on networks has revealed that network structure plays an important role in epidemic dynamics [99–104]. As discussed in Section 1.2, plant trade on both the international and domestic scale has been attributed to the spread of plant pests and diseases. Plant trade can naturally be conceptualised as a network of individual traders connected by transactions amongst one another. Therefore, analysing the network structure of plant trade will

help us understand how plant pests and diseases spread and, more importantly, how to mitigate this spread.

When modelling the spread of plant diseases in the units of individual plants, each node in the network represents a metapopulation of plants (a plant nursery, garden centre, customer, etc.). We distinguish between susceptible and infected plants by SI compartments. The connections in the network, alongside other processes, will affect the spread of disease within and throughout the network.

Harwood et al. developed an SEIS (Susceptible - Exposed - Infected - Susceptible) model that linked the plant trade network in the UK to the landscape via a grid [46]. They focused on modelling the spread of *P. ramorum* and *P. kernoviae.* Their network of plant trade was constructed using AgCensus data for England and Wales in combination with data from the HortWeek Suppliers Guide to inform the distribution of sites that grow susceptible nursery stock. It was assumed that the spatial distribution of susceptible stock of Scotland was similar to England, hence they overlaid a subset of the AgCensus data from England over Scotland. The locations of retailers were informed by the records of the Plant Health and Seeds Inspectorate. Links were established from data to connect holdings with distribution hubs and retailers with outlets. However, links between suppliers and retailers were generated stochastically due to data unavailability. The network construction yielded a small-world and scale-free network. Trade was simulated via a specified trading period, randomly selected between 1 and 60 days during the months of March and August. The authors posed skepticism on modelling the spread of disease on the plant trade network without also incorporating the spread via the landscape. Network models that wish to not also incorporate the landscape should then focus on pests and diseases which are not already established outside the network. *Harwood et al.* modelled trade flows in their plant trade model where nodes had an assumed capacity, N, and $0.6N$ plants were assumed to be traded annually. The trade of plants was divided equally over each node's sales period.

McKelvey et al. developed a bipartite network model to predict the potential spread of *P. ramorum* in the US using plant nursery trade flows [2]. The authors used a Bayesian approach, modelling probabilities of transmission of *P. ramorum* from entry to destination nodes in the network. The bipartite network described nursery stock shipments from the west coast of the US to destinations eastward where *P. ramorum* is undetected in the natural forest landscape. Most of the network construction used data from the Freight Analysis Framework Commodity Origin-Destination database (a relational database describing the movement of commodities, in tonnage and monetary value), between statistical regions. These regions acted as nodes in the model network. The probability of disease transmission was driven by trade volumes. The authors did not have access to nursery trade volumes specifically, and so multiplied the volume of movement by a referenced proportion of commodities that are nursery stock. The network approach used by *McKelvey et al.* was very general, the bipartite network consisted of sources of infection, destinations to uninfected nodes and a probability of infection from infected to those uninfected.

Bayesian analysis was used to make comparisons about the likelihoods of various paths of infection and to identify likely proximate infection sources [2]. The authors highlighted that the model can also identify new sites that are at high risk of importing disease from these proximate sources, even in absence of detection. The probability of transmission was determined by the parameter p, which defined the probability that a single unit of flow (unit of trade) results in the initial infection of the flows destination. There were limitations to the method the authors used, including the assumptions that the probability of infecting different destinations from a given source is independent and that new infections all come from one source.

The method of network construction by *McKelvey et al.* has the advantage of simplicity, a bipartite structure which only considers the links connecting infected to uninfected nodes [2]. However, this method does not consider the effects of connections amongst infected nodes. As already highlighted in this thesis, network structure has been shown to

have an effect on disease spread dynamics and this is something the model does not take into account.

Pautasso and Jeger analysed the relationship between probability of persistence and of transmission in small-sized (100 nodes) directed networks [88]. Disease spread on the network was a deterministic Susceptible-Infected-Susceptible (SIS) model, with a probability of persistence parameter P_p and probability of transmission P_t. The latter was a homogeneous edge weight to all edges in the network. The authors simulated this disease model on multiple replicates of four network structures: Local, small-world, scale-free and random. For all networks, there was a fixed number of links, set at an arbitrary value of 369 to ensure that networks on a lattice were well-connected.

Pautasso and Jeger found that the epidemic threshold (the level of disease transmission which separates the occurrence of a disease spreading or dying out) can be described by a linear relationship between the probabilities of transmission and persistence, for all network structures [88]. Small-world and scale-free networks showed a significantly lower epidemic threshold, as opposed to random and local network structures. Interestingly, epidemic threshold was not affected by the node the disease started in, although the starting node did affect the final size of disease equilibrium. Network structure was shown to influence final epidemic size from averaging final epidemic sizes over all starting nodes, with random structures showing significantly larger epidemics than small-world structures. Overall, from their analysis, the authors showed that even in small-sized networks, different network structures significantly affect the spread of disease.

Moslonka-Lefebvre et al. investigated the effect of network structure on disease spread in small-sized (100 and 500 nodes) networks [89]. The authors looked at the correlation coefficient between links to and from nodes for networks of varied size, structure, connectance and clustering. The network structures considered were local, random, small-world and three variations of scale-free (positive, negative and no correlation between in and out-degree of nodes). Disease spread was modelled using the same method as in *Pautasso and Jeger* [88]. The important difference in this investigation is that the probability of per-

sistence parameter was set to 0 and the epidemic threshold was thus calculated from the probability of transmission parameter only [89]. The authors state from their unpublished observations that epidemic threshold conditions did not change with single or multiple seeded infections [89]. Analysis of variance tests were conducted on the epidemic threshold and the correlation coefficient between in- and out-degrees for each network replicate. This was repeated for each level of connectance and network structure. Additionally multivariate regressions were conducted for the epidemic threshold against the correlation coefficient between in- and out-degree and the average clustering coefficient for different network sizes, structures and connectance levels.

The authors found that the epidemic threshold decreased significantly with increased connectance levels for all network structures and sizes [89]. They also found that the epidemic threshold is lower for scale-free networks only with a positive correlation between in- and out-degree of nodes. It has been noted that this result also holds for large networks [89,105]. The authors found that in sparsely connected networks, clustering can become an important factor for disease spread. *Moslonka-Lefebvre et al.* state that their main conclusion is that "in directed networks, analyses of the influence of clustering on the epidemic threshold can be spurious if they do not consider simultaneously the effect of the correlation coefficient between in- and out-degree" [89]. In this investigation, the authors found that the epidemic threshold is negatively related to the correlation coefficient between in- and out-degree for all network structures, irrespective of network size. For small-sized and sparse networks, however, this does not hold and clustering becomes an important factor. Clustering was found to not always have an effect on epidemic threshold on small-sized networks, which is not the case with large networks.

Pautasso et al., in a further study, investigated the role of hierarchical categories (producers, wholesalers and retailers) on small-size (100 nodes) directed networks [70]. Without data to inform network structure, the authors modelled disease spread in networks at four connectance levels (local, random, small-world and scale-free). In addition, scale-free networks were analysed with positive, no and negative correlation between in- and out-degree. Each network structure was analysed with 100, 200, 400 and 1000 links. The same dis-

ease spread model was used as in *Pautasso and Jeger*, with a probability of persistence P_p and probability of transmission P_t [88]. The value of P_t was set to the reciprocal of the leading eigenvalue of the adjacency matrix for each network realisation [70]. P_p was also the same for all nodes. In their model, all nodes were of equal capacity, and the only differences between nodes were due to their in- and out-degrees. Nodes were classified retroactively into the category producer, wholesaler and retailer according to the equation: $\Delta = (x-y)/[(x+y)/2)]\%$, where x and y represented the number of outgoing and incoming links respectively. The value of Δ was calculated for all nodes, where nodes were classified as producers with Δ at least 20%, and retailers with Δ at most -20%. All remaining nodes were classified as wholesalers. *Pautasso et al.* also varied these ranges for producers to (60%, 100%) and retailers to $(-60\%, -100\%)$. From their analysis, the authors found that variations in producers, wholesalers and retailers can affect the epidemic threshold [70].

Nelson and Bone used the classification of plant nurseries into three tiers (growers, wholesalers and retailers) [106]. Growers were defined to produce plants, trade amongst other growers and sell to wholesalers. Wholesalers bought from growers, traded with other wholesalers and sold to retailers. Retailers bought from wholesalers and sold to the public. This tiered system comes from *Pautasso et al.*, where nodes are assigned to groups based on their ratio of in-coming and out-going links [70]. Networks were generated stochastically, with an assumed scale-free structure. First a scale-free network was generated between wholesalers and growers using the "ba.game" function of the igraph package for the programming language R [106, 107]. Then intra-tier edges amongst growers and then wholesalers were added using a preferential attachment algorithm adapted from *Barábasi and Albert* [108] where the probability of adding a new edge between two nodes is proportional to the sum of the node's in- and out-degrees. Then using the same algorithm, inter-tier edges were added between wholesalers and growers. Retailers were finally added using this preferential attachment algorithm. Note that this model uses the assumption that the network is scale-free, in absence of data, and uses algorithms to produce an approximately scale-free network in this tiered system. The ratio of incoming to outgoing links that determined a node as a wholesaler, grower and retailer was pre-specified and analyses were conducted on the effect of varying the proportion of nodes in each tier [106]. *Nelson and Bone* modelled

the spread of disease on the model trade network using the same probability of transmission and persistence model as in [70, 88]. Nodes in the network had an associated level of infection, a real number in the interval $[0, 1]$ [106]. The probability of persistence was set to 0.5 by default and the probability of transmission was set for each simulated network such that the leading eigenvalue of the transition matrix was 1.1 (above 1 so that the disease spread). In their analysis, the authors found a linear relationship of the epidemic threshold between probabilities of transmission and persistence, as was found in *Pautasso et al.* [88].

We summarise the literature discussed for plant trade network modelling in the table below.

Table 1.1. Summary of plant trade network models discussed based on the network and disease models used.

Disease Transmission Model	Real/Simulated Network	Discrete/Continuous Time	Comments	Reference
SEIS	Combined real and simulated	Continuous	Simulated network for Scotland	[46]
			Simulated links to fill in data gaps	
Bayesian probabilistic model	Real	N/A	Find most likely sources of infection	[2]
			No connections between infected nodes	
SIS	Simulated	Discrete	Homogeneous edge weights	[88, 89, 106]
SIS	Simulated	Discrete	Homogeneous edge weights	[70]
			Retroactive classification of hierarchical categories	

Research on the network structure of plant trade in particular has suffered from insufficient data. In comparison to animal trade, availability of data has allowed for research highlighting the contact structure of salmon and trout in Great Britain [109], the trade of pigs [110] and the movements of livestock in Scotland [111, 112]. Data collection is needed to develop further understanding of the movement of plants via trade and the spread of plant pests and diseases. In the next section, we provide examples of the research conducted using animal trade networks.

Animal trade

There is a vast literature of the trade of animals using network theory. For example, a systematic literature search on poultry trade networks and avian influenza found 362 references to influenza spread pathways and 57 references on the human and animal movements of poultry within production and trade networks [113]. From this review, we found three common approaches in the studies that analysed poultry trade networks: surveys that

were conducted to get data on nodes and edges [114–122], networks that were constructed from well understood networks and existing data sets [122–130], or theoretic networks with some studies using existing data for model fitting [131, 132].

There has been a lot of research on trade networks of livestock with a focus on the spread of foot and mouth disease, highlighted in [133]. The increased regulation on the movements of livestock resulted in greater access to data on movements such that network analysis could be conducted [133]. Similarly, there has been substantial research on the spread of swine flu amongst pigs, with many different modelling approaches, shown in the review by *Andraud and Rose* [134]. The authors highlighted several models on pig trade networks and the spread of disease [135–137]. There has also been extensive infectious disease modelling in aquatic systems, as reviewed by *Murray and Salama* [138]. The authors highlight network model approaches in [139–142].

Jones et al. combined data from England, Wales and Scotland to produce a *multi-layered network* for the salmon and trout aquaculture industry [109]. A multi-layered network is a network which has multiple types of nodes and connections, to represent different types of relationships [143]. Each layer is a set of edges which represent a specific relationship [143]. In this network, the nodes represent fish farms and recreational fisheries. The layers include: the movement of live fish between nodes, waterborne pathogen transmission via the river system, waterborne transmission in the marine environment, and local transmission from humans or animals nearby nodes. *Jones et al.* constructed their layered network from data provided by the Centre for Environment, Fisheries and Aquaculture Science and by Marine Science Scotland, detailing live fish movements from 2009 to 2013. Such data is available for fish movements due to legislation by the European Union [109]. The authors were able to get data on river connections between sites from the European Environment Agency's *European Catchments and River Network System* and data for water currents from the Scottish Environment Protection Agency [109]. With this data, the authors were able to capture the full industry structure over a three-year period. The network was analysed using several metrics, including the in-degree, out-degree and betweenness centrality measures. More detail on the metrics used can be found in [109]. Note that the authors

found the betweenness measure to be the most effective out of many measures investigated. The authors found that most transport network nodes are reachable from cross-boarder connections, highlighting the importance of tracking live fish movements in Great Britain. They also showed that large epidemics in the salmon and trout aquaculture industry are dependent on the live fish transport network.

However, data is sometimes not widely available in animal trade. For instance in developing countries, the live movement of animals can be predominantly driven by the backyard production of animals [144]. *Kukielka et al.* in their investigation of pig trade in the country of Georgia used data from a questionnaire given to 487 farmers to inform a network model [135]. Shipment origin and destination information was collected at the geographic level of villages and thus nodes represented villages in their network. An edge existed between nodes if at least one shipment was made from one village to another. The authors could not capture the full network, as only four of nine regions of the country were included in the study. An exponential random graph model was used to analyse pig trade between villages as a function of village characteristics and network structure. Exponential random graph models are statistical models to make inferences on what drives network structure in an observed network [145]. The limitation to this method is the reliance on having access to enough data to construct an observed network. The more of a network that is modelled, the more the assumptions from that model will influence the analysis.

Lentz et al. analysed the trade of pigs in Germany in relation to the spread of infectious diseases in livestock populations [110]. To do this, they used data from the HI-Tier database [146], to construct three trade networks: a static network, a network as a time-series and a network where time and causality are taken into account. In their analysis, they highlight the connectivity differences in networks with and without directed edges. Almost all nodes were reachable from any other node with undirected edges; however, when considering directed edges this reduced to approximately a quarter of total nodes. *Lentz et al.* posit that it is common in trade networks that the "giant strongly connected component" and the "giant in-component" have high disease spread potential, whereas the "giant out-component", "tendrils" and "external nodes" do not. For brevity, we will not

define these, but full details can be found in [110]. The authors also point out that it is common in trade networks for nodes to either have very long or very short ranges (the number of nodes reachable from a particular node).

The authors calculated assortativity coefficients (the tendency for nodes to share connections with similar nodes) across different categories and found that there was high assortativity for nodes to connect within the same federal state, implying that trade restrictions across federal state borders would have a small effect on network connections yet contain spread [110]. They found that the static network can capture nodes as part of two risk classes, however more detailed understanding can be gained, such as maximum outbreak size, from the temporal network (network over snapshots of time). Interestingly, they found that the volume of trade did not alter their results, when considering weighted or unweighted networks. In their work on the temporal network, they were able to use the "unfolding accessibility" method to mimic a Susceptible-Infected model to understand how a worst-case scenario epidemic progresses and use it for causal contact tracing. Access to data in similar quantities and detail for plant trade would allow for similar investigations.

Enright and Kao used data to produce a network on the movements of cattle in Scotland [112]. Their investigation was to assess the potential epidemiological consequences of the Cattle Tracing System (CTS), which allowed the unreported movement of cattle between holdings. The authors' question was how the inclusion of these additional network links makes the system more vulnerable to the spread of disease. They analysed five undirected networks of annual trade and 36 directed networks consisting of three networks for each month of 2014 (one for recorded cattle movements, another with CTS links, and a third network which contracted all directly or indirectly connected holdings into one node). The authors assessed network susceptibility to disease spread by analysing the length of infection and infector chains. An infection chain for a node is the total number of nodes that can be directly or indirectly infected by that node [112, 133]. Conversely, an infector chain for a node is the total number of other nodes that can either directly or indirectly infect that node [112]. They found that when CTS links were included in the network, the average number of holdings a node could infect more than doubled. *Enright and Kao*

concluded that the strategic removal of a few links in the network would reduce outbreak potential and highlight the utility of monitoring the cattle trade as a network as opposed to many separate holdings [112].

Throughout this section, we see that researchers either have data on the entire network of interest and require no modelling for the network or have very little network data and resort to theoretical investigations, testing many different contact structures. This first, data-driven approach can be seen in research focused on animal trade. The data collection and legislation involved in the movements of livestock allow for entire trade networks to be analysed, without use of modelling network contact structure [109–112].

A less extreme case of this method was used for the plant trade in Great Britain in the previously mentioned study by *Harwood et al.* [46]. The authors used spatial data of nursery stock covering England and Wales and data on business sizes for growers to recreate as much of a plant trade network as possible and then modelling was used to fill in gaps. Though a significant amount of data was used to create a consistent network structure, the approach was limited regarding edge weights. This was due to the absence of detailed sales data.

The second, theoretical approach is used with no data to inform network construction and a network structure is assumed. Examples of this approach are found in *Pautasso and Jeger*, *Moslonka-Lefebvre et al.*, *Pautasso et al.* and *Nelson and Bone* [70, 88, 89, 106]. These approaches can be limiting, as no data is used to inform network structure or edge weights. Thus, we see a need for the development of network construction methods that incorporate data when availability is limited.

1.5 Thesis overview

Our objective is to model the domestic plant trade network in the UK. We draw inspiration from the approach of *Pautasso et al.* [70] and analyse sales data to gain a better understanding of how nurseries distribute their connections and classify nodes into subcategories. Thus, we develop a novel phenomenological approach to construct a model trade network that includes node subcategories as a parameter, as opposed to retroactive assignment [70].

In our disease spread model, we extend the single-nursery compartmental model by *Bate et al.* [76] to include every node in the network and the trade of infected plants between nodes. We incorporate the volume of trade to include its impact on the trade of infected plants, hence expanding upon the models that assume equal probability of transmission between nodes [86, 106]. Similar to *Harwood et al.*, we investigate the effectiveness of inspection regimes [46], considering both periodic stock inspections and inspections conducted during trade.

The thesis consists of six chapters. In Chapter 1, we introduced the key concepts for modelling the spread of disease on networks and compartmental, metapopulation models that we used throughout the thesis. We reviewed the existing literature on modelling plant trade network structures and of plant disease spread via trade.

Chapter 2 details the methods we developed to construct a general trade network model involving the assignment of nodes into subcategories. In Chapter 3, we apply these methods to construct a model representing annual trade in the UK domestic plant trade network, using plant nursery sales data. We use social network analysis to investigate the network structure, looking for weaknesses in an epidemiological context.

In Chapter 4, we construct a Susceptible-Infected compartmental metapopulation model, simulating the spread of a pest or disease on the network model from Chapter 3. We investigate the effects of different seedings (where the disease begins) and compare the effectiveness of scheduled stock inspections against trade inspections. Note that the code

used to run simulations and conduct the analysis in Chapters 2, 3 and 4 are stored on the Github repository https://github.com/rnb19177/PlantTradeModel.git.

Chapter 5 focuses on a collaborative project which started in May 2020 in response to the COVID-19 pandemic. This project develops a compartmental metapopulation model of the Scottish health board Lothian, dividing the population into care home residents, staff and the general population.

We conclude this thesis with Chapter 6, which summarises our findings and highlights areas for future research.

Chapter 2

General Network Construction

In the literature review on network construction methods from Chapter 1, we highlighted that there appeared to be two approaches which were based on the two extremes of data availability. With the data that we had access to, we were situated somewhere in between these extremes. In our data analysis (detailed in Chapter 3), we identify four distinct trading patterns of plant nurseries. To construct a network where nurseries followed these patterns, we could not use existing methods such as those seen in previous models [46, 70, 88, 89, 106]. Thus, it was necessary to design a new method to model the network for domestic plant trade in the UK. Once this network was constructed, we developed a method of adding heterogeneous weights to the edges of the network using this grouping of nodes in conjunction with sales data. We generalised this network construction approach to highlight the methods we developed and the applicability of this method to other systems.

Throughout this chapter we describe in detail the method we designed to construct model hierarchical trade networks with heterogeneous edge weights. The method is inspired by the hierarchical categories of *Pautasso et al.* [70]. We structure this chapter by first defining nodes and secondly defining edges and edge weights. We present the network construction algorithm, followed by a summary of model parameters and discussion. Throughout the chapter, for ease of interpretation, we provide illustrated examples.

2.1 Introduction and background

Our goal is to simulate the spread of a plant pest or disease on a model of domestic plant trade for the UK, informed from sales data we have access to (detailed in Chapter 3). We want to model the entire domestic plant trade network for the UK, including customers. By including customers in a network model, there is potential for assessing the exposure that customers have to a pest or disease spreading throughout the network. To the best of our knowledge, this is the first model that attempts to do this, as other models available in the literature only focus on the grower/producer, wholesaler and retailer section of trade [70, 88, 89, 106]. Furthermore, from talks at FERA (Food and Environment Research Agency), it was brought to our attention that, unlike what was seen in previous models [88, 89, 106], plant nurseries do not play discrete roles in the market. In fact, there is a spectrum of trading behaviour. We note that the changes in ranges of Δ (ratio of in- and- out degree) to retroactively classify nodes as a producer, wholesaler and retailer in *Pautasso et al.* did attempt to include this spectrum of trading behaviour, however with a view constrained to the grower/producer, wholesaler and retailer section of trade [70].

Our interest is in how plant pests and diseases spread via domestic trade alone, and so our modelling approach diverges from network models incorporating spatial spread on a landscape such as in *Harwood et al.* [46]. There is, to the best of our knowledge, no existing research that models the entire network of domestic plant trade or outlines all of the individuals involved in the plant trade industry. This is primarily due to the lack of data with enough detail to conduct these investigations.

Our ultimate goal is to construct a network which is compatible with the SI model constructed in Chapter 4. The first step in our modelling approach is to characterise the nodes. To this end, we investigate the sales data of plant nurseries to identify all of the different groups of people that buy plants in the UK. Then we look at the proportions of each nursery's sales towards these groups to understand how nurseries trade in the network. From this, we identify multiple patterns that describe how nurseries distribute their sales to nodes in the network. If we know how the proportion of a nursery's out-going edges

are distributed, we can determine how edges are assigned with a pre-specified number of customers (out-degree). From these observed patterns, it is immediate that modelling the connections between the groups of nodes, in accordance with these patterns, requires a novel network construction approach.

This method of construction allows for a network model with parameter sets describing: the initial distribution of seller and trader nodes, in-degree/out-degree distribution parameters and consignment size distribution parameters. In conjunction with this method, available data and existing literature, researchers can construct a model trade network for analysis. This method of network construction can create the landscape for which a researcher can use to further elucidate data requirements. Unlike other methods to construct a network, this method is not tied to an assumed structure as is with the methods to construct scale-free networks or random graphs. This is also one of the few methods which will model weighted trade networks and is generalised for many varieties of node categorisations. A similar approach, the weighted stochastic block model, has been used to construct weighted networks, sampling from exponential distributions [147]. Our process aims to logically construct a model network from node categories, characterised by trading behaviour inferred from data analysis, existing literature and informed assumptions.

2.2   Methods

In this section, we describe our methodology to construct a weighted trade network. We start with defining the nodes and edges in the network and then describe our approach to node subcategorisation, modelling node out-degrees/in-degrees and modelling edge weights. We then describe in detail our algorithm for constructing the network.

When constructing any model network, it is crucial to first define what the nodes and the edges between them represent. Once this is established, the next steps further characterise the nodes according to their features; for example, nodes may be labelled depending on their role within the network or edges may be assigned directions/weights. All of these

details will depend on the system being modelled and what type of analysis is being conducted on the network. Our interest is in modelling trade networks, and therefore in the following we will refer exclusively to this type of network. The networks built using this model are all weakly-connected by construction.

2.2.1 Node classification

We begin our model description by characterising nodes, i.e., agents in the network. In a general trade network, we can identify three types of nodes based on the trading behaviour of the agents they represent. In fact, we can categorise nodes as *sellers*, who solely sell goods to other nodes, *traders*, who buy from sellers and other traders and sell to others, and *buyers*, who exclusively purchase from other nodes in the network [70]. This classification on nodes naturally induces a hierarchy between them, which we schematically represent in Figure 2.1. In the figure, the arrows represent the flow of goods. The nodes at the top of the hierarchy do not buy from anyone in the network, meaning either these nodes produce the goods they sell or buy from outside the network. Similarly, the nodes at the bottom do not sell to anyone in the network, meaning they either sell outside the network or use the goods in some other way.
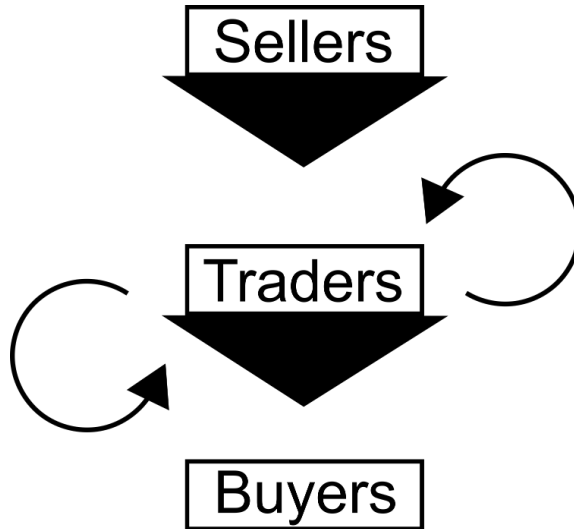
Figure 2.1. Schematic depiction of the hierarchy in trade networks.

2.2.2 Edge classification

Now that nodes have been classified, we move on to describing the edges in our network model. An edge between nodes in a trade network typically represents a trading relationship between those nodes, with the direction indicating the selling of goods from one node to another. A weight may also be assigned to the edges; this can be used to represent, for instance, the quantity of goods exchanged, the value of the transaction or the physical distance between nodes. Going forward, edges in our model network will represent selling transactions. Edges originate from the node selling the goods and point to the node purchasing them. The weight of each edge represents the quantity of products sold, which we will refer to as *consignment size*.

2.2.3 Node categorisation

So far, we have categorised nodes depending on where they sit in the trade hierarchy. Nodes are either sellers, traders or buyers. This classification induces a partition of V, the set of nodes in the network. We represent this partition by $V = V^{(s)} \cup V^{(t)} \cup V^{(b)}$. These sets can then partition further to account for subcategories:

$$V^{(s)} = \bigcup_{i=1}^{n_s} V_i^{(s)}, \quad V^{(t)} = \bigcup_{i=1}^{n_t} V_i^{(t)}, \quad V^{(b)} = \bigcup_{i=1}^{n_b} V_i^{(b)}.$$

The nodes in subcategory $v \in V$ are therefore denoted $\{N_1^v, N_2^v, \ldots, N_{|v|}^v\}$.

These subcategories can be defined based on any criterion which can distinguish between different sellers, traders and buyers. Some examples of distinguishing characteristics for subcategories are the ratio of in-going to out-going links, the ratio of in-strength/out-strength, or qualitatively distinct *customer demographics*.

Definition 2.2.1 (customer demographic)**.** Let v be a node subcategory. The *customer demographic* $\mathcal{D}^v(\omega)$ of node ω in subcategory v is:

$$\mathcal{D}^v(\omega) = (p_{t_1}, \ldots, p_{t_{n_t}}, p_{b_1}, \ldots, p_{b_{n_b}}),$$

where

$$p_i \in [0, 1], \text{ for all } i \in \{t_1, \ldots, t_{n_t}, b_1, \ldots, b_{n_b}\},$$

and

$$\sum_{i=1}^{n_t} p_{t_i} + \sum_{i=1}^{n_b} p_{b_i} = 1.$$

In this chapter, we subcategorise nodes based on their customer demographics. However, the model easily generalises to situations where the nodes are subcategorised according to other criteria.

We proceed as follows, starting at the bottom of the trading hierarchy; cf., Figure 2.1. Since we are subcategorising nodes based on their customer demographic and since buyers do not have customers by definition of our model, buyer nodes can be subcategorised following any criterion meaningful to the individual researcher. Some examples would be subcategorising buyers based on how the product is used, geographic location, volume of goods bought or distinct distribution of suppliers. Moving on to seller and trader nodes, an in-depth analysis of sales data can reveal qualitatively different distributions of customer groups. The precision in what constitutes a meaningfully different distribution of customers depends on the quality of data available. The result of this analysis is what informs the subcategorisation of nodes in the seller and trader categories. We now have subcategories of our seller, trader and buyer nodes.

Each seller and trader subcategory $\{V_1^{(s)}, \ldots, V_{n_s}^{(s)}, V_1^{(t)}, \ldots, V_{n_t}^{(t)}\}$ has an associated customer demographic, each denoted as $\{\mathcal{D}^{s_1}, \ldots, \mathcal{D}^{s_{n_s}}, \mathcal{D}^{t_1}, \ldots, \mathcal{D}^{t_{n_t}}\}$. These customer demographics can take many different forms, varying in complexity. We now describe a couple of examples. A simple example would be a node subcategory v, whose nodes only

sell to one specific trader subcategory $V_1^{(t)}$. Nodes ω in this subcategory v would then have a customer demographic of the form $\mathcal{D}^v(\omega) = \{p_{t_1}, \ldots, p_{b_{n_b}}\} = \{1, 0, \ldots, 0\}$. Customer demographics can also be stochastic, to allow variation amongst nodes in a subcategory. An example of a variable customer demographic would be if, for nodes in subcategory $v \in \{V_1^{(s)}, \ldots, V_{n_s}^{(s)}, V_1^{(t)}, \ldots, V_{n_t}^{(t)}\}$, more than 50% of their customers consisted of nodes from subcategory $V_1^{(b)}$ and an equal proportion to all other buyer subcategories. In this example, nodes in subcategory v do not sell to traders. Then, the customer demographic for nodes $\omega \in v$ would be $\mathcal{D}^v(\omega) = \{p_{t_1}, \ldots, p_{b_1}, p_{b_2} \ldots, p_{b_{n_b}}\} = \{0, \ldots, X, \frac{1-X}{n_b-1}, \frac{1-X}{n_b-1}, \ldots, \frac{1-X}{n_b-1}\}$, where X is a uniform random variable such that $X \sim U[0.5, 1]$. We conclude this section with an illustrative example.

Illustrated example: Customer demographics

Suppose that nodes in a trade network can be classified in such a way that subcategories consists of two sellers $\{V_1^{(s)}, V_2^{(s)}\}$, two traders $\{V_1^{(t)}, V_2^{(t)}\}$ and two buyers $\{V_1^{(b)}, V_2^{(b)}\}$. In this example we will focus on the customer demographic of seller subcategory $V_1^{(s)}$. In Figure 2.2 is an example where the first node of $V_1^{(s)}$, which we denote by $N_1^{s_1}$, has an out-degree of 5, for illustrative purposes. In this example, $N_1^{s_1} \in V_1^{(s)}$ has a customer demographic $\mathcal{D}^{s_1}(N_1^{s_1}) = (p_{t_1}, p_{t_2}, p_{b_1}, p_{b_2}) = (0.4, 0.2, 0.2, 0.2)$. This is shown in the histogram, where the x-axis represents each seller and trader node subcategory and the y-axis represents the proportion of out-going links $V_1^{(s)}$ nodes assign to other node subcategories. The customer demographic displayed above informs how these links should be distributed. This is then realised in the network by the grey node assigning 2 links to $V_1^{(t)}$ nodes (blue) and one link to nodes in $V_2^{(t)}$ (red), $V_1^{(b)}$ (green) and $V_2^{(b)}$ (yellow).
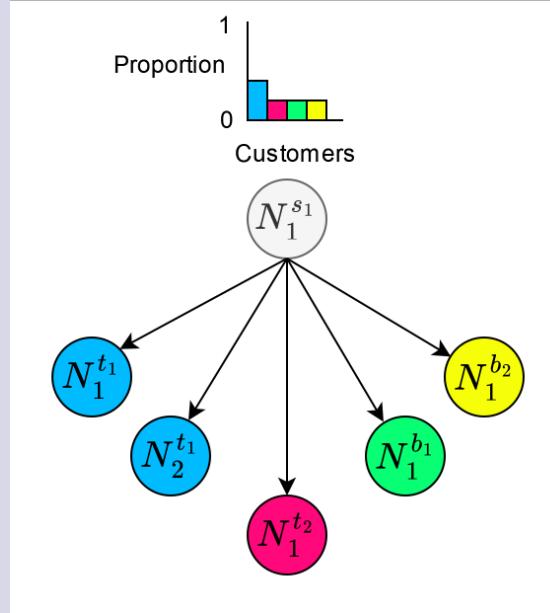
Figure 2.2. Depiction of how customer demographics correspond to the assignment of links in the network, in an example of a node with an out-degree of 5 and customer demographic $\mathcal{D}^{s_1}(N_1^{s_1})$, displayed in the histogram.

2.2.4 Out-degree modelling

The next step in our modelling process is to describe how the out-degrees of nodes should be modelled. Indeed, the customer demographic does not give us any insight in this regard, it simply tells us, once we know what the out-degree of a node should be, how to distribute the edges. Recall that the out-degree for a node is defined to be the total number of out-going links stemming from that node, or, in other words, the total number of unique customers a seller/trader has. We will again make use of subcategories and assume that nodes in the same seller/trader subcategory have out-degrees modelled from the same distribution. These distributions will be denoted by $(\mathcal{K}_{\text{out}}^{s_1}, \ldots, \mathcal{K}_{\text{out}}^{s_{n_s}}, \mathcal{K}_{\text{out}}^{t_1}, \ldots, \mathcal{K}_{\text{out}}^{t_{n_t}})$.

In practice, modelling the out-degree distribution for each seller and trader subcategory requires data on the number of customers that sellers and traders have in a given time-frame. For out-degree distributions, a discrete (or discretised) positive distribution is required, of which there are many options to choose from. Furthermore, to reject unrealistic values, this distribution will be truncated. Indeed, for a seller to be included in the network, it has to perform at least one sale and thus there is an imposed lower bound of an out-degree of 1. Out-degrees are also truncated from above for two reasons: first, to reject unrealistic values, and second, to reject assigning more edges than there can exist in the network.

Illustrated example: Out-degree distribution

Suppose that the out-degree distribution for the node subcategory $V_1^{(s)}$, $\mathcal{K}_{out}^{s_1}$ is described by a negative binomial distribution [148]. This means $\mathcal{K}_{out}^{s_1} \sim NB(\mu, \sigma^2)$, with probability mass function $P(\mathcal{K}_{out}^{s_1} = k) = \binom{k-1+\frac{\mu^2}{\sigma^2-\mu}}{k} \left(\frac{\sigma^2-\mu}{\sigma^2}\right)^k \left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2-\mu}}$. Thus, the negative binomial distribution $\mathcal{K}_{out}^{s_1}$, can be modelled with the parameters mean, μ, and variance σ^2 (Figure 2.3).

In Figure 2.3 we display a few instances of this distribution for varying values of μ and σ^2. The distribution will need to be truncated in order to ensure that generated values stay within an appropriate range. The lower truncation is set to 1 because this is the minimum number of customers a seller must have in order to be included in the network. The upper truncation, for this example has been chosen to be 400, and represents an upper bound on the maximum number of customers a node from $V_1^{(s)}$ can have. Vertical dashed lines are used in Figure 2.3 to represent this truncation which restricts the simulated values to within $[1, 400]$. For a given μ and σ^2, nodes in the subcategory $V_1^{(s)}$ will be assigned out-degrees, simulated from this truncated distribution.
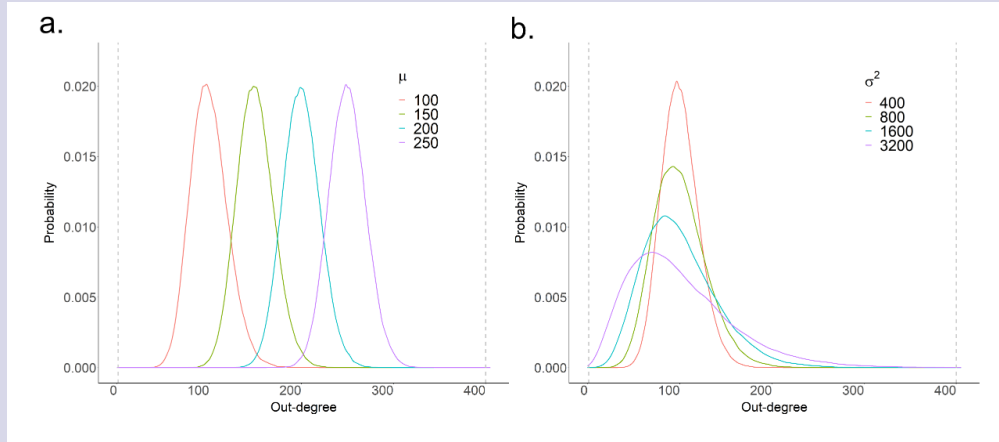
Figure 2.3. An example of a truncated negative binomial distribution and the effect that the two parameters the mean d_μ and variance σ^2 have on the shape of the curve. Changes in μ translate the position of the distribution peak ($\sigma^2 = 400$). Changes in σ^2 alter the shape of the distribution. The vertical dashed lines represent the points of truncation.

2.2.5 In-degree modelling

Now that we have successfully modelled the out-degrees, we move on to describing how in-degrees should be modelled. The in-degree describes how many nodes a given node buys from. It is worth pointing out that in-degree modelling will only be needed for traders and buyers, since sellers do not buy from anyone within the network by construction. In-degrees can be modelled in a similar way as out-degrees, by simulating from given distributions $(\mathcal{K}_{in}^{t_1}, \ldots, \mathcal{K}_{in}^{b_{n_b}})$. In-degree distributions must also generate discrete, positive values similarly to out-degree distributions. In fact, these distributions should generate a minimum value of 1, given that, by design, traders and buyers in the network must buy from someone. There will also be constraints to in-degrees regarding an upper truncation, since nodes cannot be assigned an in-degree larger than $|V^{(s)}| + |V^{(t)}|$. There may be further constraints to the maximum number of suppliers a node can have, which is idiosyncratic to the network analysed.

2.2.6 Edge weight modelling

Let us begin this subsection by recalling that, in our model, an edge between nodes represents a sale of goods, with the direction of the edge describing the movement of goods, out-going from the agent selling and in-going to the agent buying. We can further associate a weight to this edge to model metrics of interest such as the quantity of goods in the transaction or the monetary value of the transaction. We could also use some custom metric to indicate the strength of the trading relationship, such as number of transactions over a given time period. Throughout this chapter, weights will represent the quantity of goods in a given transaction. We now describe how such weights will be assigned to edges in our model.

Intuitively, we can imagine each seller/trader being assigned a *capacity* (a total amount of stock). Then, each out-going link leaving these nodes will be assigned a weight, with the constraint that the weights of all edges leaving a given node must add up to, or at minimum not exceed, the node's capacity. Approaching edge weights this way thus re-

quires modelling capacities, which in turn requires large quantities of data regarding sales volumes. However, obtaining access to such large volume of data may not always be possible, and thus other approaches need to be explored. Instead of focusing on the capacity of sellers and traders, we will focus on distributions modelling the consignment sizes of nodes in the buyer and trader categories. These distributions will also be informed by data; however, the cost of acquiring sufficient data is lower with this approach because many sales from traders and buyers can be collected from a smaller number of sellers. To be suitable for consignment sizes, a distribution must be positive, discrete/discretised and truncated to reject extreme values. The smallest lower truncation must be 1, as this is the smallest quantity of goods that can be bought. Consignment size distributions for traders and buyers are denoted by $\{\mathcal{C}^{t_1}, \ldots, \mathcal{C}^{b_{n_b}}\}$. Note that we assume that customer consignment size is independent of which seller or trader the customer buys their goods from. If there is enough data, this assumption can be avoided by assigning each node subcategory multiple consignment size distributions associated with trade from each of their suppliers. This chapter will, however, continue with the assumption of independence. We conclude this section with an illustrated example. In the next section, we will describe how all of the ideas and concepts introduced so far piece together in our trade network modelling algorithm.

Illustrated example: Consignment size distribution

Suppose that the quantity of goods that nodes from subcategory $V_1^{(b)}$ buy is described by a truncated log-normal distribution $\mathcal{C}^{b_1}$ [149]. This means that the natural log of $\mathcal{C}^{b_1}$ follows a normal distribution, $\log(\mathcal{C}^{b_1}) \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$, where $\bar{\mu}$, $\bar{\sigma}$ represent the mean and standard deviation for $\log(\mathcal{C}^{b_1})$. We parameterise this distribution by the untransformed mean (μ) and standard deviation (σ), to make interpretation of values simpler. Thus, we calculate the log-normal by $\log\left(\mu/\sqrt{1 + \sigma^2/\mu^2}\right)$ and the variance as $\log\left(1 + \sigma^2/\mu^2\right)$. The distribution in this example is truncated such that the exponential of the distribution takes values between 1 and 100. Displayed in Figure 2.4 is the exponential of the log-normal distribution. The x-axis represents the consignment size and the y-axis represents the probability for an in-going edge to a node in subcategory $V_1^{(b)}$ to be assigned that consignment size. In Figure 2.4 we display a few instances of this truncated distribution for varying parameters μ and σ. For a given μ and σ, nodes from subcategory $V_1^{(b)}$ have their in-going edges weighted by a randomly generated value from this truncated distribution. We discretise these values to represent a quantity of trade.
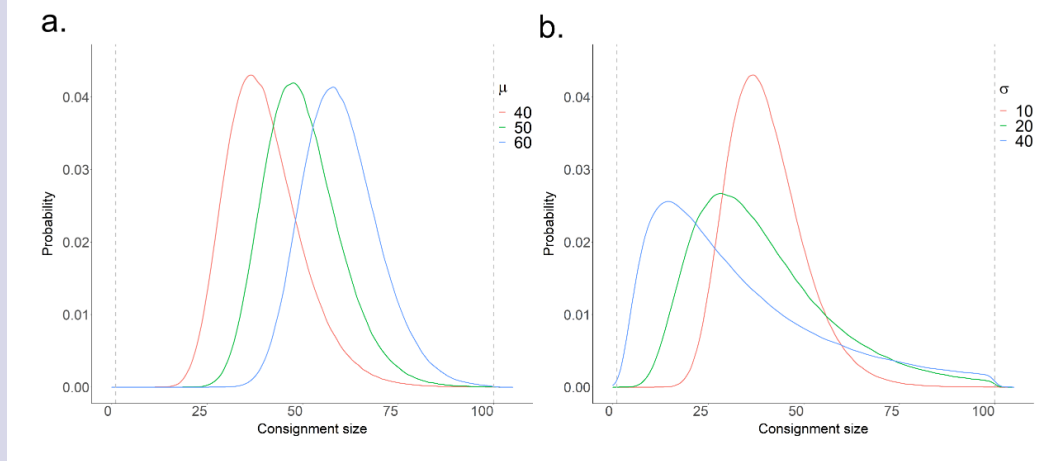
Figure 2.4. An example of the exponential of a truncated log-normal distribution, with the parameters μ and σ. The vertical dashed lines represent the points of truncation. Changes in μ translate the distribution. Changes in σ alter the shape of the distribution.

2.3 Network construction algorithm

We can now model all aspects of the network. Nodes have been defined to be members of subcategories, $\{V_1^{(s)}, \ldots, V_{n_s}^{(s)}, V_1^{(t)}, \ldots V_{n_t}^{(t)}, V_1^{(b)}, \ldots, V_{n_b}^{(b)}\}$, in a trade hierarchy of sellers, traders and buyers; see Figure 2.1. The parameters of the network have been described with customer demographics $\{\mathcal{D}^{s_1}, \ldots, \mathcal{D}^{s_{n_s}}, \mathcal{D}^{t_1}, \ldots, \mathcal{D}^{t_{n_t}}\}$, out-degree distributions $\{\mathcal{K}_{\text{out}}^{s_1}, \ldots, \mathcal{K}_{\text{out}}^{s_{n_s}}, \mathcal{K}_{\text{out}}^{t_1}, \ldots, \mathcal{K}_{\text{out}}^{t_{n_t}}\}$, in-degree distributions $\{\mathcal{K}_{\text{in}}^{t_1}, \ldots, \mathcal{K}_{\text{in}}^{t_{n_t}}, \mathcal{K}_{\text{in}}^{b_1}, \ldots, \mathcal{K}_{\text{in}}^{b_{n_b}}\}$ and consignment-size distributions $\{\mathcal{C}^{t_1}, \ldots, \mathcal{C}^{b_{n_b}}\}$. In this section, we describe the process of constructing a network from the aforementioned parameters.

Firstly, the number of nodes in each seller subcategory are required as an initial condition. Hence, to initialise the network construction, we have the parameters $\{|V^{s_1}|, \ldots, |V^{s_{n_s}}|\}$. This given number of sellers and their associated customer demographics and out-degrees, will determine all other nodes and edges, thus generating the rest of the network. For the sake of simplicity, in the following we will assume there is no cross trading between traders. If we drop this assumption, the number of traders $\{|V^{t_1}|, \ldots, |V^{t_{n_t}}|\}$ must also be specified as an initial condition to the network construction. We remark that the number of trader and buyer nodes can also be specified, thus constraining the network further to achieve a desired result. We now move on to how the trader nodes are generated.

Nodes from the seller subcategories $\{N_1^{(s_1)}, \ldots, N_{|s_1|}^{(s_1)}, \ldots, N_1^{(s_{n_s})}, \ldots, N_{|s_{n_s}|}^{(s_{n_s})})\}$ will be assigned an associated out-degree and a customer demographic generated from $\{\mathcal{K}_{\text{out}}^{s_1}, \ldots, \mathcal{K}_{\text{out}}^{s_{n_s}}\}$ and $\{\mathcal{D}^{s_1}, \ldots, \mathcal{D}^{s_{n_s}}\}$, respectively. The customer demographic of a seller node gives the proportion of all out-going edges that must be assigned to traders from that node. Multiplying each of these proportions by the seller node's out-degree will give the total number of edges that must be assigned to traders from sellers in the network. We now begin to add trader nodes to the network, with an iterative process that occurs for every trader node subcategory. Starting with subcategory $V_1^{(t)}$, suppose that there is an x number of edges assigned from seller nodes to nodes in the subcategory $V_1^{(t)}$, where $x > 0$. We generate the first node of $V_1^{(t)}$, $N_1^{t_1}$, and assign this node an in-degree, simulated from the in-degree distribution $\mathcal{K}_{\text{in}}^{t_1}$, where this distribution is truncated above by the minimum between x

and all other upper truncations on this distribution. We then generate another node from $V_1^{(t)}$, $N_2^{t_1}$, with an in-degree simulated from $\mathcal{K}_{\text{in}}^{t_1}$ where the upper truncation is now the minimum between x minus the in-degree for $N_1^{t_1}$ and all other upper truncations for this distribution. We iteratively add nodes from subcategory $V_1^{(t)}$, where for each additional node, the distribution $\mathcal{K}_{\text{in}}^{t_1}$ is truncated from above by the minimum between x minus the sum of the in-degrees of existing $V_1^{(t)}$ nodes, and all other upper truncations. This process continues until the total number of edges from sellers to $V_1^{(t)}$ nodes is equal to the sum of the in-degrees from $V_1^{(t)}$ nodes. We apply this process for all trader node subcategories. Once these processes have been conducted, we will know the total number of trader nodes in the network.

Remark: Suppose that there exists trade among trader nodes in the network. The customer demographics combined with the out-degrees impose an upper bound on each trader's out-degree, i.e., a trader cannot sell to more nodes in a particular group than the total amount of nodes which are in that group. Thus an upper bound for each trader node's out-degree will be the total number of traders divided by the individual trader's proportion of links assigned to other traders, minus one, since a node cannot sell to itself. If the total number of traders or buyers is also specified, this too will contribute to the imposed upper bound in the same way. Out-degrees can be controlled further by adding more upper bounds based on literature or data.

So far we have described how to model the number of seller and trader nodes in the network, each with associated in-degree, out-degree and customer demographic. The number of edges to be assigned to trader nodes has also been calculated. The next step is to describe how to calculate the number of buyer nodes and the number of edges to be assigned to buyers from sellers and traders.

We proceed in a similar way as we did with generating the trader nodes. From the seller and trader nodes, we calculate the number of edges to be assigned to buyers by multiplying each seller and trader nodes out-degree by the proportion of edges to be assigned to buyers (from the respective customer demographics). We calculate the total number of

buyer nodes from the given buyer nodes' in-degree distribution in the same iterative way which we used to calculate the number of traders. Thus the total number of nodes in the network is determined.

Now that all nodes have been created, the next step is to assign the edges in the network. Edges must be assigned in the network to be consistent with each node's customer demographic, in-degree and out-degree. The novel method of assigning edges to achieve this consistency is described below and illustrated in Figure 2.5.

Firstly, for each node subcategory, we compile a paired lists of nodes, entitled the "To" and "From" list. A "To" list, consists of nodes in a node subcategory, repeated for every assigned in-going link the node has. The associated "From" list consists of nodes, repeated for every out-going link they have assigned to that node subcategory. The "From" and "To" lists together represent where links start and end regarding sales to a particular node subcategory. We construct these pairs of lists for each node subcategory.

For each of these pairs of lists, we conduct an iterative process to connect nodes via edges. Recall that connections in a network can be represented in an edge-list, a list of all pairs of nodes that are connected by an edge. We construct the edge list for the network by going through each "From" and "To" list pair, assigning edges with an iterative process. For illustrative purposes, we will present an edge-list in a table with two columns where the first column represents the start of the edge and the second column represents where the edge ends. Additional columns can be added to represent attributes associated with a given edge. We add a third column to represent the size of the consignment (edge weight). We describe the iterative process below and in Figure 2.5.

1. Select one of the available "From" and "To" list pairs (top left box of Figure 2.5);

2. Select the first element in the "From" list and choose at random an element from the "To" list. This selection represents the assignment of an edge between these nodes (top right box of Figure 2.5). If the first element is repeated in the "From" list, the

selection from the "To" list must remove previous selections as candidates in the random selection process in the "To" list. If an element is in both the "From" and "To" lists, then it must also be removed as a candidate.

3. These entries are then removed from the "From" and "To" lists and added to the edge-list (bottom left box of Figure 2.5).

4. Repeat steps 2 and 3 until the currently chosen pair of "From" and "To" lists are empty (bottom right box of Figure 2.5).

5. Repeat step 1 until all "From" and "To" lists are empty, resulting in an edge-list with all links assigned corresponding to customer demographics, in-degrees and out-degrees.
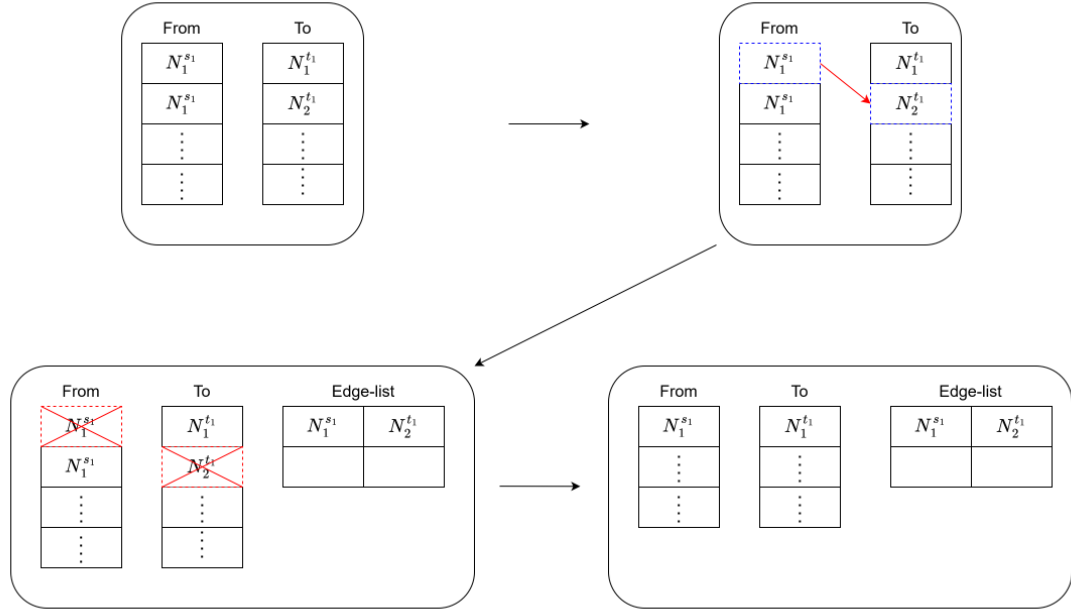
Figure 2.5. Schematic representation of the edge assignment process.

In the final step to the network construction, we assign weights to the edges, representing the volume of trade for each consignment. We add a third column to the edge-list, and fill the entries in this column by generating edge weights from the associated consignment size distributions of the nodes in the second column of the edge-list.

Thus we have constructed a model network which describes: who is in the network, how big the network is, how edges are distributed and the heterogeneous distribution of consignments within the network. This concludes the network construction process. From the steps outlined in this section, this method results in a model with 10 parameter sets, summarised in Table 2.1.

Table 2.1. A summary of the parameter sets in the model.

Parameter	Description
$\{n_s, n_t, n_b\}$	The number of seller, trader and buyer subcategories
$\{|V^{s_1}|, \ldots, |V^{s_{n_s}}|\}$	Distribution of Seller node subcategories*
$\{\mathcal{K}_{\text{out}}^{s_1}, \ldots, \mathcal{K}_{\text{out}}^{s_{n_s}}\}$	Out-degree distributions for seller subcategories
$\{\mathcal{K}_{\text{out}}^{t_1}, \ldots, \mathcal{K}_{\text{out}}^{t_{n_t}}\}$	Out-degree distributions for trader subcategories
$\{\mathcal{K}_{\text{in}}^{t_1}, \ldots, \mathcal{K}_{\text{in}}^{t_{n_t}}\}$	In-degree distributions for trader subcategories
$\{\mathcal{K}_{\text{in}}^{b_1}, \ldots, \mathcal{K}_{\text{in}}^{b_{n_b}}\}$	In-degree distributions for buyer subcategories
$\{\mathcal{C}^{t_1}, \ldots, \mathcal{C}^{t_{n_t}}\}$	Consignment size distributions for trader subcategories
$\{\mathcal{C}^{b_1}, \ldots, \mathcal{C}^{b_{n_b}}\}$	Consignment size distributions for buyer subcategories
$\{\mathcal{D}^{s_1}, \ldots, \mathcal{D}^{s_{n_s}}\}$	Customer demographics for seller subcategories
$\{\mathcal{D}^{t_1}, \ldots, \mathcal{D}^{t_{n_t}}\}$	Customer demographics for trader subcategories

* It is optional to also include the distribution of trader and/or buyer node classes to achieve a desired result.

2.3.1 Illustrated example: Model summary

Concluding the running example, we summarise what information/parameters are associated with each node subcategory (Figure 2.6). Each node subcategory is displayed with its vertical position reflecting its place in the trading hierarchy, the nodes which only sell at the top, those which only buy at the bottom and nodes which both buy and sell in between.

Node groups which sell are defined by their qualitatively distinct customer demographic, illustrated within the histograms (**a**). A customer demographic is defined by the discrete distribution of the proportions of links a seller or trader assigns to other node subcategories. Figure 2.6 shows that the node subcategory $V_1^{(s)}$ (grey) is defined by selling most of its goods to the trader subcategory $V_1^{(t)}$ (blue), whereas $V_2^{(s)}$ (orange) is defined by selling most of its goods equally amongst the node subcategories $V_2^{(t)}, V_1^{(b)}, V_2^{(b)}$ (red, green and blue).

Seller and trader node subcategories $(V_1^{(s)}, V_2^{(s)}, V_1^{(t)}, V_2^{(t)})$ have an associated out-degree distribution, as sketched in subgraphs (**b**) with the y-axis representing probability and the x-axis representing quantity. The only constraints on these distributions is that they are confined to positive and discrete values. In the example of Figure 2.6, $V_1^{(s)}$ follows a negative binomial distribution, $V_2^{(s)}$ has a bi-modal distribution of a small and large number of customers. The out-degree distribution of $V_1^{(t)}$ follows a uniform distribution and the out-degree of $V_2^{(t)}$ follows a flat distribution with a medium amount of customers.

The nodes from trader and buyer subcategories have an associated distribution from which their consignment sizes are drawn from. Again, these distributions can take any shape, examples of which are sketched in subgraphs (**c**) within Figure 2.6. The x-axis and y-axis represent quantity and probability respectively. In this example, we see nodes from $V_1^{(t)}$ order primarily "medium sized" consignments, $V_2^{(t)}$ order many small and few larger consignments. The consignments for $V_1^{(b)}$ follow a log-normal distribution and $V_2^{(b)}$ consignments are drawn from a bi-modal distribution.

Trader and buyer nodes have associated in-degree distributions, sketched in subgraphs (**d**) which represent the distribution that describes how many suppliers a node in that subcategory has. For example, the in-degree for $V_1^{(t)}$ is characterised by a power-law distribution and $V_2^{(t)}$ by a uni-modal distribution with high variance. The in-degrees for $V_1^{(b)}$ are governed by a bi-modal distribution of a small and large number of suppliers, whereas $V_2^{(b)}$ node in-degrees are heavily skewed to the right indicating a primarily smaller number of suppliers.
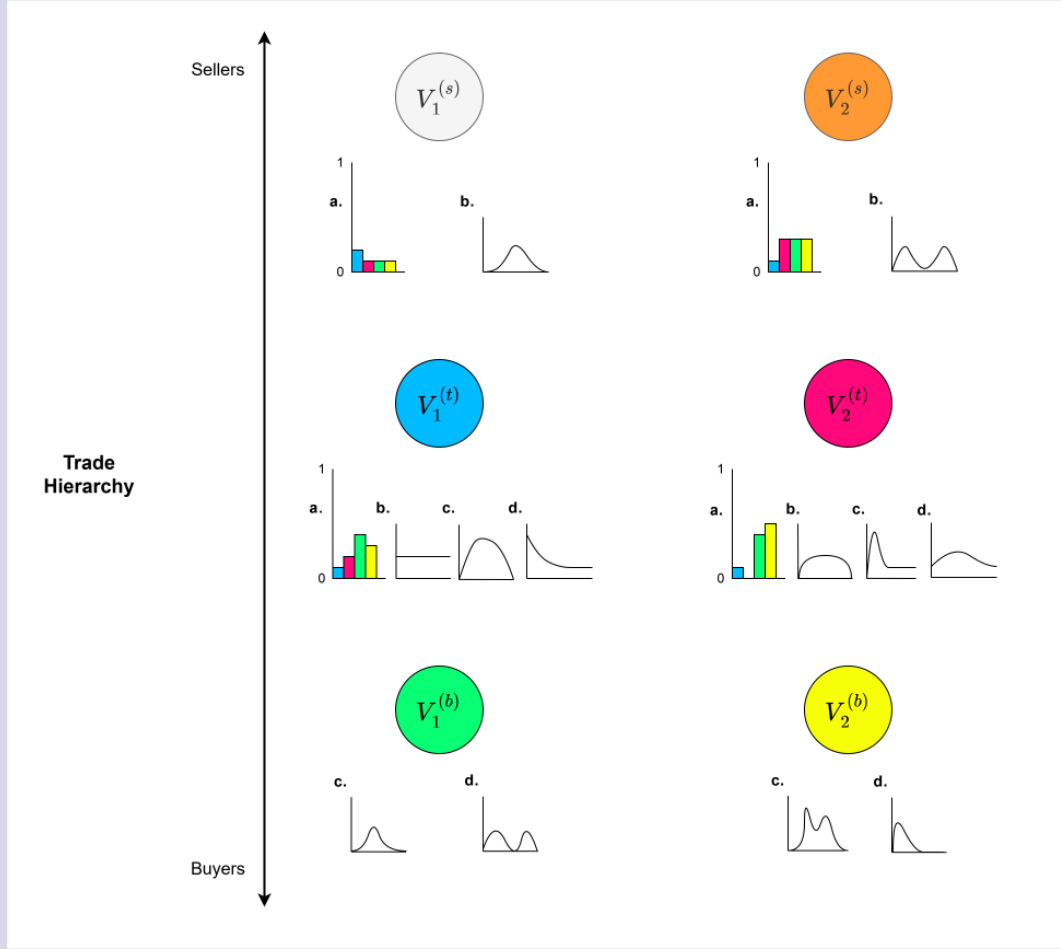
Figure 2.6. Summary of the model and what parameters are associated with each node subcategory $(V_1^{(s)}, V_2^{(s)}, V_1^{(t)}, V_2^{(t)}, V_1^{(b)}, V_2^{(b)})$. The trade hierarchy represents a spectrum with sellers at the top and buyers at the bottom. Each node category is split into two subcategories, with their respective sub-figures displayed below the given node. (**a**) The node subcategory's customer demographic. (**b**) The distribution to simulate out-degrees. (**c**) The consignment size distribution. (**d**) The in-degree distribution.

2.4 Discussion

In this chapter, we have developed and generalised a phenomenological method of network construction, which allows for the construction of both weighted and unweighted, hierarchical trade networks with parameterised node categories. Importantly, this method poses a framework for which researchers can follow to construct model trade networks and inform data requirements. This method is constructed from the basic premise that nodes in a trade network can be split into three groups (sellers, traders and buyers), similar to *Pautasso et al.* [70]. However, we partition these groups into n_s, n_t and n_b subcategories allowing for a more detailed labelling of nodes. Adding labels to nodes in networks helps in applying results from network analysis to the real-world systems they are modelling by making it easier to identify nodes of interest. In this section we discuss some of the novelties and limitations to the network construction method we have developed.

The customer demographics and out-degree distributions in our model logically correspond to the way a business can be characterised by their clientele. The in-degree and consignment size distributions are also an appropriate way to describe the demand that customer groups have for the goods they buy. In a network, the number of nodes, out-going links and in-going links are all inter-dependent. For instance, in a directed network of n nodes, there can be a maximum $n(n-1)$ number of edges (assuming no self-edges). In a similar way, the initial number of seller/trader nodes, in combination with the customer demographics, impose constraints on node in- and out-degrees, which is a significant novelty to this method.

With the characterisation of nodes into subcategories, edge weights can be modelled by consignment size distributions. Attributing subcategories to nodes allows us to use relevant information that would not usually be included in network models such as the differing demands of customer groups. The decision of modelling consignment sizes from the perspective of the customer as opposed to the supplier allows for a lesser requirement of sales data. A detailed data set from a few sellers/traders could give enough sales data to model consignment size distributions for each trader and seller subcategory. This is be signifi-

cantly easier to achieve than collecting detailed sales data from many sellers/traders to model distributions for how their consignment sizes vary.

There is a limitation in how to define the seller and trader node subcategories based on customer demographics. For example, in a market where all sellers/traders have a distinct customer distribution, this method will not perform well due to the data requirements necessary to model this accurately. Furthermore, with each added node subcategory in the model, there are an additional 2 to 5 model parameters required. With insufficient data, there will be too many parameters without estimates, making model analysis exceedingly complex. These are important factors that a researcher would need to consider when implementing this step of node subcategorisation. This method works best when there are a small number of node subcategories.

The network construction presented in this chapter does not explicitly model how buyer and trader nodes (with an in-degree larger than 1) choose who they buy from, due to the added complexity of balancing the out-degrees, customer demographics and in-degrees alongside an independently defined *supplier demographic*. Similarly to customer demographics, supplier demographics would be defined as discrete distributions of seller and trader node subcategories that describe how a node from a subcategory's in-going links are distributed. Adding this feature to the model would be a clear improvement.

An assessment of this method of network construction's performance for a known real world trade network is required. This method does provide a framework for constructing a trade network (with or without weights) and helps make clear the data requirements. However this method does not supersede the construction of a model network informed by a significant quantity of data.

Chapter 3

Application of Network Construction to Domestic Plant Trade

In this chapter, we use the method outlined in Chapter 2, in combination with the data we received from FERA to construct a model trade network to represent annual domestic plant trade in the UK. From our data analysis, we find that consignment sizes follow log-normal distributions, different for each node group, and that nurseries can be characterised by four different patterns of trading behaviour. We conduct a sensitivity analysis of the model parameters, using several network centrality measures to assess the importance of nodes. From our analysis, we find that customers do not significantly contribute to network structure and that edge weights change the relative significance of nodes. We find that the network is robust to small changes in market structure, with significant changes only seen in the extremes.

3.1   Introduction

As we mention in Chapter 1, the domestic trade of plants contributes to the spread of pests and diseases, making biosecurity efforts difficult. Very little is known about the structure of domestic trade; thus, it is our goal to construct a model trade network for the domestic

trade of plants in the UK. We generalise and present our network construction approach in Chapter 2, and in this chapter we use this method alongside our sales data to construct our model plant trade network. We analyse this model network using several network centrality measures, which are detailed in Section 3.3.4. These centrality measures are used to understand which node subcategories are most central in the network, in an epidemiological context. Our analysis comes in three parts, varying the parameters affecting node subcategory consignment sizes, out-degrees, and the initial distribution of nurseries. We assess how each node subcategory's centrality scores change as we vary the parameters. In the next section we describe the data and the analysis we conduct to inform our model.

3.2   Data sets and forms

We have been provided by FERA two sets of UK plant nursery sales data. Both of these data sets were collected by DEFRA from the nurseries directly. The data was then cleaned and organised by FERA. We were then securely transferred the data for the purpose of this research. We keep the identity of the plant nurseries that gave the data anonymous. Each of these data sets are described in this section.

The first, smaller set of sales data consists of plant sales from one UK plant nursery over the course of 2013 to 2015 and purchases from 2013 to 2016. Each row in the data represents a sale or purchase, and the columns include the delivery address (or supplier code), delivery date, plant species and quantity. For the data on purchases, the country the plants were bought from is also specified. This data consists of 40 different genera of plants, $23,772$ sales and $11,523$ purchases. Note that the delivery addresses in this data are not anonymised, which is relevant for the next data set.

The second data set is larger, consisting of sales data from four different plant nurseries, which we label $\{N_1, N_2, N_3, N_4\}$. The data for nurseries N_1 and N_3 include purchasing data and all nurseries supplied data of sales. The data from the four nurseries span different time frames: N_1 sales span all of the year 2017, N_2 sales cover the end of 2016 to the

middle of 2018, N_3 sales start from the middle of 2013 to the middle of 2016 and N_4 sales span the end of 2016 to the third quarter of 2018. This data includes $186,846$ transactions and over 900 different genera of plants. In the larger data set, the rows indicate individual purchases/sales. The columns in this data set include: "Data source", "Genus", "Date", "Receiver location", "Supplier location", "Variety", "Quantity", "Supplier Postcode", "Receiver Postcode". We note that this larger data set includes the smaller data set (Nursery N_3) but is stored in a different format. In the larger data set, the receiver locations for nursery N_3 are anonymised, however the receiver locations are included in the smaller data set. Hence, we merge the two data sets by matching the sales from the smaller set into the N_3 sales in the larger set. We achieved this by matching the dates, quantities, receiver post-codes and genus across data sets. In the instance where a sale was included in the smaller data set but not present in the larger data set, the sale would be added to the merged data set. In our analysis, we only consider the nursery sales and not the purchases. Therefore, when referring to the sales data in this chapter, we will be referring to only the sales in the merged data set. We show an anonymised example of the data structure in Figure 3.1. For clarity, we highlight the columns we used in green.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DataSource	Date	Genus	PlantCode	PotSize	Receiver_location	Supplier_location	Variety	price	quantity	value	Supplier_PostCode	Receiver_PostCode	species
2	N_1_StockIn	DD/MM/YY	XXXX	XXXX	XXXX	N_1	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
3	N_1_StockOut	DD/MM/YY	XXXX	XXXX	XXXX	Customer_1	N_1	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
4	N_2_StockOut	DD/MM/YY	XXXX	XXXX	XXXX	Customer_2	N_2	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
5	N_3_StockIn	DD/MM/YY	XXXX	XXXX	XXXX	N_3	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
6	N_3_StockOut	DD/MM/YY	XXXX	XXXX	XXXX	Customer_3	N_3	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
7	N_4_StockOut	DD/MM/YY	XXXX	XXXX	XXXX	Customer_4	N_4	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX

Figure 3.1. Screenshot of an anonymised example of the sales data from four plant nurseries supplied by FERA. The columns highlighted in green are those we used in our analysis.

The total number of nurseries in the UK is unknown, however there is an estimate in the existing literature. The Ornamental Horticulture Roundtable Group published a report analysing the 2019 Horticulture sector skills survey, which includes estimates of the total number of businesses in various sub-sectors of the Ornamental Horticultural sector [150]. In their estimations, there are 630 businesses in the UK in the Ornamental sub-sector and 6525 businesses in the "Retail" sub-sector. We use these estimates for the total number of nurseries and retailers in the UK. We can then use this to scale the consignment size and out-degree distribution parameters. We use this data to assume that for a given number of

nurseries in the network, there are 10 times more retailers. Data regarding plant nursery annual sales and number of customers would allow for constructing a network where the sizes and out-degrees of nurseries are modelled from the supply perspective.

This concludes the section describing the data we used to inform our model. In the next section we will describe our method of network construction, with specific reference to how we used our data to inform model parameters.

3.3   Methods

In this section we present our method of constructing a model network to represent domestic plant trade in the UK. Our focus in this section is to highlight how we have used our data with the methodology described in Chapter 2. We proceed in a similar fashion by first defining the nodes, then the edges, out-degrees, in-degrees, edge weights and a summary of the algorithm.

3.3.1   Node classification

There is a low resolution understanding as to who the buyers and sellers are in the network. Prior to any data analysis, we assume that plant trade consists of plant nurseries that trade amongst each other, retailers (such as garden centres) and then customers. We use the trade data to make a more detailed and informed characterisation of the nodes in the network.

Plants are used in a variety of ways, from industry use such as construction, reducing noise pollution and farming, to decorative use by shops and the general public. Hence, there is a logical case to make that different plant species will have different trade networks, thus making the plant species an important factor to consider in the analysis. Given that the focus of our research is on the spread of plant pests and diseases, we chose to subset the data and focus on a select few hosts. We chose the host relevant to OPM (Oak) because

it is a simple one host pest and is an example of a recently introduced pest to the UK via trade [35]. OPM is far from the only threat to Oak trees, there are 90 total pests that are a threat to Oak trees, identified from the UK plant health risk register [1]. Thus, we focus our analysis on the subset of the sales data that includes *Quercus* in the genus of the plant, consisting of $2,643$ sales. We repeat the data analysis for the four hosts of main concern to the UK of *Xylella fastidiosa*: Lavender, Olive, Prunus and Rosemary [151]. In the data, these five hosts altogether consist of approximately $18,000$ sales.

Our first step is to construct the node subcategories from the existing customers in the sales data. We identify customers through a manual process which involved using Google search and Google maps with the customer descriptions and postcodes. Customer descriptions for a sale, when given, take the form of a name or a business name (usually identified by the company name ending with "ltd."). In the cases where we cannot identify a customer clearly, we assign the sale into an "Unknown" category. We first assign each customer in the data to a preliminary subcategory (e.g., Farm, Contractor, Landscaper, member of the public, Garden centre, Nursery, etc.). Regarding the spread of infectious diseases on this network, our dominant concern is the further transportation of the plant. Therefore, we distill these into simpler subcategories: *Commercial*, *Consumer*, *Nursery* and *Retailer*. We define these subcategories to be simple and broad, yet identify how the plant is used once bought.

Thus our node subcategories so far in the network consist of two buyers (Commercial and Consumer), and two traders (Nursery and Retailer). We do not consider a seller subcategory in this network. We define nodes in the *Commercial* subcategory to be businesses where the plants they bought are planted elsewhere, such as landscapers and contractors. Nodes in the *Consumer* subcategory represent customers where the plants they bought are planted on their premises, such as parks, farms and private gardens. Nodes in the *Retailer* subcategory are defined to be businesses other than plant nurseries that sell plants. This subcategory primarily consists of garden centres and hardware stores. Nodes in the *Nursery* subcategory simply represent plant nurseries in the UK. However, from the sales

data, we investigate the Nursery subcategory further for qualitatively distinct customer demographics, as discussed in the previous chapter.

We find that for each plant species (Oak, Lavender, Olive, Prunus and Rosemary), each nursery (N_1, N_2, N_3, N_4) sells the majority $(> 50\%)$ of their plants to one customer subcategory, an example is shown in Figure 3.2. In this figure, the proportion of sales of Oak plants to customer groups are shown for each nursery (N_1, N_2, N_3, N_4). We measure both in terms of consignment (subgraph **A**) and by volume of plants (subgraph **B**). In most cases, the predominant customer group still holds across these different measures. This pattern breaks between these measures in the sales of Prunus for nursery N_3 (not shown). The differences between these measures suggest the existence of differences between the consignment size distributions of customer groups.
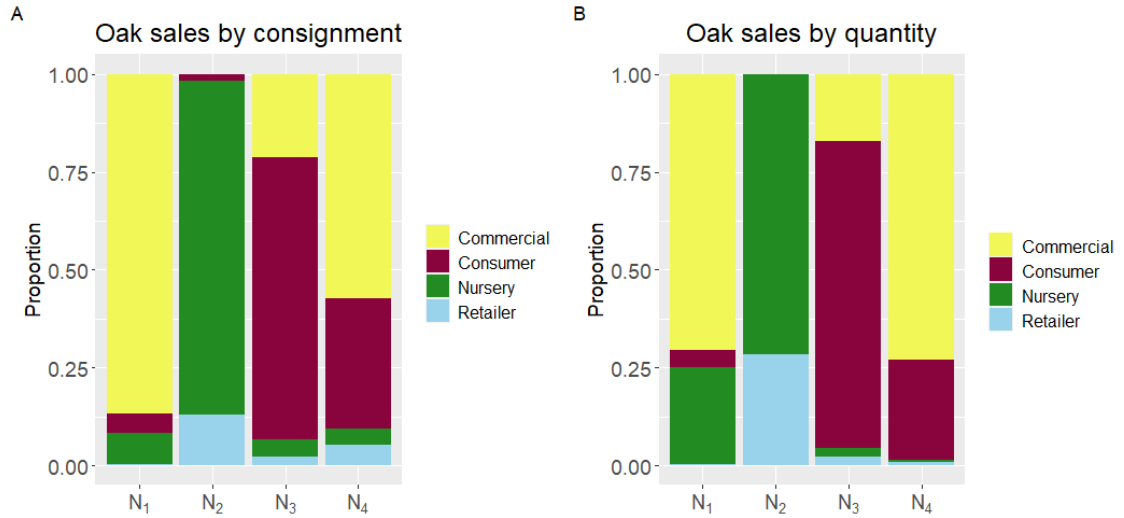
Figure 3.2. Customer demographics displayed for Oak sales of the four nurseries from our data (N_1, N_2, N_3, N_4). **A**. Measuring sales by consignment (number of orders).**B**. Measuring sales by quantity (total number of plants sold).

Since the sample size of the data is relatively small, we assume that nurseries dominated by the custom of retailers exist, despite not being seen in the data. Hence we define the node subcategories, with traders $\{N_{\mathrm{Com}}, N_{\mathrm{Cons}}, N_{\mathrm{Nur}}, N_{\mathrm{Ret}}, \mathrm{Ret}\}$ and buyers $\{\mathrm{Com}, \mathrm{Cons}\}$. The node subcategories are defined in Table 3.1.

Table 3.1. Table of node subcategories alongside their description and role in the trade hierarchy as described in Chapter 2.

Node subcategory	Description	Role
Com	Commercial customers in a profession where the plants they buy will be planted elsewhere as part of their business, e.g., Landscapers, Forestry, Contractors.	Buyer
Cons	Non-commercial customers, where the plants they buy, they will plant on their premises, e.g., A park, farm, or private garden.	Buyer
N_{Com}	Plant nurseries in the UK which sell mostly ($> 50\%$) to commercial customers	Trader
N_{Cons}	Plant nurseries in the UK which sell mostly ($> 50\%$) to consumers	Trader
N_{Nur}	Plant nurseries in the UK which sell mostly ($> 50\%$) to nurseries	Trader
N_{Ret}	Plant nurseries in the UK which sell mostly ($> 50\%$) to retailers	Trader
Ret	Retailers other than nurseries who will then sell the plants, e.g., Garden Centres.	Trader

With this characterisation of nodes in the network, trading behaviour is heterogeneous, with nurseries specialising their sales towards one specific customer subcategory, shown in Figure 3.3 by the solid black edge (a major customer). There are also dotted edges to other node subcategories (minor customers).
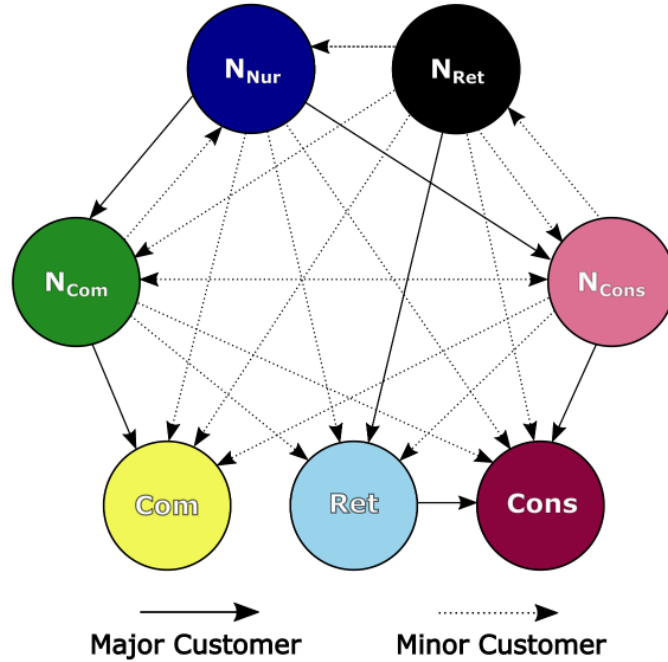
Figure 3.3. Network schematic summarising node subcategories and network structure. *Major customer* - designate more than 50% of outgoing edges to those node subcatgories. *Minor customer* - designate all other out-going links to those node subcategories.

Without access to retailer sales data, we assume that retailers sell only to the consumer group. The rationale behind this assumption is that retailers mostly consist of garden centres, which tend to sell to the general public. We would need access to sales data from garden centres to assess how appropriate this assumption is.

This concludes the subsection on defining node subcategories. In the next subsection, we move on to modelling the edges in the network.

3.3.2 Edge classification

As in Chapter 2, an edge between nodes in this network represents the sale of plants, with the direction of the edge indicating the movement of the plants. In this section, the distributions for modelled out-degrees, in-degrees and consignment sizes are presented.

Out-degree modelling

As discussed in the previous chapter, an out-degree distribution is required to be positive and the values generated from the distribution must be integers. Hence, we choose the negative binomial distribution to model the out-degrees for all trader node subcategories [148]. As in the illustrated example in Chapter 2, we parameterise the negative binomial distributions by the mean and variance, which we denote d_μ, d_{σ^2}. We truncate these distributions to a minimum value of 1, and a maximum value of 3,000. In the data, the maximum number of customers a nursery has in a given year is approximately 1,500, so we choose an arbitrary value of 3,000 for the upper truncation.

For simplicity and due to lack of data, we assume that the out-degrees for all trader nodes $(N_{\text{Com}}, N_{\text{Cons}}, N_{\text{Nur}}, N_{\text{Ret}}, \text{Ret})$ to be from the same distribution. Hence, the out-degree distribution for trader nodes and the associated parameters is denoted by $\mathcal{K}_{\text{out}}(d_\mu, d_{\sigma^2})$.

In-degree modelling

We model the in-degrees in this network from several assumptions. Firstly, nodes in the commercial and consumer node subcategories have an in-degree of 1. Secondly, trader nodes have no preference when assigning links to nodes other than the preference prescribed by their customer demographic. And finally, there can be no repeated edges.

Edge-weight modelling

For each species in the data, we plot the distribution of consignment sizes by customer group: Commercial, Consumer, Nursery and Retailer. Figure 3.4 shows the distributions for Oak sales on a $\log_{10}$ scale. We can see that the distributions for all customer groups are skewed and heavy-tailed. The consignment sizes are highly varied, spanning 5 orders of magnitude. The commercial and consumer consignment distributions are similarly shaped, with mostly smaller shipments (a peak at 0) and few large consignments. The nursery and retailer consignment distributions consist of much larger average consignment sizes (a peak at 3) with fewer small consignments. The distribution of consignment sizes per group is consistent with what we would expect from each group; the commercial and consumer groups (buyers) buying smaller consignments than the nursery and retailer groups (traders).
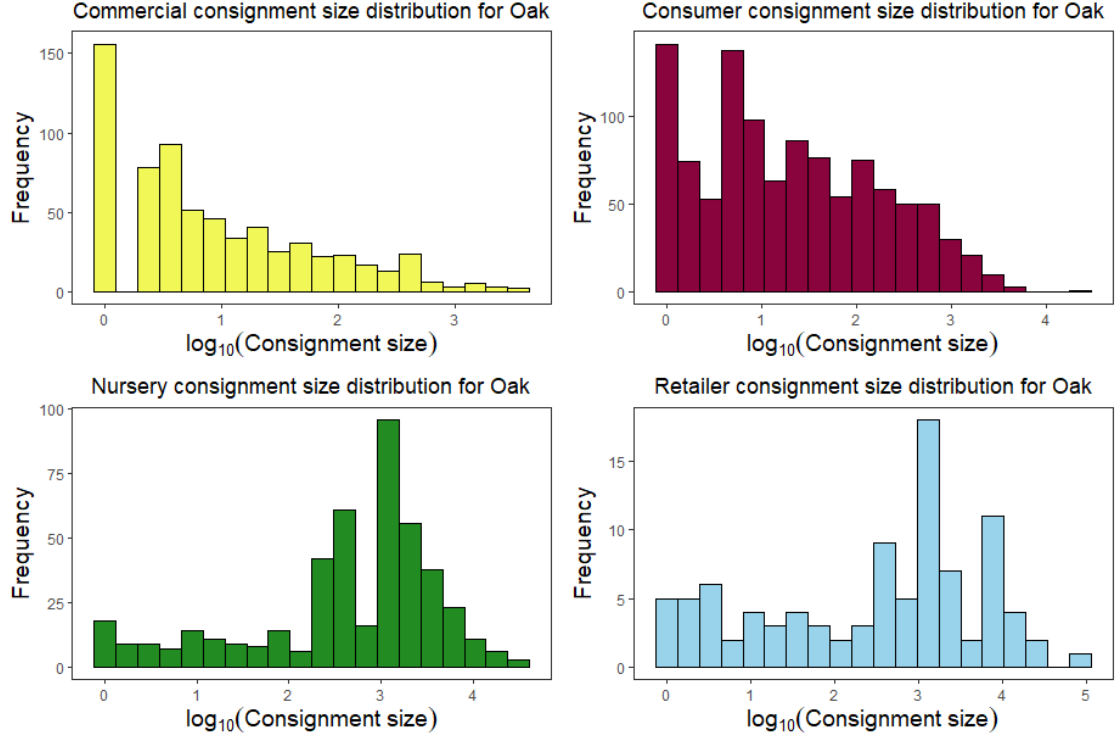
Figure 3.4. Distribution of consignment sizes for sales in Oak, separating for each customer group. We use the sales from all nurseries in our data $\{N_1, N_2, N_3, N_4\}$.

We test the skewness of these distributions against Boxcox transformations for $\lambda \in (-4, 4)$ using increments of 0.1 [152]. We find that the values of λ which minimise skewness were close to zero (for Oak sales, minimum skewness is achieved at $-0.3, -0.1, 0.2, 0.1$ for commercial, consumer, nursery and retailer groups respectively). An optimal λ of 0 indicates a log-normal distribution. Thus, we model consignments with an assumed log-normal distribution, parameterised by a mean and standard deviation. From the data and Figure 3.4, the truncations for each node subcategory can be given. Again, the minimum value for all distributions will be 1. For commercial and consumer nodes, the upper truncation will be 10^4 and for nurseries and retailers, it will be 10^5.

We are unable to classify the purchases from other nurseries (identified in the data) into one of our subcategories ($N_{\text{Com}}, N_{\text{Cons}}, N_{\text{Nur}}, N_{\text{Ret}}$). It is for this reason that consignment sizes for nurseries are to be modelled from one distribution. Hence, we denote the consignment size distributions $\left(\mathcal{C}_{Com}(y_\mu^{Com}, y_\sigma^{Com}), \mathcal{C}_{Cons}(y_\mu^{Cons}, y_\sigma^{Cons}), \mathcal{C}_{Nur}(y_\mu^{Nur}, y_\sigma^{Nur}), \mathcal{C}_{Ret}(y_\mu^{Ret}, y_\sigma^{Ret})\right)$.

This concludes the section on modelling edges and edge weights. In the next section we outline our algorithm to construct the network.

Network construction algorithm

We use the same steps outlined in Chapter 2 to construct the trade network. Firstly, we input a distribution of nurseries and retailers $(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|, |\text{Ret}|)$, where the nodes each have an assigned out-degree and customer demographic.

Recalling that nursery subcategories are defined by assigning more than 50% of their outgoing links to one customer group (either commercial, consumer, nursery or retailer), we describe the method of constructing the customer demographics for nursery nodes in the following way. The proportion of a nursery's links to its dominant customer, p_1, is sampled from a uniform distribution between 0.5 and 1. The second proportion, p_2, is then sampled from a uniform distribution between 0 and $(1 - p_1)$. p_3 is sampled uniformly between 0 and $(1 - p_1 - p_2)$. In order for all the proportions to sum to 1, p_4 is determined, given by $(1 - p_1 - p_2 - p_3)$. The proportions p_2, p_3 and p_4 are assigned randomly to the non-dominant customers for each nursery. This results in each nursery having a unique customer demographic that follows the rule we identified in our data analysis. The retailer customer demographic is simple as we assume that retailer nodes sell only to consumers.

As in Chapter 2, our initial distribution of nurseries and retailers, customer demographics and out-degrees all impose an upper bound on each nursery's out-degree, i.e., a nursery cannot sell to more nurseries than there exists in the network. We also generate retailer out-degrees this way, with simply the initial upper bound of $3,000$ because the number of consumers is not pre-specified. We assume that the commercial and consumer nodes have an in-degree of 1, the number of these nodes are calculated by the total number of outgoing links to commercial and consumer nodes from the nurseries and retailers.

We then implement the edges assignment process as in the previous chapter. Assigning edges uniformly does not guarantee that all retailer nodes will have an in-going edge. We assume that the domestic trade network consists of one weakly-connected component, hence any retailers at the end of the edge assignment process with an in-degree of 0, are removed from the network, along with their associated customers.

Finally, we assign weights to the edges in the network. For each edge, we generate edge weights via the buyers associated consignment size distribution and add these weights to the third column of the edge-list.

Thus, we have our model which constructs a plant trade network where we input the number of nurseries and retailers alongside out-degree and consignment distribution parameters.

3.3.3 Model summary and parameter values

As previously stated, the network represents the annual trade of a given plant species. In our analysis we use as a baseline the data estimates for the trade of Oak plants. Thus, for each of the nurseries in the data, we calculate the total number of unique customers per year. Similarly, for the consignment sizes, we calculate the mean and variance of the sale quantity of Oak plants for commercial, consumer, nurseries and retailers. We show a summary of the model parameters and our baseline values in Table 3.2.

A network of 632 nurseries (158 nurseries of each type) and 6320 retailers, results in a network of approximately 2.5 million nodes and edges. The calculations we use in our analysis do not perform well with very large networks, especially directed networks where the adjacency matrices are asymmetric. Thus, we analyse a smaller network with 160 total nurseries, with 40 nurseries of each type as a base case (Table 3.2). We make the simplifying assumption that the number of retailers in the network is 10 times the number of nurseries, as the base case. The estimates for the out-degree and consignment size distributions come from the full-sized network with an estimated number of 630 nurseries and

6525 retailers. Hence, we scale the means and standard deviations in our model by $\frac{160}{630}$ and the variances by $\left(\frac{160}{630}\right)^2$. We show this explicitly in Table 3.2. The network simulations using these parameters result in networks of approximately 150000 nodes and edges, which are small enough to conduct our analysis.

Table 3.2. Parameters for network construction and base case values considered.

Parameter	Description	Baseline value	Source
$|N_{\text{Com}}|$	Number of commercial dominated nurseries	40	Simplifying Assumption
$|N_{\text{Cons}}|$	Number of consumer dominated nurseries	40	Simplifying Assumption
$|N_{\text{Nur}}|$	Number of nursery dominated nurseries	40	Simplifying Assumption
$|N_{\text{Ret}}|$	Number of retailer dominated nurseries	40	Simplifying Assumption
$|\text{Ret}|$	Number of retailers	$1,600$	Assumption
d_μ^{out}	Nursery and retailer out-degree mean	$380 \times (160/630)$	Scaled data estimate
$d_{\sigma^2}^{\text{out}}$	Nursery and retailer out-degree variance	$114,587 \times (160/630)^2$	Scaled data estimate
y_μ^{Com}	Commercial consignment-size mean	$79 \times (160/630)$	Scaled data estimate
y_σ^{Com}	Commercial consignment-size standard deviation	$297 \times (160/630)$	Scaled data estimate
y_μ^{Cons}	Consumer consignment-size mean	$185 \times (160/630)$	Scaled data estimate
y_σ^{Cons}	Consumer consignment-size standard deviation	$822 \times (160/630)$	Scaled data estimate
y_μ^{Nur}	Nursery consignment-size mean	$1844 \times (160/630)$	Scaled data estimate
y_σ^{Nur}	Nursery consignment-size standard deviation	$3567 \times (160/630)$	Scaled data estimate
y_μ^{Ret}	Retailer consignment-size mean	$3497 \times (160/630)$	Scaled data estimate
y_σ^{Ret}	Retailer consignment-size standard deviation	$9588 \times (160/630)$	Scaled data estimate

3.3.4 Methods of analysis

In this section we describe the methods used to analyse the network structure. As mentioned in Chapter 1, network centrality measures help quantify how nodes contribute to the structure of a network. In this chapter, we choose a range of network centrality measures from an epidemiological perspective. The centrality measures we use include: out-degree, in-degree, out-strength, in-strength, betweenness, hub and authority scores. We refer back to Chapter 1 for the definitions of these measures.

We look at the in-degree because this is a simple measure which gives an indication of a node's level of exposure to sources of infection. The out-degree is similarly chosen to give an indication of a node's ability to directly infect others. The in- and out- strength have

similar interpretations to in- and out- degree and are a means of comparing the effect that edge weights have in the model.

The betweenness centrality score measures how often a node appears in the shortest paths amongst all pairs of nodes and thus gives an indication of how much a node raises the rate of disease transmission across the entire network. If we remove a node with high betweenness, this increases the distance that a pest/disease is required to travel in order to spread throughout the network. Thus, quarantines or trade restrictions are more effective at nodes with high betweenness scores. In weighted networks, edge weights are interpreted as distances for this measure. As edge weights in our network represent the volume of trade, a larger amount of trade is more indicative to a closer relationship between nodes. Thus, for the betweenness measure, we transform edge weights to their reciprocal so that more plants traded between nodes represents a closer relationship. We also consider the unweighted version of this measure.

The hub and authority scores assign importance to nodes based on the importance of their neighbours. Hubs scores are tied to out-going edges and authority scores with in-going edges, thus the hubs act as important sources of infection and authorities as destinations. Thus, investigation into the nodes with the highest hubs and authority scores can give us insight into the significant pathways of disease spread. We consider unweighted versions of both the hub and authority scores.

To give an intuition of the overall network structure, we visualise a smaller version of the network. We then continue our analysis with the baseline parameter values in Table 3.2. To further understand the structure of our network, we simulate 100 networks and calculate the out-degree and in-degree for each node subcategory.

Furthermore, in our analysis, we want to see how changing the model parameters affects the network structure; in particular, the average influence on network structure between node subcategories. Due to the small data set this model is based on and the variability in each simulation of the network, we focus on identifying general patterns through aver-

aging many network simulations. We order our analysis into three parts: the distribution of nurseries ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$), the nursery and retailer degree distribution parameters (d_μ^{out}, $d_{\sigma^2}^{out}$), and the consignment size distribution parameters (y_μ^{Com}, y_σ^{Com}, y_μ^{Cons}, y_σ^{Cons}, y_μ^{Nur}, y_σ^{Nur}, y_μ^{Ret}, y_σ^{Ret}).

The distribution of nurseries

We expect the distribution of nurseries ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) to be different depending on the plant species traded, as each species will move within its own trade network. It is logical to assume that the trade networks for some species will vary greatly for plant species which serve different functions, such as horticultural and ornamental plants. Hence, in our analysis, we consider a fixed total number of nurseries (160) and retailers (1,600) and analyse different initial distributions of nurseries ($N_{\text{Com}}, N_{\text{Cons}}, N_{\text{Nur}}, N_{\text{Ret}}$) to investigate how this may change the network structure. The distribution of nurseries can be described by a weak composition of four nursery subcategories (a sum of four non-negative integers). This corresponds to $\binom{160+4-1}{4-1} = 708,561$ distributions of nurseries to investigate, if we consider all possible combinations. To simplify and make the analysis computationally feasible, we focus on specific scenarios.

S1: Uniform distribution ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (40, 40, 40, 40)

S2: Predominantly Commercial ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (80, 20, 40, 20)

S3: Predominantly Consumer ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (20, 50, 40, 50)

S4: N_{Com} dominant ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (130, 10, 10, 10)

S5: N_{Cons} dominant ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (10, 130, 10, 10)

S6: N_{Nur} dominant ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (10, 10, 130, 10)

S7: N_{Ret} dominant ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$) = (10, 10, 10, 130)

As a baseline ($S1$), we consider a uniform distribution of nurseries as we have no data to estimate the true distribution. This allows us to see the effect of changing the distribution of nurseries more clearly.

In $S2$, we look at a nursery distribution for the trade of commercial plants, e.g., horticultural plants. In this network, most nurseries will sell predominantly to commercial customers (in this instance half of all nurseries, chosen arbitrarily). With this increase in $|N_{\text{Com}}|$, to keep the total number of nurseries constant, other nurseries must decrease. We assume that a plant used commercially has less non-commercial demand, and so $|N_{\text{Cons}}|$ and $|N_{\text{Ret}}|$ will decrease accordingly. We do not change $|N_{\text{Nur}}|$ in this scenario.

In $S3$, we look at plants traded in a predominantly consumer (non-commercial) network. Similarly as in scenario 2, this translates to an increase in $|N_{\text{Cons}}|$ and $|N_{\text{Ret}}|$ and a respective decrease in $|N_{\text{Com}}|$. Again, we do not change $|N_{\text{Nur}}|$.

For scenarios $4-7$, we look at extreme cases where nursery distributions consist of mostly one nursery type. In these scenarios, we arbitrarily set the nursery type of interest to 130 and all others to 10. We investigate these extremes to see how significant changes to the core of inner trading amongst nurseries affects the overall network structure.

We note that the final number of retailers in the network will not necessarily equal the initially specified 1600, due to the deleting of retailer nodes with an in-degree of zero. Throughout these scenarios, the final number of retailer nodes may change as a result of the change in the distribution of nurseries.

Given that other network models on plant trade do not include the customers, for each scenario, we also calculate network measures on the network consisting of only the nurseries and retailers, to see whether there are any significant differences when including the customers in the trade network. For each of the scenarios, the values of all other parameters are set to baseline as in Table 3.2.

The analysis for each scenario is conducted by simulating the network 100 times and for each simulation, taking the mean of the centrality score for each node subcategory. In this set of analysis, we consider all aforementioned centrality measures. We present the 100 means for each node subcategory in box plots.

Out-degree mean and variance

The out-degree distribution $\mathcal{K}_{\text{out}}(d_\mu^{out}, d_{\sigma^2}^{out})$ models the out-degrees for all nursery and retailer nodes. Hence changing the parameters d_μ^{out} and $d_{\sigma^2}^{out}$ will have an effect on all node types in the network. By varying the out-degree parameters, we can assess how changes in network density affect which node groups are most influential.

The maximum out-degree in the data was 1500, which gives us some indication of what ranges to consider when varying the out-degree parameters for a full-sized network. The network we consider is roughly a quarter of the size (160 nurseries), and so for our analysis, we consider d_μ^{out} in the sequence $(100, 200, ..., 1000)$ and for $d_{\sigma^2}^{out}$, we consider the sequence $(2000, 4000, ..., 20000)$.

Given that commercial and consumer nodes have an assumed in-degree of 1 and an out-degree of 0, the centrality scores of these nodes are not of interest. However, the commercial and consumer nodes may have some effect on the centrality scores of the nurseries and retailers. Hence such nodes are retained in the network when calculating these scores. Thus, we constrain our analysis to the nursery and retailer node subcategories.

We calculate the mean network centrality measure for each nursery and retailer node subcategory. We repeat this for 100 network simulations, to address the variability in each network simulation for the same parameter set. Therefore, for each parameter combination, we calculate the average network centrality score per node subcategory 100 times and then we average over these 100 values. Thus, we calculate this average for each combination of parameter values for d_μ^{out} and $d_{\sigma^2}^{out}$, and for each centrality measure. We present our

results in the form of heatmaps. In this set of analysis, we do not consider the in-degree, out-degree, in-strength and out-strength measures.

Customer consignment distributions

The final set of parameters we analyse are the customer consignment distribution parameters $\{y_\mu^{Com}, y_\sigma^{Com}, y_\mu^{Cons}, y_\sigma^{Cons}, y_\mu^{Nur}, y_\sigma^{Nur}, y_\mu^{Ret}, y_\sigma^{Ret}\}$. We are interested in how changes in customer demand affect the network structure. The range of values we consider are given in Table 3.3. We construct heatmaps in our analysis for the customer consignment distribution parameters in the same way as the out-degree distribution parameters (Section 3.3.4). We do not calculate unweighted betweenness, hub and authority measures for variations in the consignment size parameters because differences are only seen when considering edge weights. In addition, we do not consider the in-degree, out-degree, in-strength and out-strength measures.

Table 3.3. Parameters that model customer sales distributions and ranges of values considered.

Parameter	Description	Baseline value	Range considered
y_μ^{Com}	Commercial consignment-size mean	$79 \times (160/630)$	$(10, 100, 200, ..., 1000)$
y_σ^{Com}	Commercial consignment-size standard deviation	$297 \times (160/630)$	$(100, 1000, 1500, ..., 5000)$
y_μ^{Cons}	Consumer consignment-size mean	$185 \times (160/630)$	$(10, 100, 200, ..., 1000)$
y_σ^{Cons}	Consumer consignment-size standard deviation	$822 \times (160/630)$	$(100, 1000, 1500, ..., 5000)$
y_μ^{Nur}	Nursery consignment-size mean	$1,844 \times (160/630)$	$(10, 100, 200, ..., 1000)$
y_σ^{Nur}	Nursery consignment-size standard deviation	$3567 \times (160/630)$	$(100, 1000, 1500, ..., 5000)$
y_μ^{Ret}	Retailer consignment-size mean	$3497 \times (160/630)$	$(10, 100, 200, ..., 1000)$
y_σ^{Ret}	Retailer consignment-size standard deviation	$9588 \times (160/630)$	$(100, 1000, 1500, ..., 5000)$

3.4   Results

In this section, we present the results from the analysis on our network model. We first show a smaller version of the trade network to visualise the network structure to gain an intuition on the output of our model.

We then show our analysis using our baseline parameter values. We analyse the out-degree and in-degree distributions of the network and we present these distributions for each node subcategory. We move on to our sensitivity analysis, changing the initial distribution of nursery subcategories, the out-degree mean and variance for nursery and retailer nodes and the consignment size distribution parameters.

3.4.1 Network structure

Network visualisation

To gain an intuition on how our network model is structured, we show in Figure 3.5 a smaller version of the network with 32 nurseries, 8 of each nursery subcategory and 320 retailers. This network consists of $2,500$ nodes and edges and, with this visualisation, the hierarchical structure of trade can be seen more clearly. We see that there is an inner core of trading between nurseries at the centre, with trade expanding outwards to retailers, who have their own set of customers. From a network of this size, we can see that this network has a similar network topology to a tree.

32 Nurseries Network for Oak
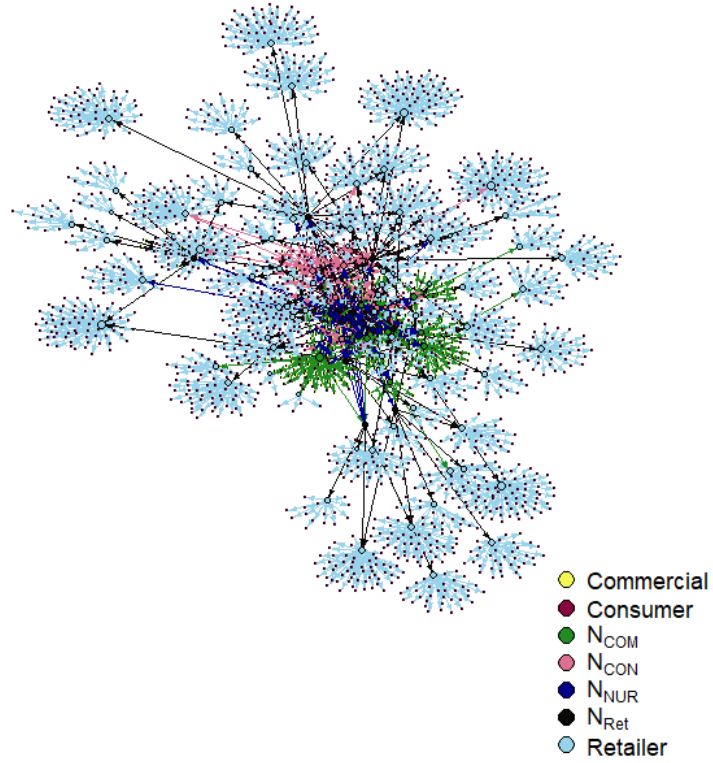
Figure 3.5. Visualisation of network for Oak trade with 32 nurseries and 320 retailers. Node colour represents node subcategory. Edge colour corresponds to the selling node subcategory.

We then proceed with the network with parameter values as shown in Table 3.2.

Degree distributions

The out-degree distribution for each nursery and retailer subcategory for the 100 network simulations is shown in Figure 3.6. From this figure we can see that all nurseries and retailers have similar out-degree distributions, which is expected from our modelling assumptions. The peak of all nursery and retailer distributions occurs at 2, which is consistent with our parameter values. We also see the effect of truncating the distributions for each node subcategory. The out-degrees for N_{Nur} nodes are constrained the most and so in subgraph (c) we see this distribution has the smallest maximum out-degree of 2.6. Whereas, the $N_{\mathrm{Com}}, N_{\mathrm{Cons}}$ and N_{Ret} nodes all have similar maximum out-degrees of 2.8 and 3 (sub-

graphs $(a), (b), (d)$. The retailer nodes have the longest tail due to the only constraint on their out-degrees being the initial upper truncation of 3.47 (3000 on a $\log_{10}$ scale) to reject extreme values. The short right tail at approximately 3.1 also indicates that our initial upper truncation was appropriately chosen to not cause bunching up at this value. From these graphs we see that nurseries and retailers on average can directly infect 100 other nodes but a superspreader has the potential to directly infect up to 1000 other nodes.
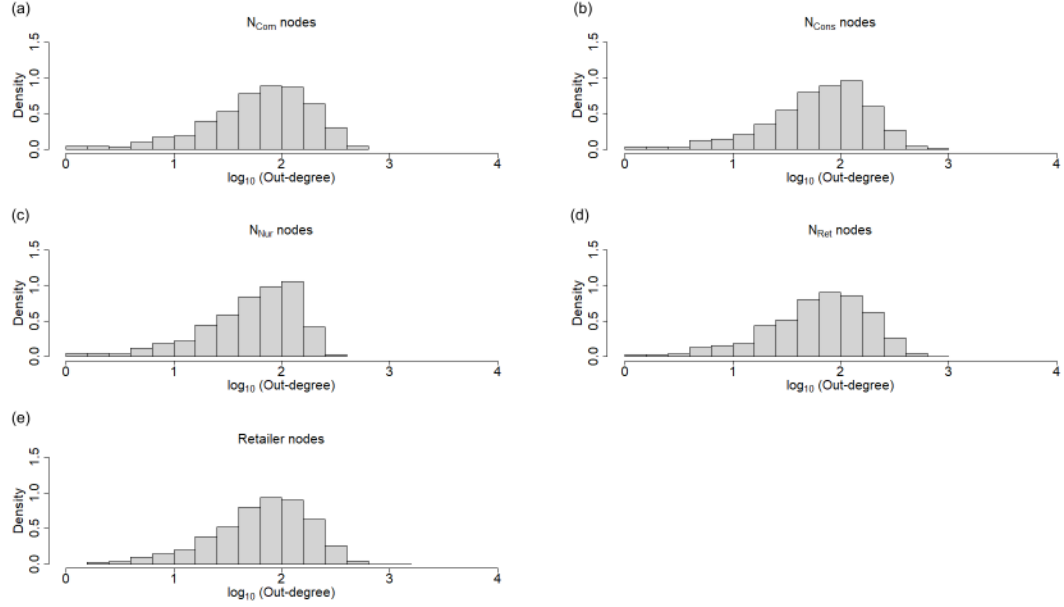
Figure 3.6. Out-degree distribution shown per node subcategory (excluding commercial and consumer nodes) for 100 network simulations. The parameter values used are shown in Table 3.2.

We see the in-degree distribution for each nursery and retailer subcategory in Figure 3.7. Nursery nodes (subgraphs $(a) - (d)$) all have the same distribution, with in-degrees between 0.8 and 1.7 and a peak value at approximately 1.4. Thus nurseries have the potential to be directly infected by between 10 to 30 other nurseries. Retailer nodes have relatively smaller in-degrees which fall in the range of 0 to 1, with a peak at around 0.5. These results suggest that retailers are at a considerably lower risk of exposure to disease compared to nurseries.
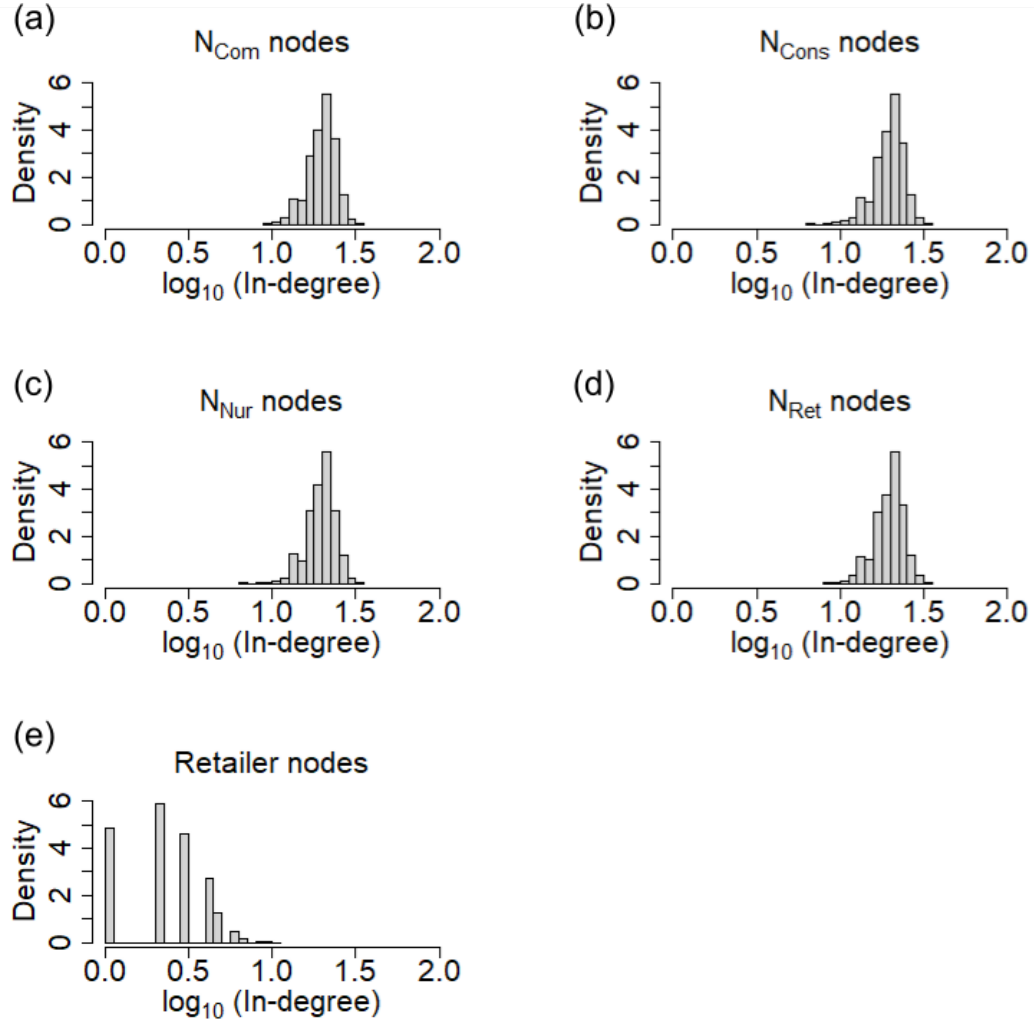
Figure 3.7. In-degree distribution shown per nursery and retailer node subcategory for 100 network simulations. The parameters for the network constructions are shown in Table 3.2.

3.4.2 The distribution of nurseries

We expect this analysis to give insight into the effect that a change in the plant trade market may have on network structure. As described in the methods of analysis section, we consider seven different scenarios of nursery subcategory distributions ($|N_{\text{Com}}|$, $|N_{\text{Cons}}|$, $|N_{\text{Nur}}|$, $|N_{\text{Ret}}|$). For each scenario, we see how network measures change for each node subcategory. We note that all of the figures in this section are displayed on a $\log_{10}(1 + \text{data})$ scale.

We calculate centrality measures for both the network with and without customers (commercial and consumer nodes). For all centrality measures and scenarios (except out-degree), the relative significance of each node subcategory does not change when using the network subset. For brevity, we show an example below in Figure 3.8 for scenario 1 and the weighted betweenness measure. All comparisons for the other scenarios and centrality measures can be found in Appendix B.
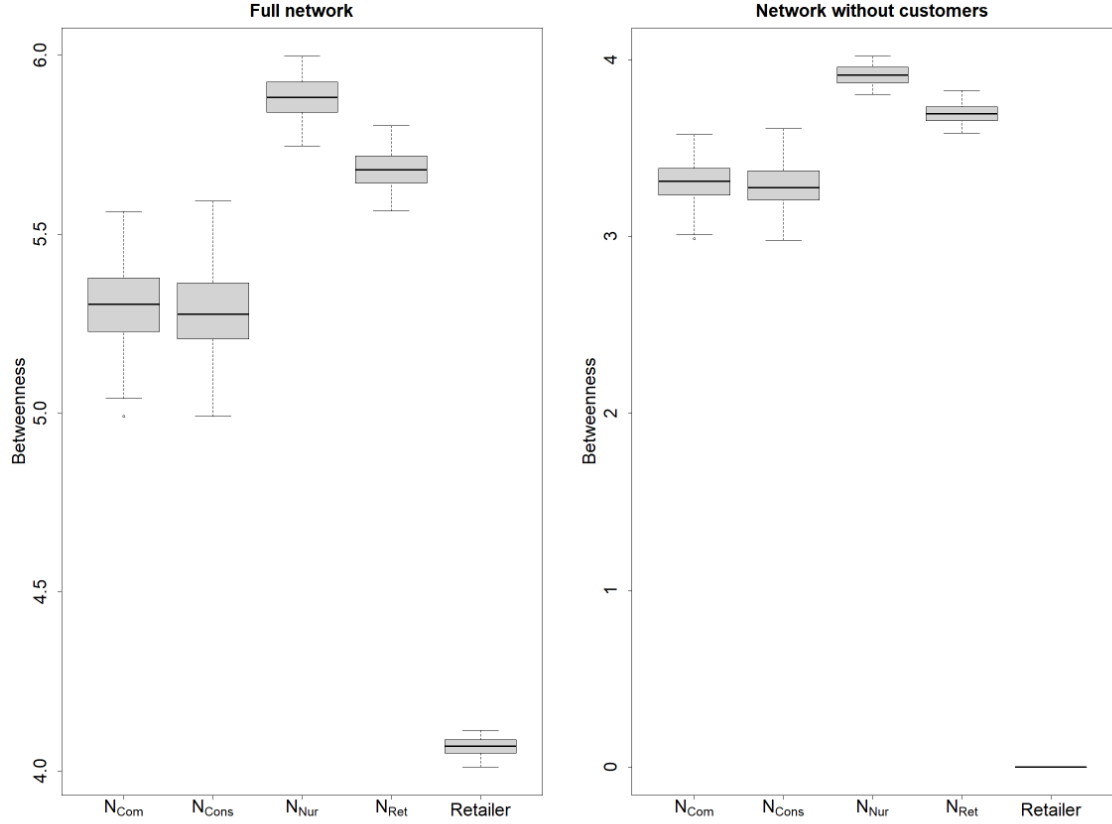
Figure 3.8. Box plots of the average betweenness scores per node subcategory for 100 networks. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows betweenness scores calculated on the entire network, the plot on the right shows betweenness scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes. Parameters values used are shown in Table 3.2.

Removing the customer nodes changes the relative out-degrees of nursery and retailer nodes; see Figure 3.9. In the full network, all nurseries and retailers have similar scores, except N_{Nur} nodes with smaller out-degrees. When removing the customers, N_{Nur} and N_{Ret} nodes have the largest out-degree scores, as expected. If using out-degree to infer

which nodes have the greatest outbreak potential, the networks which ignore customers and only consider the nursery and retail sector may misinterpret those most central nodes. It is interesting however, that for all other centrality measures we use, we see no difference in relative node significance when removing customers.
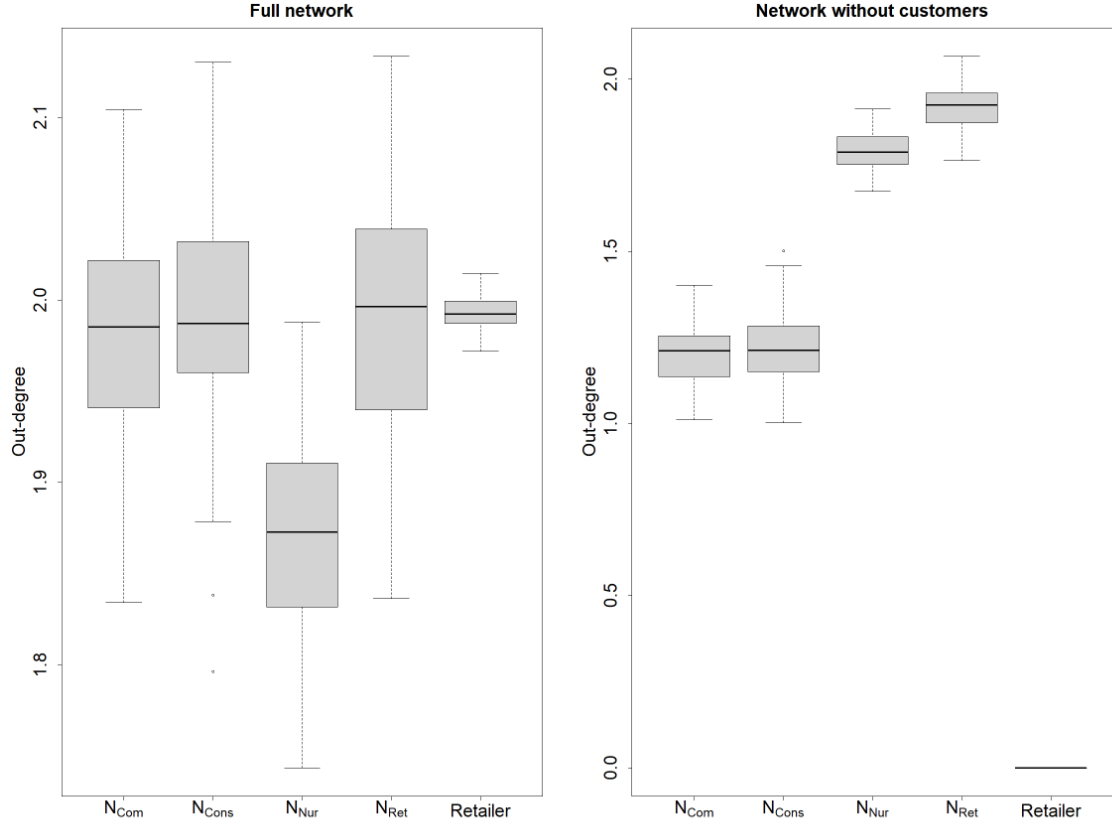
Figure 3.9. Box plots of the average out-degree score per node subcategory for 100 networks. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes. Parameters values used are shown in Table 3.2.

These are important differences to highlight given that other model plant trade networks do not include customers [2, 46, 70, 88, 89, 106]. All subsequent figures in this section are shown for centrality measures calculated on the entire network.

Considering the average weighted betweenness scores across the seven scenarios in Figure 3.10, retailers consistently have low scores which do not vary across scenarios, and so are omitted. Nurseries across scenarios have mostly similar betweenness, with N_{Nur} nodes achieving the highest scores, N_{Ret} nodes the second highest and N_{Com} and N_{Cons} nodes consistently joint with the lowest scores. However, in scenario 6 (networks primarily consisting of N_{Nur} nodes), the node subcategory with the highest betweenness score changes to N_{Ret}. In scenario 7 (networks primarily consisting of N_{Nur} nodes), the scores for N_{Ret} decrease and become equal to the scores for N_{Com} and N_{Cons}. With the baseline model parameters (Table 3.2), the network structure regarding average weighted betweenness scores appears robust to changes in nursery distribution, with qualitative differences only seen when the network consists of mostly N_{Nur} or N_{Ret} nurseries (scenario 6 and scenario 7).

We then consider the average unweighted betweenness scores (Figure 3.11). In comparison with the weighted betweenness measure, the node subcategory with the highest score changes from N_{Nur} to N_{Ret}. Otherwise, the prevailing pattern remains the same, with N_{Nur} and N_{Ret} nodes having higher betweenness scores than N_{Com} and N_{Cons}. In scenarios 1 to 5, N_{Nur} and N_{Ret} nodes are mostly similar in their betweenness scores, and significant differences only occur in scenarios 6 and 7. In scenario 6 (mostly N_{Nur} nurseries), the betweenness scores for N_{Nur} nurseries decrease to similar levels as N_{Com} and N_{Cons} nodes. In scenario 7 (mostly N_{Ret} nurseries), the unweighted betweenness scores for N_{Ret} decrease such that N_{Nur} nurseries have the highest scores.

This analysis shows that trade restrictions/quarantines would be most effective at N_{Nur} nodes. We see that, using the unweighted betweenness measure, this changes to N_{Ret} nodes. Thus, depending on how the volume of trade affects disease transmission, we see different recommended control measures. These results are robust to small changes in the market, and do not change unless there is a much larger number of N_{Nur} or N_{Ret} nodes (scenarios 6 and 7). Interestingly, from both measures and all scenarios, nurseries are a better choice than retailers to focus trade restrictions or quarantines.
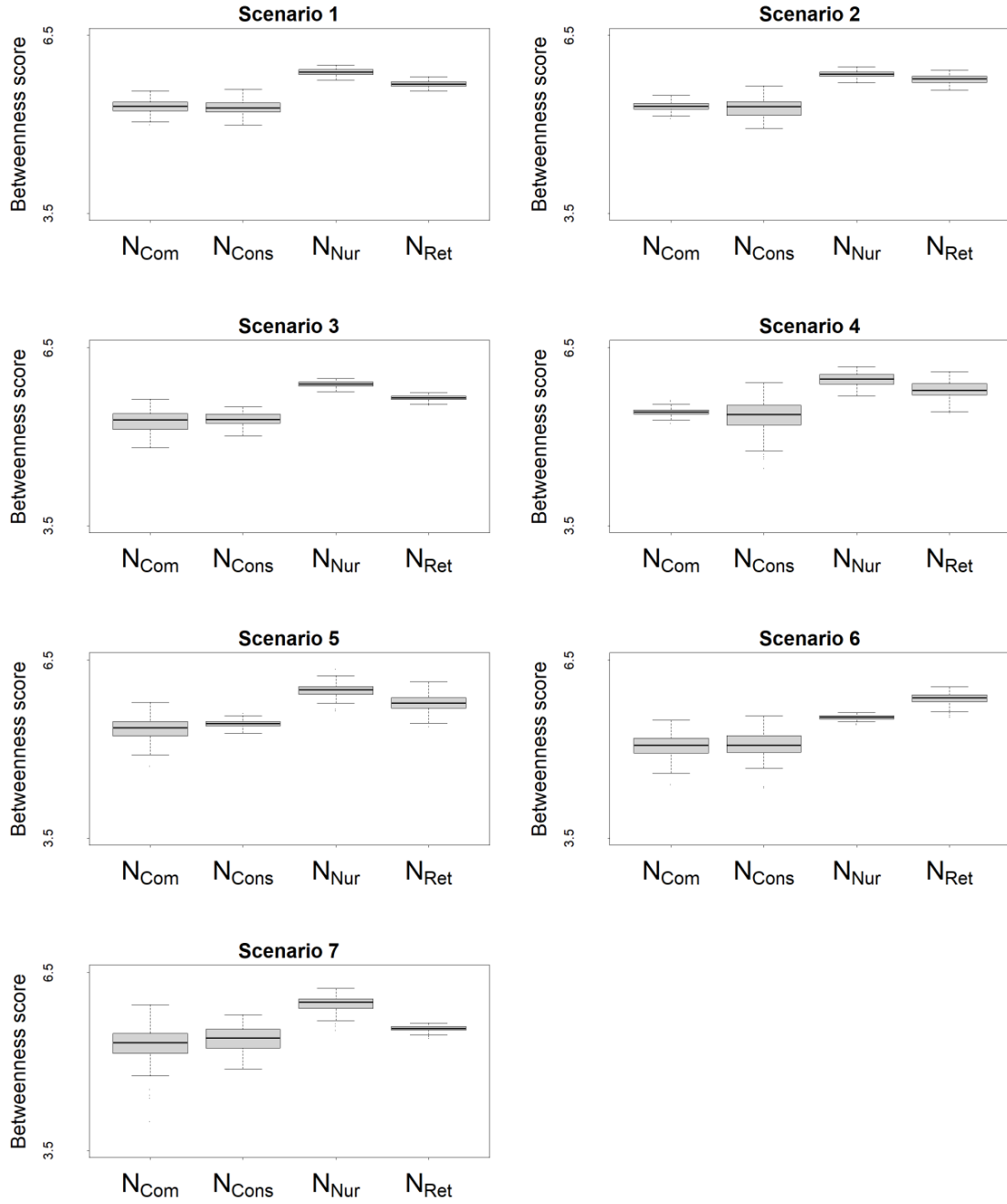
Figure 3.10. Weighted betweenness scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average betweenness scores per nursery node subcategory. Parameters values used are shown in Table 3.2.
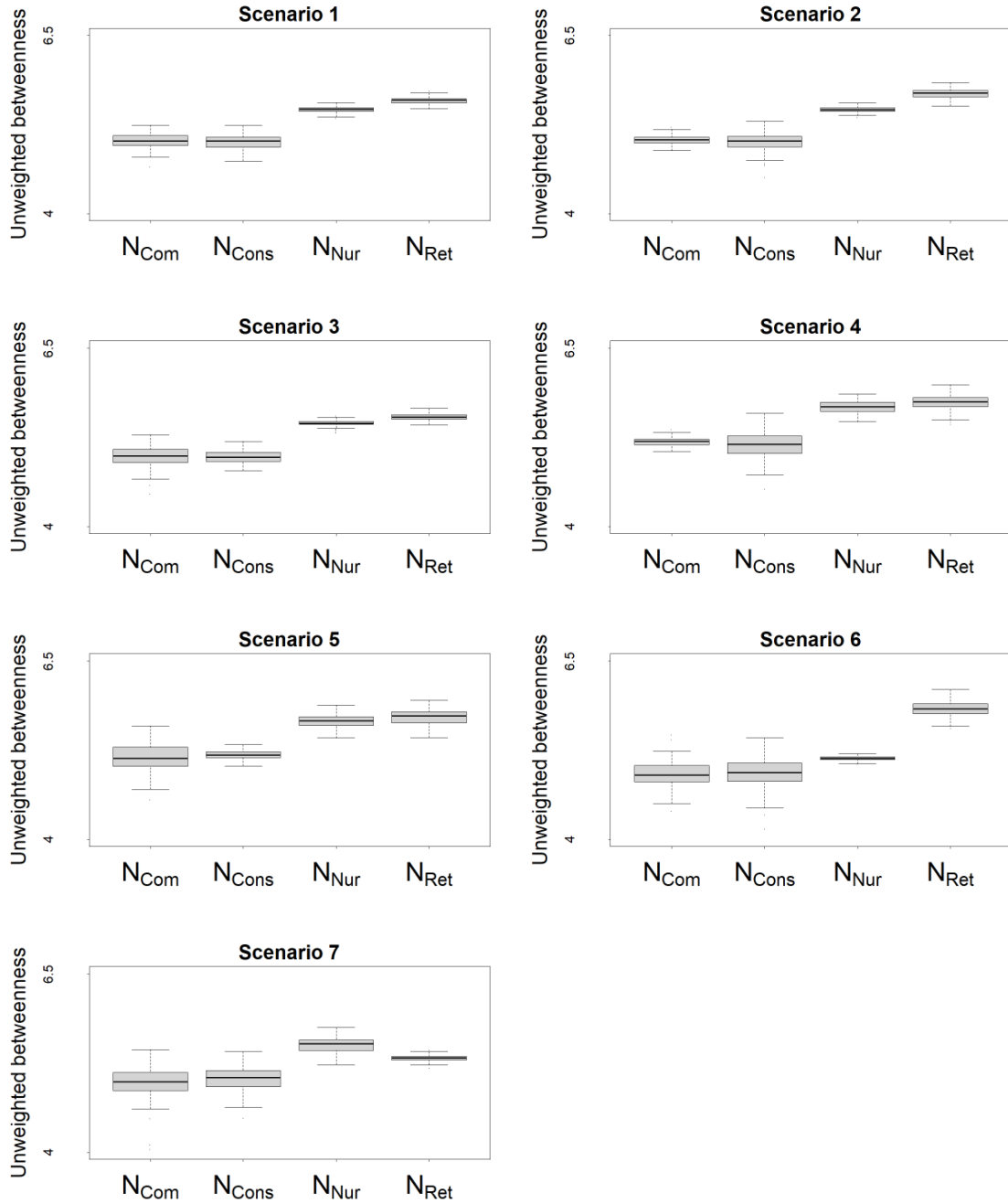
Figure 3.11. The unweighted implementation of the betweenness centrality measure, shown across all 7 considered scenarios. The scores are transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average unweighted betweenness scores per nursery node subcategory. Parameters values used are shown in Table 3.2.

The average in-strength scores show a consistent pattern for scenarios 1 through 6, with all nurseries having similar scores and retailers attaining the lowest scores (Figure 3.12). The relative average in-strength between nursery node subcategories does not change across scenarios because nursery subcategories, by model assumption, have no preference to which nursery subcategory they sell to. It is therefore understandable that nurseries have a larger average in-strength when there are more links being assigned to nurseries (scenario 6). In scenario 7, retailer nodes average in-strength becomes similar to nurseries. This is expected in this scenario as the majority of nurseries in the network are subcategory N_{Ret}. Thus, in scenario 7, organising inspections at retailers as well as nurseries should be a consideration, given their increased significance.
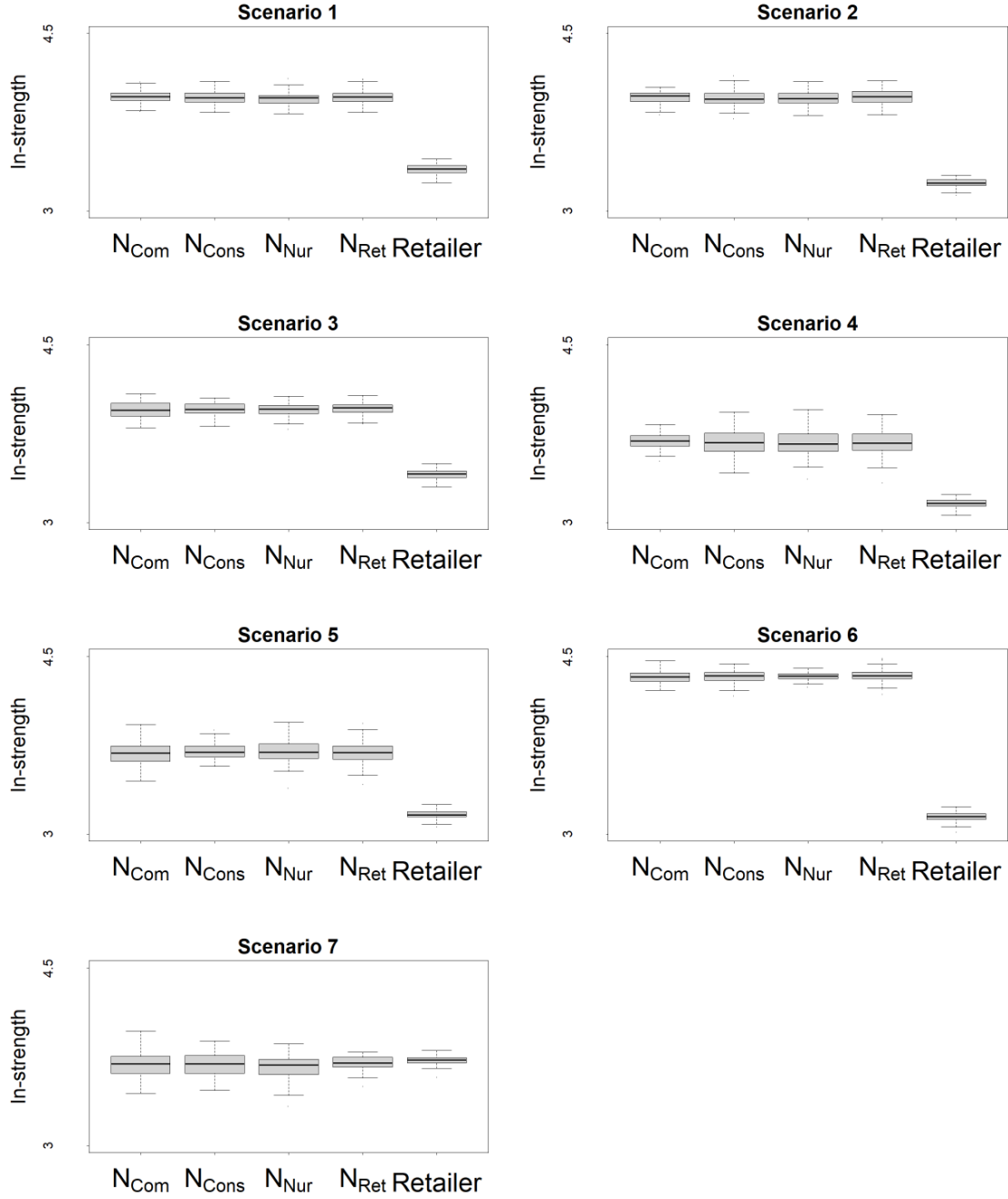
Figure 3.12. In-strength scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average in-strength scores per nursery and retailer subcategory. Parameters values used are shown in Table 3.2.

The in-degree scores show mostly the same pattern as we observe with the out-strength scores. The only difference we see is that the in-degree scores for retailers do not equal nursery scores in scenario 7. Therefore, without access to the edge weights in the network,

we would not capture retailer nodes' increased significance in markets dominated by N_{Ret} nodes and thus misplace judgement on where to focus inspections.
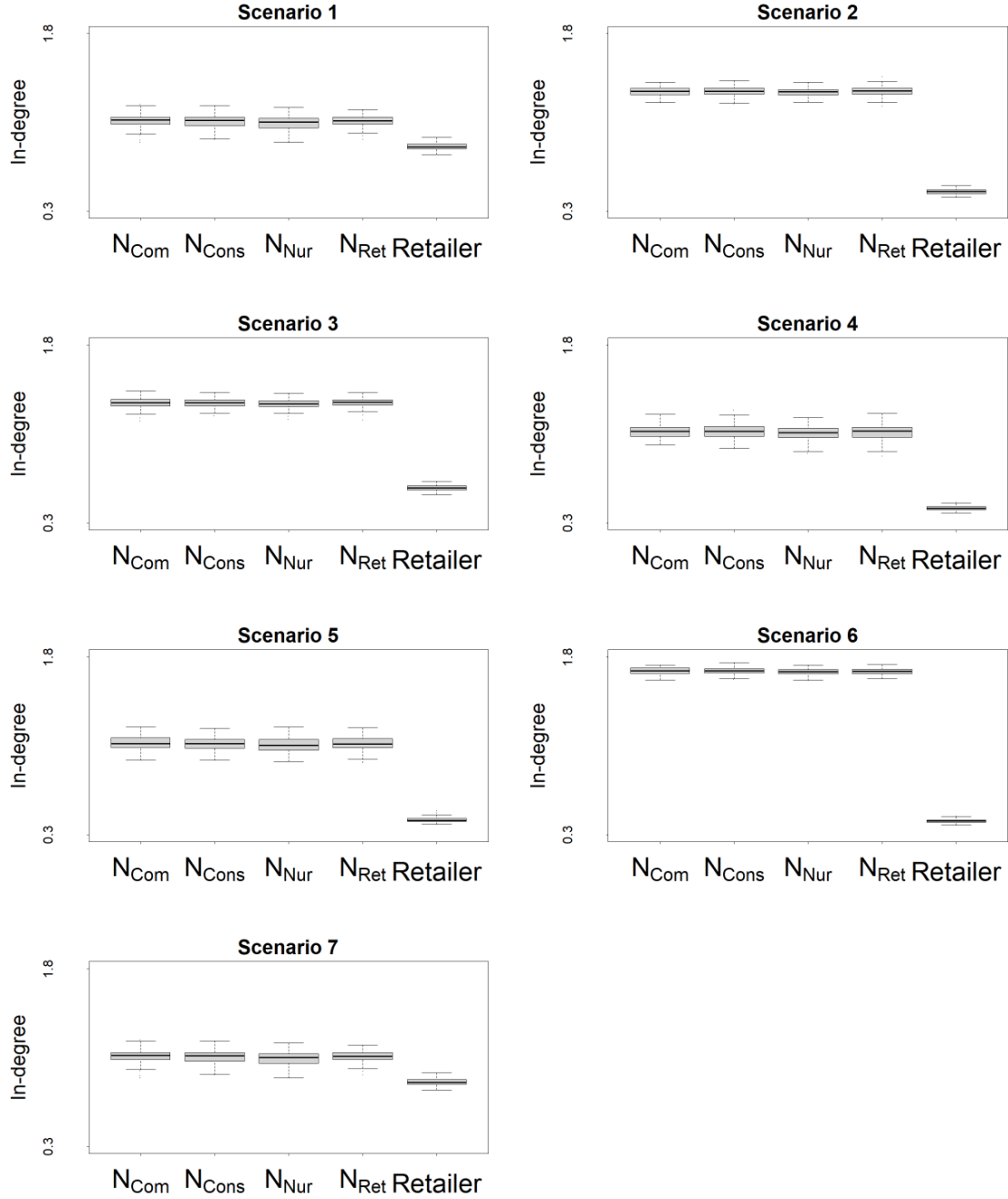
Figure 3.13. In-degree scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average in-degree scores per nursery and retailer subcategory. Parameters values used are shown in Table 3.2.

Out-strength scores show a consistent pattern across all 7 scenarios (Figure 3.14). Retailers have the lowest out-strength scores which do not vary across scenarios. N_{Ret} nodes have the largest average out-strength and N_{Nur} nodes have slightly lower out-strength. N_{Com} and N_{Cons} nodes have similar out-strength scores, lower than N_{Nur} nodes. This is a direct result from the consignment size parameters and the out-degree distribution parameters seen in Table 3.2.
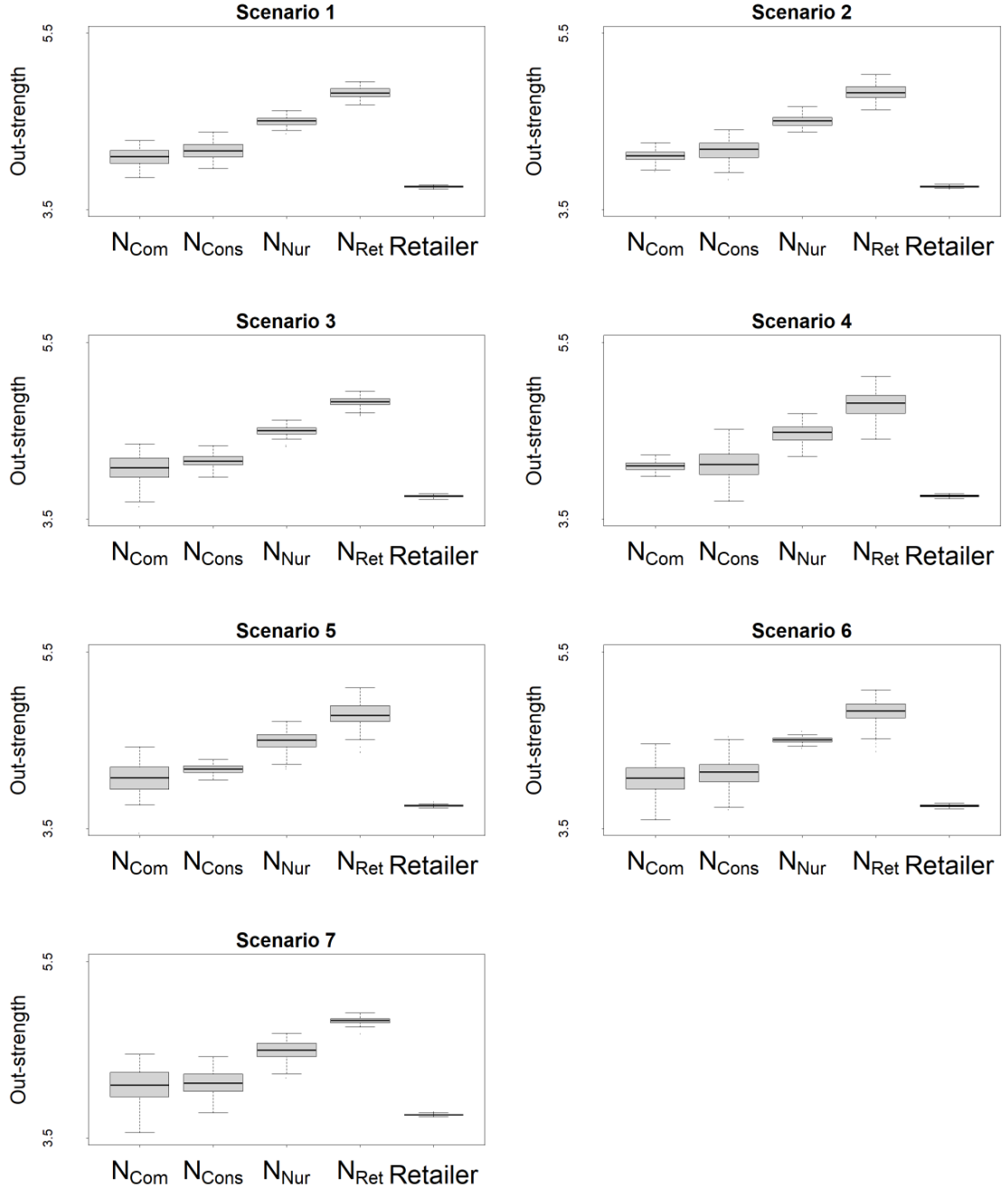
Figure 3.14. Out-strength scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average out-strength scores per nursery and retailer subcategory. Parameters values used are shown in Table 3.2.

The relative out-degree scores across nurseries and retailers are also robust to changes in nursery scenario (Figure 3.15). Out-degrees for all nursery and retailer node subcategories are similar, with N_{Nur} scores slightly smaller. This is a direct result from our assump-

tion that out-degrees for these nodes are modelled from the same distribution. Without knowledge of the out-strength scores, we would infer that all nurseries and retailers are of equal capability to infect other nodes. Combining the results from the out-strength and out-degree measures (Figure 3.14, Figure 3.15) we see that N_{Ret} and N_{Nur} nodes pose the greatest spreading potential.
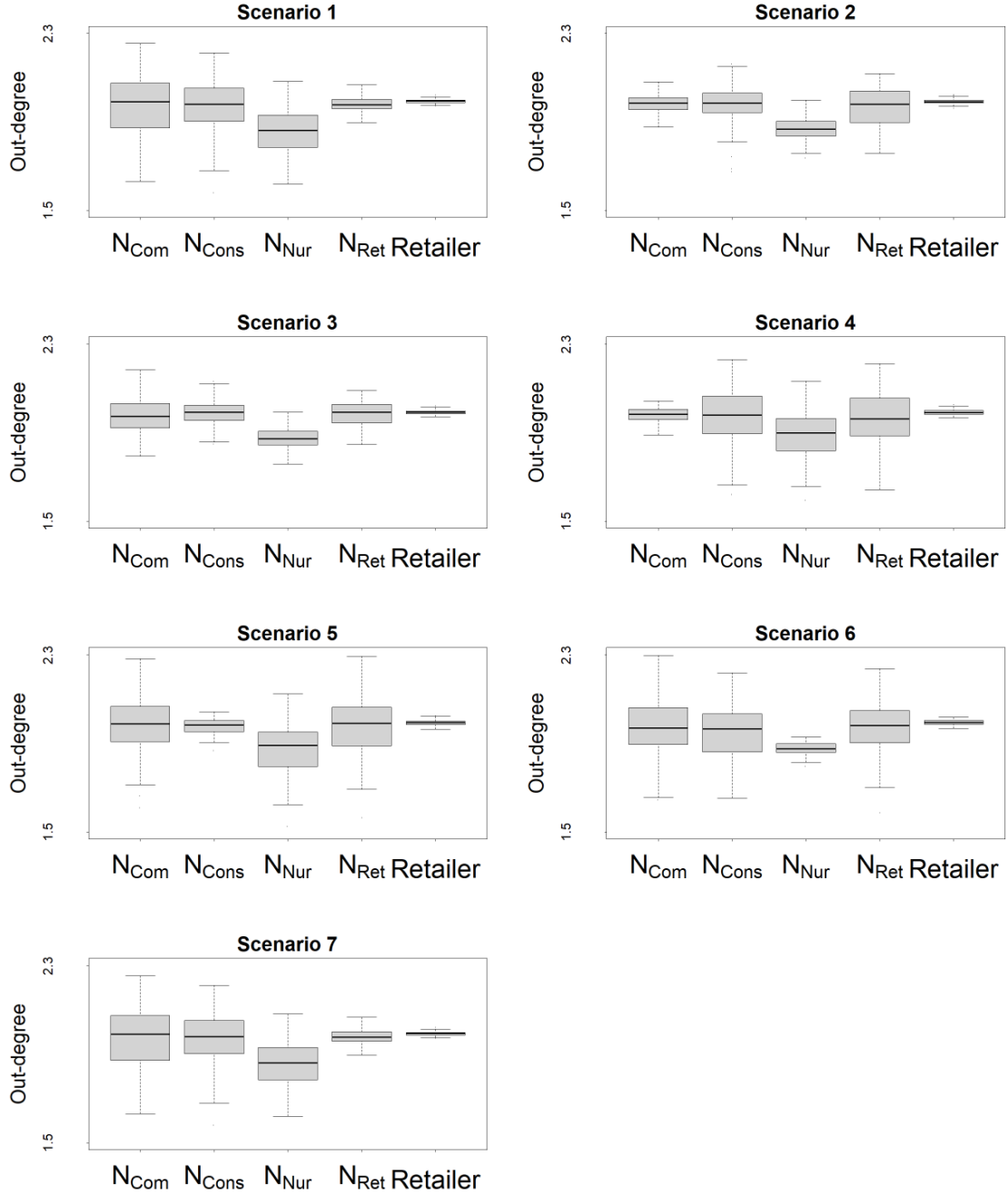
Figure 3.15. Out-degree scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Each box plot shows the distribution of 100 average out-degree scores per nursery and retailer subcategory. Parameters values used are shown in Table 3.2.

The average weighted hub scores across all scenarios follow the same pattern (Figure 3.16). We remove the retailer nodes in this analysis for clarity. Scores across scenarios are in the range between 0 and 0.1. We see that N_{Ret} nodes have the highest average hub scores

with the most variability whereas all other node subcategories hub scores are close to 0. In scenarios $4, 5$ and 6, the lower quartile for the N_{Ret} scores decrease to overlap with all other nurseries' scores. In scenarios with fewer N_{Ret} nodes than baseline (scenarios $2, 4, 5$ and 6) we see the variance in the average hub scores for N_{Ret} increases significantly. When the number of N_{Nur} nodes increases in scenario 6, we see the hub scores for these nodes increase. Thus, from this graph we see that disease control measures would best be focused on N_{Ret} nodes.

The average weighted authority scores (Figure 3.17) show that all nurseries and retailers have similar scores across the first five scenarios. The relative significance of node subcategories change in scenario 7, where retailers have the highest authority scores. In scenario 6 (mainly N_{Nur}), we see a tenfold increase in all scores. The authority scores imply that for most scenarios, there is no preferred node subcategory to focus inspections for disease. In markets dominated by N_{Ret} nurseries, there is a slight preference to retailer nodes for inspections.

For the average unweighted hub scores (Figure 3.18), N_{Nur} nodes have the highest scores in all scenarios. In scenarios $1 - 6$ we see that all other nurseries have equal scores. In scenario 7, N_{Ret} and N_{Nur} nodes are the largest hubs in the network. In scenario 4 (mainly N_{Com}) and scenario 5 (mainly N_{Cons}), the lower quartile drops to 0 to intersect with the scores for other nurseries. Thus, when not considering edge weights, we see a change from N_{Ret} to N_{Nur} nurseries as the recommended node subcategory to target control measures.

For the average unweighted authority scores, we see that all nurseries have equally higher scores than retailers across all scenarios (Figure 3.19). This is the opposite result to the weighted authority scores. For scenario 4 and 5, the lower quartile for nursery scores decreases to 0, intersecting with the retailer nodes scores. The relative scores across node subcategories do not change as the scenario changes. We only see changes in the scores when considering a composition of nurseries consisting of mostly one type (scenarios 4 and

5). The unweighted hub and authority scores suggest that, when ignoring edge weights, the most significant pathway for spread is from N_{Nur} nodes to all nursery nodes.
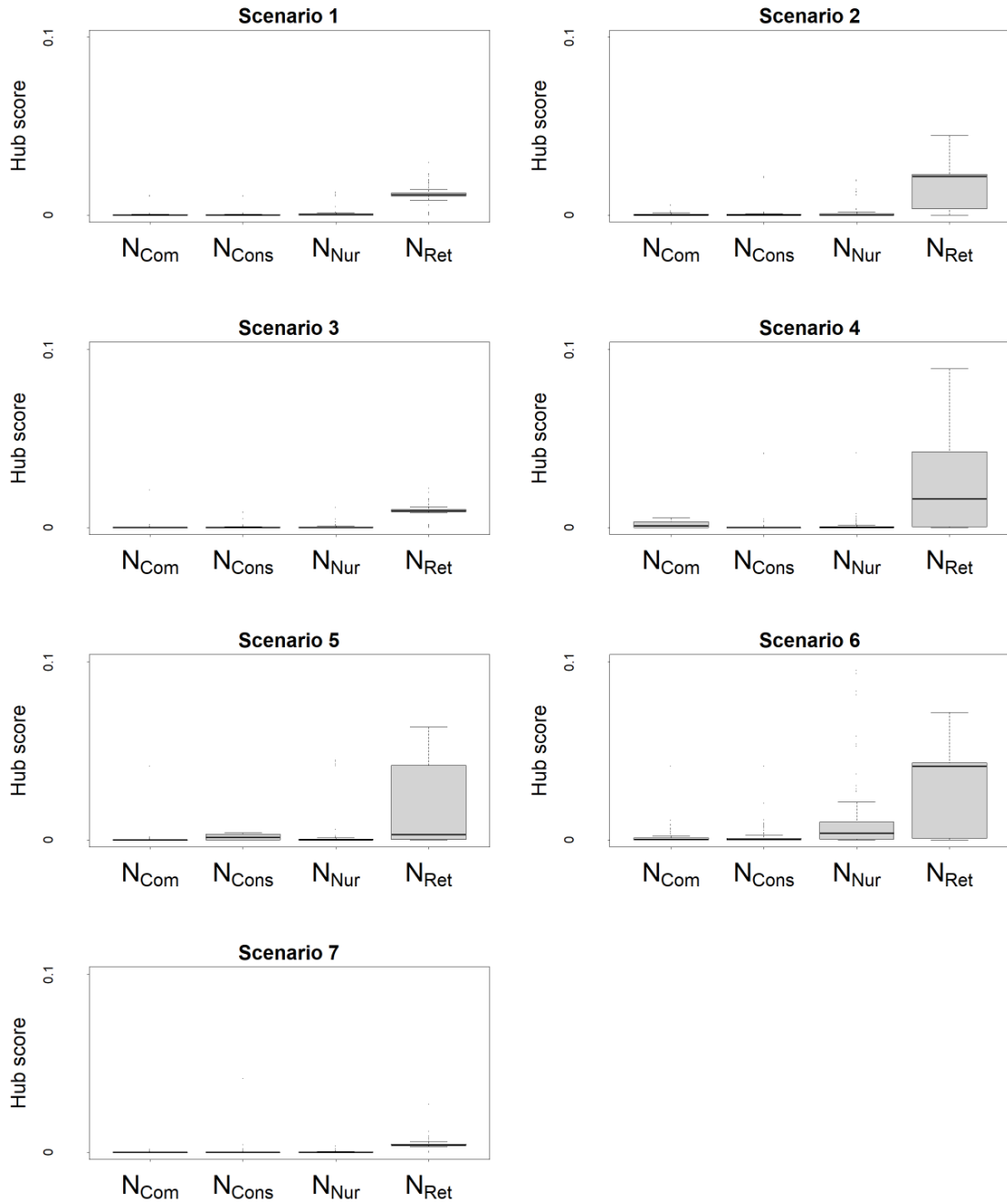
Figure 3.16. Average weighted hub scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$.
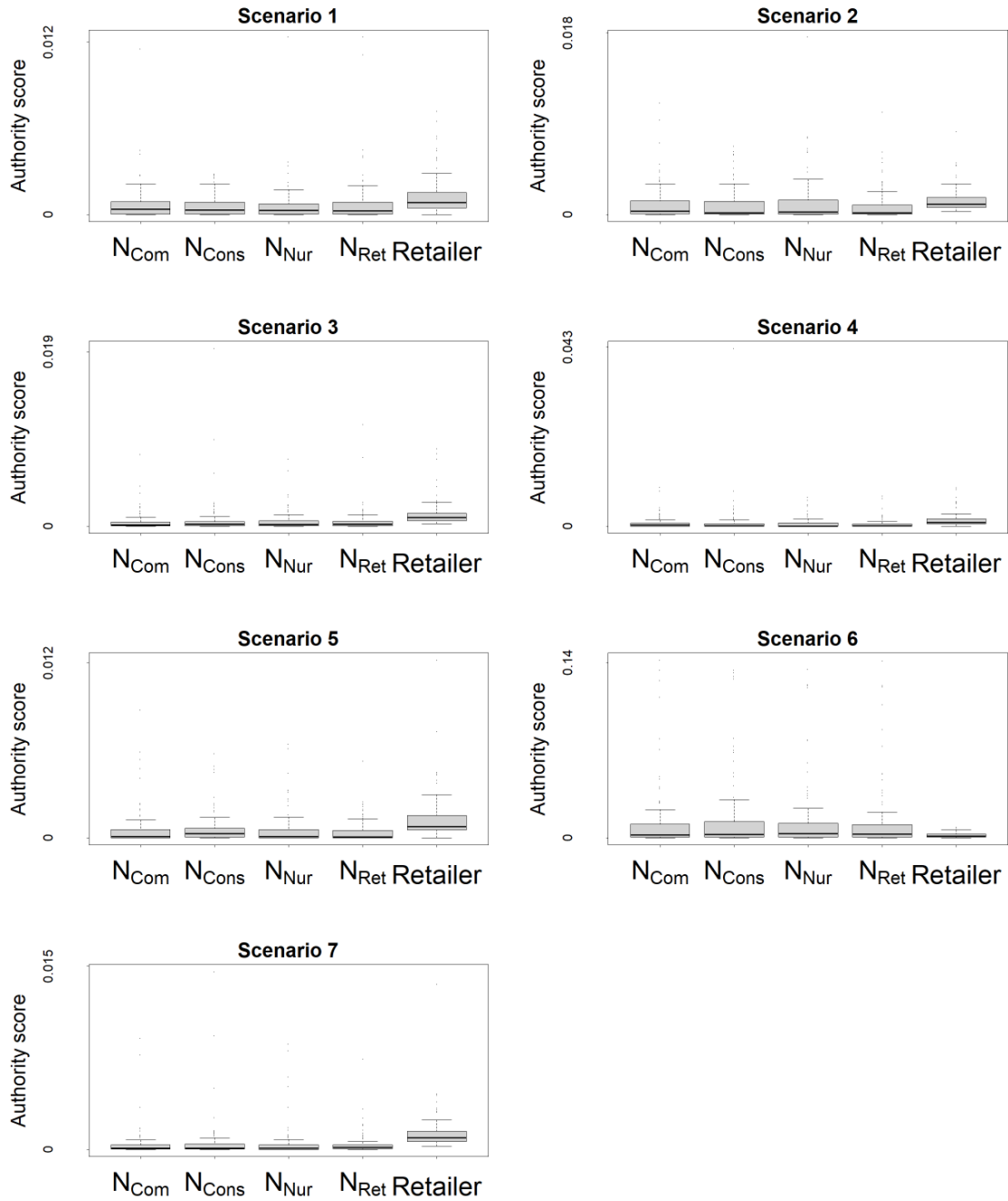
Figure 3.17. Average weighted authority scores shown across all 7 considered scenarios, with data transformed by $\log_{10}(1 + \text{data})$. Note the differences in the scale of the y-axis for different scenarios.
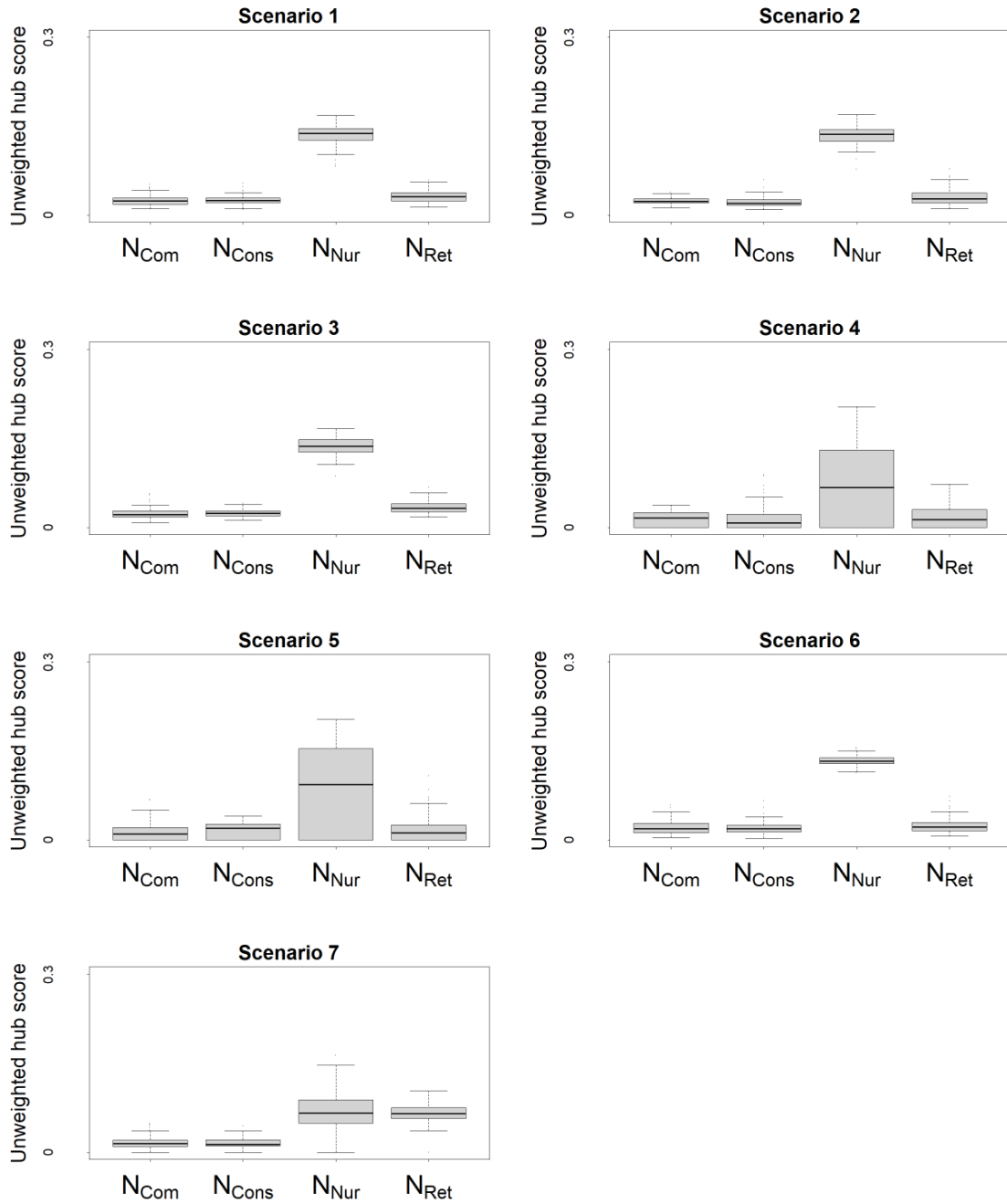
Figure 3.18. Average unweighted hub scores per nursery group with data transformed by $\log_{10}(1 + \text{data})$.
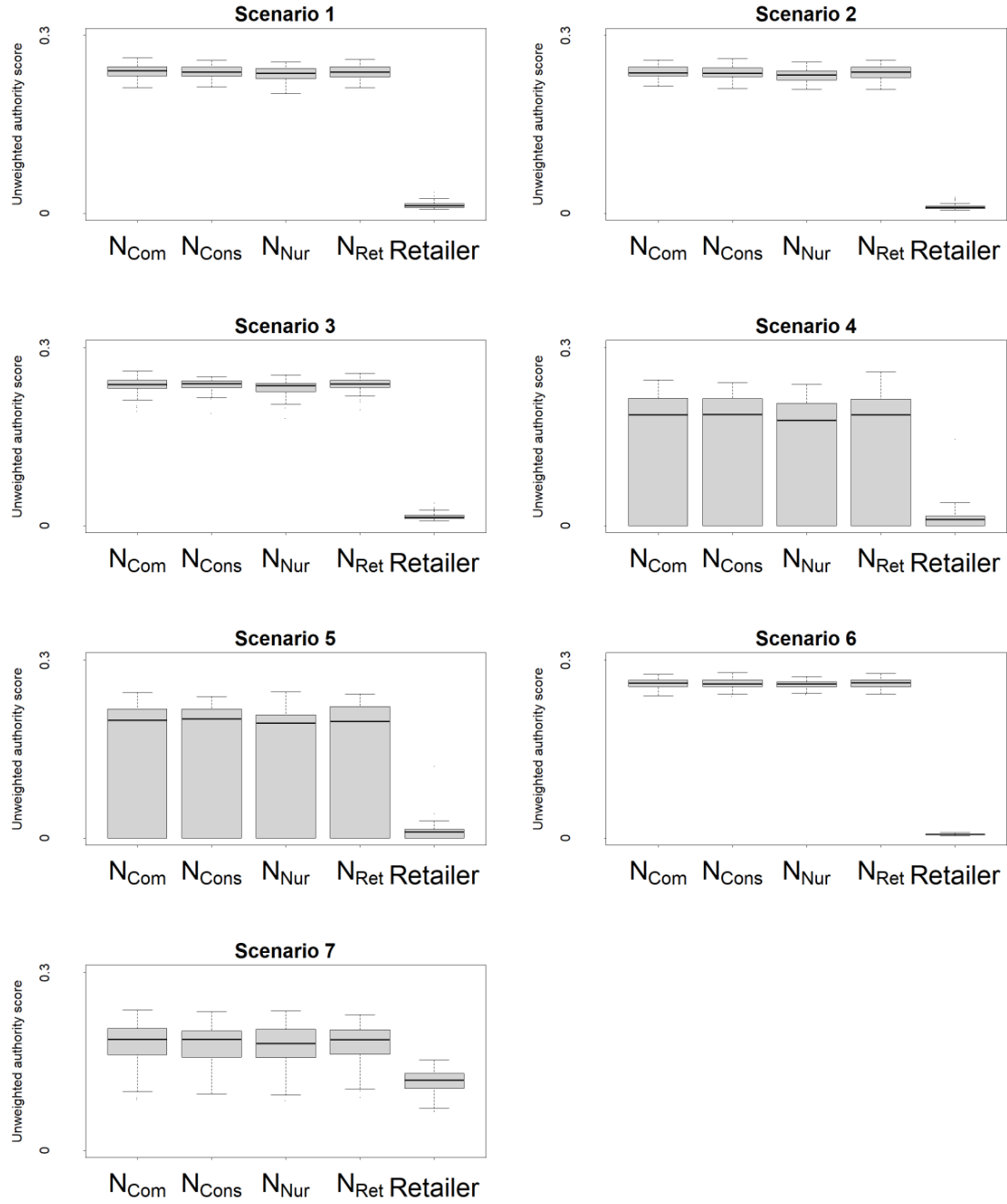
Figure 3.19. Average unweighted authority scores per nursery group with data transformed by $\log_{10}(1 + \mathrm{data})$.

This concludes our analysis of the distribution of nursery subcategories. In the next section, we vary the out-degree parameters and see how this affects network structure.

3.4.3 Out-degree mean and variance

Recalling that in our model, we have one distribution that governs the out-degrees for nurseries and retailers $\mathcal{K}_{\text{out}}(d_\mu^{out}, d_{\sigma^2}^{out})$. In this section we analyse the average centrality score per node subcategory for varying the out-degree mean and variance together. Our methods and the values we consider are as described in Section 3.3.4.

The average authority scores for each nursery node subcategory do not change significantly as d_μ^{out} and $d_{\sigma^2}^{out}$ increase (Figure 3.20). The scores are very small, so we take a $\log_{10}$ transform for clarity. The transformed authority scores for nurseries range from -3 to -1. Retailer nodes have the highest authority scores on average. Authority scores increase as d_μ^{out} increases for all nodes. Authority scores show no dependence on $d_{\sigma^2}^{out}$ for nurseries and retailers, for $d_\mu^{out} > 200$. For low out-degree mean values (250 and below), nurseries and retailers have similar authority scores. When out-degree mean increases from 500, retailer nodes scores become larger than nursery nodes.

The average hub scores for each node subcategory are displayed on a $\log_{10}$ scale (Figure 3.21), ranging between -4 and 0. Retailer nodes have consistently low, constant values and so are not displayed. We see that that scores are independent to changes in $d_{\sigma^2}^{out}$ for $d_\mu^{out} > 200$. Increases in hub scores are seen through increases in d_μ^{out} for all nodes. The node subcategory with the highest hub scores is N_{Ret}. We see the hub scores for N_{Ret} are also the most responsive to increases in the out-degree mean.

A combination of factors contributes to retailers having the largest authority scores: retailers consignment mean is the largest (Table 3.2) and there are more retailer than nursery nodes. Retailers having the largest authority scores implies that the node with the biggest hub scores is N_{Ret}, as seen in Figure 3.21. For larger out-degree means, N_{Ret} nodes become more highly connected to retailers which have higher authority scores. This results in N_{Ret} nodes getting larger hub scores.
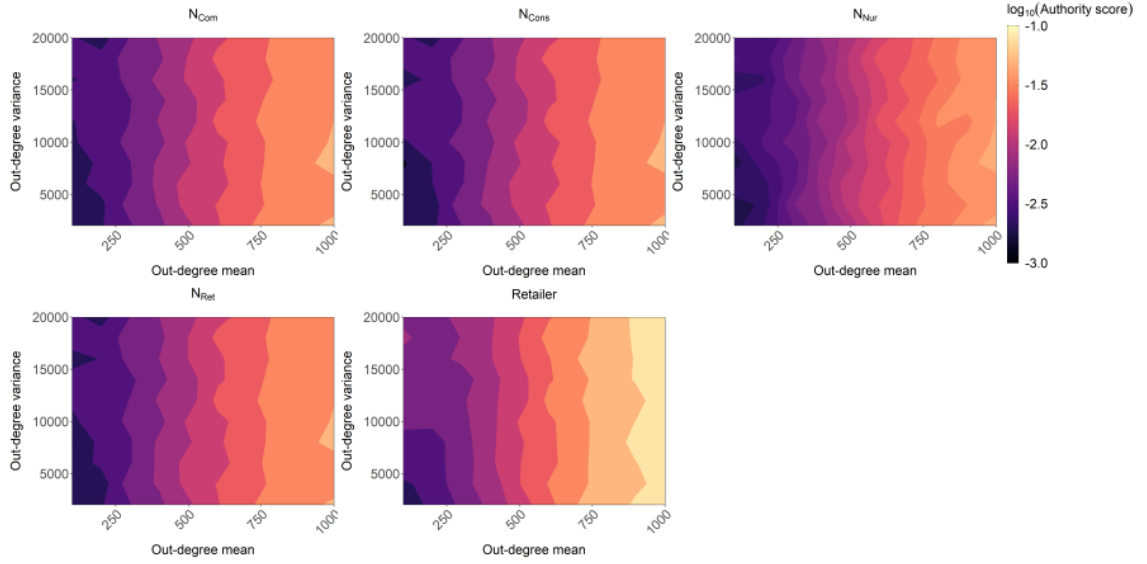
Figure 3.20. Average weighted authority score per node subcategory for 100 networks, computed for a grid of out-degree mean d_μ^{out} and variance $d_{\sigma^2}^{out}$ values. The values shown are on a $\log_{10}$ scale.
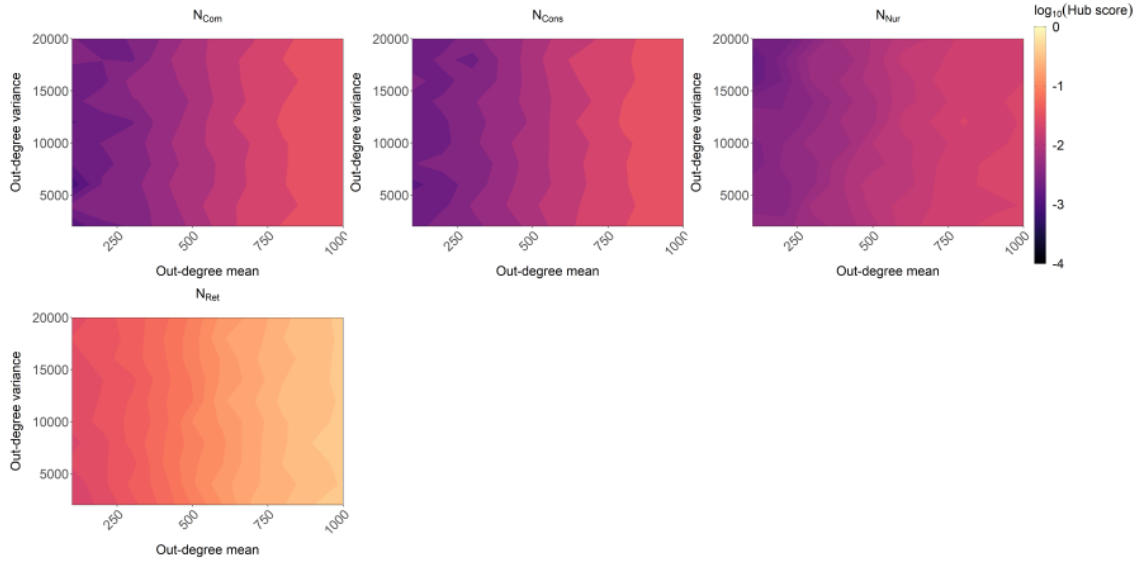
Figure 3.21. Average weighted hub score per node subcategory for 100 networks, computed for a grid of out-degree mean d_μ^{out} and variance $d_{\sigma^2}^{out}$ values. The values shown are on a $\log_{10}$ scale.

When considering the unweighted measures, we see a change in which node subcategories have the largest hub and authority scores (Figure 3.22 and Figure 3.23). The unweighted authority scores for all nursery subcategories show a similar nonlinear pattern, achieving increasing average authority scores either side of an out-degree mean of 500, reaching a peak score of 1. Increases in out-degree variance for nursery groups has a small, decreasing

effect on average authority scores for most values of out-degree mean. Average authority scores for retailers have the smallest values. These scores are independent to changes in out-degree variance and increase with out-degree mean. Due to there being 160 nurseries in the network, out-degree mean values from 300 result in almost all nurseries connected to each other and high authority scores.

Each nursery node's maximum out-degree is limited from above by the total number of nurseries minus one, divided by the proportion of sales directed to nurseries. The average proportion of edges assigned to nurseries in the network is 25%, therefore the average nursery's maximum out-degree is $159/0.25 = 636$. Hence for out-degree mean values greater than 700, in Figure 3.22 we see retailer nodes average authority score starts to increase, as nursery nodes are on average not becoming more connected whereas retailer nodes are.

For the unweighted hub scores in Figure 3.23, a nonlinear relationship emerges between out-degree mean and hub scores for the nursery subcategories. N_{Com} and N_{Cons} nodes receive low hub scores, which peak at an out-degree mean of 700 and decrease with deviations from that value. For all nursery groups, increases in out-degree variance have a weak decreasing effect on hub scores. N_{Nur} nodes achieve the largest scores, peaking at 1 when $d_\mu = 300$ and $d_{\sigma^2} = 2000$. Deviations from these values result in lower hub scores for N_{Nur} nodes. Unweighted hub scores for N_{Ret} nodes are independent to changes in out-degree variance for $d_\mu < 800$. N_{Ret} scores increase with the out-degree mean, getting the highest scores for $d_\mu \geq 700$.

The nonlinear relationship between the unweighted hub scores and d_μ for N_{Nur} nodes emerges as a direct result of the definition of nursery subcategories. As the out-degree mean increases, all nodes are more connected, and so N_{Nur} nodes become more connected to well-connected nodes. However, N_{Nur} nodes only have 159 other nurseries to assign links to and so by definition the least upper bound on their out-degrees is 320. Hence, out-degree means greater than 320 only increases the connectivity of other nursery subcategories and as a result N_{Nur} nodes become less connected by comparison, lowering their hub score.

The results from the hub and authority scores highlights that for larger and more dense networks, retailer nodes become increasingly the preferred node subcategory for inspections and N_{Ret} nodes most preferred for control measures. For the unweighted case, all nurseries remain the preferred choice for inspections as we vary network size and density. However, network size appears to have a nonlinear effect on the preferred node subcategory choice for control measures. For smaller networks, N_{Nur} nodes are the preferred choice, but as networks get larger, this is replaced by N_{Ret} nodes.
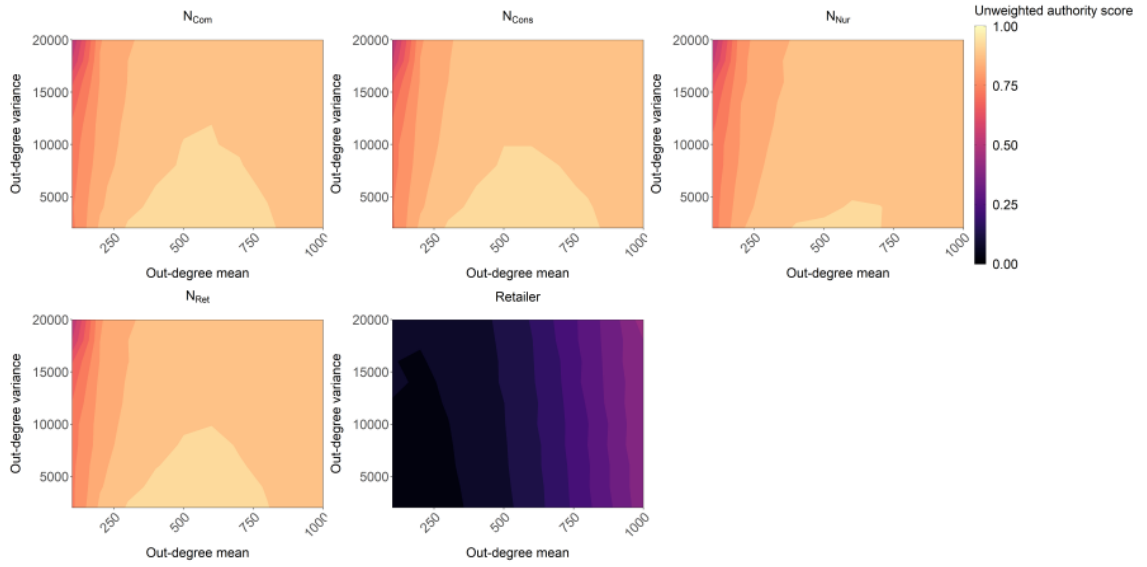
Figure 3.22. Average unweighted authority score per node subcategory for 100 networks, computed for a grid of out-degree mean d_{μ}^{out} and variance $d_{\sigma^2}^{out}$ values.
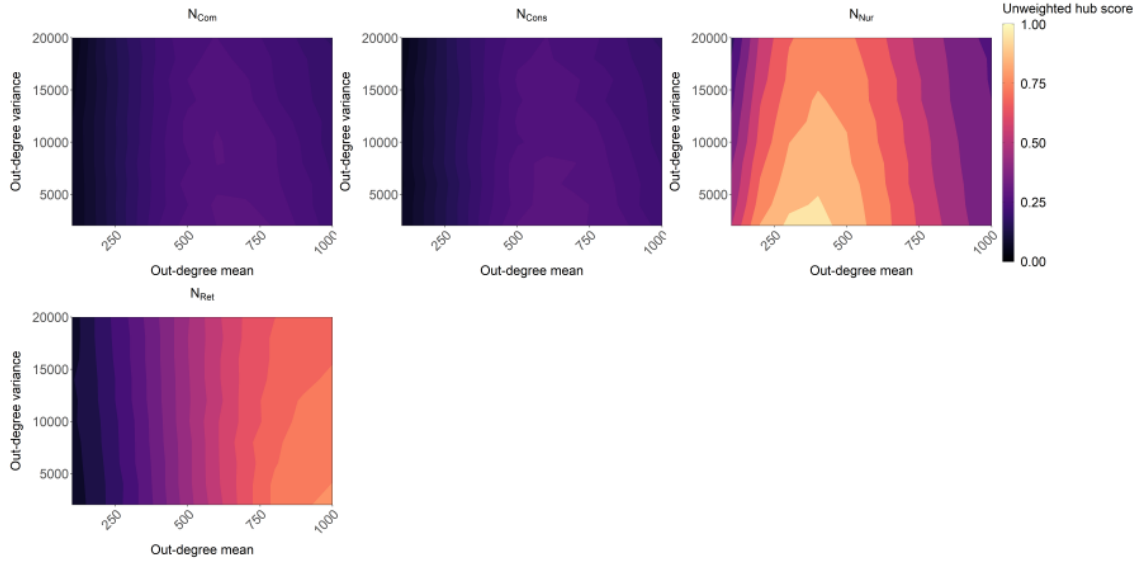
Figure 3.23. Average unweighted hub score per node subcategory for 100 networks, computed for a grid of out-degree mean d_μ^{out} and variance $d_{\sigma^2}^{out}$ values.

We show the average weighted betweenness scores for each node subcategory in Figure 3.24, on a $\log_{10}$ scale. We see independence to changes in $d_{\sigma^2}^{out}$ for each node subcategory. Increases in d_μ^{out} for all nodes results in increases of betweenness. The nodes with the largest betweenness scores are N_{Nur} and N_{Ret}. For $d_\mu > 700$, the scores for N_{Ret} nodes become the largest. Retailer nodes consistently have the smallest betweenness scores.
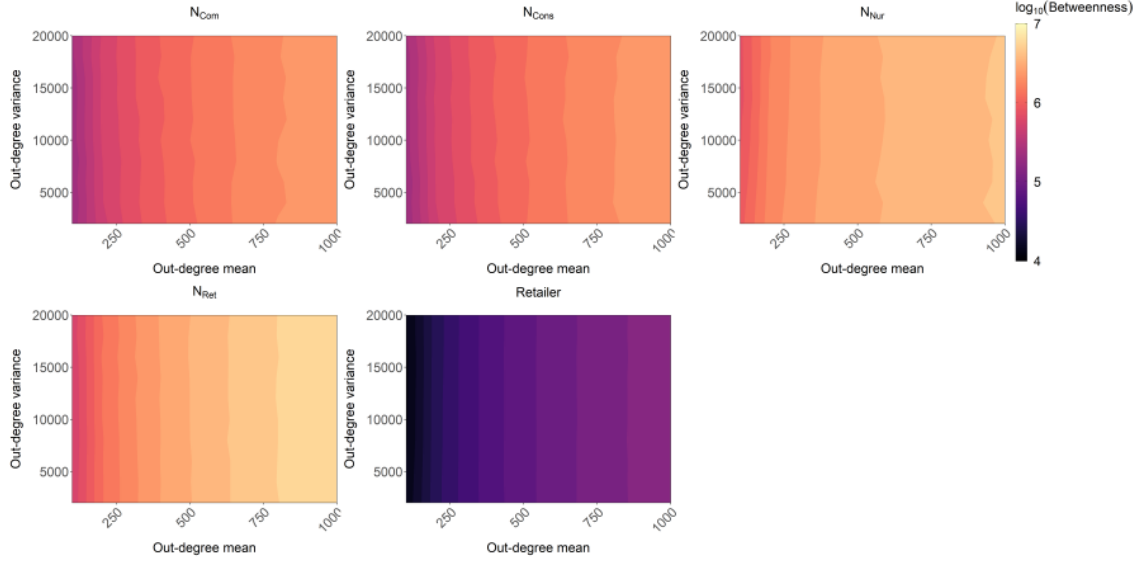
Figure 3.24. Average weighted betweenness score per node subcategory for 100 networks, computed for a grid of out-degree mean d_{μ}^{out} and variance $d_{\sigma^2}^{out}$ values. Results are shown on a $\log_{10}$ scale.

For the unweighted betweenness measure, we see the same pattern for N_{Com}, N_{Cons}, N_{Ret} and retailer nodes, with independence to out-degree variance (Figure 3.25). The scores for N_{Com} N_{Cons} and N_{Nur} are smaller than the weighted measure. For N_{Nur} nodes, the relationship between out-degree mean and unweighted betweenness is nonlinear, where scores peak at an out-degree mean of 200 to 300 and out-degree variance of 2000 to 4000, though this effect is small. This result again comes from the previously mentioned maximum out-degree of N_{Nur} nurseries. Most N_{Nur} nodes out-degrees do not change for $d_{\mu} > 320$, resulting in N_{Nur} nodes appearing in fewer of the shortest paths.
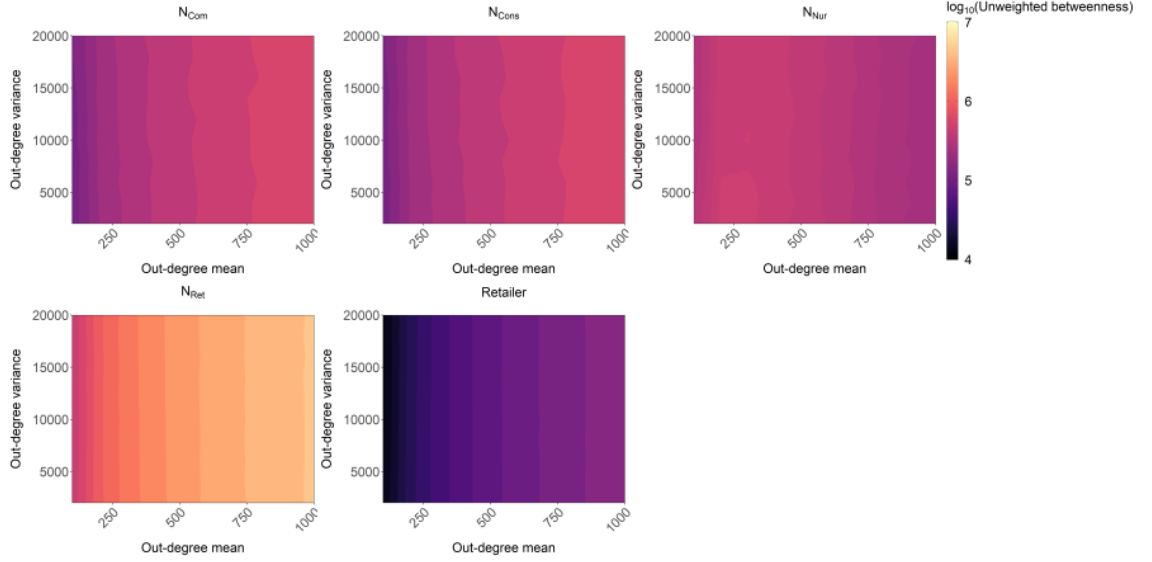
Figure 3.25. Average unweighted betweenness score per node subcategory for 100 networks, computed for a grid of out-degree mean d_μ^{out} and variance $d_{\sigma^2}^{out}$ values. Results are shown on a $\log_{10}$ scale.

Thus, nurseries remain the preferred choice to focus trade restrictions and quarantines compared to retailers as the network changes in size and density, regardless of edge weights. For a disease where the volume of trade is relevant, N_{Nur} and N_{Ret} nodes are the preferred choice to focus trade restrictions. This preference increases for larger and more dense networks. When trade volume is not relevant, for larger networks, there is a very strong preference to focusing trade restrictions at N_{Ret} nodes.

This concludes our analysis on the out-degree parameters. In the next section we analyse each pair of consignment size distribution parameters.

3.4.4 Customer consignment distributions

For each pair of consignment-size distribution parameters, we calculate centrality scores for each nursery and retailer group. Details of our methods can be found in Section 3.3.4.

We note that changing the commercial and consumer consignment size parameters has an insignificant effect on the centrality scores considered for nurseries and retailers, and so we do not investigate those parameters further.

We show the average authority scores for changes in nursery consignment size parameters, on a $\log_{10}$ scale (Figure 3.26). Scores for all nursery types increase when both the nursery consignment size mean and standard deviation increase. The reverse effect is seen for retailer nodes as authority scores increase for smaller nursery mean and standard deviations. Thus, as nursery nodes buy plants in larger quantities, they all become better candidates for inspections.

We also show the average hub scores on a $\log_{10}$ scale (Figure 3.27). Average hub scores for N_{Com} and N_{Cons} nodes are small and do not vary significantly for the nursery consignment size parameters. We do not show the scores for retailer nodes because they are very close to 0 and do not vary. For N_{Nur} nodes, the average hub scores increase for larger nursery consignment size mean and standard deviation values. The scores for N_{Ret} nodes decrease with increases in nursery consignment mean and standard deviation. As nursery consignment size mean and standard deviation increase, the volume of plants that nurseries receive increases, raising the hub score of nurseries. Because nurseries only receive links from other nurseries, this larger hub score affects the authority scores for all nurseries because these nodes are receiving larger volumes of trade.

Therefore Figure 3.27 and Figure 3.26 show that when nursery consignments are small, retailer and N_{Ret} nodes become preferred candidates for targeted inspections and control measures, respectively. When nursery consignments are large, inspections should be targeted at all nurseries and control measures at N_{Nur} and N_{Ret} nodes.
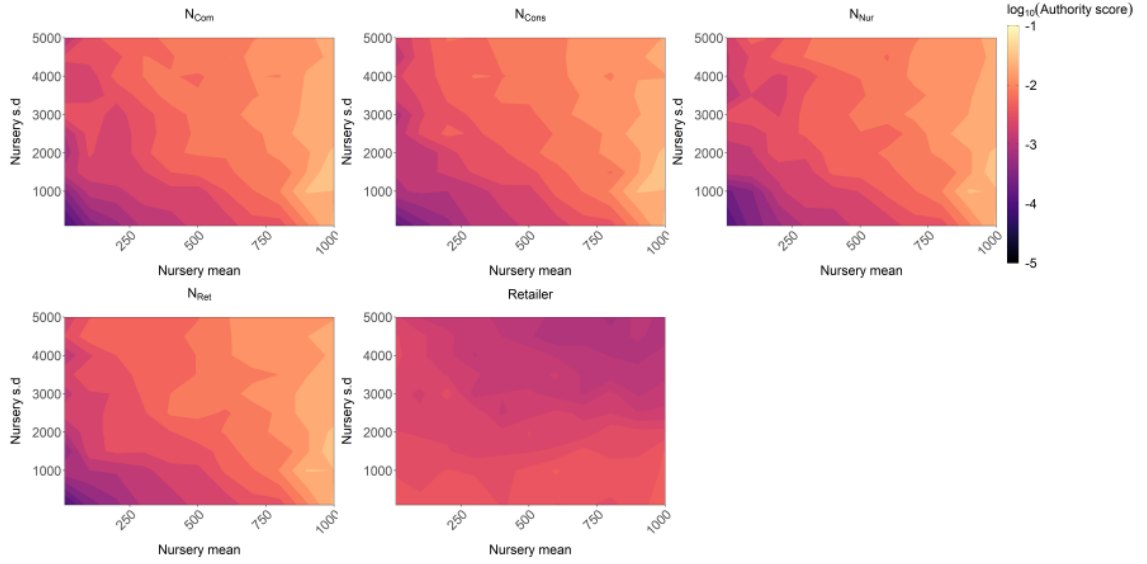
Figure 3.26. Changes to each node subcategories average weighted authority score in response to changes in the nursery consignment size mean and standard deviation parameters. Values are shown on a $\log_{10}$ scale.
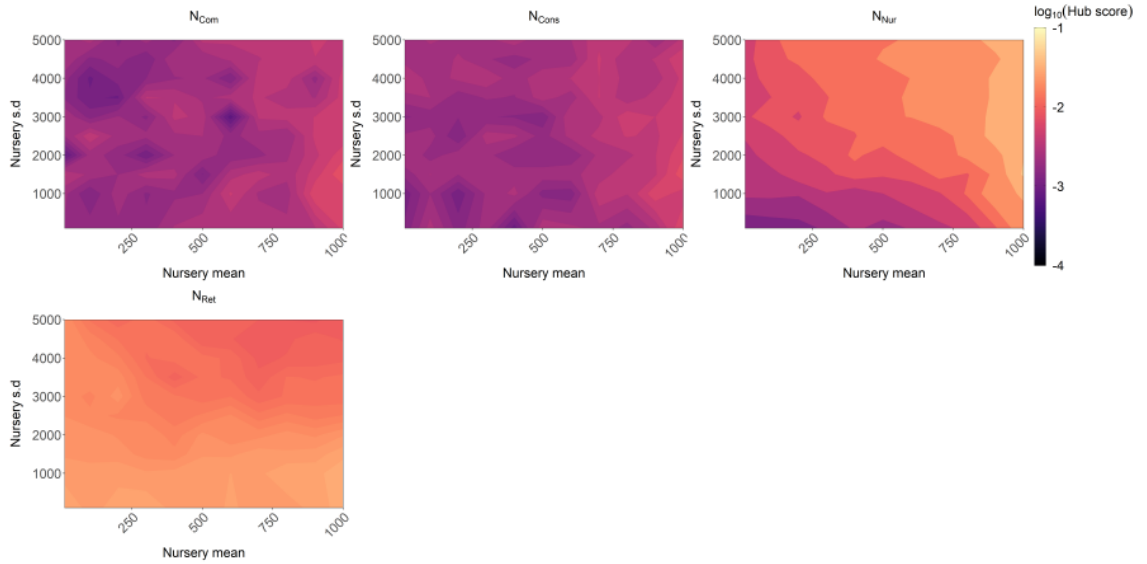
Figure 3.27. Changes to each node subcategories average weighted hub score in response to changes in the nursery consignment size mean and standard deviation parameters. Values are shown on a $\log_{10}$ scale.

The average betweenness scores (on a $\log_{10}$ scale) for all nursery types except N_{Nur} do not change significantly for the nursery consignment parameters considered (Figure 3.28). Retailer scores do not change and so we remove these from our figures for clarity. All nurseries average betweenness scores are dependent on the nursery consignment size mean for values

less than 300. As nursery consignment mean values increase above 300, the betweenness scores become more dependent on the nursery consignment size standard deviation. N_{Nur} nodes scores are the highest for most values considered. N_{Ret} nodes average betweenness scores do not vary much and are the second highest for all values we consider. For very low nursery consignment standard deviation and mean values above 200, N_{Ret} nodes become the node subcategory with the highest average betweenness in the network as opposed to N_{Nur} nodes.

Thus, for most values we consider, varying the consignment sizes of nurseries does not change which nursery subcategory to target trade restrictions. These results suggest to focus trade restrictions/quarantines on N_{Nur} and N_{Ret} nurseries when the volume of trade affects disease spread.
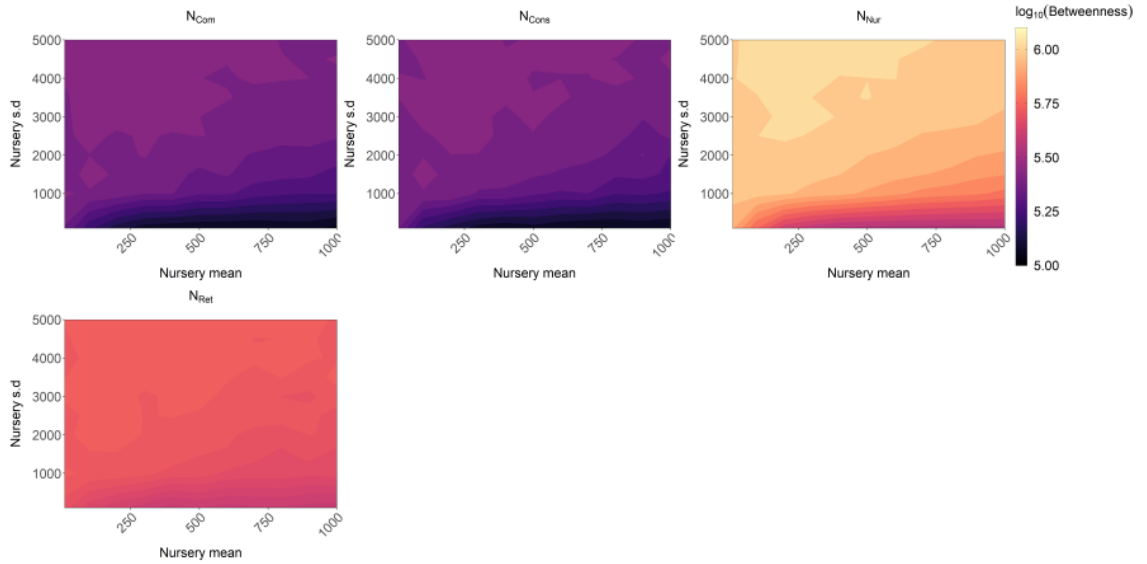
Figure 3.28. Changes to each node subcategories average weighted betweenness score in response to changes in the nursery consignment size mean and standard deviation parameters. Values are shown on a $\log_{10}$ scale.

Figure 3.29 shows the average authority scores for nurseries and retailers in response to changes in retailer consignment size parameters, on a $\log_{10}$ scale. Scores for nursery subcategories show a nonlinear dependence on retailer consignment size mean and standard deviation. For low retailer mean values (< 200), nursery scores are independent to retailer

consignment standard deviation. For small retailer standard deviation (< 1000) nursery scores become independent to changes in the retailer consignment size mean. Increases in both retailer consignment mean and standard deviation result in decreased nursery scores. As retailer mean increases, retailer scores increase and have a growing dependence on retailer standard deviation. For larger retailer consignment mean values, increases in the retailer standard deviation result in a decrease to all nursery and retailer authority scores. Scores for all nursery and retailer nodes become similar for retailer standard deviation greater than 2000. This also occurs for retailer consignment mean values close to 1000.
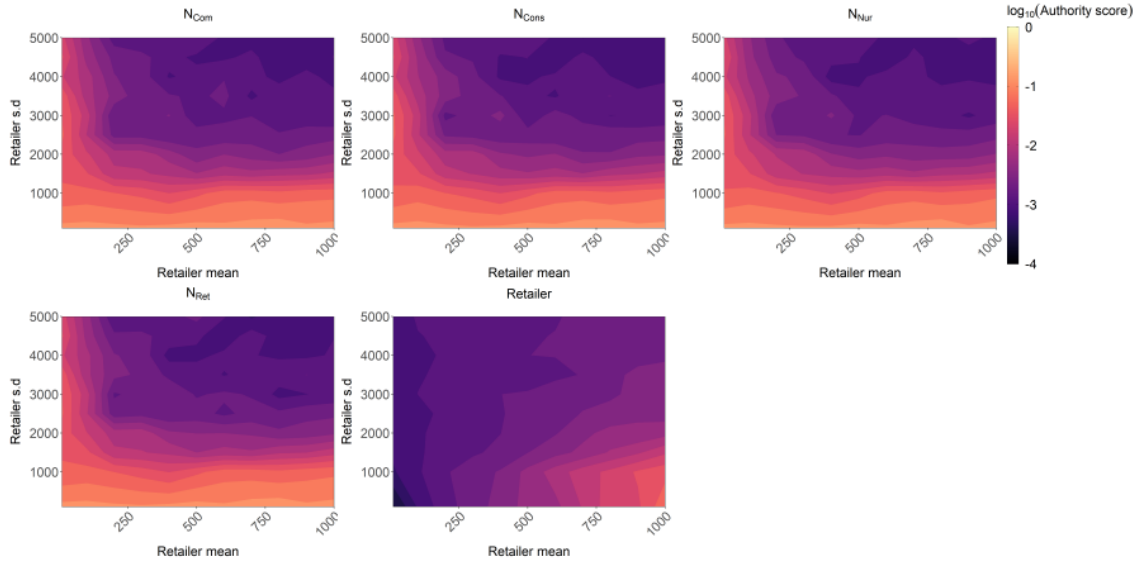
Figure 3.29. Changes to each node subcategories average weighted authority score in response to changes in the retailer consignment size mean and standard deviation parameters. Values are shown on a $\log_{10}$ scale.

Average hub scores for changes in retailer consignment size mean and standard deviation do not vary for retailer nodes and so are not displayed Figure 3.30. Average hub scores for N_{Com}, N_{Cons} and N_{Nur} show the same relationship with changes in retailer consignment size parameters as the average authority scores in Figure 3.29. N_{Com} and N_{Cons} nodes show lower average hub scores. Interestingly, N_{Ret} nodes hub scores show the same pattern as retailer nodes authority scores. When retailer consignment standard deviation is below 2000, N_{Nur} nodes are consistently the largest hubs in the network, even when the retailer consignment mean is greater than the nursery consignment mean of 500 (Table 3.2). For

large values of retailer mean and standard deviation, N_{Ret} nodes become the largest hubs in the network by a factor of 10. This result impacts which node type to focus control measures on.
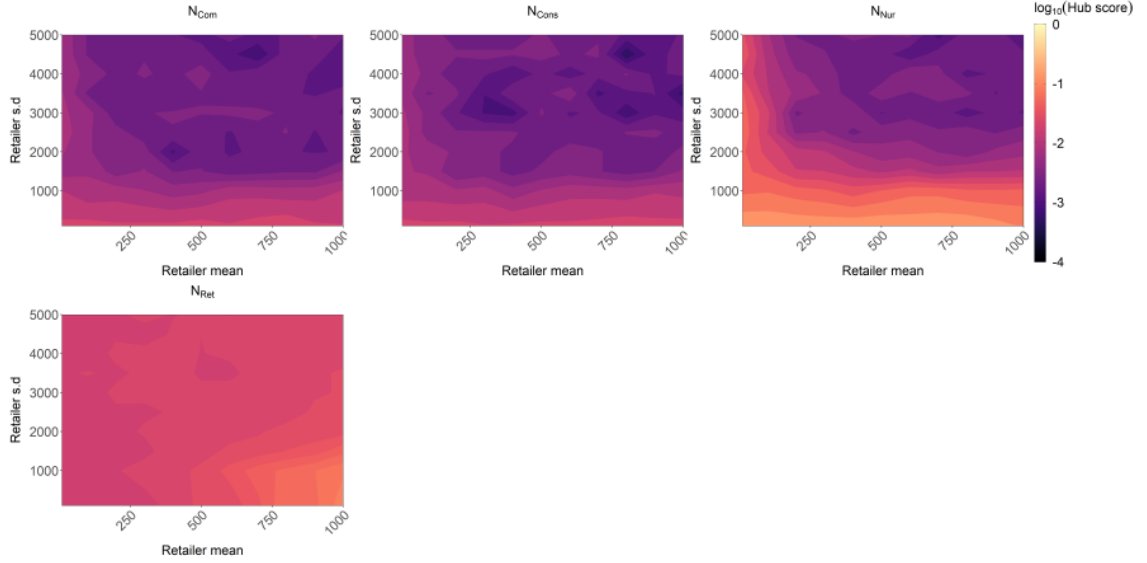
Figure 3.30. Changes to each node subcategories average weighted hub score in response to changes in the retailer consignment size mean and standard deviation parameters. Values are shown on a $\log_{10}$ scale.

We therefore see that the retailer consignment size parameters can effect whether retailers are good candidates for inspections and where to focus control measures. We find that if retailer consignments have standard deviation greater than 2000, and are not very small (greater than 200), then inspections should be at both nurseries and retailers, with control measures targeted at N_{Nur} and N_{Ret}. This is also the case when retailer consignments are on average very large. Otherwise, inspections at nurseries and control measures at N_{Ret} nodes are preferred.

Betweenness scores do not vary for nurseries or retailers with changes to the retailer consignment parameters (Figure 3.31). N_{Nur} nodes are best for targeted trade restrictions, followed closely by N_{Ret} nodes.
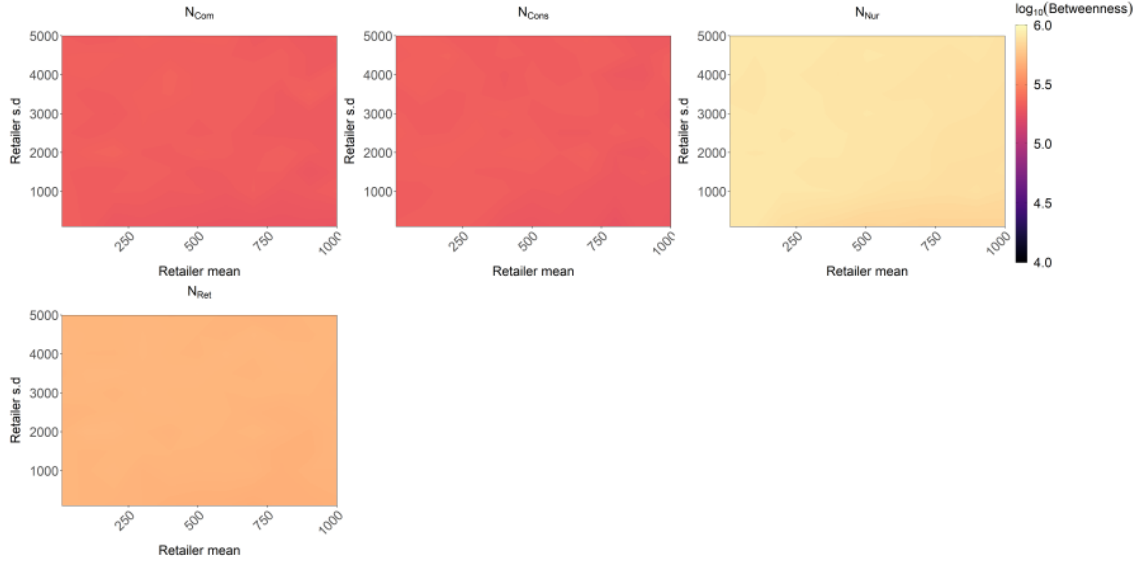
Figure 3.31. Changes to each node subcategories average weighted betweenness score in response to changes in the retailer consignment size mean and standard deviation parameters.

This concludes our sensitivity analysis. In the next section, we discuss our methodology and results. We summarise what we have learned and highlight limitations to our approach.

3.5 Discussion

Our algorithm yields a directed and weighted network representing annual Oak trade in the UK, where nurseries are characterised as specialising their sales towards one of four groups: commercial customers, consumers, nurseries and retailers. We assume that out-degrees follow a negative binomial distribution and we find that consignment sizes follow log-normal distributions which are different for nurseries, retailers and customers. The network topology closely resembles a tree, with dense trading amongst nurseries and retailers in the centre, branching out to customers. We analyse the network structure with several weighted and unweighted centrality measures and vary network construction parameters in this chapter.

We find in our analysis that edge weights affect which node subcategories are considered most central to the network structure. For the unweighted measures, N_{Nur} nurseries are the largest hubs, all nurseries are authorities and N_{Ret} nurseries have the highest betweenness. When weights are considered, N_{Ret} nurseries become hubs, retailers become authorities and N_{Nur} nurseries have the highest betweenness scores. Putting this in the context of disease control, this corresponds to different choices in where to focus control measures, inspections and trade restrictions/quarantines respectively. These results highlight the importance of modelling edge weights in trade networks because of the consequential changes in the epidemiological relevance of nodes. Thus, depending on the pest or disease at hand, it is important to determine if the volume of trade is relevant to spread as this will influence which nodes to focus our biosecurity efforts.

Betweenness scores for retailers are consistently lower than nurseries through all three sets of analysis, indicating that inter-nursery trading is more significant to disease transmission across the network. N_{Nur} and N_{Ret} nurseries consistently have the highest betweenness scores in all of our analysis. Relative betweenness scores do not change in response to changes in the out-degree parameters. Though we see in some extreme cases that nursery consignment sizes can affect the relative betweenness scores of node subcategories, the consignment size parameters mostly do not have an effect. For the unweighted betweenness scores, this only changes the node subcategory with the highest score from N_{Nur} to N_{Ret}. Thus, in a world of finite resources, targeting N_{Nur} and N_{Ret} nurseries with trade restrictions/quarantines would be most effective at reducing spread during an outbreak. This result holds throughout all of our analysis.

The network structure is robust to changes in the distribution of nursery subcategories. Only in cases where nurseries consist predominantly of one subcategory do we see changes in the relative significance of node subcategories for centrality measures. This implies that our results may generalise to significant market-changing events such as the UK's exit from the EU.

The relative importance of node subcategories remains consistent as the network becomes larger and more dense, as shown in our analysis of the out-degree mean and variance parameters. Thus, for larger networks with more annual trading, our results will remain the same. The only exception is that, when the volume of trade is irrelevant, N_{Ret} replace N_{Nur} nurseries as the most suitable target for control measures.

Analysis of the nursery consignment size parameters reveals that in networks where nurseries buy more plants, all nurseries become more likely to receive infection and N_{Nur} nodes have greater potential for larger outbreaks.

The size and variation of retailer consignment sizes affect the most significant pathways for spread in the model. We identify regions of retailer consignment size parameters where the primary pathway for spread can change from between N_{Ret} and retailers, to N_{Nur} and all other nurseries. There is also a small region for high retailer consignments with low variation where these pathways are equally significant. Therefore, data on retailer consignments are important to collect for the model, with consequences to disease control strategies.

The relative importance of node subcategories for most centrality scores does not change when removing customers from the network. The only exception is for the out-degree. When customers are not considered, N_{Nur} and N_{Ret} nurseries are most central in terms of out-degree. Whereas in the full network, N_{Nur} nurseries are least central, with retailers and all other nurseries equally more central. This is an interesting result considering that no previous model trade network includes customers. The relative out-degree of the node subcategories in the network without customers more closely agrees with the relative significance seen from the other centrality measures on the entire network.

Though this network model is constructed with the spread of disease in mind, this network would be useful in other applications such as bio-economic modelling. The inclusion of weights into the edges of the network construction also makes this model useful for modelling the distribution of produce on a nation-wide scale.

Remark: In the calculations for the consignment sizes, the minimum consignment size was taken to be 3 instead of 1 for all consignment size distributions. This was due to a coding error when taking the exponential of a truncated log-normal distribution with the lower truncation set to 1 instead of 0. We do not expect this to change the results given that the analysis focuses on general trends and patterns, looking at the relative significance of node subcategories across several centrality measures. The consignment size means for nurseries and retailers were large (Table 3.2) and so the change of a minimum value from 3 to 1 will have an insignificant effect on the analysis in this chapter. The commercial and consumer consignment size means were the smallest, with a value of approximately 6. The change in minimum value from 3 to 1 would therefore have a significant effect on these distributions. However, commercial and consumer nodes were found to contribute insignificantly to network structure, and thus lowering the minimum consignment size for these nodes would not change our results.

3.5.1 Limitations

Our network is aggregated to represent annual plant sales and is assumed to be static. This has limitations to analysing network structure as paths in an aggregated network may not exist in a network which changes over time [110]. Though, as plant trade is seasonal, aggregating trade for over a year captures the entire trade cycle and thus the volume of plants traded and the number of customers should be appropriate.

It is important to note that the sales information given in the data only gives the billing addresses rather than the shipping address and so this leads to, inevitably, some inaccuracy in calculations of typography.

Another limitation of the data set is that it is practically impossible to know for certain the source location of a plant. In the sales data, there may be *rebadging*: the process of a plant having its country of origin relabelled when moving from one country to another [16]. This uncertainty in the data induces limitations when trying to incorporate international trade with domestic trade and especially for incorporating the risks of different plant suppliers as there may be hidden, compounded levels of risk.

In our analysis, we focus on identifying general patterns in changes to network structure. Hence, when analysing the effect of changes in model parameters, we average centrality scores for each node subcategory and repeat this for 100 network simulations. There are limitations to this approach. By taking averages of centrality scores per node subcategory in each network construction, the information regarding the distribution of these values per network is lost. With access to more data, the analysis of the network structure could then rely on centrality scores of individual nodes to be more representative of reality.

In order to simplify the model and focus the analysis, all nurseries and retailers out-degrees are modelled with the same negative binomial distribution. This may be limiting in the analysis of the out-degree parameters. However, there are some differences in the out-degree distributions as each node has a different customer demographic, which imposes distinct upper bounds for each nodes out-degree. This results in some variation of the out-degree distributions across node subcategories. Access to more data would enable the parameterisation of out-degree distributions for each node subcategory, allowing a more detailed analysis.

3.6 Conclusions

In this chapter we apply the method of network construction described in Chapter 2 to the UK's domestic plant trade network using a combination of existing public data and sales data provided by FERA. This is to our knowledge the first model network for domestic

plant trade which considers the heterogeneous trade of plants and also includes customers as nodes in the network.

The results from our analysis highlight that edge weights have a significant impact on network structure and nodes of epidemiological interest. As networks become larger and more connected, all nurseries and retailers become more infectious and prone to infection. Changes in the distribution of nursery subcategories, a proxy to structural changes in the market, only significantly affect network structure in extreme cases, suggesting a degree of structural robustness. The consignment sizes of retailers significantly affect the pathways of transmission across the network, calling for the collection of data in this sector of trade. Customers do not contribute significantly to the structure of the network, thus supporting the exclusion of customers in previous network models of plant trade.

In the next chapter, using this network, we explicitly model the spread of pests and diseases throughout the UK's domestic plant trade in a compartmental metapopulation framework.

Chapter 4

Epidemiological Model on a Static Network

4.1 Introduction

In this chapter we use the trade network model from Chapter 3 to simulate the spread of a generic plant pest/disease in a compartmental, metapopulation framework. Our aim for this chapter is to understand how pests and diseases spread throughout the network and to investigate the effects of different inspection implementations.

We begin by describing the modelling approach for plant movement across the network over time, identifying all potential sources and sinks of the system. Features to introduce to our model include disease compartments, inter- and intra- node disease spread dynamics and death rates. Control measures take two forms in our model: inspections that occur during trade and scheduled inspections of all nursery and retailer node stock. From our model equations we calculate the sources and sinks to maintain a constant population size for nurseries and retailers.

Our analysis focuses on disease spread amongst nurseries, first simulating the model with no control measures and varying the disease transmission and death rate parameters. We conduct a sensitivity analysis on the inspection parameters and then end our analysis by

comparing the costs and benefits of different frequencies of scheduled inspections. In our sensitivity analysis, we consider scenarios 1 to 3 of nursery distributions from Chapter 3.

The results from our analysis show that a pest or disease starting in a N_{Nur} nursery, i.e., a nursery that mostly sells to other nurseries, greatly increases the probability of an outbreak and is far more difficult to control with inspections. Large outbreaks occur if the transmission rate is at least 2, which roughly translates to the disease tripling every month. From a single imported infection, disease can persist at a low prevalence for between 1 and 3 years, implying a great difficulty at tracing the initial infection event. Similar to our results in Chapter 3, the results are robust to small changes in the market (distributions of nurseries).

Our analysis continues, identifying conditions for when inspecting purchases in addition to sales is effective. The two types of inspection we consider are close substitutes; however, increasing the frequency of scheduled inspections can render trade inspections unnecessary. Finally, we identify when it is economical to increase the frequency of scheduled inspections, depending on the price of inspection relative to the cost of the plant.

4.2   Network model as a metapopulation of plants

In this section we adapt our network from Chapter 3 to a compartmental, metapopulation framework. We first identify the sources, sinks and movements of plants in this system. Next, we define the timescale and time-step in our model, introduce metapopulation dynamics and then the compartmental part of the model which describes the disease spread dynamics.

We use the same network from Chapter 3 which represents the annual trade of Oak plants in the UK. In this chapter, we represent the trade network as a weighted $M \times M$ adjacency matrix $\boldsymbol{A}$, where M is the number of nodes in the network. Each element of $\boldsymbol{A}$, $A_{i,j} \in \mathbb{R}^+$ for $i, j \in \{1, 2, ..., M\}$, represents the number of plants sold from node i to node j. The total number of plants a node sells (out-strength) is $\sum_{k=1}^{M} A_{i,k}$; similarly, the total number

of plants a node buys (in-strength) is $\sum_{k=1}^{M} A_{k,i}$. These represent the total trade output and input, respectively.

In our model, the network does not change over time and thus we constrain the timescale of our analysis to the short term (3 years). Time, t, is measured in months, with step size $\Delta t = 1$. We make a simplifying assumption to evenly distribute the annual volume of trade across each month. Thus, we divide the entries of the adjacency matrix $\boldsymbol{A}$ by 12.

We model the movement of individual plants, with nodes in the network acting as sub-populations. Recall from Chapter 3 that the network consists of nurseries (N_{Com}, N_{Cons}, N_{Ret} and N_{Nur}), retailers and customers (commercial and consumer). Figure 4.1 summarises how plants move throughout this network, depicting the metapopulation structure by node groups of nurseries, retailers and customers ($\boldsymbol{N}$, $\boldsymbol{R}$, $\boldsymbol{C}$). The possible sources of plants into the network are imports to nurseries and retailers, and plants grown by nurseries. Plants move throughout the network via inter-nursery trade and via sales from nurseries to retailers and customers or from retailers to customers. The sinks we consider in the system are exports overseas from nurseries and retailers as well as a removal term for nurseries, retailers and customers. We assume for nurseries and retailers that plants are thrown away when they are no longer in a saleable condition and this occurs before the plant dies. Therefore, throughout this chapter, we refer to this form of removal as *plant death*. Additionally, removal for nurseries and retailers includes *control measures*. Similar to *Bate et al.*, we consider control measures in the form of restriction and removal [76]. Restriction takes the form of inspections of consignments (from seller and/or buyer) which we refer to as *trade inspections*. We model removal from *scheduled inspections*, where for nurseries and retailers, all stock on the premises is periodically inspected for disease. We assume that infected plants caught in inspections are removed and we do not consider false positive outcomes from inspections. For customers, the removal term includes the death of both healthy and infected plants.
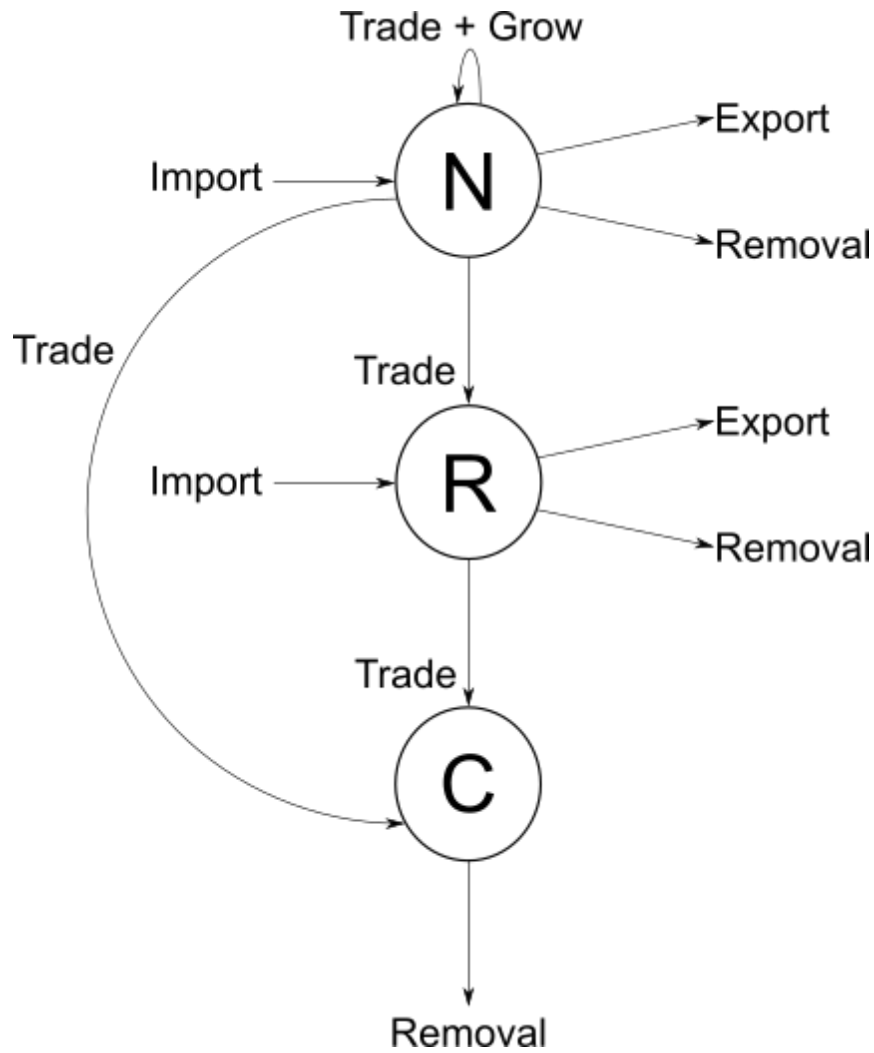
Figure 4.1. Schematic representation of how plants move throughout the network. The groups $\boldsymbol{N}$, $\boldsymbol{R}$ and $\boldsymbol{C}$ each represent the nursery, retailer and customer nodes respectively. Plants enter the network (sources) via imports from overseas to nurseries and retailers, and nurseries which grow some of their own plants. Nurseries trade amongst themselves, retailers and customers, with retailers trading to customers. Plants leave the network (sinks) via exports overseas and a removal term, which describes plant death and infected plants removed via inspection.

Before considering the spread of disease, we will first describe how the number of plants, N_i, at each node i in the network is modelled over time. The total number of plants that a customer has changes over time. For time t and customer $i \in \boldsymbol{C}$, this is denoted by $N_i(t)$. For simplicity we do not model the number of plants each customer already has, i.e., for

$i \in \boldsymbol{C}$, $N_i(0) = 0$. Instead, we model the number of plants that each customer buys and for $t > 0$ we apply a death rate, d, to all plants they have previously bought up to that point $(N_i(t))$. Hence, we have

$$N_i(t+1) = N_i(t) - dN_i(t) + \sum_{k=1}^{M} A_{k,i}, \qquad \text{for } i \in \boldsymbol{C}$$

For nursery and retailer nodes ($i \in \boldsymbol{N} \cup \boldsymbol{R}$), we assume that the total number of plants on the premises is fixed over time, i.e., $N_i(t) = N_i$, for all $t \geq 0$. We assume this to avoid an unrealistic accumulation or depletion of plants over the course of simulations. Therefore, we define the *source*, $G_i(t) \geq 0$, and *sink*, $E_i(t) \geq 0$, terms to keep the total number of plants at each nursery and retailer constant. Each nursery and retailer node at every time step throw away a proportion of their plants that have died. Nurseries and retailers also gain the plants they buy from other nurseries and lose the plants they sell. Hence, we have

$$N_i = N_i - dN_i + \sum_{k=1}^{M} A_{k,i} - \sum_{k=1}^{M} A_{i,k} + G_i(t) - E_i(t), \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R}$$

where

$$G_i(t), E_i(t) \geq 0 \qquad \forall t \geq 0.$$

We have now described how plants are traded across the network and introduced the role of sources and sinks in the model. In the next section we will introduce disease dynamics, to place this trade model into a Susceptible-Infected compartmental meta-population framework.

4.3 Susceptible-Infected compartmental dynamics

In this section we detail the equations for the spread of a generic pest/disease following a Susceptible-Infected compartmental meta-population model. We model the spread of a generic pest/disease on a trade network over time. Other epidemic models have relied on a generic probability of transmission between connected nodes, on homogeneously weighted networks, constructed with an assumed degree distribution and assumed class of nodes [70, 106]. Weights on the edges of a network affect the probability of transmission and introduce a heterogeneity that will influence the dynamics of disease spread throughout the network, resulting in different recommended control measures. Our network model allows us to incorporate heterogeneous weights into the probability of transmission.

Figure 4.2 illustrates the forces acting on the susceptible and infected plants in a nursery. Nurseries can import plants from overseas and buy plants from other nurseries, where there is a risk of receiving infected plants. Nurseries can grow their own plants, which are assumed to move only into the susceptible compartment. Disease then spreads via a force of infection between the susceptible and infected compartments. Both susceptible and infected plants can leave the nursery through multiple sinks: sales overseas (exports), sales to other nodes and death. Additionally, infected plants can be removed via scheduled inspections or inspections through trade. For retailers, the process is the same except we assume they do not grow their own plants. We do not consider sources of disease from the natural environment.
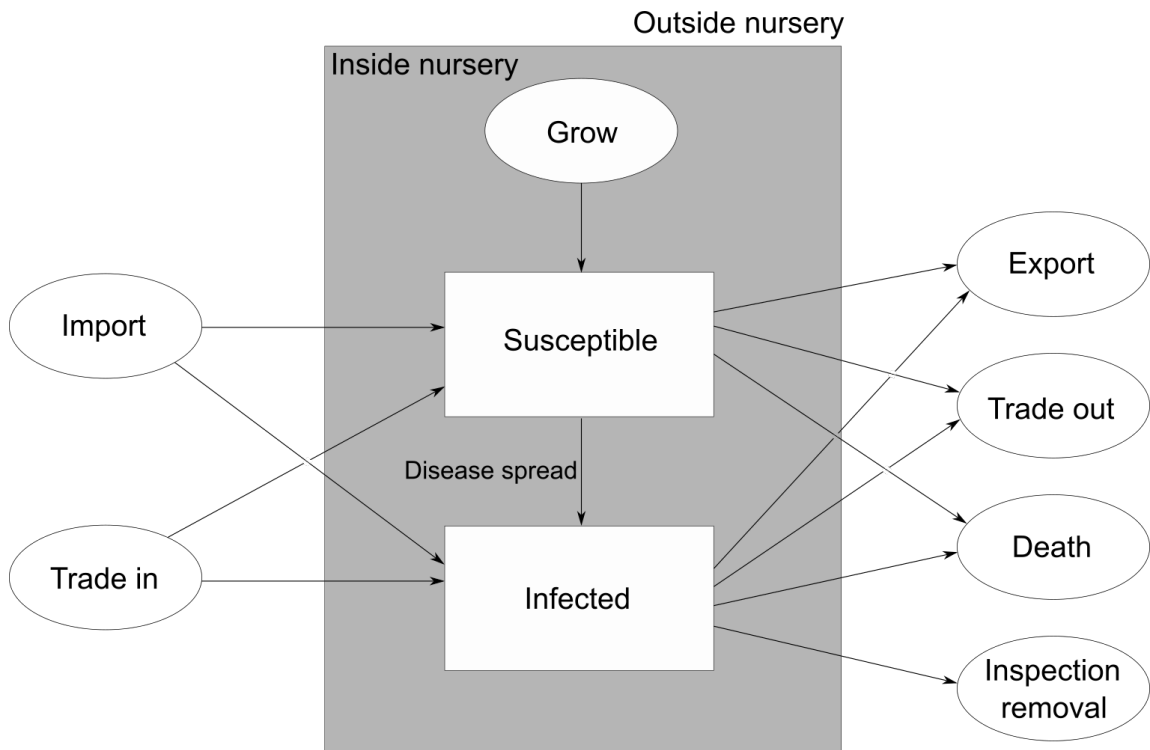
Figure 4.2. Representation of how the infection process is modelled for an individual nursery in the Susceptible-Infected compartmental system. Plants that the nursery grows are assumed to only move into the susceptible compartment. Plants also enter a nursery through imports and domestic trade, each carrying the potential to bring in infection. Disease spreads from within the nursery between the infected and susceptible compartments. Both susceptible and infected plants can leave the nursery through exports, domestic trade and death. Infected plants also leave the network through inspections which occur when preparing consignments and regular scheduled inspections. For retailers, this process is the same except they can not grow their own plants.

For customers, we do not model the spread of disease given the difficulty of assuming how many other plants a customer has and how these plants are stored in relation to others. Instead, we will count how many infected plants are received. This will serve as an indicator to the level of further spread the disease/pest has outside the trade network.

Keeping the flow of trade in the network fixed over time requires addressing the imbalances that lie in the total number of plants sold and bought for each trader in the network. If a trader buys more plants than they sell, then iterating trade over time will result in unbounded accumulation of plants for that trader, or, in the opposite situation, it will result in the trader selling more plants than they stock. These imbalances can be accounted for

by incorporating imports from overseas and own-grown plants as sources and considering exports overseas as sinks in the system. In the model, we will select $G_i(t)$ and $E_i(t)$ so that these imbalances do not arise.

4.4 Model development

In this section we develop the model equations for a generic pest or disease spreading across the network in a Susceptible-Infected metapopulation framework. The number of *susceptible* and *infected* plants in subpopulation i at time t are denoted $S_i(t)$ and $I_i(t)$, respectively. We first present the equations for each nursery and retailer node. For clarity, we show each component of the model in words in Equation 4.1 and Equation 4.2. Therefore, for all $i \in \boldsymbol{N} \cup \boldsymbol{R}$, we have

$$
\begin{aligned}
S_i(t+1) =\ & S_i(t) - \big(\textbf{Susceptible death}\big) \\
& + \big(\textbf{Total trade input} - \textbf{Infected trade input}\big) \\
& - \big(\textbf{Total trade output} - \textbf{Infected trade output}\big) \\
& + \big(\textbf{Total grown/imported} - \textbf{Infected imported}\big) \\
& - \big(\textbf{Total exported} - \textbf{Infected exported}\big) \\
& - \big(\textbf{Internal disease spread}\big),
\end{aligned}
\tag{4.1}
$$

$$
\begin{aligned}
I_i(t+1) =\ & I_i(t) - \big(\textbf{Infected death}\big) \\
& + \big(\textbf{Infected trade input}\big) \\
& - \big(\textbf{Infected trade output}\big) \\
& + \big(\textbf{Infected imported}\big) \\
& - \big(\textbf{Infected exported}\big) \\
& + \big(\textbf{Internal disease spread}\big) \\
& - \big(\textbf{Infected removal}\big).
\end{aligned}
\tag{4.2}
$$

We now consider the equations for commercial and consumer nodes $i \in \boldsymbol{C}$. Recall that commercial and consumer nodes have a time-dependent population size $N_i(t)$, where $S_i(t) + I_i(t) = N_i(t)$ and $N_i(0) = 0$. For customers, we do not model disease spread, but only the number of infected plants received at time t. Therefore, the equations for customers are much simpler and for all $i \in \boldsymbol{C}$ we have

$$S_i(t+1) = S_i(t) - \left(\textbf{Susceptible death}\right)$$
$$+ \left(\textbf{Total trade input} - \textbf{Infected trade input}\right), \tag{4.3}$$
$$I_i(t+1) = I_i(t) - \left(\textbf{Infected death}\right)$$
$$+ \left(\textbf{Infected trade input}\right) \tag{4.4}$$

Now our system of equations is described in Equation 4.1 to Equation 4.4. In subsequent sections, we detail how each aspect of these equations are developed, highlight the assumptions that we make and then present the equations fully in their mathematical form (Equations 4.16-4.19).

4.4.1 Internal disease spread

We model disease spread within nursery and retailer nodes with a force of infection. For this model we assume that the force of infection is frequency dependent and therefore we have for each nursery and retailer node $i \in \boldsymbol{N} \cup \boldsymbol{R}$

$$\left(\textbf{Internal disease spread}\right) = \beta \frac{S_i(t)I_i(t)}{N_i}, \tag{4.5}$$

where $\beta > 0$, the *transmission rate*, is assumed to be the same for all nurseries and retailers.

4.4.2 Death rates for susceptible and infected deaths

We have four different death rates in our model: the death rate for susceptible and infected plants for customers (d_{S_c}, d_{I_c}) and for nurseries and retailers (d_S, d_I). Relating this to Equations 4.1 to 4.4, we simply have

$$(\textbf{Susceptible death}) = d_S S_i(t), \qquad \text{for } i \in \textbf{N} \cup \textbf{R},$$

$$(\textbf{Infected death}) = d_I I_i(t), \qquad \text{for } i \in \textbf{N} \cup \textbf{R},$$

$$(\textbf{Susceptible death}) = d_{S_c} S_i(t), \qquad \text{for } i \in \textbf{C},$$

$$(\textbf{Infected death}) = d_{I_c} I_i(t), \qquad \text{for } i \in \textbf{C}.$$

The death rates for customers will not affect disease dynamics in our model, thus we do not investigate different values for d_{S_c}, d_{I_c} and so for simplicity we set these parameters to 0 (Table 4.1). In our analysis we vary the death rates for nurseries and retailers (d_S, d_I), considering death rates equivalent to all plants removed after 1, 6 and 12 months. We also consider the case of no death rate. We assume that $d_I \geq d_S$. Since our time step is 1 month, we calculate monthly death rates from the estimated life span λ in months, using the following formula $d = 1 - (10)^{-7/\lambda}$. This formula is derived below.

First consider the simple equation for the number of susceptibles $S(t)$, where time, t, is measured in months and a general death rate d:

$$S(t + 1) = (1 - d)S(t).$$

Considering a lifespan of λ months, we have:

$$S(\lambda) = (1 - d)S(\lambda - 1) = \cdots = (1 - d)^{\lambda} S(0) = 0,$$

$$(1 - d)^{\lambda} = 0. \tag{4.6}$$

Assuming that $S(0) > 0$ and solving for d we get $d = 1$ as the solution to Equation 4.6. This is problematic, given that any lifespan of λ months would result in $d = 1$. This problem occurs because this is the only way in this discrete time framework to have exactly 0 susceptible after λ months. To solve this issue, we specify a given tolerance, instead of 0. The largest number of susceptibles we have in this model is in the order of 10^6 and so the tolerance we use will be 10^{-7}. Returning to Equation 4.6, replacing 0 with 10^{-7}, we have:

$$(1 - d)^\lambda = 10^{-7}.$$ (4.7)

Solving Equation 4.7 for d we get:

$$d = 1 - 10^{-7/\lambda}.$$

We consider lifespans of 1, 6 and 12 months. We get respective death rates 1, 0.9 and 0.7 (rounding to the first significant figure). For an indefinite life span, we get a death rate of 0.

4.4.3 Modelling the trade of consignments

The total number of plants from incoming and outgoing trade for a node is simple to calculate in our model, and is shown below

$$\left(\textbf{Total trade input}\right) = \sum_{k=1}^{M} A_{k,i}, \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R} \cup \boldsymbol{C}$$

$$\left(\textbf{Total trade output}\right) = \sum_{k=1}^{M} A_{i,k}, \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R}.$$

In the rest of this section, we describe how we model the trade inspection process and the number of infected plants included in a given consignment $A_{i,j}$. This will allow us to describe the expressions for $\left(\textbf{Infected trade input}\right)$ and $\left(\textbf{Infected trade output}\right)$.

We are first required to make assumptions about the process in which consignments are prepared. For a sale from node i to j, the seller selects an $A_{i,j}$ number of plants from their stock, we assume one by one. Ignoring inspections, this is done without replacement. The seller's stock consists of a combination of susceptible and infected plants. The probability of selecting an infected plant, with this formulation, follows a *hyper-geometric* distribution [153]. We use this distribution to assume that for each consignment, the seller will select the expected value of infected plants. For a consignment of size $A_{i,j}$, this will be $\frac{A_{i,j} I_i(t)}{N_i}$, where $I_i(t)$ is the number of infected plants the seller has at time t and N_i is the total number of plants in stock. We assume that customers do not inspect the plants they receive.

Nurseries and retailers can apply their own control measures by inspecting plants they have sold on their way out of their premises and/or plants they have bought on their way in. These inspections will affect how many infected plants are removed from a seller's premises and how many infected plants are received by the buyer. In the next section we will detail the initial approach and the subsequent model for these control measures.

Trade inspections

We assume that both nurseries and retailers can inspect outgoing and incoming consignments. For a node $i \in \boldsymbol{N} \cup \boldsymbol{R}$, outgoing and incoming inspections have a *success rate* of α^{out} and α^{in}, respectively. The process of inspecting consignments will reduce the number of infected plants traded.

Alternatively, we can model a repeated inspection process. In this formulation, detected plants are re-selected and undergo the same inspection process to fulfil the buyer's original consignment. This is an iterative process of inspecting an ever-decreasing number of plants, until the original consignment of plants has been fulfilled.

Figure 4.3 illustrates the inspection process of the buyer and seller for a given consignment $A_{i,j}$. Firstly, the seller, i, selects plants to a total $A_{i,j}$ and inspects for signs of infection. We assume susceptible plants all pass through inspections. The infected plants selected fail inspection at a given rate α^{out}. The infected plants that fail inspection are then removed and the seller selects replacements. This is an iterative process which converges, except for perfect inspection ($\alpha^{out} = 1$). For this reason, and that perfect inspection is unrealistic, we do not consider the case where $\alpha^{out} = 1$, and thus assume that $\alpha^{out} \in [0, 1)$.

Once $A_{i,j}$ plants have passed through inspection, the plants are sent to the buyer, j. Assuming that buyers also inspect plants on their way in ($\alpha^{in} > 0$), this introduces another iterative process. The buyer begins their inspection where infected plants are found with a given rate α^{in}. The buyer accepts the plants that pass inspection. The infected plants that are found are removed (from the seller's compartment) and the buyer orders replacements from the same seller. The buyer repeatedly orders plants until the buyer accepts the number of plants equal to the initial consignment. For $\alpha^{in} < 1$, this process will converge. Hence we also do not consider $\alpha^{in} = 1$, and thus assume that $\alpha^{in} \in [0, 1)$.
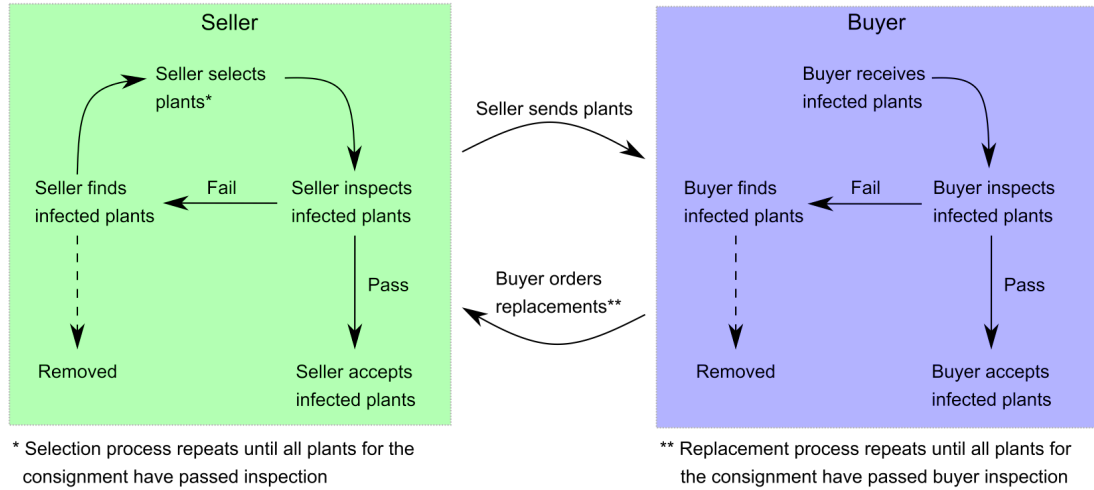
Figure 4.3. Schematic representation of the trade inspection process for a given consignment from a seller to a buyer.

We are interested in two quantities; the number of infected plants caught and missed from both the buyer and seller inspections. Thus there are three iterative processes to model: the seller's preparation of a given consignment to a buyer, the total number of plants re-ordered by the buyer and the total number of plants sent to the buyer for inspection from the seller.

The number of plants leaving the seller

Consider the consignment $A_{i,j}$ at time t, where the seller node i has $I_i(t)$ infected and N_i total plants. Before any plants reach the buyer, the seller undergoes a number of selections, picking plants out, inspecting and removing infected plants and then re-selecting.

1. From the assumed hyper-geometric distribution for preparing a consignment, the expected number of infected plants included in this consignment is $\frac{I_i(t)}{N_i} A_{i,j}$.

2. The number of plants for the second selection is simply the number of infected plants caught in the first selection, which is $\alpha^{out} \frac{I_i(t)}{N_i} A_{i,j}$. We expect $\alpha^{out} \left(\frac{I_i(t)}{N_i}\right)^2 A_{i,j}$ infected plants included in the selection.

3. The number of plants in the third selection will be the number of infected plants that were caught in the previous selection, $(\alpha^{out})\left(\frac{I_i(t)}{N_i}\right)^2 A_{i,j}$. Hence there will be $(\alpha^{out})^2 \left(\frac{I_i(t)}{N_i}\right)^3 A_{i,j}$ infected in this selection.

4. And so on...

5. Therefore the total number of infected plants in each selection for the consignment of size $A_{i,j}$ can be described by the infinite series:

$$\frac{I_i(t)}{N_i} A_{i,j} + \alpha^{out} \left(\frac{I_i(t)}{N_i}\right)^2 A_{i,j} + \cdots = A_{i,j} \sum_{k=1}^{\infty} (\alpha^{out})^{k-1} \left(\frac{I_i(t)}{N_i}\right)^k.$$

6. This is a convergent geometric series ($\alpha^{out} \frac{I_i(t)}{N_i} < 1$) and so we can evaluate the sum:

$$A_{i,j} \sum_{k=1}^{\infty} (\alpha^{out})^{k-1} \left(\frac{I_i(t)}{N_i}\right)^k = \frac{I_i(t)}{N_i} A_{i,j} \sum_{k=1}^{\infty} (\alpha^{out})^{k-1} \left(\frac{I_i(t)}{N_i}\right)^{k-1}$$

$$= \frac{I_i(t)}{N_i} A_{i,j} \left(1 - \alpha^{out} \frac{I_i(t)}{N_i}\right)^{-1}$$

$$= \frac{I_i(t)}{N_i - \alpha^{out} I_i(t)} A_{i,j}.$$

7. Therefore, to calculate how many infected plants are removed in this process, we multiply by α^{out} as this is the proportion of plants found in every step:

$$\alpha^{out} \frac{I_i(t)}{N_i - \alpha^{out} I_i(t)} A_{i,j}.$$

8. To calculate how many infected plants that are subsequently sent to the buyer, we multiply by $(1 - \alpha^{out})$, as this is the proportion of plants missed in every step:

$$\gamma_1 = (1 - \alpha^{out}) \frac{I_i(t)}{N_i - \alpha^{out} I_i(t)} A_{i,j}.$$

We now know how many infected plants are initially sent to the buyer for the first consignment preparation. The next step is to describe how the buyer inspects plants, how many plants are accepted by the buyer and how many are rejected with subsequent replacements ordered.

Buyer inspections

We describe the number of infected plants that the buyer receives for inspection, similar to how we approached the seller inspections:

1. Firstly, the buyer receives $\gamma_1 = (1 - \alpha^{out}) \frac{I_i(t)}{N_i - \alpha^{out} I_i(t)} A_{i,j}$ infected plants from their initial order of $A_{i,j}$ plants. The buyer will catch $\alpha^{in} \gamma_1$ of these in inspections and re-order another $\alpha^{in} \gamma_1$ plants.

2. The iterative process for the seller inspections will then repeat, but for a consignment size of $\alpha^{in}\gamma_1$ plants. The buyer will then receive $\gamma_2 = (1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\alpha^{in}\gamma_1$ number of infected plants in the second selection. $\alpha^{in}\gamma_2$ infected plants will be caught in inspections and so the buyer will re-order $\alpha^{in}\gamma_2$ plants.

3. The buyer receives $\gamma_3 = (1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\alpha^{in}\gamma_2$ infected plants from the seller. $\alpha^{in}\gamma_3$ infected plants will be caught through inspections, and so the buyer will re-order $\alpha^{in}\gamma_3$ plants.

4. And so on...

5. The number of infected plants that are sent to the buyer for inspection are therefore modelled by the infinite sum:

$$\gamma_1 + \gamma_2 + \gamma_3 + \cdots = \sum_{k=1}^{\infty} \gamma_k.$$

6. Notice that $\gamma_k = (1-\alpha^{out})\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}\alpha^{in}\gamma_{k-1}$ and we know that $\gamma_1 = (1-\alpha^{out})\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}A_{i,j}$, therefore we can evaluate γ_k:

$$\begin{aligned}
\gamma_k &= (1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\alpha^{in}\gamma_{k-1} \\
&= \left((1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\alpha^{in}\right)^2 \gamma_{k-2} \\
&= \cdots \\
&= \left((1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\alpha^{in}\right)^{k-1}\gamma_1 \\
&= \left((1 - \alpha^{out})\frac{I_i(t)}{N_i - \alpha^{out}I_i(t)}\right)^k (\alpha^{in})^{k-1}A_{i,j}.
\end{aligned}$$

7. Therefore, the number of infected plants sent to the buyer for inspection are:

$$\sum_{k=1}^{\infty} \gamma_k = \sum_{k=1}^{\infty} \left((1 - \alpha^{out}) \frac{I_i(t)}{N_i - \alpha^{out} I_i(t)} \right)^k (\alpha^{in})^{k-1} A_{i,j}$$

$$= A_{i,j} \left(\frac{(1 - \alpha^{out}) I_i(t)}{N_i - \alpha^{out} I_i(t)} \right) \sum_{k=1}^{\infty} \left(\frac{(1 - \alpha^{out}) \alpha^{in} I_i(t)}{N_i - \alpha^{out} I_i(t)} \right)^{k-1}.$$

8. To prove this series converges we require $\frac{(1-\alpha^{out})\alpha^{in} I_i(t)}{N_i - \alpha^{out} I_i(t)} < 1$. We prove this by using the fact that $I_i(t) \leq N_i$:

$$\frac{(1 - \alpha^{out}) \alpha^{in} I_i(t)}{N_i - \alpha^{out} I_i(t)} \leq \frac{(1 - \alpha^{out}) \alpha^{in} N_i}{N_i - \alpha^{out} I_i(t)}$$

$$\leq \frac{(1 - \alpha^{out}) \alpha^{in} N_i}{N_i - \alpha^{out} N_i}$$

$$= \frac{(1 - \alpha^{out}) \alpha^{in}}{(1 - \alpha^{out})}$$

$$= \alpha^{in}$$

$$< 1.$$

9. Therefore the series converges and we can calculate the sum:

$$\sum_{k=1}^{\infty} \gamma_k = A_{i,j} \left(\frac{(1 - \alpha^{out}) I_i(t)}{N_i - \alpha^{out} I_i(t)} \right) \sum_{k=1}^{\infty} \left(\frac{(1 - \alpha^{out}) \alpha^{in} I_i(t)}{N_i - \alpha^{out} I_i(t)} \right)^{k-1}$$

$$= A_{i,j} \left(\frac{(1 - \alpha^{out}) I_i(t)}{N_i - \alpha^{out} I_i(t)} \right) \left(1 - \frac{(1 - \alpha^{out}) \alpha^{in} I_i(t)}{N_i - \alpha^{out} I_i(t)} \right)^{-1}$$

$$= A_{i,j} \left(\frac{(1 - \alpha^{out}) I_i(t)}{N_i - \alpha^{out} I_i(t)} \right) \left(\frac{N_i - \alpha^{out} I_i(t)}{N_i - \alpha^{out} I_i(t) - (1 - \alpha^{out}) \alpha^{in} I_i(t)} \right)$$

$$= A_{i,j} \left(\frac{(1 - \alpha^{out}) I_i(t)}{N_i - \alpha^{out} I_i(t) - (1 - \alpha^{out}) \alpha^{in} I_i(t)} \right).$$

10. Thus, the number of infected plants the buyer receives at time t is

$$F_i(\alpha^{out}, \alpha^{in}, t) A_{i,j},$$

where

$$F_i(\alpha^{out}, \alpha^{in}, t) = \left(\frac{(1 - \alpha^{out})(1 - \alpha^{in})I_i(t)}{N_i - \alpha^{out}I_i(t) - (1 - \alpha^{out})\alpha^{in}I_i(t)} \right) \qquad \text{for } i = 1, ..., M,$$

represents the proportion of a consignment that includes infected plants which are not found in trade inspections.

In this section we have calculated the number of infected plants sent to the buyer and described how the seller prepares and inspects plants for a consignment $A_{i,j}$. Therefore, the (**Infected trade input**) and (**Infected trade output**) terms from Equations 4.1 to 4.4 are expressed below:

$$(\textbf{Infected trade input}) = \sum_{k=1}^{M} F_k(\alpha^{out}, \alpha^{in}, t) A_{k,i}, \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R},$$

$$(\textbf{Infected trade input}) = \sum_{k=1}^{M} F_k(\alpha^{out}, 0, t) A_{k,i}, \qquad \text{for } i \in \boldsymbol{C},$$

$$(\textbf{Infected trade output}) = \sum_{k=1}^{M} F_i(\alpha^{out}, \alpha^{in}, t) A_{i,k}, \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R}.$$

Note that for customer nodes, $i \in \boldsymbol{C}$, we assume they do not inspect their purchases, equivalent to $\alpha^{in} = 0$ for these nodes.

4.4.4 Sources and sinks

As previously mentioned, in order to keep the populations of each nursery and retailer constant $(S_i(t) + I_i(t) = N_i)$, we introduce the respective source and sink terms $G_i(t) \geq 0$ and $E_i(t) \geq 0$, for $i \in \boldsymbol{N} \cup \boldsymbol{R}$, which model plants entering/leaving the system. Plants that enter the system are interpreted as those imported from international trade or grown by the nursery. Plants that leave the system are interpreted as exports via international trade. Thus, we have

$$(\textbf{Total grown/imported}) = G_i(t),$$

$$(\textbf{Total exported}) = E_i(t).$$

Next, we need to consider the infected component of sources and sinks, i.e., the terms for **Infected imported** and **Infected exported**.

Importing infected plants

Nurseries and retailers who import plants from overseas are at risk of receiving infected plants. The event of a pest or disease successfully being established into a nursery or retailer node will be rare. Thus, we can model this process by periodically seeding an infected plant into a randomly selected nursery or retailer node that imports plants. We choose to seed one node with one infected plant for simplicity. We can model the number of infected plants imported from international trade for a given nursery or retailer node i at time t, from the discrete random variable $X(t) \sim U(i \in \boldsymbol{N} \cup \boldsymbol{R} : G_i(t) > 0)$ and the function $\nu_i(t, \tau^{import}, P^{import})$, defined below:

$$\nu_i(t, \tau^{import}, P^{import}) = P^{import} \, \delta(X(t) - i) \sum_{k \in \mathbb{N}} \delta(t - k\tau^{import}). \tag{4.8}$$

The parameter $P^{import} \in \{0, 1\}$ is used to control whether further seedings are considered in the model. The function $\delta(x)$ represents the delta function, which for $x \in \mathbb{R}$:

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0. \end{cases}$$

For simplicity of the analysis, we do not consider further incursions from the international trade network. Thus, in Equation 4.1 and Equation 4.2, we have

$$\left(\textbf{Infected imported}\right) = \nu_i(t, \tau^{import}, P^{import}), \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R}.$$

Exporting infected plants

We model infected plants included in exports in the same way as domestic trade (Section 4.4.3). The minor difference is that we consider exported plants as one consignment $E_i(t)$, with the expected number of infected plants included in exports given by $\frac{E_i(t)I_i(t)}{N_i}$. We also summarise the incoming trade inspections for foreign trade into one parameter α^{export}. Thus, when considering inspections, the number of infected plants exported for each trader in the system is

$$\left(\textbf{Infected exported}\right) = F_i(\alpha^{out}, \alpha^{export}, t)E_i(t), \qquad \text{for } i \in \boldsymbol{N} \cup \boldsymbol{R}. \tag{4.9}$$

4.4.5 Infected removal

In this section we calculate the number of infected plants that are removed from nursery and retailer nodes via control measures. These control measures include the inspections of consignments and the scheduled inspections of stock.

Infected plants removed from trade inspections

The total number of infected plants removed from trade inspections for a given consignment $A_{i,j}$ consists of the total number of infected plants removed from the buyer and the total removed from the seller. In Section 4.4.3 we found that the total number of infected plants that are sent to a buyer for inspection is

$$\left(\frac{(1 - \alpha^{out})I_i(t)}{N_i - \alpha^{out}I_i(t) - (1 - \alpha^{out})\alpha^{in}I_i(t)}\right)A_{i,j}.$$

Thus, multiplying by α^{in}, the total number of infected plants removed by the buyer's inspections can be given by

$$\left(\frac{(1-\alpha^{out})\alpha^{in}I_i(t)}{N_i-\alpha^{out}I_i(t)-(1-\alpha^{out})\alpha^{in}I_i(t)}\right)A_{i,j} = \frac{\alpha^{in}}{(1-\alpha^{in})}F_i(\alpha^{out},\alpha^{in},t)A_{i,j}. \qquad (4.10)$$

Next, we model the total number of infected plants that the seller removes until the consignment has been fulfilled. Finally, we account for the further re-orders from the buyer's inspections.

1. As previously calculated for the consignment $A_{i,j}$, a total $\alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}A_{i,j}$ are removed by the seller (Section 4.4.3).

2. The buyer receives $\gamma_1 = (1-\alpha^{out})\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}A_{i,j}$ infected plants to inspect, ordering a total $\alpha^{in}\gamma_1$ more replacements. Hence, a total $\alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}\alpha^{in}\gamma_1$ will be removed from the seller's inspections for this second order.

3. And so on...

4. Therefore, the total number removed from the seller's inspections is given by

$$\alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}A_{i,j} + \sum_{k=1}^{\infty}\alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}\alpha^{in}\gamma_k$$

$$= \alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}A_{i,j} + \alpha^{out}\frac{I_i(t)}{N_i-\alpha^{out}I_i(t)}\alpha^{in}\sum_{k=1}^{\infty}\gamma_k$$

$$= \frac{\alpha^{out}I_i(t)}{N_i-\alpha^{out}I_i(t)}\left(1 + \frac{(1-\alpha^{out})\alpha^{in}I_i(t)}{N_i-\alpha^{out}I_i(t)-(1-\alpha^{out})\alpha^{in}I_i(t)}\right)A_{i,j}$$

$$= \frac{\alpha^{out}I_i(t)}{N_i-\alpha^{out}I_i(t)}\left(\frac{N_i-\alpha^{out}I_i(t)}{N_i-\alpha^{out}I_i(t)-(1-\alpha^{out})\alpha^{in}I_i(t)}\right)A_{i,j}$$

$$= \left(\frac{\alpha^{out}I_i(t)}{N_i-\alpha^{out}I_i(t)-(1-\alpha^{out})\alpha^{in}I_i(t)}\right)A_{i,j}$$

$$= \frac{\alpha^{out}}{(1-\alpha^{out})(1-\alpha^{in})}F_i(\alpha^{in},\alpha^{out},t)A_{i,j}. \qquad (4.11)$$

We can now calculate the total number of infected plants removed from a consignment via trade inspections by summing Equation 4.11 and Equation 4.10

$$\frac{\alpha^{out}}{(1-\alpha^{out})(1-\alpha^{in})}F_i(\alpha^{in},\alpha^{out},t)A_{i,j} + \frac{\alpha^{in}}{1-\alpha^{in}}F_i(\alpha^{out},\alpha^{in},t)A_{i,j}$$
$$= \frac{\alpha^{out}+\alpha^{in}(1-\alpha^{out})}{(1-\alpha^{out})(1-\alpha^{in})}F_i(\alpha^{out},\alpha^{in},t)A_{i,j}$$
$$= R_i^{trade}(\alpha^{out},\alpha^{in},t)A_{i,j}. \tag{4.12}$$

We condense this term with the introduction of $R_i^{\text{trade}}(\alpha^{out},\alpha^{in},t)$ which represents the proportion of consignment $A_{i,j}$ that are removed from the consignment inspection process

$$R_i^{\text{trade}}(\alpha^{out},\alpha^{in},t) = \frac{\alpha^{out}+\alpha^{in}(1-\alpha^{out})}{(1-\alpha^{out})(1-\alpha^{in})}F_i(\alpha^{out},\alpha^{in},t).$$

Thus, the total number of infected plants removed from the consignment inspection process for all consignments is given by

$$R_i^{trade}(\alpha^{out},\alpha^{in},t)\sum_{k=1}^{M}A_{i,k}. \tag{4.13}$$

Recall that we consider exports $E_i(t)$ to be a consignment that is also subject to inspections (Section 4.4.4). Thus, for each nursery and retailer, $i \in \mathbf{N} \cup \mathbf{R}$, the total number of infected plants removed from the inspections of exports is given by

$$R_i^{trade}(\alpha^{out},\alpha^{export},t)E_i(t). \tag{4.14}$$

Scheduled inspections from external sources

The other control measure we consider for nurseries and retailers is inspection for disease by an external body on a periodic basis. These scheduled inspections may be more effective due to the inspector having access to information, techniques and technology that the nursery/ retailer does not have access to. To give an example, this access may be restricted through the high cost of training staff or the cost of expensive equipment. Importantly,

when these inspections take place, we assume that it is an inspection of all plants the nursery or retailer has, in contrast to the trade inspections which are just for the consignments.

We model scheduled inspections for nurseries and retailers by a *frequency of inspections* $\tau^{insp} \in \mathbb{N}$, an *inspection effectiveness* $r \in [0,1]$ and a *time-point of first inspection* $z \in \mathbb{N}$. We model this with the function $\theta(t, z, \tau^{insp}, r)$ which we define below

$$\theta(t, z, \tau^{insp}, r) = r \sum_{k \in \mathbb{N}} \delta(t - z - k\tau^{insp}).$$

Therefore, the number of infected plants removed from node i at time t from stock inspections is

$$\theta(t, z, \tau^{insp}, r)I_i(t). \tag{4.15}$$

Thus, our (**Infected removal**) term from Equation 4.2 is given by the addition of expressions 4.13, 4.14 and 4.15 as shown below:

$$
\begin{aligned}
(\textbf{Infected removal}) = {}& R_i^{trade}(\alpha^{out}, \alpha^{in}, t) \sum_{k=1}^{M} A_{i,k} \\
& + R_i^{trade}(\alpha^{out}, \alpha^{export}, t)E_i(t) \\
& + \theta(t, z, \tau^{insp}, r)I_i(t).
\end{aligned}
$$

4.4.6 Model equations

Now that all aspects of Equations 4.1 to 4.4 have been defined, we can present the equations in their mathematical representation. For nursery and retailer nodes $i \in \boldsymbol{N} \cup \boldsymbol{R}$, we have

$$S_i(t+1) = S_i(t) - d_S S_i(t)$$

$$+ \sum_{k=1}^{M}(A_{k,i} - F_k(\alpha^{out}, \alpha^{in}, t)A_{k,i})$$

$$- \sum_{k=1}^{M}(A_{i,k} - F_i(\alpha^{out}, \alpha^{in}, t)A_{i,k})$$

$$+ G_i(t) - \nu_i(t, \tau^{import}, P^{import})$$

$$- \left(E_i(t) - F_i(\alpha^{out}, \alpha^{\text{export}}, t)E_i(t)\right)$$

$$- \beta \frac{S_i(t)I_i(t)}{N_i}, \tag{4.16}$$

$$I_i(t+1) = I_i(t) - d_I I_i(t)$$

$$+ \sum_{k=1}^{M} F_k(\alpha^{out}, \alpha^{in}, t)A_{k,i}$$

$$- \sum_{k=1}^{M} F_i(\alpha^{out}, \alpha^{in}, t)A_{i,k}$$

$$+ \nu_i(t, \tau^{import}, P^{import})$$

$$- F_i(\alpha^{out}, \alpha^{\text{export}}, t)E_i(t)$$

$$+ \beta \frac{S_i(t)I_i(t)}{N_i}$$

$$- \left(R_i^{trade}(\alpha^{out}, \alpha^{in}, t) \sum_{k=1}^{M} A_{i,k} \right.$$

$$+ R_i^{trade}(\alpha^{out}, \alpha^{export}, t)E_i(t)$$

$$\left. + \theta(t, z, \tau^{insp}, r)I_i(t) \right). \tag{4.17}$$

Similarly, the equations for each customer node $i \in \boldsymbol{C}$ are given by

$$S_i(t+1) = S_i(t) - d_{S_c}S_i(t) + \sum_{k=1}^{M}(A_{k,i} - F_k(\alpha^{out}, 0, t)A_{k,i}), \tag{4.18}$$

$$I_i(t+1) = I_i(t) - d_{I_c}I_i(t) + \sum_{k=1}^{M} F_k(\alpha^{out}, 0, t)A_{k,i}. \tag{4.19}$$

We have now defined all aspects of our model equations (Equations 4.16-4.19). Below, we summarise the model parameters in Table 4.1 alongside their baseline values.

Table 4.1. Description of model parameters and their baseline values.

Parameter	Description	Units	Value
α^{out}	Inspection effort of nursery and retailer nodes for outgoing plants		0
α^{in}	Inspection effort of nursery and retailer nodes for incoming plants		0
α^{export}	Mean inspection effort for buyers from the international trade network		0
β	Transmission rate of pest within node	$[\text{Month}]^{-1}$	2
d_S	Death rate of susceptible plants for nurseries and retailers	$[\text{Month}]^{-1}$	0.7
d_I	Death rate of infected plants for nurseries and retailers	$[\text{Month}]^{-1}$	1
d_{S_c}	Death rate of susceptible plants for commercial and consumer customers	$[\text{Month}]^{-1}$	0
d_{I_c}	Death rate of susceptible plants for commercial and consumer customers	$[\text{Month}]^{-1}$	0
r	Removal rate of infected plants within nurseries from scheduled inspections	$[\text{Month}]^{-1}$	0
τ	Frequency of importing disease	$[\text{Month}]$	0
τ^{insp}	Frequency of scheduled inspections	$[\text{Month}]$	6
z	Time of first scheduled inspection	$[\text{Month}]$	6
P^{import}	Binary indicator for subsequent infected imports		0
$|N_{\text{Com}}|$	Number of commercial dominated nurseries		40
$|N_{\text{Cons}}|$	Number of consumer dominated nurseries		40
$|N_{\text{Nur}}|$	Number of nursery dominated nurseries		40
$|N_{\text{Ret}}|$	Number of retailer dominated nurseries		40
$|\text{Ret}|$	Number of retailers		1,600
d_μ^{out}	Nursery and retailer out-degree mean	$[\text{Plants}]$	$380 \times (160/630)$
$d_{\sigma^2}^{\text{out}}$	Nursery and retailer out-degree variance	$[\text{Plants}]^2$	$114,587 \times (160/630)^2$
y_μ^{Com}	Commercial consignment-size mean	$[\text{Plants}]$	$79 \times (160/630)$
y_σ^{Com}	Commercial consignment-size standard deviation	$[\text{Plants}]$	$297 \times (160/630)$
y_μ^{Cons}	Consumer consignment-size mean	$[\text{Plants}]$	$185 \times (160/630)$
y_σ^{Cons}	Consumer consignment-size standard deviation	$[\text{Plants}]$	$822 \times (160/630)$
y_μ^{Nur}	Nursery consignment-size mean	$[\text{Plants}]$	$1844 \times (160/630)$
y_σ^{Nur}	Nursery consignment-size standard deviation	$[\text{Plants}]$	$3567 \times (160/630)$
y_μ^{Ret}	Retailer consignment-size mean	$[\text{Plants}]$	$3497 \times (160/630)$
y_σ^{Ret}	Retailer consignment-size standard deviation	$[\text{Plants}]$	$9588 \times (160/630)$

4.4.7 Explicit expressions for $G_i(t)$ and $E_i(t)$

Now that our equations are fully defined, we can explicitly calculate the source and sink terms in our model $(G_i(t), E_i(t))$. Sources and sinks are measured in units of plants; thus we also have the requirement that $G_i(t) \geq 0$ and $E_i(t) \geq 0$. With the assumption that for

each nursery and retailer $S_i(t) + I_i(t) = N_i$, we can calculate the source term $G_i(t)$ and sink term $E_i(t)$ for a nursery or retailer node $i \in \boldsymbol{N} \cup \boldsymbol{R}$. Summing Equations 4.16 and 4.17 yields

$$G_i(t) - cE_i(t) = b - a,$$

where a, b, and c are defined as

$$a = \sum_{k=1}^{M}(A_{k,i} - A_{i,k}),$$

$$b = d_S S_i(t) + (d_I + \theta(t, z, \tau^{insp}, r))I_i(t) + R_i^{trade}(\alpha^{out}, \alpha^{in}, t)\sum_{k=1}^{M} A_{i,k},$$

$$c = 1 + R_i^{trade}(\alpha^{out}, \alpha^{export}, t).$$

Therefore, we define both $G_i(t)$ and $E_i(t)$ in a piece-wise fashion as

$$G_i(t) = \begin{cases} b - a & \text{if } b - a \geq 0 \\ 0 & \text{if } b - a < 0, \end{cases} \tag{4.20}$$

$$E_i(t) = \begin{cases} \frac{a-b}{c} & \text{if } b - a \leq 0 \\ 0 & \text{if } b - a > 0. \end{cases} \tag{4.21}$$

Hence, if a node sells and throws away more than it buys, there must be a source, $G_i(t)$, representing a combination of importing or growing to keep the stock level at N_i. Conversely, if a node buys more than it both sells and throws away, then it must have a sink, E_i, representing exports, scaled to take into account for the number of infected plants thrown away by inspections.

4.5 Methods of analysis

In this section we describe how we conduct our analysis on the model. We are interested in how disease spreads throughout the network and how effective different interventions can be at reducing spread. As previously mentioned, we restrict our simulation period to 36 months to account for our network structure not changing over time. Most of our analysis is restricted to the nursery nodes in our model.

When seeding disease in the network, we choose a nursery, i, at random and set $I_i(0) = 1$. For each set of analysis, we look at the effects of seeding the disease in each nursery subcategory $\{N_{Com}, N_{Cons}, N_{Nur}, N_{Ret}\}$. We repeat the simulations 100 times for 100 network replicates (10000 simulations total), to average over the variability in each network construction for each set of parameter values. Additionally, we investigate the effects of changing the nursery distribution, considering scenarios 1, 2 and 3 from the previous chapter. For convenience, we recall these in Table 4.2.

Table 4.2. Description of the scenarios we consider for the distribution of nursery subcategories in our model.

Scenario	Description	Value
S1	Uniform distribution of nurseries	$(|N_{Com}|, |N_{Cons}|, |N_{Nur}|, |N_{Ret}|\,) = (40, 40, 40, 40)$
S2	Predominantly Commercial plants	$(|N_{Com}|, |N_{Cons}|, |N_{Nur}|, |N_{Ret}|) = (80, 20, 40, 20)$
S3	Predominantly Consumer plants	$(|N_{Com}|, |N_{Cons}|, |N_{Nur}|, |N_{Ret}|) = (20, 50, 40, 50)$

The metrics we use to analyse the spread of disease include the proportion of infected plants per node subcategory at the end of the simulation ($t = 36$ months) and the time t when 20% of plants in nurseries are infected. The latter metric is used because this model network does not respond to the disease outbreak. Above a level of prevalence of a pest or disease, the network structure would change in response, for example with quarantines and the rewiring of edges [106]. This assumption of a fixed flow of trade in the network thus limits our analysis to the short term.

In our first set of analysis, we want to understand how disease spreads throughout the network when no control measures are in place. Thus, the only form of removal we consider are the death rate parameters for nurseries and retailers (d_S, d_I). Four time-series are presented for seedings in each nursery subcategory showing the proportion of infection over time for each node subcategory, with baseline parameters considered (Table 4.1). For the commercial and consumer node subcategories, we calculate the proportion of infected plants they receive at the current time point as opposed to the cumulative proportion of infection over time. We vary the transmission rate parameter β across values $\{0.5, 1, ..., 3\}$ and the death rate parameters d_S and d_I. When varying the death rate parameters, we assume that $d_I \geq d_S$. As we outline in Section 4.4.2, we consider death rates of $\{0, 0.7, 0.9, 1\}$.

In our next set of analysis, we introduce scheduled inspections. We fix the frequency of inspection at a baseline of $\tau^{insp} = 6$ months and vary the removal rate from inspections $r \in \{0, 0.1, ..., 1\}$. We then look at varying the trade inspection parameters α^{in} and α^{out}. We assess the effectiveness of only inspecting outgoing trade and compare this with inspections on both incoming and outgoing trade. First we vary $\alpha^{out} \in \{0, 0.1, ..., 0.9\}$ with $\alpha^{in} = 0$. Then, we consider $\alpha^{in} = \alpha^{out}$ in the range $\{0, 0.1, ..., 0.9\}$.

Then we consider the parameters for both trade inspections and scheduled inspections. We vary $\alpha^{out} \in \{0, 0.1, ..., 0.9\}$ and $r \in \{0, 0.1, ..., 1\}$ for $\alpha^{in} = 0$ and $\alpha^{in} = \alpha^{out}$. We analyse the results to compare the trade-offs and effectiveness of the two different methods of inspection. This comparison of scheduled inspections ($r \in \{0, 0.2, ..., 1\}$) against trade inspections ($\alpha^{in} = \alpha^{out} \in \{0, 0.1, ..., 0.9\}$) is repeated for different combinations of the scheduled inspection frequency $\tau^{insp} \in \{1, 3, 6\}$ and the simulation time point of first inspection $z \in \{1, 3\}$. We only consider the cases where $\tau^{insp} \geq z$. The results are analysed to see how increasing the frequency of scheduled inspections improves inspection effectiveness against disease spread. Additionally, we want to see how the time point of the first scheduled inspection after the initial incursion affects inspection effectiveness.

Finally, we consider the cost of inspections and replacing plants lost to infection to determine an optimal inspection frequency. We compare inspecting every 3 or 6 months and inspecting every month or every 3 months. We assume cost to consist of the monetary value lost to nurseries from infected plants at $t = 36$ and the price of all the scheduled inspections that occurred over the simulation. The cost incurred by all nurseries can therefore be calculated for a given cost of a plant (X_1) and cost of inspection (X_2) with the equation below:

$$\text{Cost}(X_1, X_2) = X_1 \sum_{i \in \boldsymbol{N}} I_i(36) + 160 \text{ floor}(\frac{36 - z}{\tau^{insp}}) X_2.$$

The equation is simplified by expressing the total cost as a function of the price of an inspection relative to the price of a plant $X_3 = X_2/X_1$. We use the simplified formula below:

$$\text{Cost}(X_3) = \sum_{i \in \boldsymbol{N}} I_i(36) + 160 \text{ floor}(\frac{36 - z}{\tau^{insp}}) X_3.$$

For the $\sum_{i \in \boldsymbol{N}} I_i(36)$ term, we calculate the total number of infected plants in nurseries at $t = 36$ for the 10000 simulations and then use the median value. To account for the variability in the timing of the first scheduled inspection, we average over the different timing of first inspection (z) for each τ^{insp}.

4.6   Results

The results of the analysis are presented in the same order as described in Section 4.5. Unless otherwise specified, all figures in our analysis have been generated using the values in Table 4.1.

4.6.1 No inspections

In this section, we see how disease travels across the network without any intervention. The simulations run over different nursery and retailer death rates and transmission rates.

We display time series for the proportion of infected plants per node subcategory (Figures 4.4 - 4.7). These figures display the median, 2.5% and 97.5% quantiles for each node subcategory at every time-point t for the 10000 simulations.

Considering the seeding of a pest/disease in node subcategory N_{Com}, we see the median proportion of infected plants remains low for the first 24 months before epidemic spread begins (Figure 4.4). From month 24 to 36, the median proportion of infected plants in nurseries changes from less than 5% to $\sim$ 45%. The rate of spread for different nursery subcategories is the same and slower for retailers. Nurseries reach median 20% infected by $t = 31$, whereas retailers reach a median 20% infected by $t = 35$, which makes the rate of spread for retailers slower than nurseries by approximately 4 months.
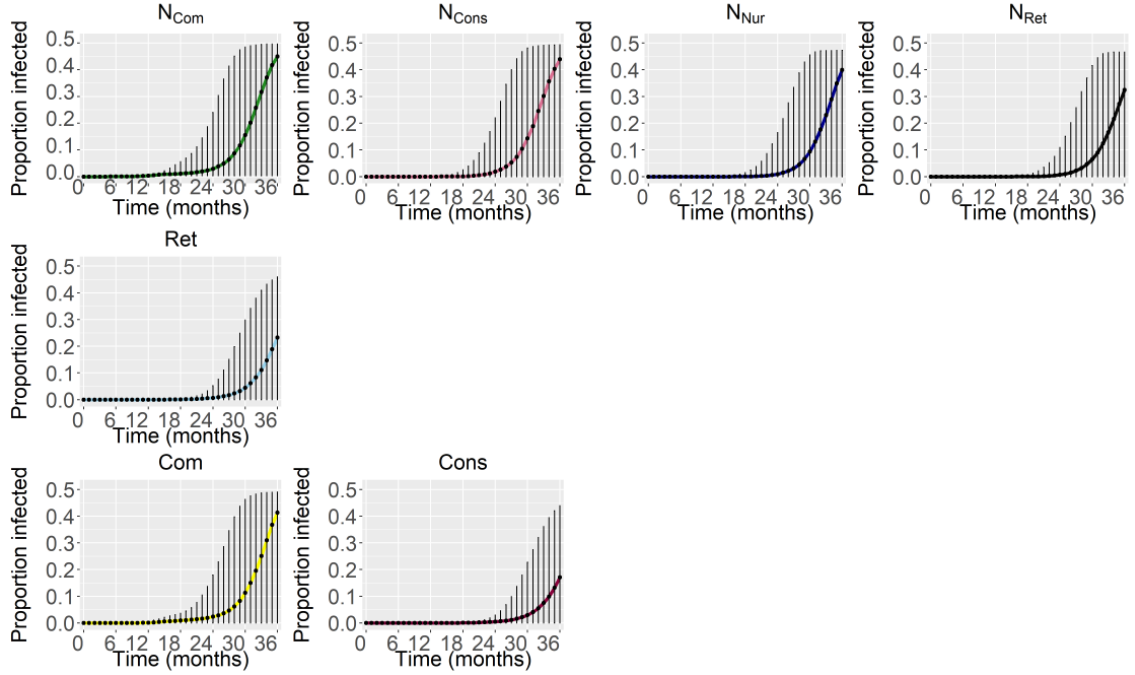
Figure 4.4. Time-series for the proportion of plants infected in each node subcategory at each time-point in the simulation. For commercial and consumer nodes, points represent the proportion of infected plants received from plants bought at the current time step. Points displayed are the median values of 100 simulations for 100 network constructions. The error bars represent the 2.5% and 97.5% quantiles. The parameter values for this result are at baseline (Table 4.1), with the pest seeded in node subcategory N_{Com}.

Recall that for customers (commercial and consumer nodes), we calculate the proportion of infected plants each customer has received. By $t = 36$, the median number of infected plants received by commercial nodes is $\sim 40\%$ and by consumer nodes is $\sim 16\%$ (Figure 4.4). The proportion of infected plants that commercial nodes receive follows the prevalence of disease within nurseries, given that we are not considering inspections. The proportion of infected plants that consumer nodes receive follows the prevalence of retailers, given that the majority of links to consumer nodes come from retailer nodes instead of nurseries in this model.

For node subcategories N_{Com}, N_{Cons}, N_{Nur} and N_{Ret} there is no variability in the simulations for the first 14, 16, 16 and 18 time-points respectively (Figure 4.4). Similarly, there is no variance for the commercial, retailer and consumer nodes for the first 12, 20 and 21 time-points. A single introductory event of a pest or disease into the network at

$t = 0$ and no subsequent introductions can result in epidemic spread one to two years later. From the upper quantiles of the nursery node subcategories, we see the rate of spread flattening off to approximately 50% prevalence. The lower quantile for all time-points and node subcategories in Figure 4.4 is 0. Therefore, depending on the node seeded and the exact network structure, an initial disease introduction can either fail to establish or reach maximum prevalence.

The time series for seedings in N_{Com}, N_{Cons}, N_{Nur} and N_{Ret} are shown in Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7, respectively. The results for seeding in N_{Cons} are almost identical to seeding in N_{Com} (Figure 4.4 and Figure 4.5). It is seeding in N_{Nur} where we see the largest difference in disease spread (Figure 4.6). With this seeding, epidemic spread begins in nurseries after 18 months, compared to 24 months for other seedings. The rate of spread is higher in all node subcategories with this seeding. To illustrate this, we compare the infection curves of N_{Com} for seedings in N_{Nur} and N_{Cons} (Figure 4.6 Figure 4.5), because the results for N_{Com}, N_{Cons} and N_{Ret} seedings are similar and N_{Com} is not a seeded node subcategory in these instances. In the N_{Cons} seeding (Figure 4.5), the final time point of 0 median infection occurs at $t = 17$ for N_{Com}. By $t = 36$, the median proportion infected is 0.45. For the N_{Nur} seeding (Figure 4.6), the final time point of 0 median infection is $t = 15$ and the time point to reach 0.45 proportion infected is $t = 28$. Therefore, for seeding in N_{Nur}, infection spreads faster by 6 months.

Important to note with the N_{Nur} seeding (Figure 4.6) is that the lower quantiles for all node subcategories are not constant at 0 throughout the time-series. This result shows that, at the 95% level, epidemic spread will occur after approximately 24 months of a pest or disease seeded in N_{Nur}. We also see the difference between the upper and lower quantiles narrows significantly at $t = 36$ for the nursery groups N_{Com}, N_{Cons} and N_{Nur}. By $t = 36$ at the 95% significance level, the nursery groups N_{Com}, N_{Cons} and N_{Nur} are at or close to 50% infection prevalence.

Seeding the pest or disease in node subcategory N_{Ret} (Figure 4.6) results in fewer infections overall and slightly slower spread than seeding in N_{Com} or N_{Cons} (Figure 4.4 and Figure 4.5). This is likely due to the network parameters given that retailer nodes buy the largest consignments, see parameter y_μ^{Ret} in Table 4.1. Thus N_{Ret} nodes are likely selling more plants than they buy, resulting in a larger $G_i(t)$ term for N_{Ret} nodes. Thus, as we do not consider additional sources of infected plants from $G_i(t)$, the model dilutes the number of infected plants that can be traded from N_{Ret} nodes in particular.

From all seedings considered (Figure 4.4 to Figure 4.7), the infection curves for the commercial and consumer nodes follow the infection curves for nursery nodes and retailer nodes, respectively. This result suggests that the prevalence of disease in nurseries insignificantly affects consumer nodes. Infected plants arrive in significant numbers to customers once the disease has spread throughout the nurseries and retailers. This indicates that the rate of entry of infected plants to customers is not affected significantly by the initial nursery infected but rather the prevalence of disease in nurseries and retailers.
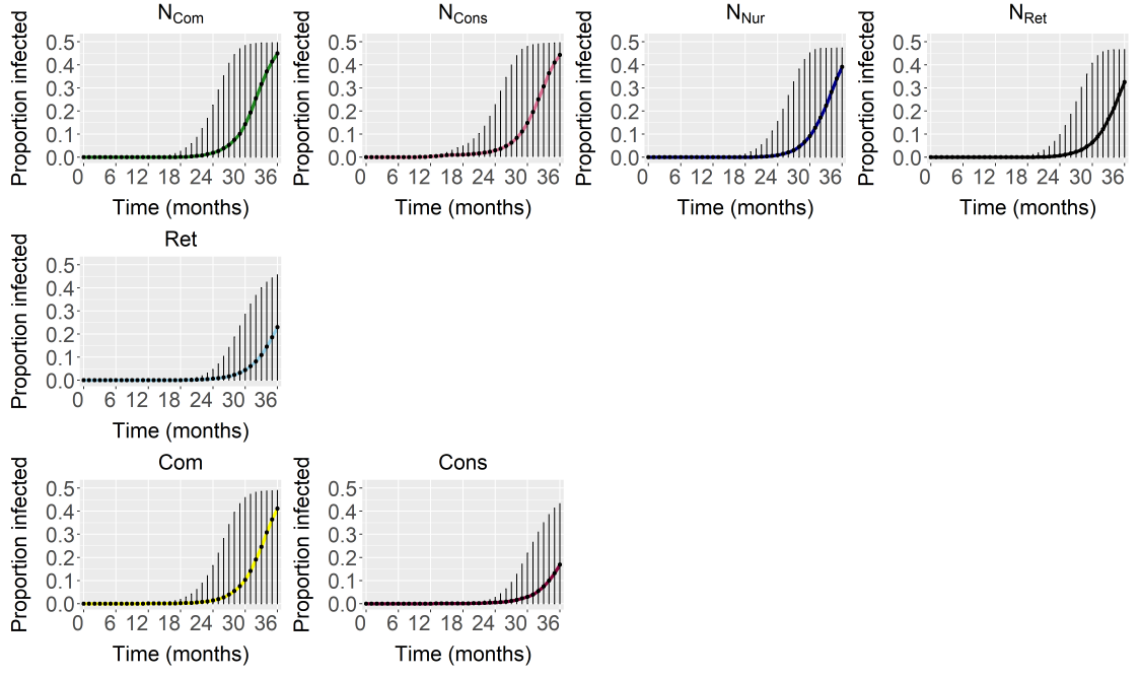
Figure 4.5. Time-series for the proportion of plants infected in each node subcategory at each time-point in the simulation. For commercial and consumer nodes, points represent the proportion of infected plants received from plants bought at the current time step. Points displayed are the median values of 100 simulations for 100 network constructions. The error bars represent the 2.5% and 97.5% quantiles. The parameter values for this result are at baseline (Table 4.1), with the pest seeded in node subcategory N_{Cons}.
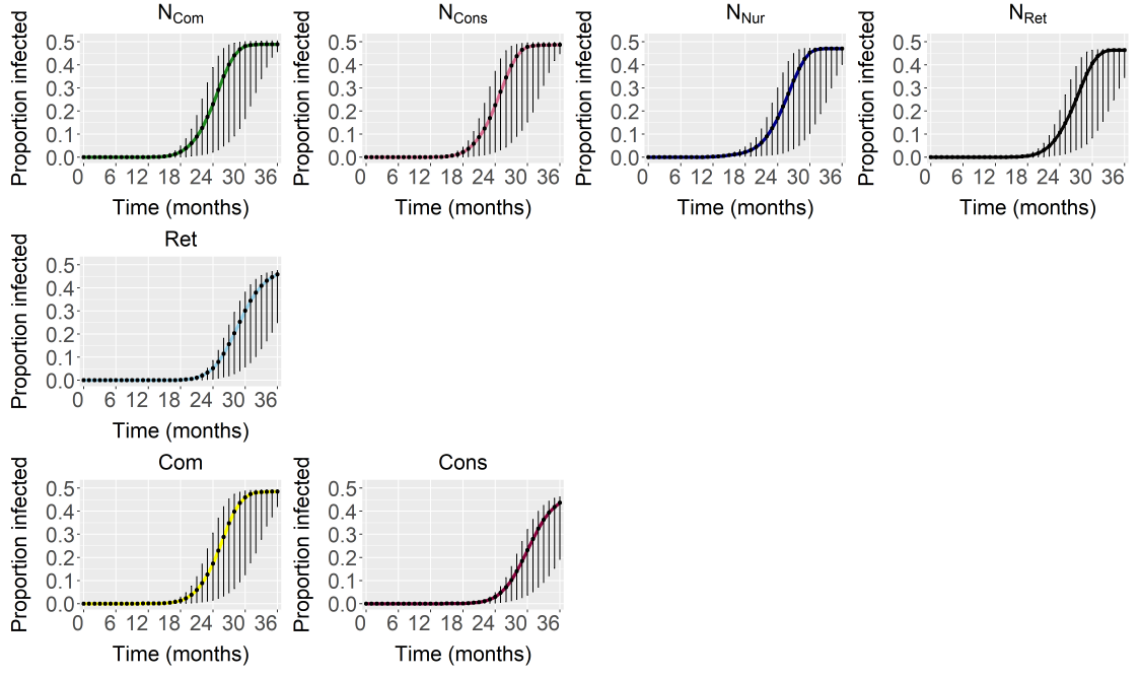
Figure 4.6. Time-series for the proportion of plants infected in each node subcategory at each time-point in the simulation. For commercial and consumer nodes, points represent the proportion of infected plants received from plants bought at the current time step. Points displayed are the median values of 100 simulations for 100 network constructions. The error bars represent the 2.5% and 97.5% quantiles. The parameter values are at baseline (Table 4.1), with the pest seeded in node subcategory N_{Nur}.
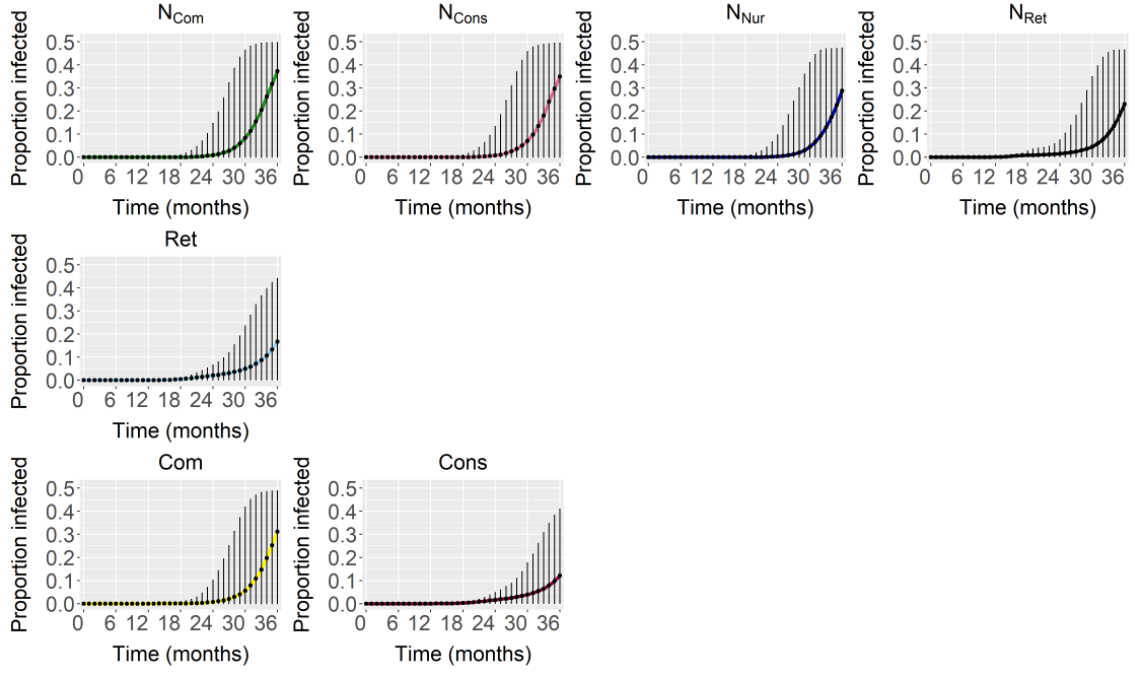
Figure 4.7. Time-series for the proportion of plants infected in each node subcategory at each time-point in the simulation. For commercial and consumer nodes, points represent the proportion of infected plants received from plants bought at the current time step. Points displayed are the median values of 100 simulations for 100 network constructions. The error bars represent the 2.5% and 97.5% quantiles. The parameter values for this result are at baseline (Table 4.1), with the pest seeded in node subcategory N_{Ret}.

We see that the size of the error bars for seedings in N_{Com}, N_{Cons} or N_{Ret} nurseries are much larger compared to N_{Nur} seedings, due to the lower quantiles being near zero (Figure 4.4 to Figure 4.7). Most of the outgoing links from N_{Com}, N_{Cons}, and N_{Ret} nodes are towards either Commercial, Consumer or Retailer nodes who do not contribute to further spread throughout the network. With the variation in simulations of the network, for N_{Com}, N_{Cons} or N_{Ret} seedings, there will be seeded nodes who have no outgoing connections to other nurseries, resulting in the near-zero lower quantiles (Figure 4.4, Figure 4.5 and Figure 4.7). Whereas, most of the outgoing links from N_{Nur} nodes are to other nurseries who can spread the disease further across the network. Thus, we do not see this near-zero lower quantile for N_{Nur} seedings in Figure 4.6.

In this section we have gained an understanding of how a generic pest or disease spreads throughout our network model, with our baseline parameter values (Table 4.1). In the next section we investigate how changing the disease transmission parameter β affects the disease spread dynamics in our model.

Varying the rate of disease transmission

In this section we vary the disease transmission parameter β and assess the effect that this has on disease spread in the network. Figure 4.8 shows the proportion of infected plants in nurseries by $t = 36$ (y-axis) for changes in β (x-axis). This plot also shows the results for each different seeding, and subgraphs (S1), (S2), and (S3) show the results for nursery scenarios 1, 2 and 3 respectively (Table 4.2). In each figure, we show the median, 25% and 75% quantiles for the 10000 simulations with a given parameter set. Note that the median value for N_{Com} seeding in these plots is almost identical to N_{Cons} and so is not visible for most points in Figure 4.8. We see that the proportion infected in nurseries does not change significantly for different nursery scenarios (Figure 4.8), hence we focus the rest of our analysis of this figure on subgraph (S1), representing scenario 1.

Immediately, we see that the disease dies out for $\beta \leq 1$ and large outbreaks occur only when $\beta \geq 2$. For N_{Com}, N_{Cons} and N_{Ret} seedings, the error bars all overlap for fixed values of β and so are not significantly different. For $\beta > 1$, the error bars for seeding in N_{Nur} do not overlap with the other seedings, showing a significant difference in the final proportion infected in nurseries for this seeding. As we saw in the time series analysis, the variability for N_{Nur} seedings is much lower than all other seedings. In addition, for $\beta \geq 2$, the variability shrinks for the N_{Nur} seeding such that the upper and lower quartile are equal to the median. Thus, the probability of an epidemic occurring is almost certain for a seeding in N_{Nur}, as long as the disease is growing. For other seedings, the large variability implies that the probability of an epidemic is far lower.

Figure 4.8. Points represent the median proportion of nurseries infected when $t = 36$, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given β value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1)

We also see the effect that varying β has on the time t until 20% of the plants in nurseries are infected (Figure 4.9). Similar to the previous graph, the median values are displayed in addition to the 25% and 75% quantiles. The subgraphs (S1), (S2) and (S3) display results for the different nursery scenarios and we show the results for different seedings within each subgraph. Firstly, there are no points plotted for $\beta \in \{0.5, 1, 1.5\}$ because in these simulations, nurseries do not reach 20% infected. Again, most plotted points for the N_{Com} seeding are identical to the N_{Cons} seeding and so are not visible in Figure 4.9. Across scenarios, there are few differences in the skewness of the results and a small divergence in the N_{Com} seeding from N_{Cons} in scenario 3 for $\beta = 2.5$ (Figure 4.9). Otherwise, the results are qualitatively the same and thus we focus our analysis on subgraph (S1).

Intuitively, the time until 20% of nurseries are infected decreases for increasing values of $\beta \geq 2$. We see that a change from $\beta = 2$ to $\beta = 3$ can cause an outbreak to occur 9 months faster. Again, we do not see a significant difference between N_{Com}, N_{Cons} and N_{Ret} seedings for the same value of β. We also see that N_{Nur} seedings are significantly different from all other seedings. For seeding in N_{Nur}, the error bars do not overlap for

changes in β, showing a significant effect on the time until 20% infection in nurseries. For N_{Com}, N_{Cons} and N_{Ret} seedings, error bars do not overlap for $\beta = 2$ and $\beta \in \{2.5, 3\}$, however, error bars do overlap for $\beta = 2.5$ and $\beta = 3$. This indicates that a change from $\beta = 2$ to $\beta = 2.5$ has a significant effect, but the change from $\beta = 2.5$ to $\beta = 3$ is insignificant. These results suggest a limit to how β affects the rate at which the disease spreads throughout the network.

Figure 4.9. Points represent the median time until 20% of plants across nurseries are infected, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given β value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

Varying death rates

Next we investigate changes in the death rates for susceptible and infected plants in nurseries and retailers. As described in Sections 4.4.2 and 4.5, we vary combinations of $d_S \in \{0, 0.7, 0.9, 1\}$ and $d_I \in \{0, 0.7, 0.9, 1\}$ with the added constraint $d_S \leq d_I$. For each death rate combination, we simulate 100 times for 100 networks and repeat this for the 4 different seedings. We average the results over the different seedings.

We see the proportion infected by $t = 36$, focusing on the nursery node subcategories (Figure 4.10-Figure 4.12). The combinations of death rates are displayed in a grid, with the columns varying d_I and the rows varying d_S. Figure 4.10 shows the results for varying death rates with nursery scenario 1. Varying d_S does not have an effect on the model, a direct result from no additional sources of infection. Intuitively, increasing d_I has a decreasing effect on the level of infection in nurseries. The only significant change we see for scenario 1 is a change from $d_I = 0$. Thus, for the parameter values we consider, there is an insignificant difference in disease spread for pests or diseases that kill the plant in one year or less.
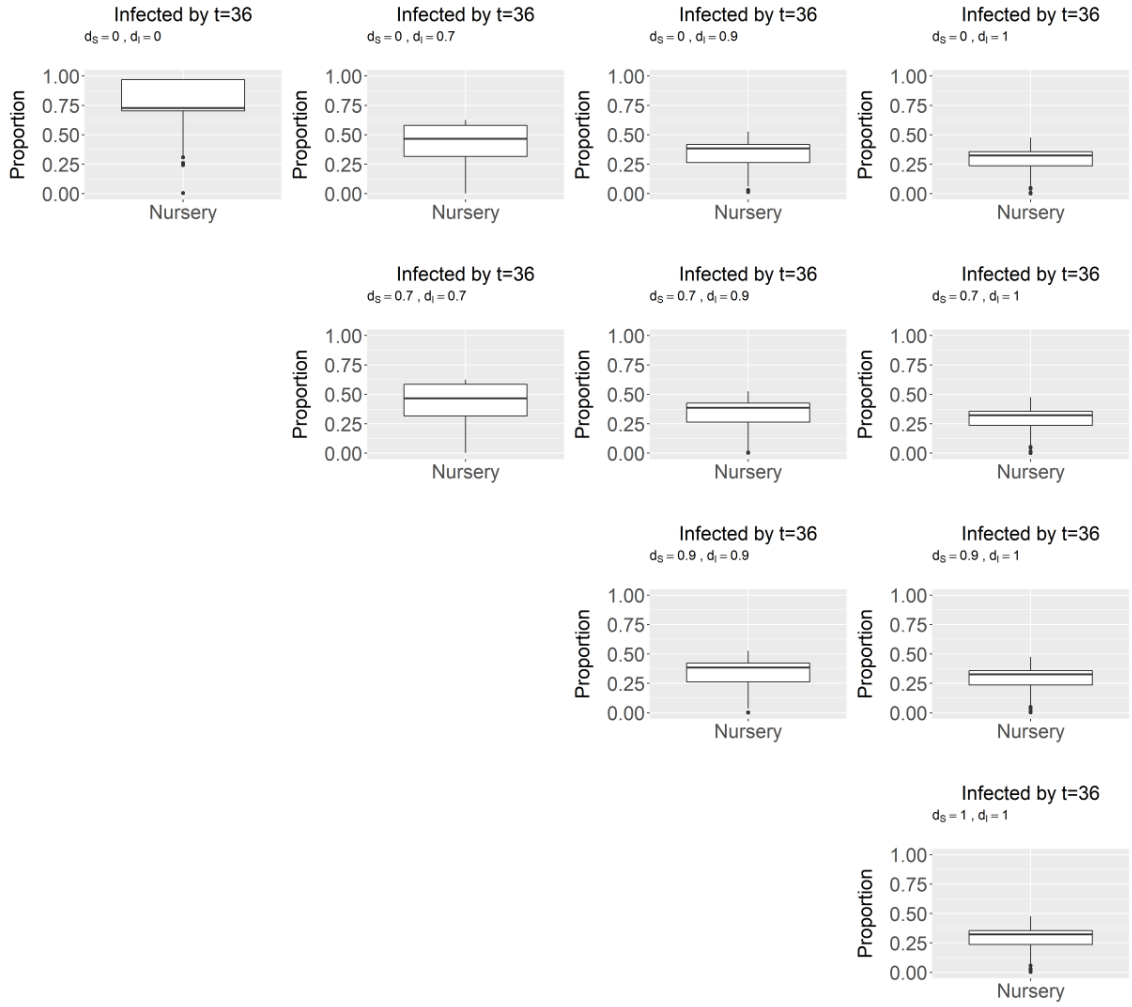
Figure 4.10. Boxplots of the proportion of infected plants by time $t = 36$ for nursery nodes. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. For this boxplot, we averaged the results for all 4 seedings. This is for nursery distribution scenario 1 and all other parameters are at baseline (Table 4.1).

The results for scenarios 2 and 3 (Figure 4.11 Figure 4.12) are mostly the same as scenario 1, except for one key difference. For both of these figures, we see that the upper and lower quartiles for $d_S = d_I = 0$ overlap with the boxplots for $d_I = 0.7$. Thus, a change in d_I from 0 is only seen to be significant in these scenarios for $d_I \in \{0.9, 1\}$. This implies that in a market focused towards commercial (scenario 2) or consumers (scenario 3), disease spread is only significantly different with pests or diseases that reduce the life expectancy to shorter than 6 months ($d_I \geq 0.9$).
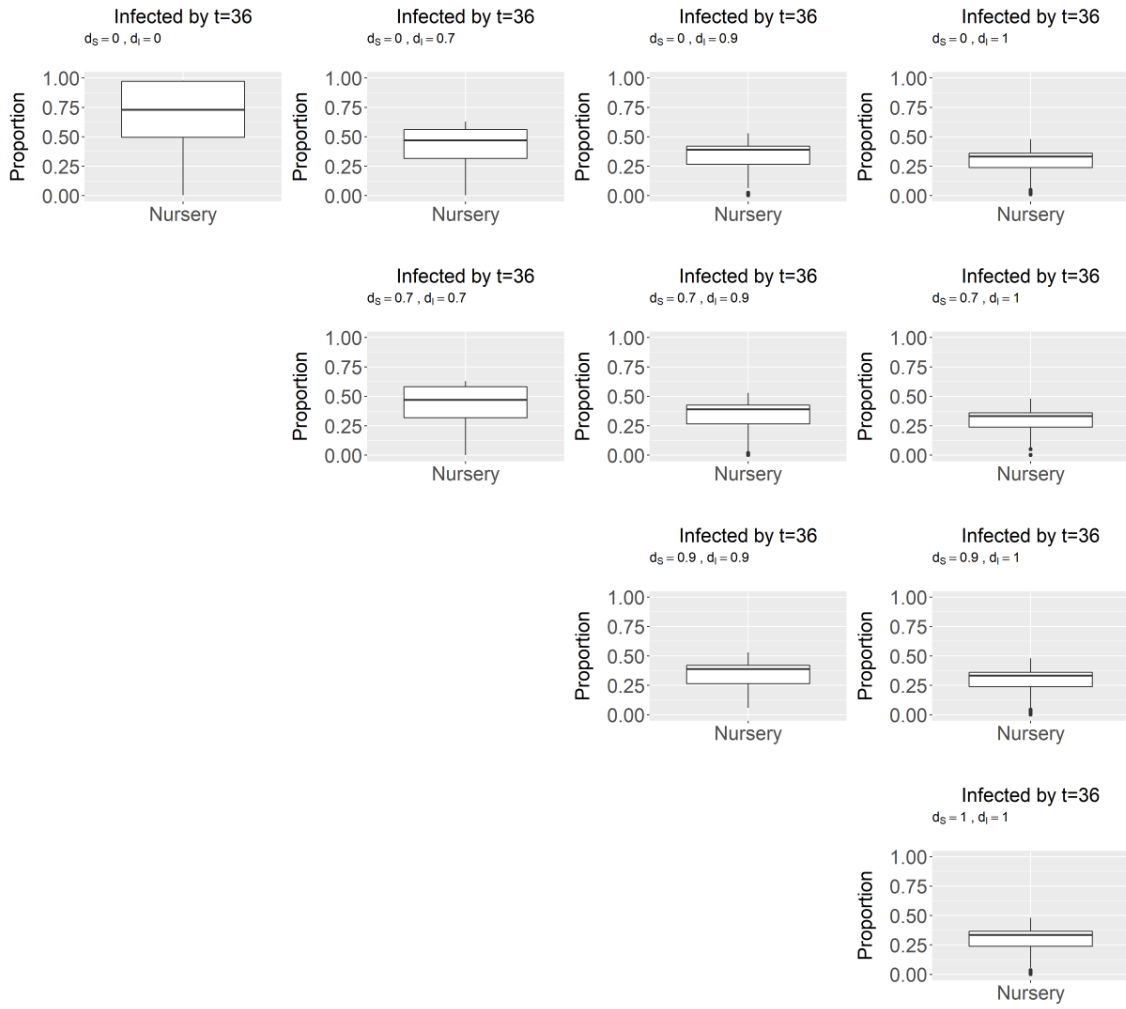
Figure 4.11. Boxplots of the proportion of infected plants by time $t = 36$ for nursery nodes. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. We average the results for all 4 seedings. This is for nursery distribution scenario 2 and all other parameters are at baseline (Table 4.1).
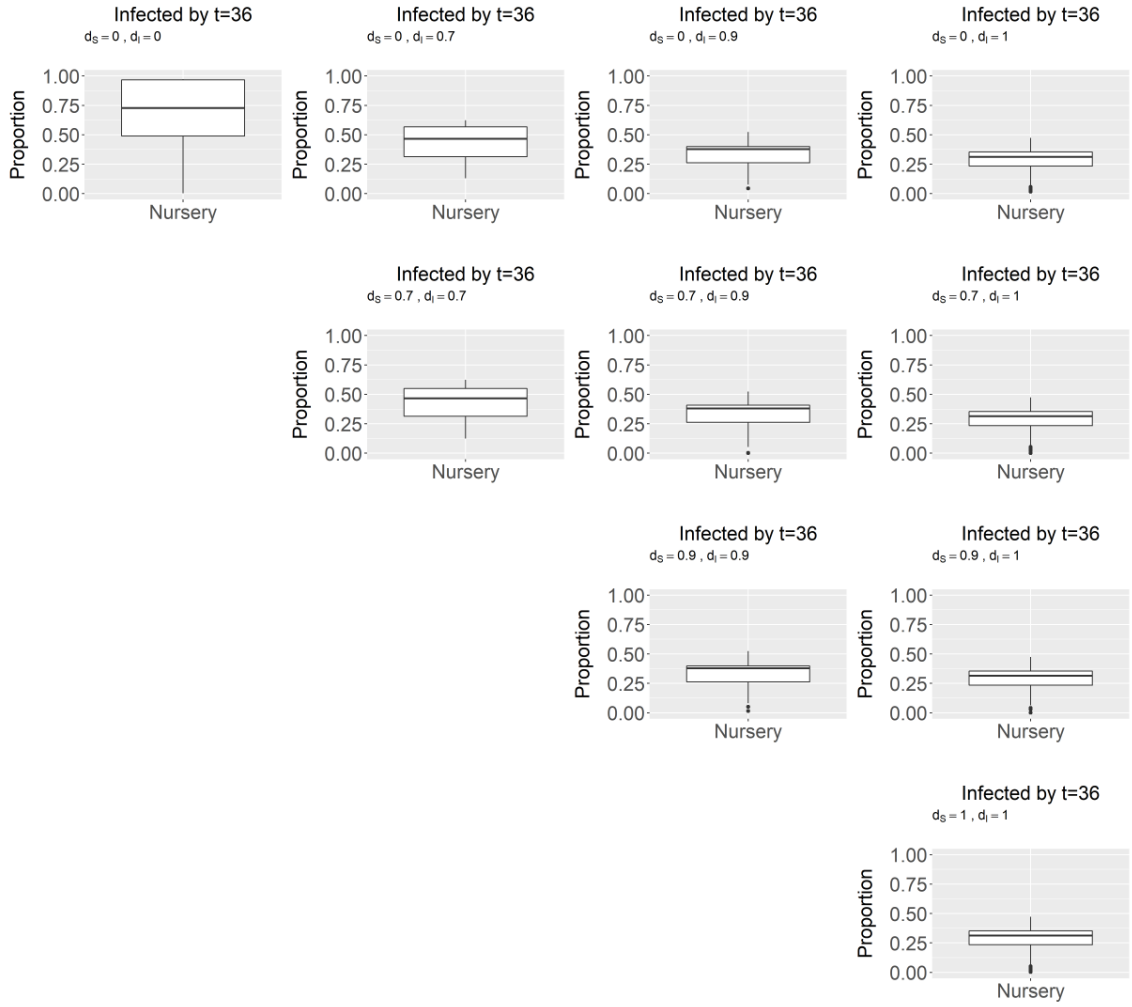
Figure 4.12. Boxplots of the proportion of infected plants by time $t = 36$ for nursery nodes. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. For this boxplot, we averaged the results for all 4 seedings. This is for nursery distribution scenario 3 and all other parameters are at baseline (Table 4.1).

Figure 4.13, Figure 4.14 and Figure 4.15 display the time until 20% of nurseries are infected for each scenario. The upper and lower quartiles for all 3 scenarios overlap and so we can say there is no significant difference between scenarios for a given combination of d_S and d_I.

Figure 4.13 shows the effect of varying d_S and d_I for nursery scenario 1. For increases in d_I, the time until 20% infected increases (Figure 4.13). The upper and lower quartiles do not overlap when changing $d_I \in \{0, 0.7, 0.9\}$ and are thus significant. The upper and lower quartiles for an increase of d_I from 0.9 to 1 overlap and so are not significantly different in

Figure 4.13. It is surprising that a disease that kills a plant in 6 months does not spread much faster than a disease that kills the plant in 1 month. The significant difference in disease spread appears to be between diseases which kill the plant in a year or longer and 6 months or less. An interesting result, for scenario 1, changing the death rate d_I from 0.7 to 0.9 significantly affects the time until 20% infected but insignificantly affects the final level of infection by $t = 36$. This implies that the pests or diseases that kill the plant quickly will be slower to begin epidemic spread but not have smaller epidemic sizes.

Again, increases in d_S do not change the results, for a fixed d_I (Figure 4.13). The higher d_S is, the more plants are removed; however, these removed are then replaced by susceptible plants. Therefore, in terms of the proportion of infected plants there is no change. In the same way, the model is more sensitive to the changes in d_I because all infected plants are replaced by susceptibles.
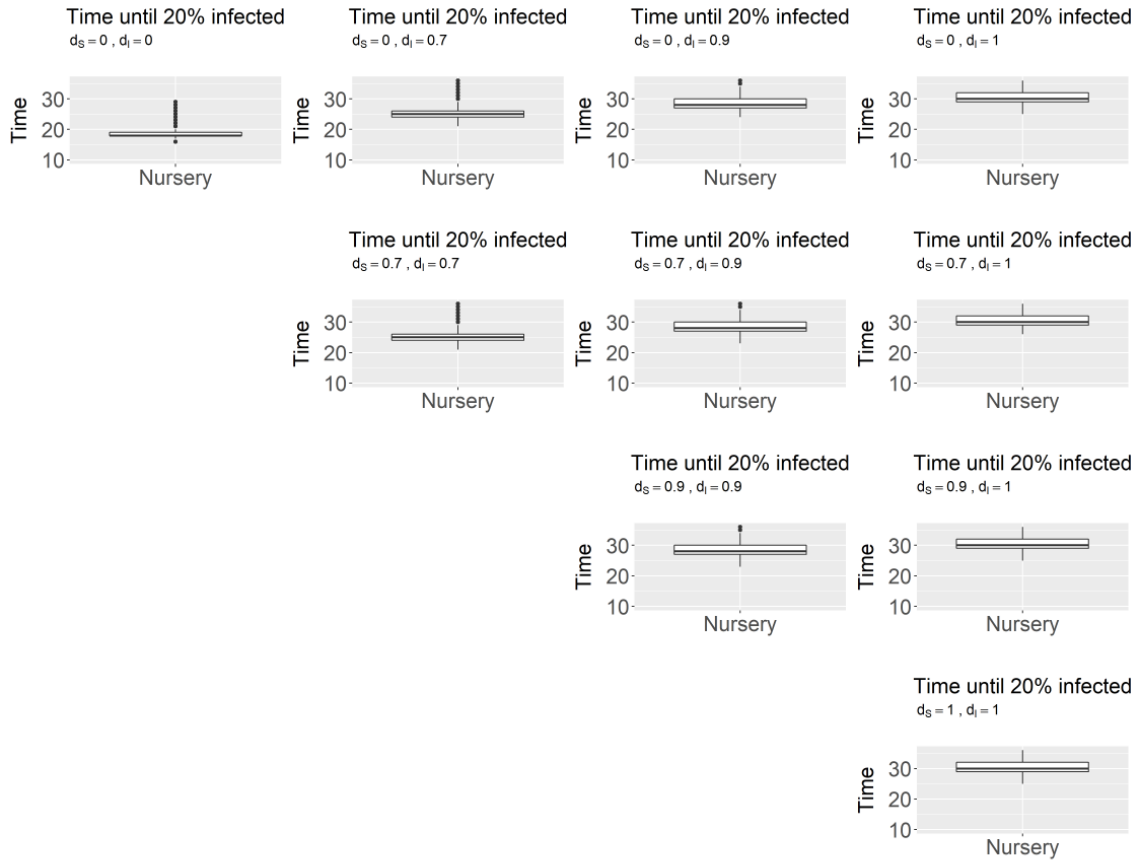
Figure 4.13. Boxplots for the time t until 20% of the plants in nurseries are infected. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. This is for nursery distribution scenario 1.
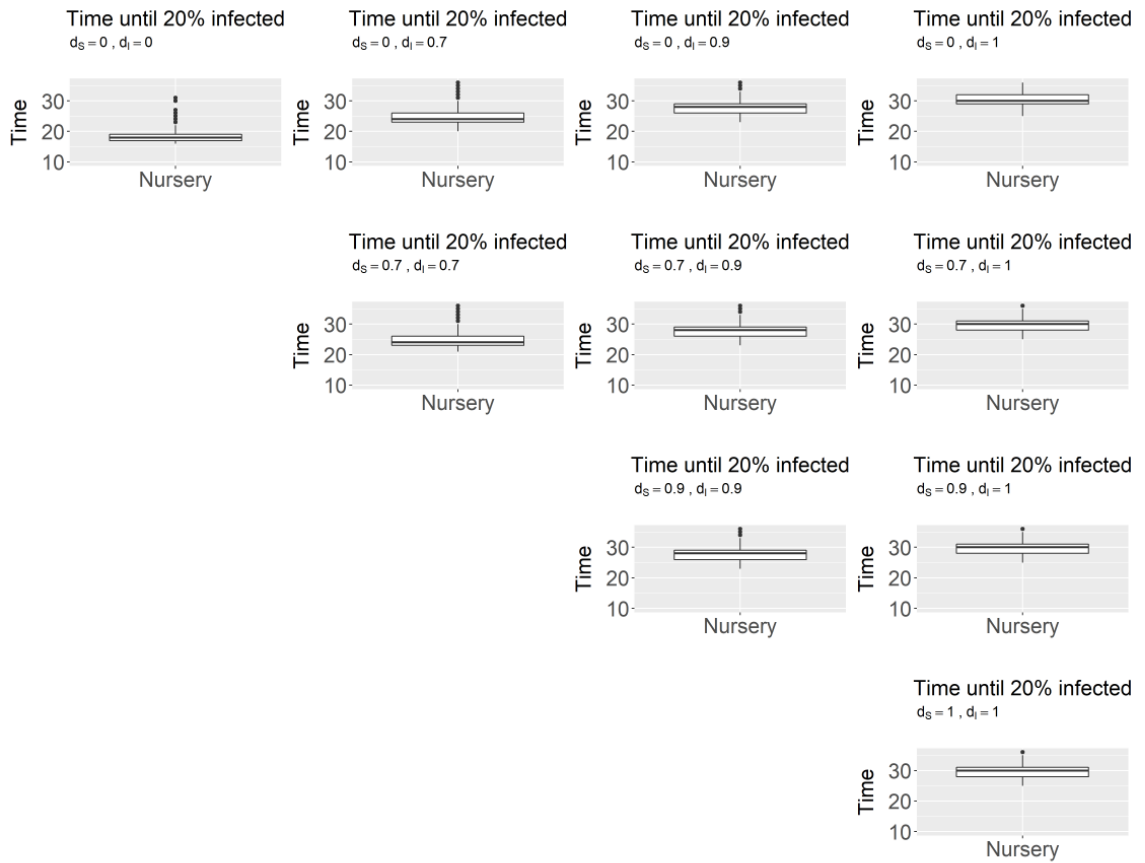
Figure 4.14. Boxplots for the time t until 20% of the plants in nurseries are infected. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. This is for nursery distribution scenario 2.
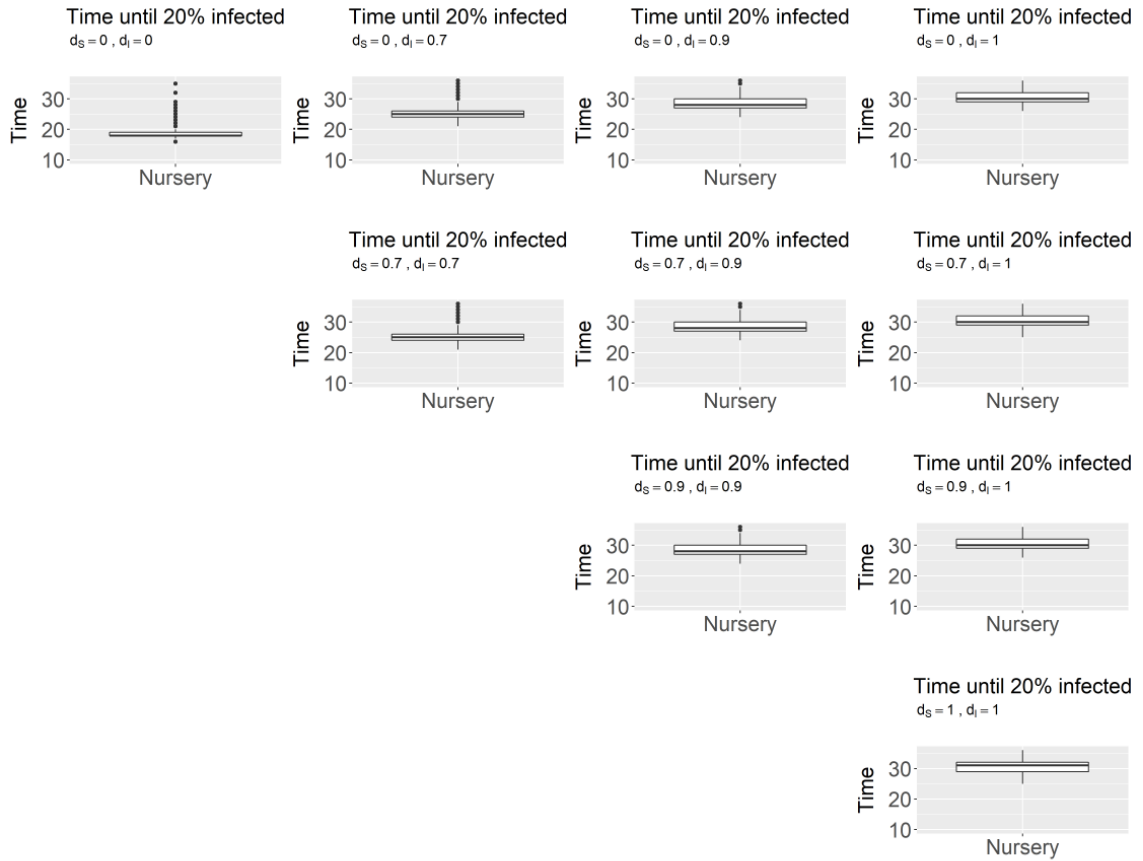
Figure 4.15. Boxplots for the time t until 20% of the plants in nurseries are infected. The boxplots are displayed on a grid, for each considered combination of death rates d_S and d_I. This is for nursery distribution scenario 3.

This ends the analysis of disease spread in a network without any intervention. From the results so far, we see that the pest spreads throughout the network from starting at a nursery, by spreading amongst nurseries before reaching retailers. Commercial nodes receive infected plants in tandem with nurseries and in proportion to the current prevalence of disease within nurseries. Once there is enough disease within nurseries, disease begins to spread to and within retailers. Consumer nodes receive infected plants from retailers, in accordance to the proportion of infected plants within retailers. In agreement with our results from Chapter 3, there is little difference when varying the nursery distribution in scenarios 1, 2 and 3. We do see a difference when changing which nursery group is seeded with the pest or disease. Seeding in N_{Nur} nurseries, i.e., nurseries that sell mostly to other nurseries, is consistently qualitatively different than N_{Com}, N_{Cons} and N_{Ret} seedings.

Seeding in N_{Nur} results in higher infection levels with less variability compared to all other seedings. We see that seeding in N_{Ret}, i.e., nurseries that mostly sell to retailers, appears to result in the slowest disease spread; however, this may be influenced from model assumptions. Variation of the parameter β shows that the disease spreads for $\beta > 1$, with further increases in $\beta > 2$ having a smaller effect on the model. We found that our results do not change for variations of d_S, stemming from model assumptions, and that increases in d_I slow down disease spread.

Next we investigate the addition of interventions to disease spread, starting with scheduled inspections that remove a proportion of infected plants in each nursery and retailer.

4.6.2   Introducing scheduled inspections

In this section, we varied the parameter $r \in \{0, 0.1, 0.2, \cdots, 1\}$, which controls the effectiveness of scheduled inspections. As a baseline, we fixed the parameter which controls the frequency of these scheduled inspections $\tau^{insp} = 6$. The time-point of first inspection is also fixed at $z = 6$.

We can see that increases in inspection effectiveness, r, has a decreasing effect on the proportion of infected plants in nurseries by $t = 36$ (Figure 4.16). This effect is seen for all seedings and nursery scenarios. There is no significant difference when changing nursery scenario. Therefore, we will focus our analysis on scenario 1.

Due to the high variability in the simulations for N_{Com}, N_{Cons} and N_{Ret} seedings, we see the error bars (25% and 75% quantiles) overlap. Therefore for these seedings, we see no significant difference in the final proportion of infected in nurseries for $0 \leq r \leq 1$. For values of $r > 0$ we see that N_{Nur} seedings are significantly different from all other seedings. For seedings in N_{Nur} there is no significant reduction in infection for $r \leq 0.8$. However, we do see a significant reduction for $r \leq 0.6$ and $r \geq 0.9$. There is also a significant reduction in infection for $r = 1$ and all other values, for N_{Nur} seedings. Therefore, we see that there are only significant differences in the final proportion of infection in nurseries

when changing r for seedings in N_{Nur} nurseries and large values of $r \geq 0.9$. Though the effect is significant, scheduled inspections do not have a large effect on reducing infection in nurseries. For example, changing r from 0 to 1 reduces the median proportion infected for N_{Com} and N_{Cons} seedings by just over 50%. This reducing effect is lowest for seedings in N_{Nur} with a change of r from 0 to 1 reducing the median proportion infected by 20%.
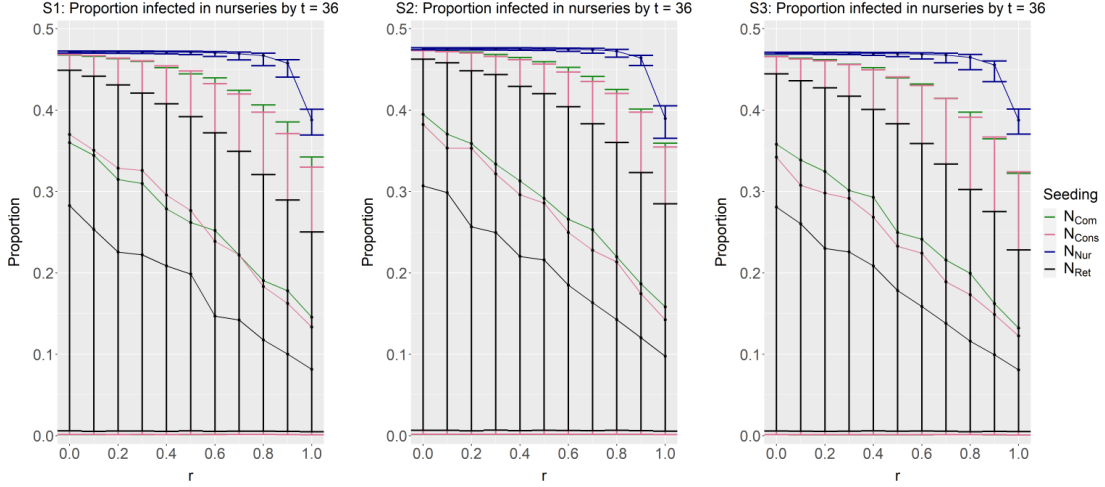
Figure 4.16. Points represent the median proportion of nurseries infected when $t = 36$, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given r value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1)

We also measure the time until 20% of plants in nurseries are infected for the same changes in r (Figure 4.17). Firstly, across all nursery scenarios and seedings we consider, increases in r delays the time t until 20% of plants in nurseries are infected. Increasing the efficacy of scheduled inspections appears not to be very effective, delaying the spread of infection by at most 9 months. Comparing across nursery scenarios, we see mostly small changes in skewness and median values and the overall trend remains the same for changes in r, per seeding. We do see some small overlaps in error bars for N_{Nur} seedings across scenarios, for example, in scenario 1 we see an overlap of the error bars between N_{Nur} and other seedings for $r \in \{0.8, 1\}$. However, for most values of r considered, we see a significant difference in the time until 20% of plants in nurseries are infected between N_{Nur} and all other seedings. For scenario 1 in Figure 4.17, we see that changing r from 0 to 1 reduces

the median time for N_{Com} and N_{Cons} seedings by approximately 16%. For N_{Ret} and N_{Nur} seedings, this reduction in the median is approximately 13% and 20% respectively. Therefore, we see that increases in r have a smaller effect on the speed in which the disease spreads (Figure 4.17) compared to reducing the final epidemic size (Figure 4.16). This is intuitive, as the scheduled inspections only affect the model every 6 time-steps, leaving the disease to spread uninterrupted at all other times.
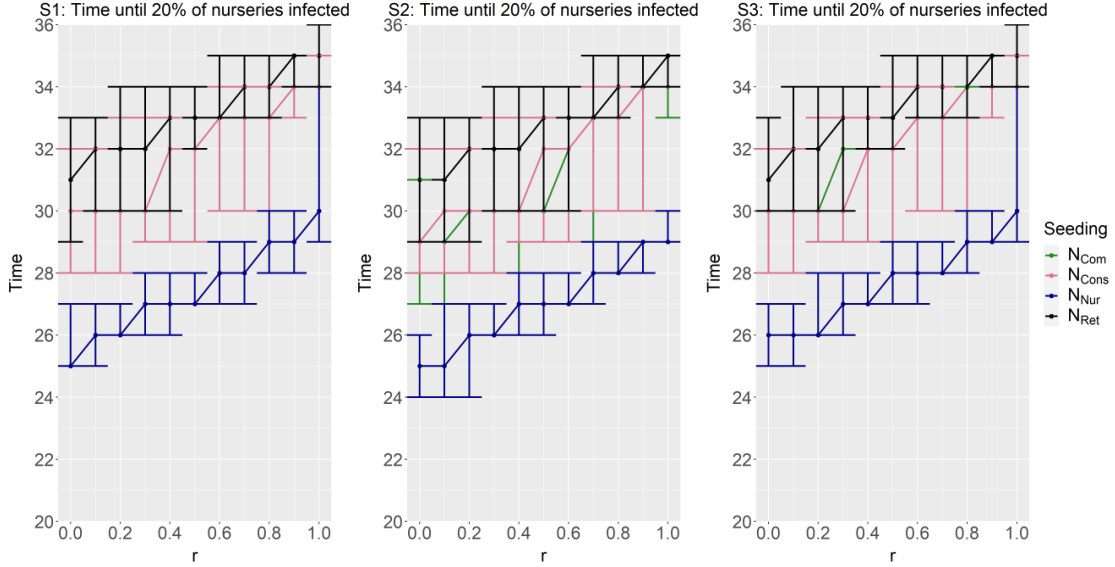
Figure 4.17. Points represent the median time until 20% of plants across nurseries are infected, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given r value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (a), (b) and (c) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

The results from these figures suggest that inspections are not an effective means of controlling disease spread, especially when the disease is seeded in N_{Nur} nurseries (Figure 4.16 and Figure 4.17). This may be due to the first inspection happening too late ($z = 6$) or the inspections not occurring frequently enough ($\tau^{insp} = 6$). This will be seen in a future section.

Next, we consider in isolation, the restrictive control measure of trade inspections.

4.6.3   Introducing trade inspections

In this section we look at two different variations of trade inspections as a form of disease control. Firstly, we investigate the case where plants are only inspected by the seller and not the buyer, i.e., $\alpha^{out} > 0$ and $\alpha^{in} = 0$, for nurseries and retailers. The second case is when inspections are conducted by both the buyer and the seller. For simplicity, we only consider when incoming inspections are as effective as outgoing inspections, i.e., $\alpha^{out} = \alpha^{in}$ for nurseries and retailers. For both cases, we vary the inspection parameters α^{out} and α^{in} in the range $\{0, 0.1, ..., 0.9\}$.

Outgoing inspections only

Considering only outward trade inspections, there are no significant changes to the proportion infected in nurseries by $t = 36$ across the 3 nursery scenarios for changes in α^{out} (Figure 4.18). Thus we focus our analysis of Figure 4.18 on nursery scenario 1. We see an overall decrease in infection as we increase α^{out}. Due to the overlapping quartiles, varying α^{out} does not have a significant effect on the proportion of infected plants in nurseries for seedings in N_{Com}, N_{Cons} and N_{Ret} (Figure 4.18).

However, changes in α^{out} for seedings in N_{Nur}, do have a significant effect. For $0.1 \leq \alpha^{out} < 0.5$, the reduction in infection is significant, though small. For $\alpha^{out} \geq 0.5$, a larger increase in α^{out} is required to have a significant effect, i.e., for $\alpha^{out} \in \{0.5, 0.7, 0.9\}$ we see non-overlapping quartiles. For $\alpha^{out} > 0.5$, infection decreases further but the variability in simulations also increases. Interestingly, as α^{out} increases, the variability in simulations increases for N_{Nur} seedings but decreases for all others. For $\alpha^{out} = 0$ increasing to 0.9, the median value for N_{Nur} seeding is reduced by approximately 20% whereas the other seedings are reduced by approximately 85%. Therefore, we see that outward trade inspections only are far more effective when the disease does not begin in a N_{Nur} nursery.
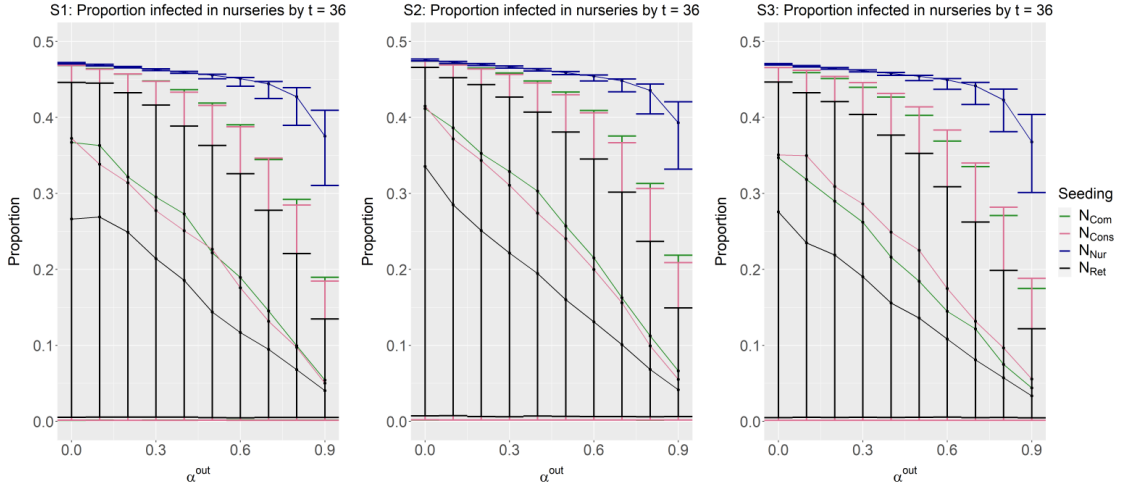
Figure 4.18. Points represent the median proportion of nurseries infected when $t = 36$, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given α^{out} value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

For each scenario, we see the same trend of increases in α^{out} causing a delay to the time until 20% infection in nurseries, with some minor changes in medians and quartiles (Figure 4.19). Similar to the previous measure, we do not see any significant difference between nursery scenarios and so we continue with our focus on scenario 1. Again, there is a significant difference between N_{Nur} and all other seedings. For seedings in N_{Com}, N_{Cons} and N_{Ret}, we only see α^{out} have a significant effect for $\alpha^{out} \leq 0.1$ and $\alpha^{out} = 0.9$, where the quartiles do not overlap. Thus, α^{out} significantly lowers the rate of spread amongst nurseries for these seedings for only very high inspection efficacy. However, this lower rate of spread will only delay the infection reaching 20% by at most 7 months.

Similarly for seeding in N_{Nur}, we only see significant effects of α^{out} for large values, for example, the quartiles for $\alpha^{out} \in \{0, 0.1, 0.2, 0.3, 0.4\}$ do not overlap with $\alpha^{out} = 0.9$. Thus if a disease starts in a N_{Nur} nursery, improving the efficacy of outgoing trade inspections is more effective, yet the disease still spreads faster and further than if the disease started elsewhere. In contrast with the previous measure, as α^{out} increases, we see the gap between N_{Nur} and other seedings narrow and converge such that for $\alpha^{out} = 0.9$ we see overlapping

quartiles. Therefore, increased efficacy of outgoing trade inspections increases the differences between seedings for final infection prevalence but decreases the differences for the rate at which the disease spreads across the network.
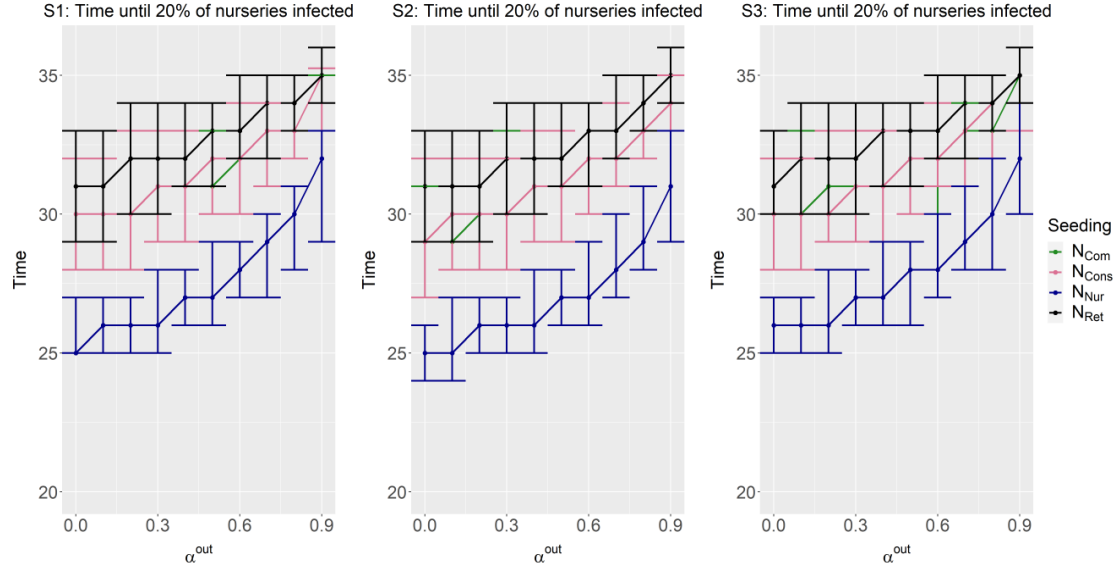
Figure 4.19. Points represent the median time until 20% of plants across nurseries are infected, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given α^{out} value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

So far, we see that outgoing trade inspections are slightly more effective at reducing disease prevalence than scheduled inspections, with the current baseline parameter values (Table 4.1). It is surprising that for inspection efficacy as high as 90% and 100%, trade inspections do not have a greater effect on controlling disease spread in the network. In the next section, we consider additional trade inspections, where nurseries and retailers inspect purchases in addition to sales.

Equal ingoing and outgoing inspections

When introducing both ingoing and outgoing inspections, we see the same patterns as we did for scheduled inspections and outgoing trade inspections only (Figure 4.20 and Figure 4.21). Increases in inspection efficacy lead to less disease prevalence by $t = 36$ (Figure 4.20) and a slowing down of disease spread (Figure 4.21). We also do not see significant differences across scenarios, thus we focus our analysis on scenario 1 for these figures.

Again, we find no difference between N_{Com}, N_{Cons} and N_{Ret} seedings and N_{Nur} seedings are different from all others (Figure 4.20). The quartiles for N_{Nur} do not overlap with other seedings, indicating a significant difference. For N_{Nur} seedings there is a significant effect in the final proportion infected in nurseries for $\alpha^{out} = \alpha^{in} \leq 0.4$. For increases in inspection efficacy $\alpha^{out} = \alpha^{in} > 0.5$, the effect of trade inspections increases for N_{Nur} seedings but decreases for all other seedings.

For α^{in}, for N_{Com}, N_{Cons} and N_{Ret} seedings, there is a non-linear effect on final infection prevalence by $t = 36$ for increases in $\alpha^{out} = \alpha^{in}$, when compared to the linear effect we saw in Figure 4.18. For larger inspection efficacy ($\alpha^{out} = \alpha^{in} > 0.6$), we see a smaller reduction in the final proportion infected decrease. For N_{Nur} seedings, we still see the nonlinear effect of increased efficacy of trade inspections; however, this increased reduction in final proportion infected is now stronger than in the previous case with only outgoing inspections. For a pest or disease starting in a N_{Com}, N_{Cons} or N_{Ret} nursery, increases in $\alpha^{in} = \alpha^{out}$ from 0 to 0.6 have the largest reducing effect on the final proportion infected in nurseries; however, when a pest or disease starts in a N_{Nur} nursery, we only see similarly sizes reductions on the final proportion infected in nurseries for $\alpha^{in} = \alpha^{out} > 0.5$.
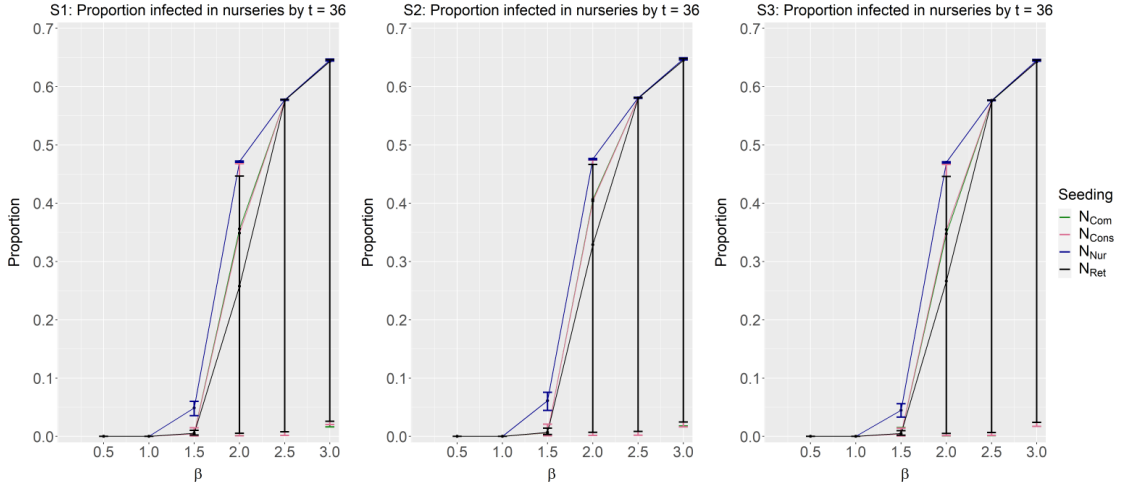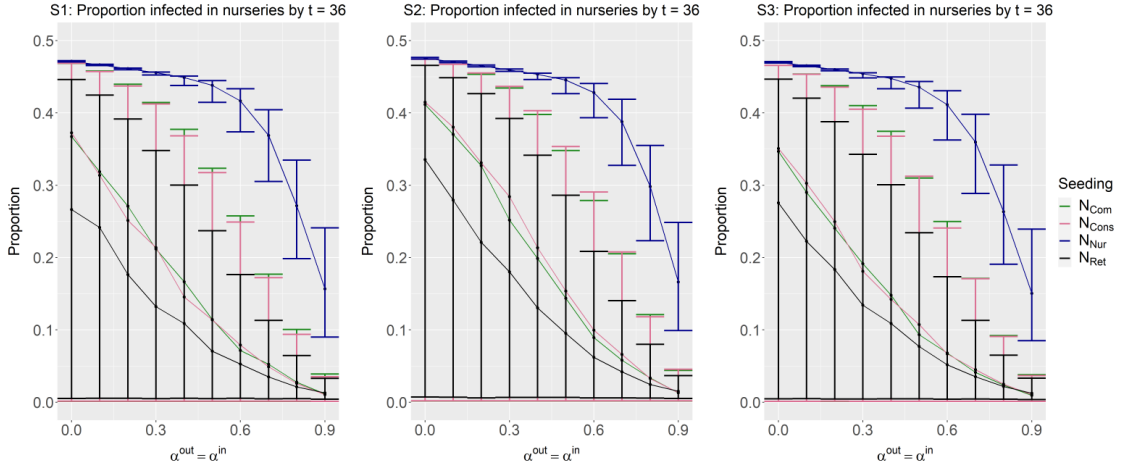
Figure 4.20. Points represent the median proportion of nurseries infected when $t = 36$, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given $\alpha^{in} = \alpha^{out}$ value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

The only differences in seeding for changes in $\alpha^{in} = \alpha^{out}$ regarding the time until 20% of nurseries are infected are in N_{Nur} nurseries and all other seedings for $\alpha^{in} = \alpha^{out} \leq 0.3$ (Figure 4.21). For $\alpha^{in} = \alpha^{out} > 0.3$, all seedings converge together. With the addition of ingoing inspections (Figure 4.21), there is less dependence on where the disease starts in regard to the rate of infection, compared to outgoing inspections only (Figure 4.19). However, the added ingoing inspection does not slow the rate of infection significantly, for a given success rate. For example, the introduction of α^{in} of equal value to α^{out} slows the median time until 20% infection in nurseries by at most 3 months.
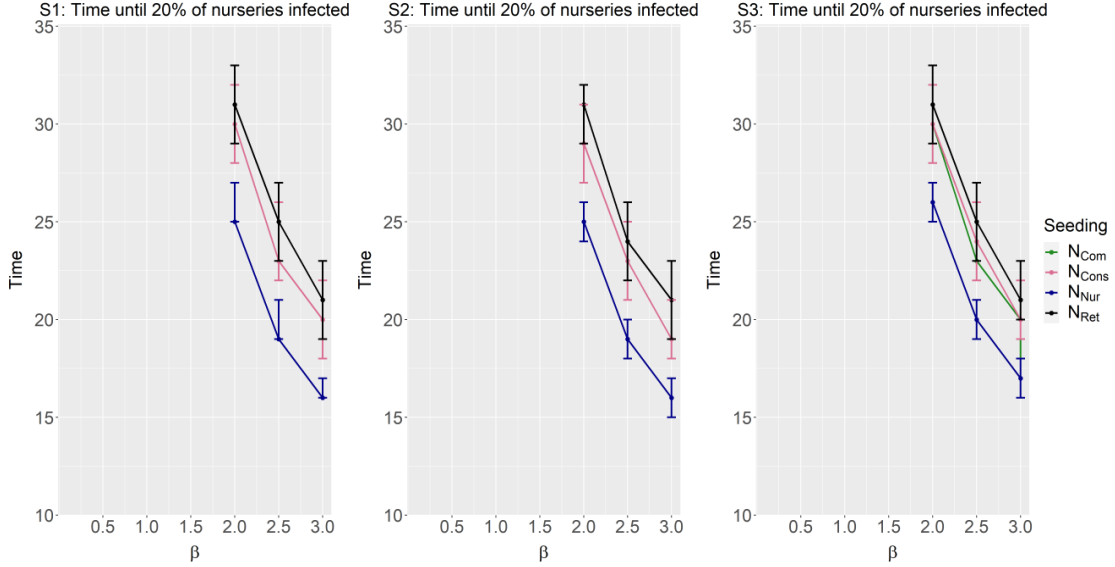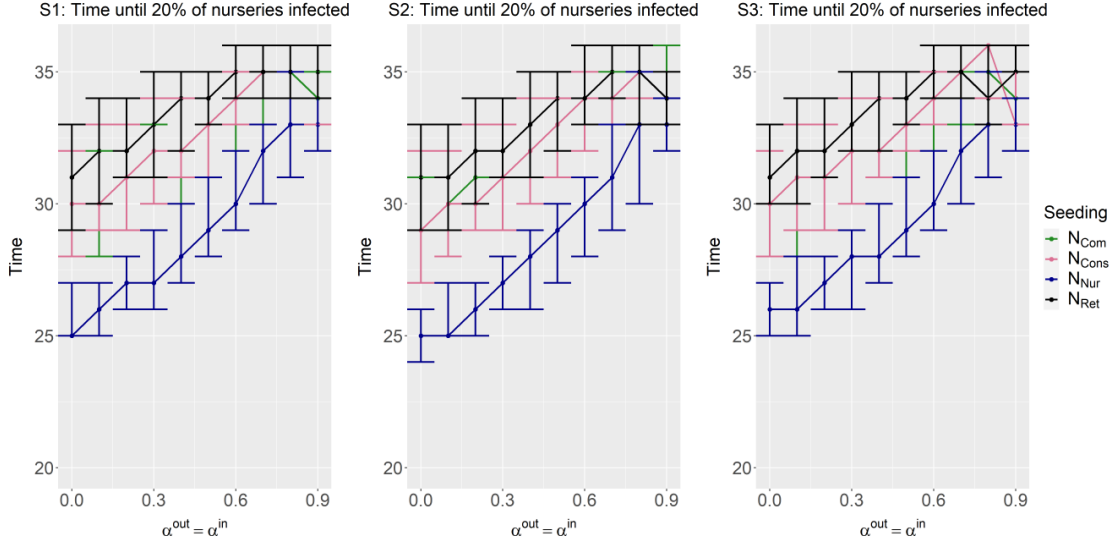
Figure 4.21. Points represent the median time until 20% of plants across nurseries are infected, calculated from 100 simulations repeated for 100 network simulations. The x-axis shows the given $\alpha^{in} = \alpha^{out}$ value for nurseries and retailers used in the simulations. The y-axis represents the proportion of plants in nurseries that are infected. Error bars indicate the 25% and 75% quantiles and each colour represents which nursery group the disease is seeded in at $t = 0$. Subgraphs (S1), (S2) and (S3) represent the different nursery distributions considered, scenarios 1, 2, and 3 respectively. All other parameter values are at baseline (Table 4.1).

We see that the addition of inspections from buyers increases the effectiveness of trade inspections (Figure 4.20 and Figure 4.21), when compared to only having inspections for outgoing consignments (Figure 4.18 and Figure 4.19). Trade inspections are also more effective at controlling and slowing disease spread than our baseline implementation of scheduled infections. However, it still appears unrealistic that either of these types of inspection, on their own, could successfully control an outbreak. Therefore, in the next section, we consider both trade and scheduled inspections, at varying efficacy combinations.

4.6.4 Comparing scheduled and trade inspections

In this section, we look at implementing both trade inspections and scheduled inspections to see the effect they both have on disease spread. We look at the two different implementations of trade inspections: outgoing inspections only and both outgoing and ingoing inspections. We vary the trade inspection parameters $(\alpha^{in}, \alpha^{out})$ in the range $\{0, 0.1, ..., 0.9\}$ and

the scheduled inspections parameter (r) in the range $\{0, 0.1, ..., 1\}$. Considering each combination of values for these parameters, we construct heat maps using each of our measures.

First we consider outgoing inspections only and look at the final proportion of infected in nurseries by $t = 36$ (Figure 4.22). There is a nonlinear trade-off for increasing α^{out} against r for reducing the final proportion of infection in nurseries. The two control measures appear to act as substitutes. For larger values of r, an increase in α^{out} has smaller reductions in the final proportion infected. This is seen for all seedings, however for N_{Nur} seedings, values of $r \geq 0.9$ or $0.6 \leq \alpha^{out} \leq 0.9$ are required for a noticeable decrease in the final proportion infected. For N_{Com}, N_{Cons} and N_{Ret} seedings, we see that for very high trade inspection efficacy ($\alpha^{out} \in \{0.7, 0.8, 0.9\}$) and most values for scheduled inspection efficacy ($r \geq 0.2$), the disease has been eradicated in nurseries by $t = 36$.
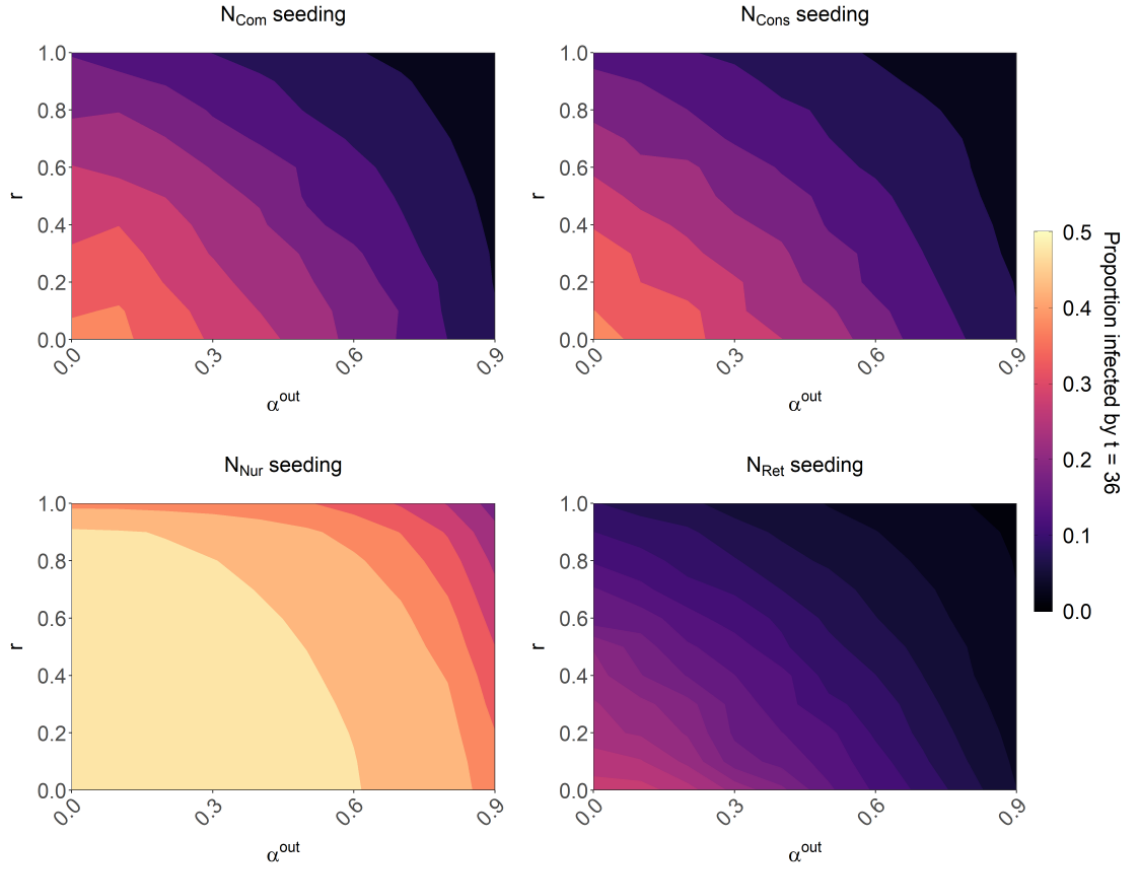
Figure 4.22. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, 0.1, ..., 1\}$. All other parameters are at baseline Table 4.1

Whereas, for the time until 20% of nurseries are infected, there is a linear trade-off between α^{out} and r (Figure 4.23). The isoclines are negatively sloped and shallow for all seedings, meaning that scheduled inspections are slightly more effective than trade inspections at slowing down rate of spread throughout nurseries. Again, we have that for N_{Nur} seedings, we require very large values for α^{out} and r to slow the rate of spread to a similar level as other seedings. The white regions in Figure 4.23 show parameter combinations where in all simulations, the disease never reaches 20% disease prevalence in nurseries. Thus, we see the disease remains at very low prevalence for $r \in \{0.8, 0.9, 1\}$ and $\alpha^{out} \in \{0.8, 0.9\}$.
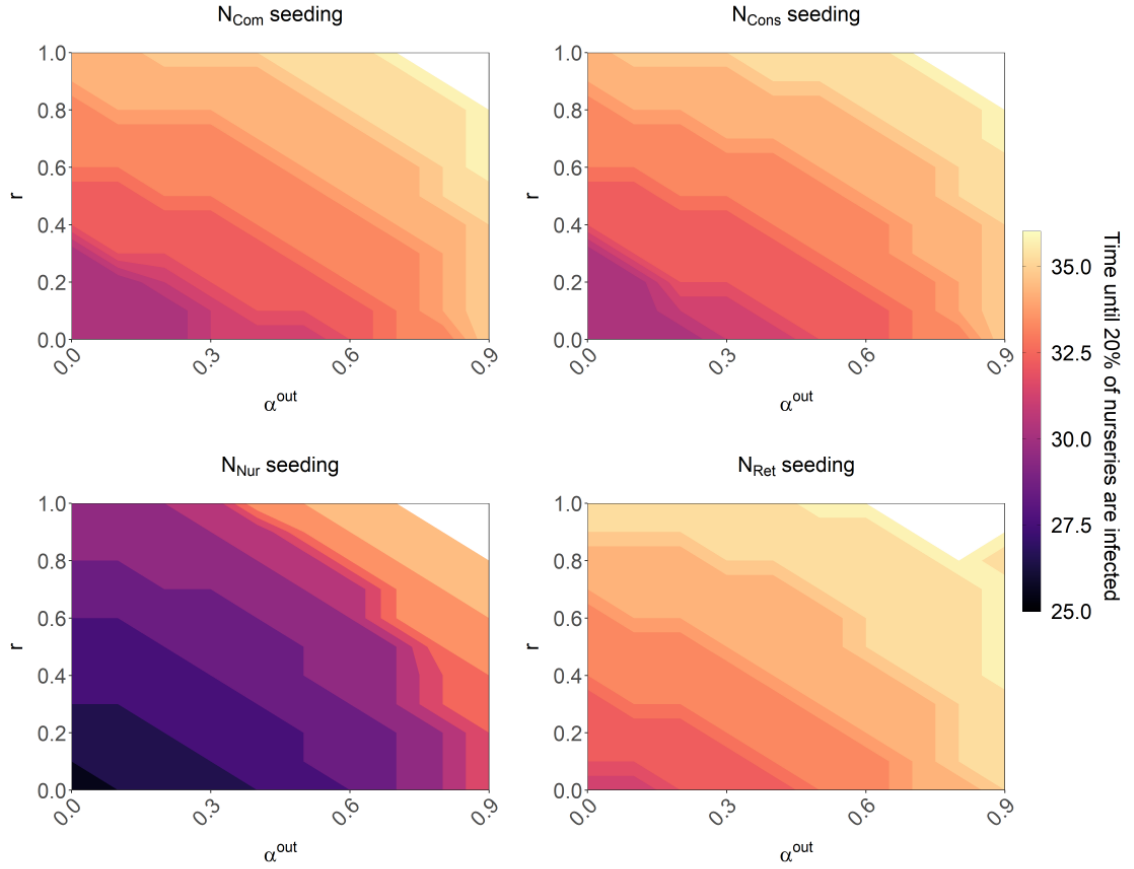
Figure 4.23. Heat map showing the median value for the time, t, when the proportion of infected plants in nurseries reaches 20%, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, 0.1, ..., 1\}$. Note the white sections of the heat map show the parameter combinations where the model does not reach 20% infected in nurseries by $t = 36$. All other parameters are at baseline Table 4.1

Thus, we see that trade inspections are slightly more effective than scheduled inspections at controlling disease spread, with little gain from implementing both control measures. Outbreaks are also much harder to control if starting in a N_{Nur} nursery. Though we identify some parameter combinations that eradicate disease in nurseries, these efficacy levels are very high and unlikely to be achievable in practice. Thus we repeat this analysis with trade inspections for both sales and purchases.

With equal ingoing and outgoing inspections, the nonlinear trade-off between trade inspections against scheduled inspections persists, though the isoclines are steeper, indicating a stronger effect of trade inspections (Figure 4.24). Disease-free regions still do not exist for N_{Nur} seedings, though the increases in $\alpha^{in} = \alpha^{out} > 0.3$ have a larger effect on disease reduction, with the greatest increases when $\alpha^{in} = \alpha^{out} > 0.6$. For N_{Com}, N_{Cons} and N_{Ret} seedings, the disease-free regions are larger. These disease-free states can be achieved without scheduled inspections, with $\alpha^{in} = \alpha^{out} \geq 0.7$. By including scheduled inspections, disease-free states can be achieved with $r = 1$ and $\alpha^{in} = \alpha^{out} = 0.4$.
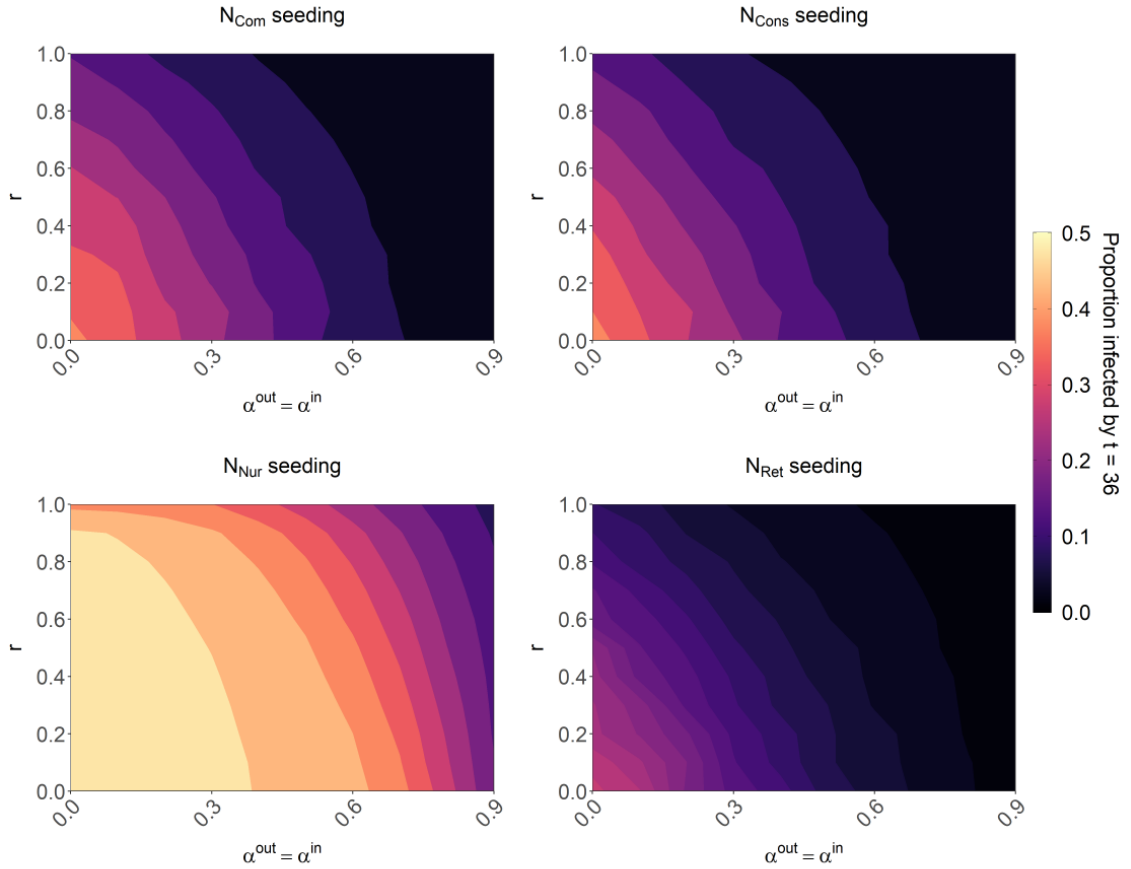
Figure 4.24. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, 0.1, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. All other parameters are at baseline Table 4.1

175

For the time until 20% of nurseries are infected, the addition of ingoing trade inspections renders these inspections more effective than scheduled inspections (Figure 4.25). There are also more parameter combinations where the disease remains below 20% prevalence for all simulations (white regions).
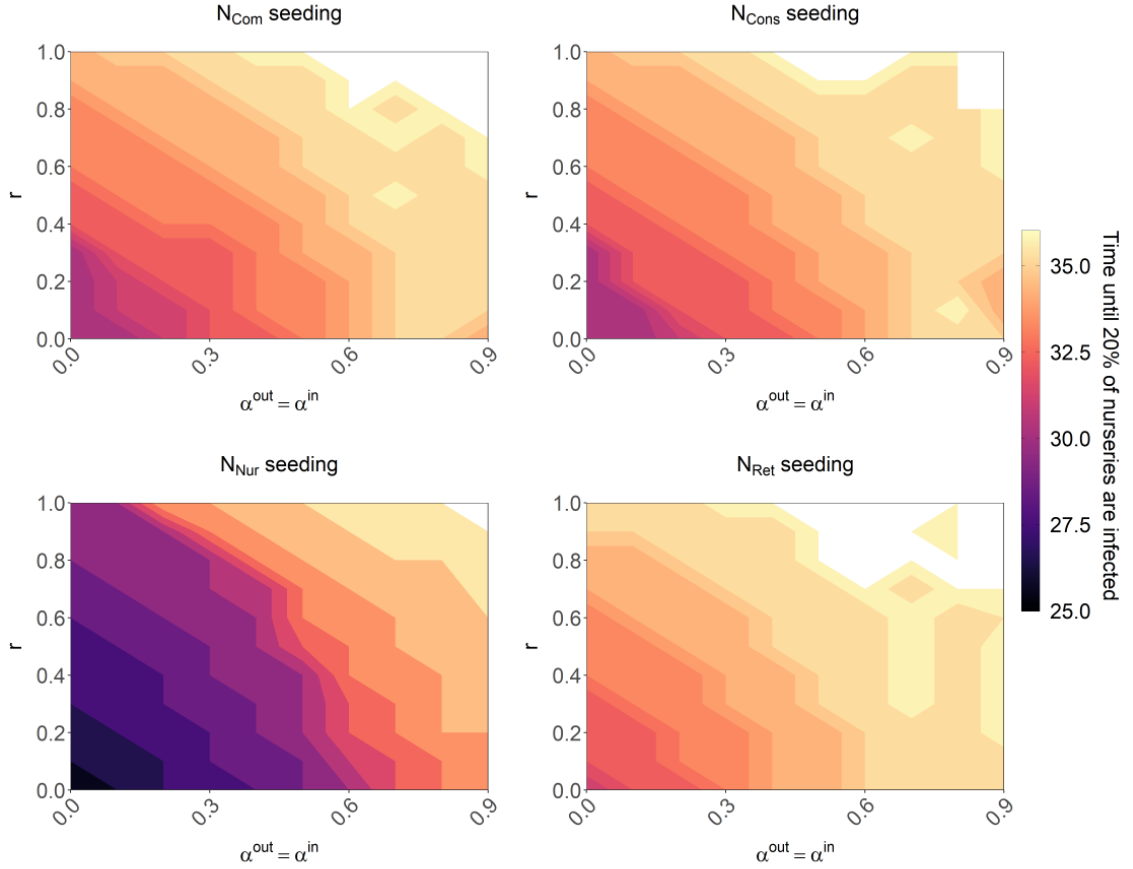
Figure 4.25. Heat map showing the median value for the time, t, when the proportion of infected plants in nurseries reaches 20%, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, 0.1, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. Note the white sections of the heat map show the parameter combinations where the model does not reach 20% infected in nurseries by $t = 36$. All other parameters are at baseline Table 4.1

Comparing the trade and scheduled inspections we find a trade-off between the two, where they act as substitute control measures. It is unrealistic to control disease spread without ingoing trade inspections due to the very high inspection efficacy required. Thus, the addition of inspecting purchases as well as sales during trade is a more achievable disease

control strategy. We see that controlling disease spread when the disease begins in a N_{Nur} nursery is much more challenging than other seedings for all combinations of inspection parameters. Thus, scheduled inspections that occur every 6 months, with the first inspection occurring at $t = 6$ are not effective at reducing spread.

In the next section, we investigate varying the frequency and time-point of first scheduled inspection to see how this may improve efforts to lower infection.

4.6.5 Varying inspection frequency and timing of first scheduled inspection

Initially, we fix the inspection frequency and change the first inspection from $z = 6$ to $z = 3$ in Figure 4.26. The general pattern remains the same as in Figure 4.24. Scheduled inspections are more effective at reducing spread for all seedings, most notably in N_{Nur}, where increases in $r > 0.4$ lowers the final proportion infected in nurseries by $t = 36$.
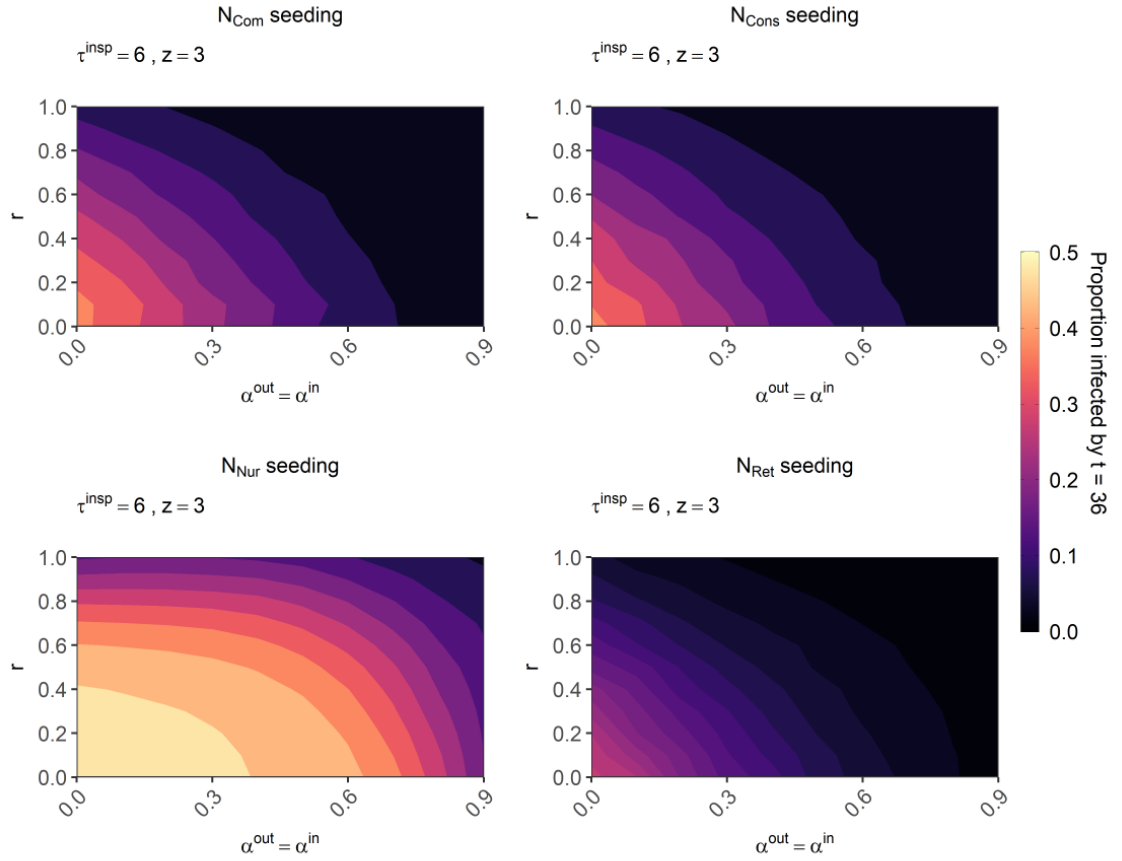
Figure 4.26. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. In addition, scheduled inspections occur every 6 months, with the first inspection occurring at $t = 1$. All other parameters are at baseline Table 4.1

Setting $z = 1$, we see an initially unintuitive result for the proportion infected in nurseries by $t = 36$ (Figure 4.27). This result is almost the same as $z = 6$ (Figure 4.24). This is because for $z = 6$, inspections occur at time-points $\{6, 12, 18, 24, 30, 36\}$ and for $z = 1$, inspections occur at $\{1, 7, 13, 19, 25, 31\}$. Since the simulations end at $t = 36$, we do not see the effect of the final inspection for $z = 6$, hence the last inspection that takes effect occurs at $t = 30$, which is very close to the final inspection for $z = 1$. Similarly with $z = 3$, with the measure we use there is effectively one more inspection taking place when compared to $z = 6$.
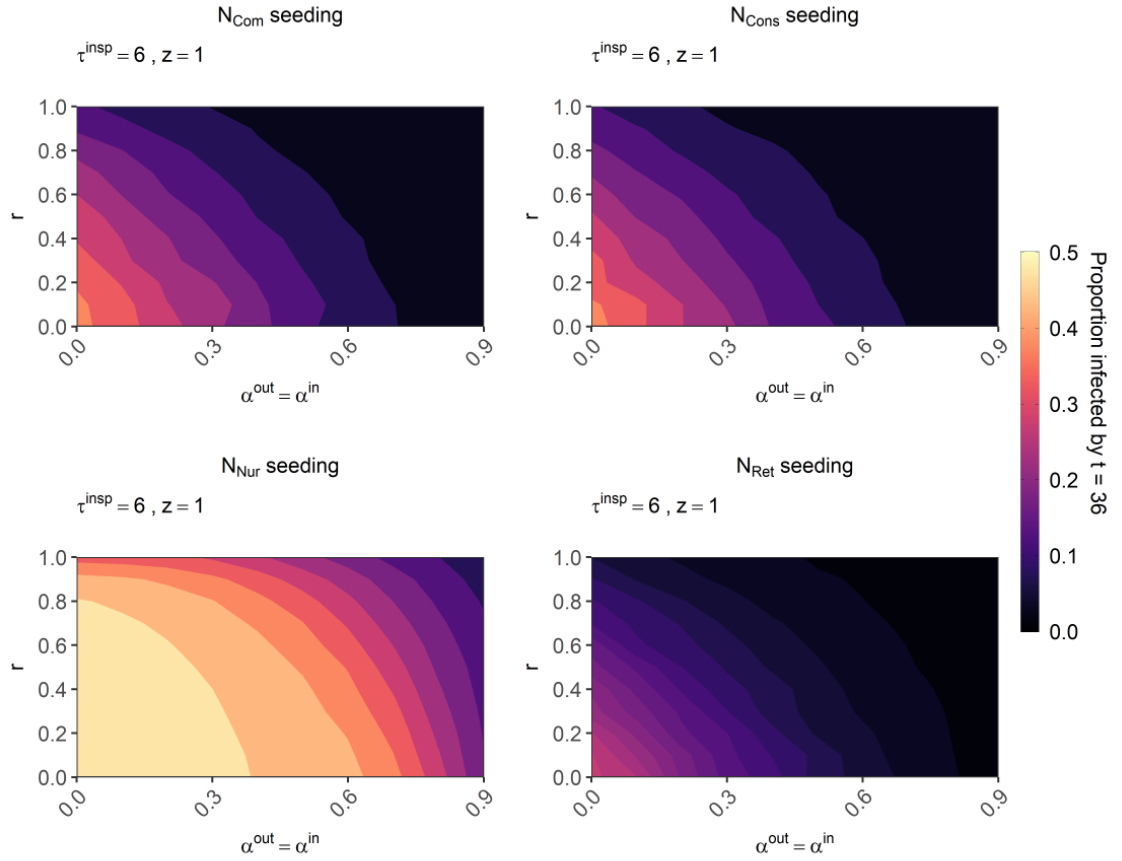
Figure 4.27. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. In addition, scheduled inspections occur every 6 months, with the first inspection occurring at $t = 1$. All other parameters are at baseline Table 4.1

Changing scheduled inspections to every 3 months ($\tau^{insp} = 3$), and first inspection to $z = 3$, increases in r have a greater reduction on the final proportion infected in nurseries (Figure 4.28), in comparison to $\tau^{insp} = 6$ (Figure 4.24). There is more of a linear trade-off between inspection parameters, when compared with $\tau = 6$ and $z = 3$ (Figure 4.26). We see that scheduled inspections are almost interchangeable with trade inspections (Figure 4.28). Bringing the time of the first inspection earlier to $z = 1$, there is little effect on N_{Com}, N_{Cons} and N_{Ret} seedings (Figure 4.29). However, for N_{Nur} seedings, scheduled inspections become more effective at lowering infection than trade inspections. For example, a combination of $[r = 0.4, \alpha^{out} = \alpha^{in} = 0]$ has the same outcome as $[r = 0, \alpha^{out} = \alpha^{in} = 0.7]$.
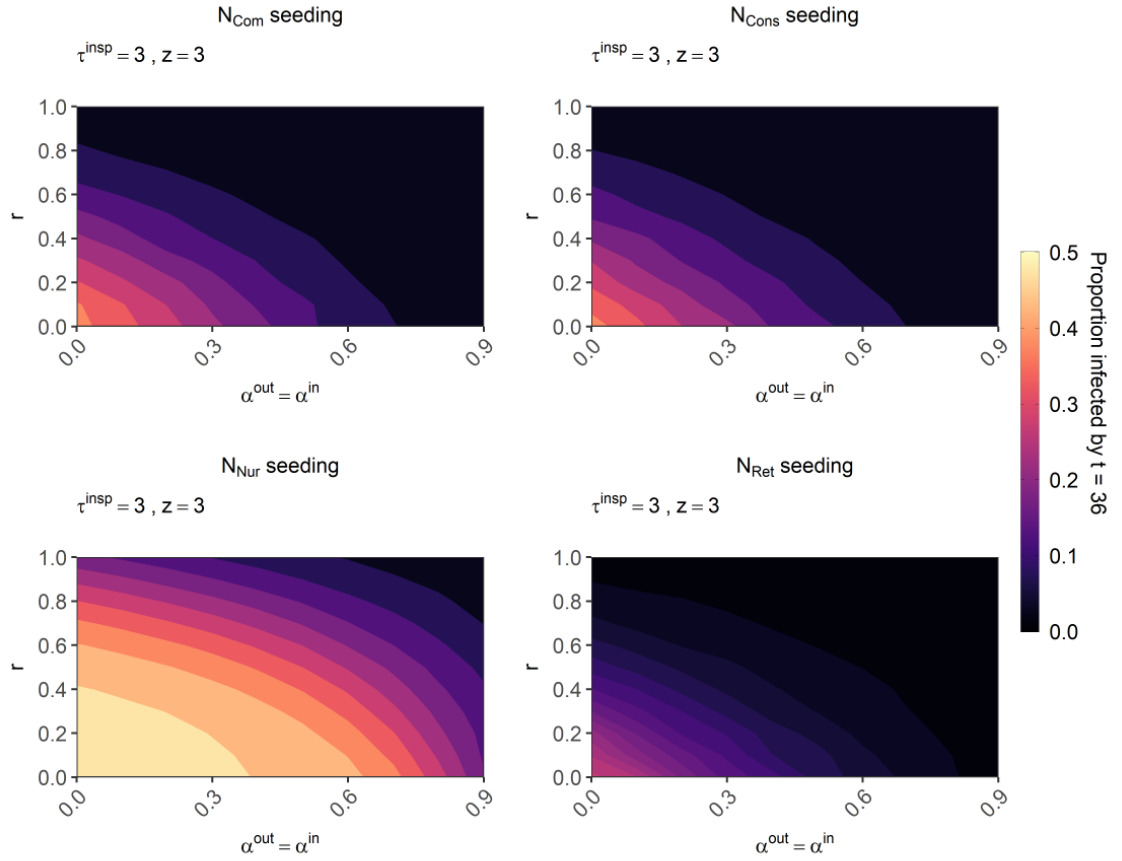
Figure 4.28. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. In addition, scheduled inspections occur every 3 months, with the first inspection occurring at $t = 3$. All other parameters are at baseline Table 4.1
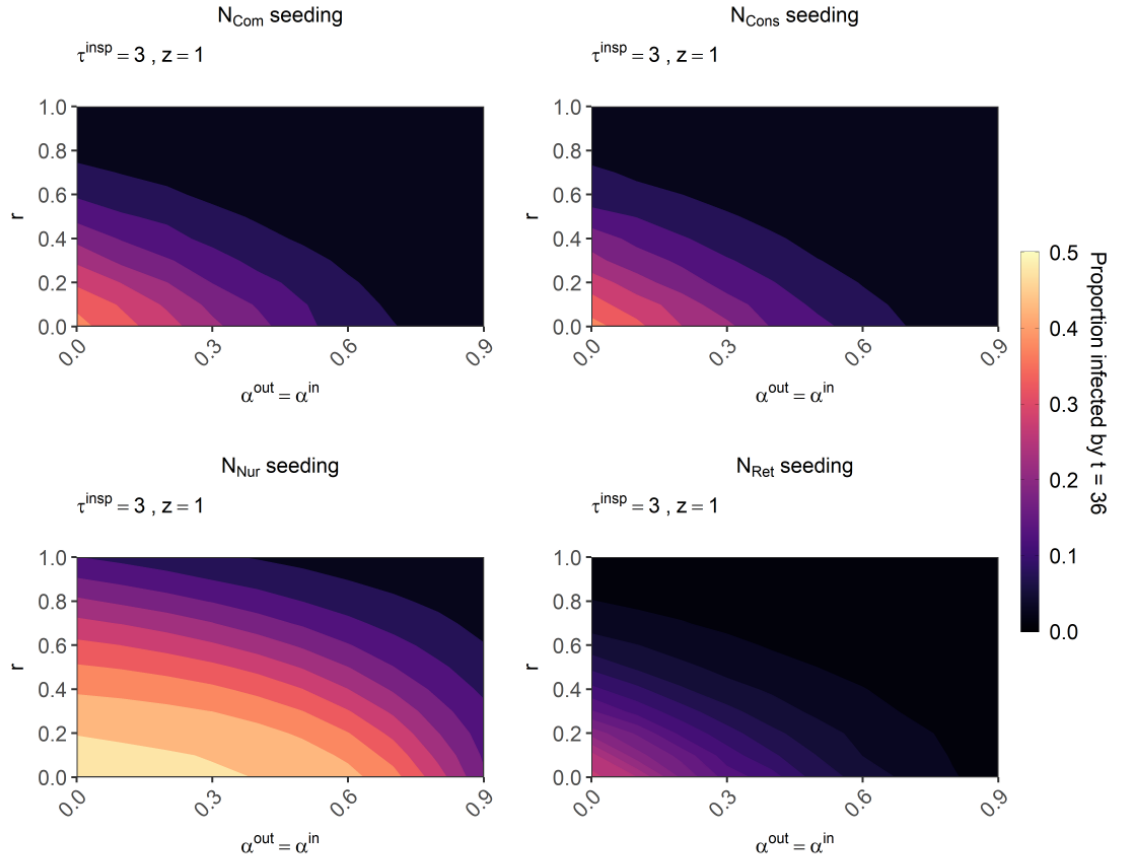
Figure 4.29. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. In addition, scheduled inspections occur every 3 months, with the first inspection occurring at $t = 1$. All other parameters are at baseline Table 4.1

When scheduled inspections occur every month, increasing the value of r becomes much more effective than trade inspections for reducing the final proportion of infected in nurseries (Figure 4.30). For seedings in N_{Nur} nurseries, a disease prevalence of approximately 0 can be achieved with $r \geq 0.5$ and no trade inspections. For all other seedings, this can be achieved for $r > 0.3$. There is now a very small effect for increases in $\alpha^{in} = \alpha^{out}$. For N_{Com}, N_{Cons} and N_{Ret} seedings, a disease prevalence of 0 can be achieved with $\alpha^{in} = \alpha^{out} > 0.6$. For N_{Nur} seedings, trade inspections alone cannot eradicate disease.
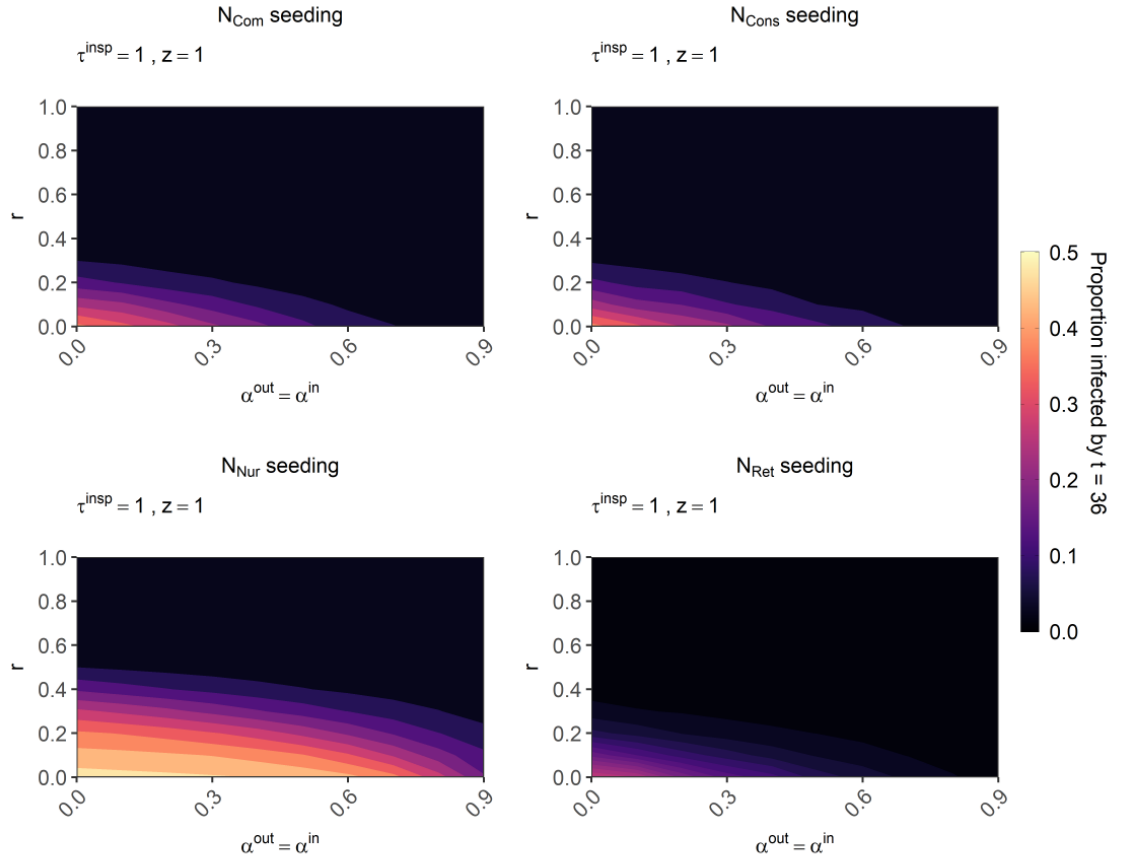
Figure 4.30. Heat map showing the median value for the proportion of infected plants in nurseries by $t = 36$, for 100 model simulations across 100 network simulations. The medians are calculated for changes in the values of trade inspection parameters $\alpha^{out}, \alpha^{in} \in \{0, 0.1, ..., 0.9\}$ and scheduled inspection parameter $r \in \{0, ..., 1\}$. Here we apply the assumption $\alpha^{out} = \alpha^{in}$. In addition, scheduled inspections occur every month, with the first inspection occurring at $t = 1$. All other parameters are at baseline Table 4.1

In this section we found that the measure we use (final proportion infected in nurseries) makes it difficult to determine how the timing of the first inspection affects the spread of disease. We showed this in our comparison of $[\tau^{insp} = 6, z = 6]$ and $[\tau^{insp} = 6, z = 1]$ (Figure 4.24 and Figure 4.27). With this in mind, we do see that increasing the frequency of inspections from every 6 months to every 3 months may not be beneficial, depending on how quickly the disease is spotted (Figure 4.26 and Figure 4.28). Unsurprisingly, scheduled inspections every month are the most effective form of control (Figure 4.30). Though, this control is so effective as to render trade inspections almost unnecessary. This result

suggests that it would be a more effective for plant nurseries to spend resources in order to inspect all nursery stock monthly, even at the cost of abandoning trade inspections.

Inspecting more frequently, though reducing disease spread, of course comes with an increase in costs to businesses. With limited resources, inspecting the entire nursery stock every month may not be feasible. Thus, in the next section we aim to take into account the cost of both disease and inspections.

4.6.6 Cost-benefit analysis of increasing scheduled inspection frequency

As detailed in Section 4.5, we calculate cost by:

$$\text{Cost}(X_3) = \sum_{i \in \boldsymbol{N}} I_i(36) + 160 \text{ floor}\Big(\frac{36 - z}{\tau^{insp}}\Big) X_3,$$

where X_3 is the cost of an inspection relative to the price of a plant. We compare $\tau^{insp} \in \{3, 6\}$ and $\tau^{insp} \in \{1, 3\}$ and find when inspections become too expensive to increase their frequency. Note that we average over the different z values for each value of τ^{insp}.

We compare the cost of inspection frequencies of 3 and 6 months, initially with an inspection cost of $X_3 = 100$ times the cost of a plant (Figure 4.31). The grey areas indicate a negative cost, thus the region where it is cheaper for nurseries to inspect every 3 months. For N_{Com} seedings, the black region covered by $r \leq 0.2$ and $\alpha^{out} = \alpha^{in} \leq 0.3$ indicates when inspections are not removing enough disease to justify inspecting more frequently. The black region covered by $0 \leq r \leq 1$ and $\alpha^{out} = \alpha^{in} \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$ indicates when inspections are removing enough disease so that it is still economical for nurseries to inspect every 6 months. When $\alpha^{out} = \alpha^{in} > 0.7$, the black region represents the case when trade inspections remove enough infection not to warrant increasing the frequency of scheduled inspections. The results for N_{Cons} seedings are very similar to N_{Com}. N_{Ret} seedings show similar results as well, with a smaller grey area which spans $0.1 \leq r \leq 1$ and

$\alpha^{out} = \alpha^{in} \leq 0.5$. For N_{Nur} seedings, the grey area is much larger, spanning almost all parameter combinations, covered by $r \geq 0.2$ and all values for $\alpha^{out} = \alpha^{in}$.
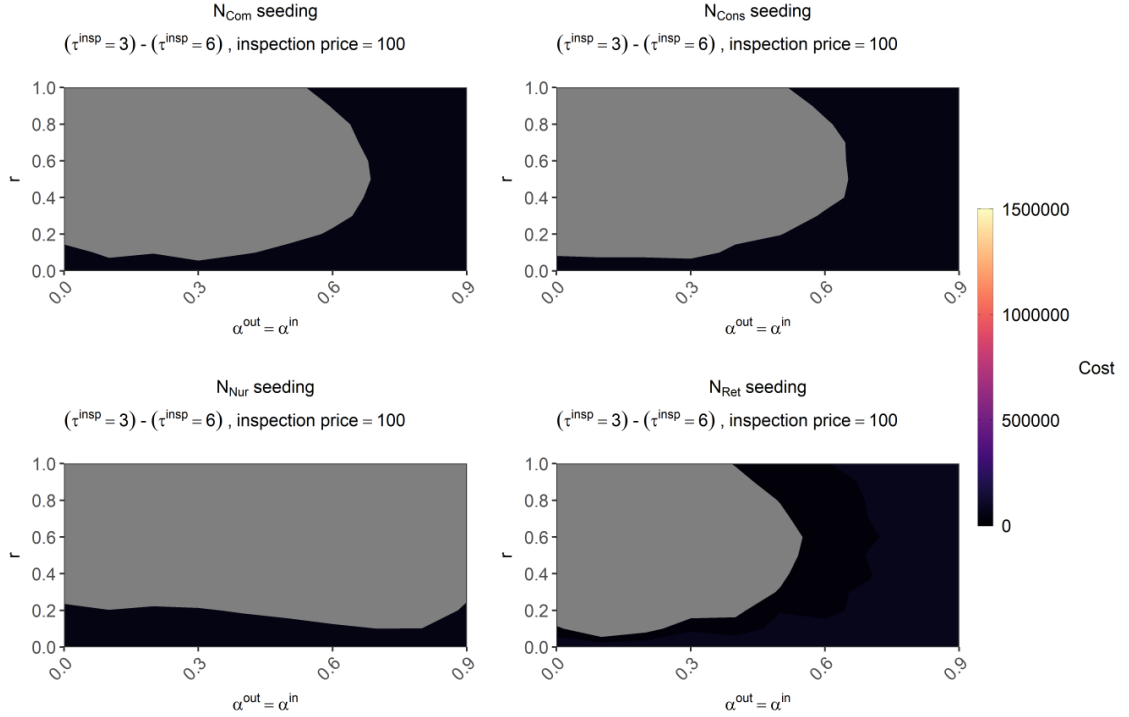
Figure 4.31. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every 3 months minus every 6 months. For $\tau^{insp} = 3$, we took the average of the median costs for $z = 3$ and $z = 1$. For $\tau^{insp} = 6$, we took the average of the median costs for $z = 6$, $z = 3$ and $z = 1$. This is for an inspection cost of 100 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. The areas in grey indicate parameter combinations which have a negative cost (where scheduled inspections every 3 months are cheaper than every 6 months). All other parameters are at baseline Table 4.1.

Increasing the cost of inspection to 400 times the cost of a plant, the grey areas shrink in size for all seedings (Figure 4.32). For N_{Com} and N_{Cons} seedings, the grey area has now been confined to $r \geq 0.2$ and $\alpha^{out} = \alpha^{in} \leq 0.4$. The grey area for N_{Ret} seedings has almost disappeared, bound by $0.3 \leq r \leq 0.9$ and $\alpha^{out} = \alpha^{in} \leq 0.2$. The grey area for N_{Nur} seedings remains large, however this is now confined to $r > 0.4$. For an inspection cost of 1140, all grey areas disappear (Figure 4.33). And thus for costs of 1140 and above, it is never economical to inspect every 3 months as opposed to every 6 months.
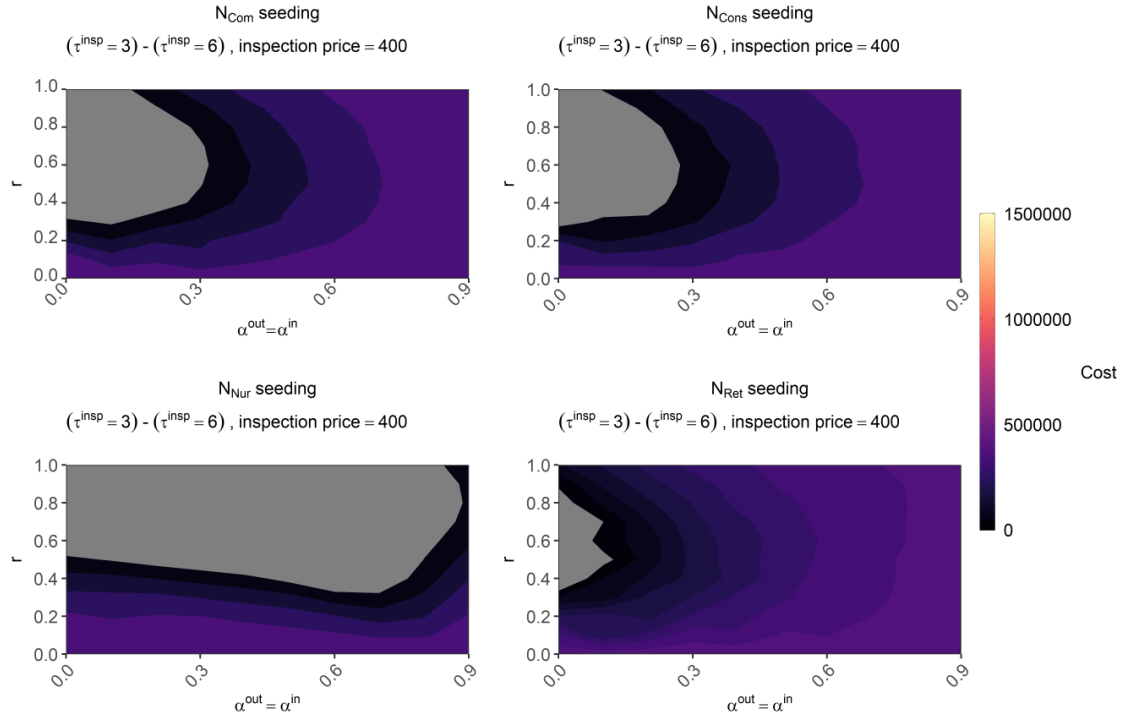
Figure 4.32. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every 3 months minus every 6 months. For $\tau^{insp} = 3$, we took the average of the median costs for $z = 3$ and $z = 1$. For $\tau^{insp} = 6$, we took the average of the median costs for $z = 6$, $z = 3$ and $z = 1$. This is for an inspection cost of 400 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. The areas in grey indicate parameter combinations which have a negative cost (where scheduled inspections every 3 months are cheaper than every 6 months). All other parameters are at baseline Table 4.1.

Figure 4.33. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every 3 months minus every 6 months. This is for an inspection cost of 1140 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. All other parameters are at baseline Table 4.1.

Next we compare the costs of inspecting every month against every 3 months for an inspection cost of 100 (Figure 4.34). We see a similar pattern compared to the 6 against 3 month figures, though the grey regions are smaller. The grey region for N_{Nur} seedings remains large, spanning almost all parameter combinations. An important difference is, for monthly inspections to be cheaper than inspecting every 3 months, we require $r < 0.6$ for N_{Com}, N_{Cons} and N_{Ret} seedings.
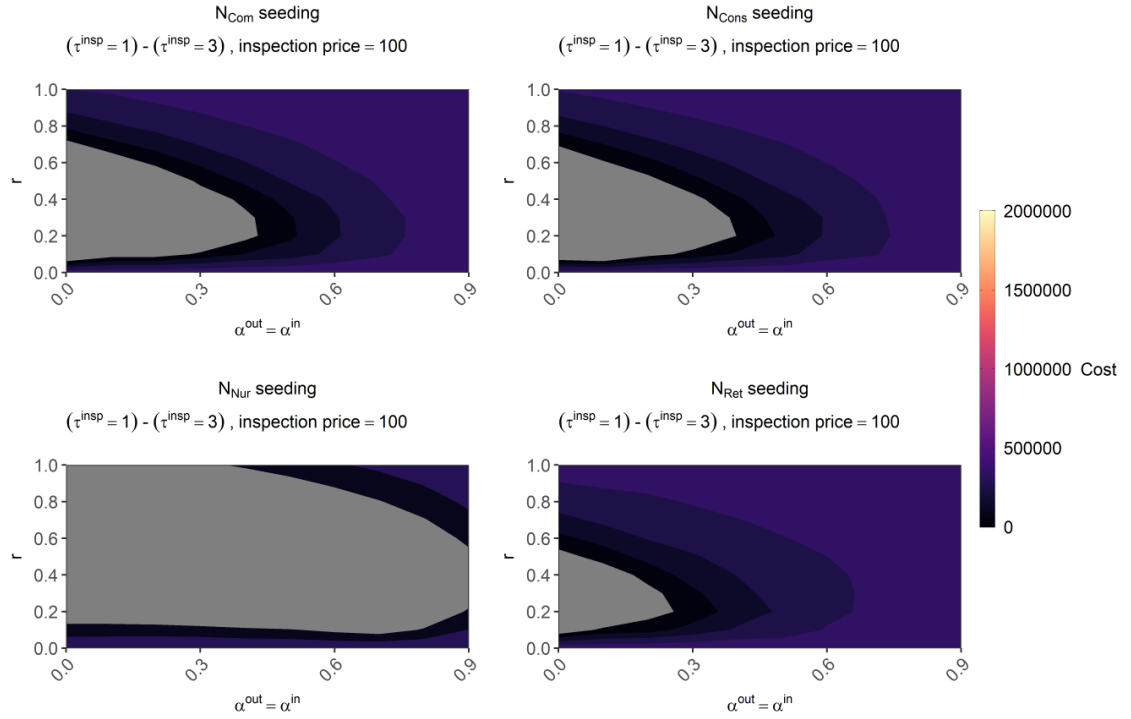
Figure 4.34. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every month minus every 3 months. For $\tau^{insp} = 3$, we took the average of the median costs for $z = 3$ and $z = 1$. This is for an inspection cost of 100 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. The areas in grey indicate parameter combinations which have a negative cost (where scheduled inspections every month are cheaper than every 3 months). All other parameters are at baseline Table 4.1.

Increasing the cost of inspection to 400, the grey regions for all seedings except N_{Nur} disappear (Figure 4.35). This region is small, confined to $0.4 \leq r \leq 0.8$ and $\alpha^{in} = \alpha^{out} \leq 0.5$. All grey regions disappear for an inspection cost of 489 and above (Figure 4.35).
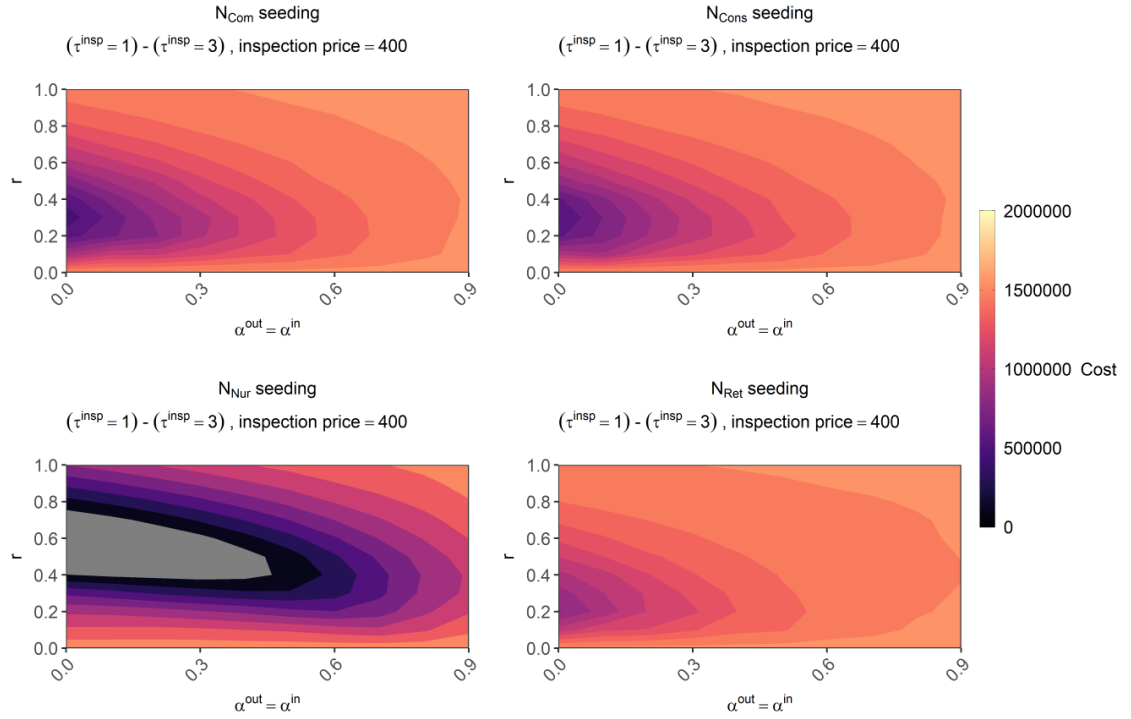
Figure 4.35. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every month minus every 3 months. For $\tau^{insp} = 3$, we took the average of the median costs for $z = 3$ and $z = 1$. This is for an inspection cost of 400 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. The areas in grey indicate parameter combinations which have a negative cost (where scheduled inspections every month are cheaper than every 3 months). All other parameters are at baseline Table 4.1.
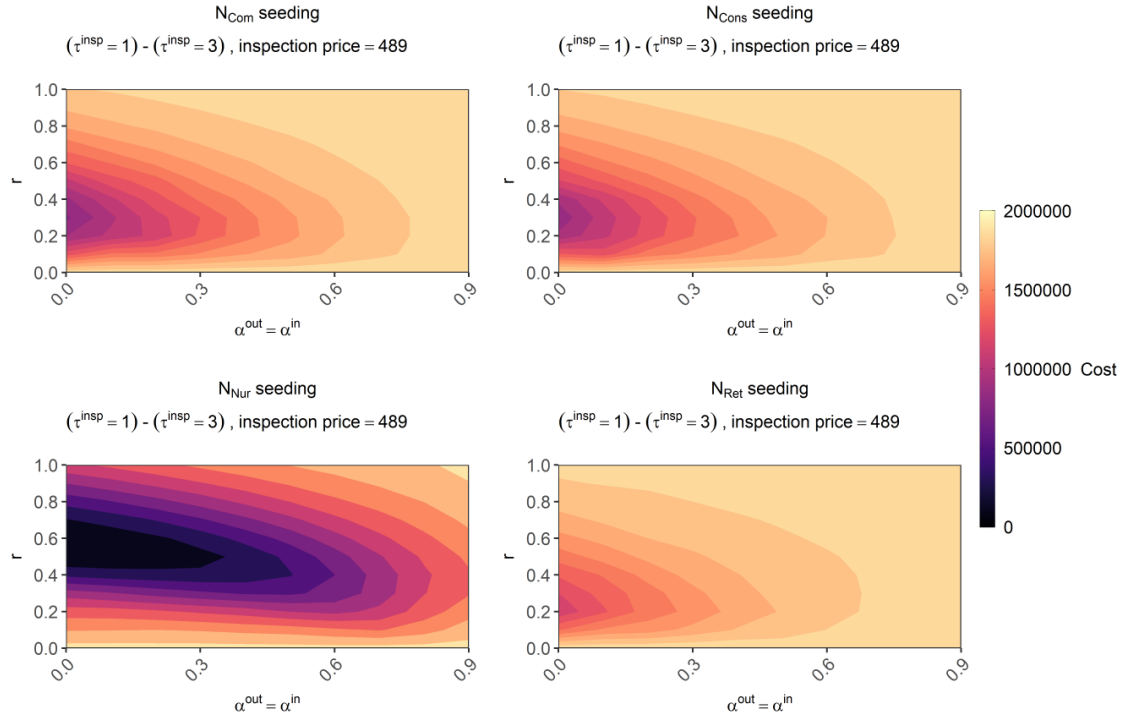
Figure 4.36. Heat map showing the median cost incurred by all nurseries for scheduled inspections occurring every month minus every 3 months. For $\tau^{insp} = 3$, we took the average of the median costs for $z = 3$ and $z = 1$. This is for an inspection cost of 432 times the cost of a plant. This is shown for combinations of inspection parameters r and $\alpha^{in} = \alpha^{out}$. All other parameters are at baseline Table 4.1.

Thus, when inspections are less than 489 times the cost of the plant, there are combinations of inspection parameters which suggest it is cheaper to have scheduled inspections every month in nurseries. For inspection costs between 489 and 1140, there are regions to prefer inspections every 3 months as opposed to every 6 months. For costs higher than 1140, nurseries are better off inspecting every 6 months. Our previous analysis shows that conducting scheduled inspections every month with an efficacy of at least 50% is adequate to control disease spread in all seedings (Figure 4.30). Thus, our cost-benefit analysis shows that the cheaper the inspection relative to the plant, the more feasible this adequate strategy becomes. Interestingly, trade inspections are counter-productive at making this control strategy feasible, revealing the potential conflicts of interest in nursery managers wanting to minimise costs and the national goal to minimise disease spread.

4.7 Discussion

In this chapter, we extended our network model into a compartmental metapopulation framework for a generic pest/disease spreading throughout the UK plant trade network. We modelled the trade of infected plants from an assumed hyper-geometric distribution and consignment inspection process. Sources and sinks were chosen to keep the population sizes constant for all nursery and retailer nodes. Our analysis focused on how disease spread differs when a pest/disease is seeded in a different nursery subcategory and on comparing the effect of different control measures.

Our analysis consistently showed that outbreaks are more likely to occur and more difficult to control if a pest is introduced to a nursery that predominantly sells to other nurseries. We also found that, once introduced, a pest or disease can remain at a low prevalence in the network for a long time period (almost three years) before epidemic spread begins. This result highlights the difficulty in tracing the initial introduction of a pest or disease and could act as an estimate for how long movements of plants within the UK should be documented. Interestingly, this coincides with the Plant Healthy scheme regulatory requirements assessment criteria of keeping plant passports for three years [154].

More frequent scheduled inspections were increasingly effective at controlling disease in comparison to trade inspections, such that for monthly scheduled inspections, trade inspections were ineffective. However, for scheduled inspections every six months, inspecting consignments became a more effective control strategy. For trade inspection efficacy of 40% and above, the addition of inspecting purchases had a noticeable effect at lowering disease spread. The most effective control strategy we identified was to inspect all nursery stock every month with at least 50% efficacy, even if consignments are no longer inspected.

From our cost-benefit analysis, we identified efficacy regions of the inspection parameters where it is in the nurseries' financial interest to inspect plants more frequently. This also revealed when potential conflicts of interest appeared, regarding the choice between minimising costs and minimising disease. Lowering inspection costs or inspecting nurseries in

a cheaper and less effective way (at a minimum efficacy of 50%) would aid in improving biosecurity in the plant trade sector of the UK.

The changes to the rate of transmission parameter suggested that infection is needed to more than double every month for a large epidemic to occur within three years. However, this requirement could be lower if infected plants remain in nurseries/retailers for longer than one month, our baseline value.

The results from our analysis were robust to changes in the distribution of nursery types in the network. This is in agreement with our results from Chapter 3. Due to the computational intensity, we could not investigate all seven scenarios from the previous chapter and only consider scenarios 1, 2 and 3 to represent a uniform distribution of nurseries and distributions skewed towards commercial and consumers, respectively (Table 4.2).

Our model provided a good starting point to modelling the spread of a general pest or disease through domestic trade. To maintain this generality, we chose to model disease spread in a susceptible-infected framework. However, there are specific stages of infection for certain pests/diseases that would improve the model for an approach focused on a specific pest/disease. Examples include the addition of compartments to represent exposed but not yet infected plants or a compartment for vaccinated/treated plants.

Note that our model equations 4.16 and 4.17 are synchronous. The number of susceptible and infected plants in each node at time $t + 1$ are calculated entirely from the state of the system at time t, and are all updated at the same time. Thus, the disease that spreads within the node at time t $(\beta \frac{S_i(t)I_i(t)}{N_i})$ occurs in synchrony with the movements of infected plants in or out of the node. Therefore in our calculations, we do not compute the spread, death or removal of infected plants before we compute the trade of plants. This will result in a limiting effect of the control measures as the disease will persist within the node through internal spread. However, with this limitation noted, it would be unrealistic to assume that all disease spread occurs before any trading, as our time-step is one month. Thus, with our chosen time-step, we are limited with either approach. Our methodology

would be improved with a smaller time-step of one day, and in our model simulations, an update to disease spread before plants are traded throughout the network.

In our analysis, we did not consider any other source of infection into the system after the seeding at $t = 0$. This was done to simplify the analysis, though this comes with limitations. In reality, nurseries which grow their own plants would be exposed to a threat of infection which is dependent on the prevalence of disease within the nursery. There are also threats of infection coming from imported plants overseas, though this will be a rare event. Since all subsequent sources of plants into the system were susceptible, the death rate d_I limits the maximum number of infected plants in the model at any one time. Additionally, the lack of sources of infection is what contributed to the suppression of infection in N_{Ret} nodes in our model. From the parameter values we used for Oak sales, retailer nodes consignment sizes were the largest, almost twice those of nurseries. This caused the nurseries that sell mostly to retailers to sell more than they buy. Therefore, due to our sources consisting of only susceptible plants, disease was artificially suppressed in our model for this node subcategory. We suspect that this is the reason that N_{Ret} nurseries were not as influential in the network as suggested from our analysis in Chapter 3. An alteration to the model that could more accurately describe sources and threats of disease would be an improvement.

In our model, inspection parameters were the same for all nurseries and retailers. An interesting extension to this model would be to have inspection parameters differ between nodes, especially given that seedings in N_{Nur} nurseries were so significant in our analysis. One example would be to only have inspections in the largest nurseries or those of a specific subcategory.

We assumed the population size of nurseries and retailers to be one year's worth of plants. This is calculated by the maximum between total bought or sold for each nursery and retailer node. A nursery's and retailer's capacity for stocking a plant would in practice vary depending on many factors such as supply, demand and the physical space the nurs-

ery/retailer has for the specific plant. In future work, this would be an important addition to consider.

We averaged trade across the year, to keep our model general. To include seasonality in the model, similar to *Harwood et al.*, we could partition the year into seasons of trade and no trade [46]. Since each plant species would have differing trading seasons, focusing on the trade of specific plant hosts or plant species would be required to make this investigation feasible. It would be interesting to see how periods of no movement across the network would affect the results from this chapter.

In our model, the force of infection was present every time-step. For a pest that multiplies over periods of time longer than a month, a continuous force of infection overestimates this process. To model the spread of a pest more accurately, we would need to alter the transmission rate parameter to be periodic, matching the time between generations of a particular pest. We expect a combination of the size of seeding, removal rate within nurseries, pest generation length and transmission rate would affect the disease spread dynamics significantly for this change in the model.

We focused the majority of our analysis on nurseries. A clear extension of this work would be to investigate the disease spread dynamics amongst the retailers and customers. To maintain generality, in addition to the many uncertain parameters in our model, we chose not to incorporate seasonality. Access to more data, or a more focused investigation where we can fix many parameter values, would allow for adding seasonality to the model. It would be interesting to develop a disease spread model on an unweighted version of the model network and compare the results with the work conducted in this chapter.

We could extend the model to include further incursions of disease from the international trade network and expose plants grown by nurseries to a force of infection. In reality, the trade network would change in response to an epidemic and with our model, trade is fixed over time. We could make the network dynamic and include quarantines for nodes with a given prevalence to simulate the model for longer time spans, similar to work conducted

in [106]. To simplify analysis, we fixed our network model parameters on the values we obtained for Oak trade. We could vary these network parameters for another host or via a sensitivity analysis with our disease spread model.

An interesting point to note about the inspection parameters (r, α^{out} and α^{in}), is that their value will be determined by the pest or disease in question as well as the capabilities of the nurseries/inspectors themselves. Some pests/diseases may be spotted with a cursory glance whilst others may require extensive testing. In addition, the values of the inspection parameters represent an efficacy, but they can equally be interpreted as a proportion of plants inspected if the inspection is 100% effective. Given these many factors, an investigation into a specific pest or disease is required to consider what values of the inspection parameters would be considered realistic.

Note that there can appear some inconsistency for our results between the two measures we use to analyse disease spread: proportion infected by 36 months and time until 20% infected in nurseries. The reason for this is that the first measure uses all of the simulations for a set of parameter values, while, for the second measure we subset to only those simulations in which infection in nurseries reaches 20% prevalence.

Due to the stochasticity of the network construction and the many simulations to analyse the model, we model disease spread in a deterministic system for simplicity and to ease computational cost. This is potentially not very limiting as disease spread in large populations can lead to deterministic behaviour [96].

From the implementation of scheduled inspections, infected plants were still traded before the inspection is performed, and thus not all plants were inspected. This explained the limited effect that scheduled inspections had on the model. An alternative implementation could be where the inspection takes place before any trade, and the number of infected in each nursery and retailer is reduced by a proportion r and then trade resumes. Modelling inspections in a system with simultaneous trade and disease spread brings in this problem of the ordering of processes. For trade inspections, we did not have this issue as the order

of operations were built into the way trade inspections function. However, with modelling an inspection of the nursery stock at a given time-step, it is important to specify if the inspection happens before any trade has occurred. Tied to this, is our chosen time-step and unit of time, measured in months. If we change our time-step in the model to days, this could improve the effectiveness of scheduled inspections as less plants will be missed.

Initially, this trade inspection process was modelled by assuming all infected plants that were caught in initial inspections from both the buyer and seller were then replaced by susceptible plants. To ensure that each seller had the capacity to replace infected plants with susceptible, the probability of inspection success was multiplied by $\frac{S_i(t)}{N_i}$, which represents the number of susceptible plants divided by the total number of plants of subpopulation i. Hence, using this method, the number of infected plants in a given consignment was calculated by $(1 - \alpha^{out} \frac{S_i(t)}{N_i})(1 - \alpha^{in} \frac{S_i(t)}{N_i}) A_{i,j} \frac{I_i(t)}{N_i}$. Recall that $A_{i,j}$ represents the number of plants traded from node i to j, and α^{out} and α^{in} represent the inspection efficacy parameters for sales and purchases respectively. We compared the two different methods of modelling trade inspections, by considering an illustrated example: $A_{i,j} = 100$, $N_i = 10^4, S_i(t) = 5000, I_i(t) = 5000, \alpha^{out} = 0.5, \alpha^{in} = 0.5$. Using the initial method we have the number of infected plants buyer j receives $(1 - \alpha^{out} \frac{S_i(t)}{N_i})(1 - \alpha^{in} \frac{S_i(t)}{N_i}) A_{i,j} = 28.125$. Whereas, using the second method, we have $F_i(\alpha^{out}, \alpha^{in}, t) A_{i,j} = 20$. There is approximately 40% less infected plants being traded in comparison with the first method. Although the initial method proposed is simpler in both concept and computation, the 40% difference incurred from ignoring subsequent inspections underestimates the effectiveness of trade inspections. Therefore, we chose to use the second method of modelling trade inspections.

The simulations for this chapter were computationally intensive, simulating the model 10000 times for each parameter combination. Initially this model was programmed using R (version 4.2), using a Windows 10 computer with an Intel Core i9-10850K CPU. The best performance we could achieve for one simulation of the model was 14 seconds, meaning a 39 day run-time for 10000 simulations. We had been attempting to reduce run-time with parallel processing, however for this version of the code there were high diminishing returns

using greater than 10 cores, achieving at best making the code run five times quicker. An eight day run-time for one parameter set would still limit how many simulations we could conduct and so we had to find a way to reduce run-time further. We then rewrote the code from R into C++ using R packages Rcpp and RcppArmadillo [155,156]. This improved the run-time of the code greatly, resulting in one simulation of the model taking 1 second. We were in addition able to run in parallel by turning the code into an R package. Using this code and the ARCHIE WeSt supercomputer (a Linux system using Intel Xeon Gold 6138 Skylake processors), we were able to run 10000 simulations over 4 seedings in 4 to 12 hours. With our old implementation, this would have taken 32 days, a minimum improvement of 56700%. These changes in our code implementation were vital to conducting the analysis in this chapter.

In hindsight, we could have simplified the analysis for each parameter combination to run fewer simulations. Instead of simulating the model for 100 random seedings per nursery group for each generated network, we could have fixed the seeding, once for each nursery in the network (160 total nurseries). This would have reduced computational load for each parameter combination by 60% (from 400×100 to 160×100).

Chapter 5

Developing a Network Epidemiological Model with Application to COVID-19

In May 2020, we began a collaborative research project in response to the ongoing COVID-19 pandemic. The authors of this work are: Matthew Baister, Ewan McTaggart, Paul McMenemy, Itamar Megiddo and Adam Kleczkowski. COVID-19 was spreading far and wide across the care home community during the first wave of the pandemic, leading to many deaths. We wanted to investigate the different routes COVID-19 was taking into care homes to cause this. Thus, we developed a compartmental metapopulation model describing the population of the Scottish health board Lothian, splitting the populace into care home residents, care home staff and the general population. This work spanned May 2020 to July 2021.

Adam Kleczkowski and Itamar Megiddo played supervisory roles. Paul McMenemy also supervised but also provided a foundational code to simulate the model. Ewan McTaggart and Matthew Baister worked in tandem on: constructing the model equations, building on from the code written by Paul McMenemy (data visualisation, data fitting and sensitivity analysis) and writing the manuscript. Ewan McTaggart reviewed the relevant literature and constructed the model schematics and Matthew Baister collected data sources for pa-

rameter estimates. A table summarising the contributions of each author can be found in Appendix C.

COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff

Matthew Baister [1]☯, Ewan McTaggart[1]☯, Paul McMenemy,[2], Itamar Megiddo[3], Adam Kleczkowski[1],

1 Department of Mathematics & Statistics, University of Strathclyde, Glasgow, UK
2 Department of Computing Science and Mathematics, University of Stirling, Stirling, UK
3 Department of Management Science, University of Strathclyde, Glasgow, UK

☯These authors contributed equally to this work.

Abstract

Care homes in the UK were disproportionately affected by the first wave of the COVID-19 pandemic, accounting for almost half of COVID-19 deaths over the course of the period from 6th March – 15th June 2020. Understanding how infectious diseases establish themselves throughout vulnerable communities is crucial for minimising deaths and lowering the total stress on the National Health Service (NHS Scotland). We model the spread of COVID-19 in the health-board of NHS Lothian, Scotland over the course of the first wave of the pandemic with a compartmental Susceptible - Exposed - Infected reported - Infected unreported - Recovered - Dead (**SEIARD**), metapopulation model. Care home residents, care home workers and the rest of the population are modelled as subpopulations, interacting on a network describing their mixing habits. We explicitly model the outbreak's reproduction rate and care home visitation level over time for each subpopulation, and execute a data fit and sensitivity analysis, focusing on parameters responsible for intra-subpopulation mixing: staff sharing, staff shift patterns and visitation. The results suggest that hospital discharges were not predominantly responsible for the early outbreak

in care homes, and that only a few such cases led to infection seeding in care homes by the 6th of March Sensitivity analysis show the main mode of entry into care homes are infections by staff interacting with the general population. Visitation (before cancellation) and staff sharing were less significant in affecting outbreak size. Our model suggests that focusing on the protection and monitoring of staff, followed by reductions in staff sharing and quick cancellations of visitation can significantly reduce future infection attack rates of COVID-19 in care homes.

Introduction

The outbreak of the SARS-CoV-2 induced disease (COVID-19) pandemic has had a profound impact, causing 3.7 million deaths by early June 2021 and economic hardship globally [157]. The care home population suffered a disproportionate amount of COVID-19 related deaths. From the week ending 13th March 2020 to the week ending 26th June 2020 (the "first wave"), 54,510 deaths were associated with COVID-19 in the UK, 40% of which were among care home residents [158]. The COVID-19 pandemic has highlighted the vulnerability of care homes, as their resident population is elderly and often suffers from several co-morbidities [159], their systems have not been developed with infection prevention and control (IPC) in mind, and their IPC guidelines have been borrowed from hospitals - a completely different setting [160]. Networks of care homes are an ecosystem connected by staff working across facilities. These connections increase the risk of COVID-19 ingress into care homes, and to protect their vulnerable community, we need to understand the ecosystem dynamics.

Care homes and their residents are enclosed societies, isolated to some extent from the general population. Their connection to broader society primarily consists of interaction with staff and visits from the general population. Care home staff potentially play a vital role in COVID-19 introduction and spread throughout the care home population. Firstly, staff exposure to infection from the general population can establish an outbreak in a home. Secondly, some staff work across multiple homes - a concept we refer to as staff sharing. Staff acting as a bridge between care homes and the general population and staff sharing induces a network, connecting care homes in a given community via their workers. This creates the potential for COVID-19 to spread from one home to another; hence, investigation of this pathway is important. We find it natural to describe this using a heterogeneous patch size metapopulation model framework.

Very few models explore COVID-19 transmission at a community level and explicitly include the unique dynamics in care homes. For example, in [161, 162] agent-based models (ABMs) of single homes are used to investigate the impact of testing strategies on the disease burden. A report by Nyguen et al. [163] uses an ABM to investigate the impact on

201

care home residents of various vaccine coverage, and reducing the weekly testing of staff. [30] However, the models in [162, 163] do have an external force of infection (FOI) from the [31] community, based on prevalence data, representing staff interaction with the community [32] and visitors. These models [161–163] do not assess the relative impact of the different [33] COVID-19 pathways into care homes. Nguyen et al. [164] extend [162, 163], using a hybrid [34] ABM-System Dynamics model, to explore the conditions under which visitation, hetero- [35] geneous care homes sizes, and the cohorting of residents impacts COVID-19 outbreak [36] severity. [37]

Rosello et al. [165] model an individual care home with a stochastic compartmental [38] model, using multiple FOI's to capture COVID-19 pathways, including visitors, hospital [39] discharges, staff working at other homes, and staff infections from the community. They [40] find that importations of infections by staff from the community are the main driver of [41] outbreaks, and importation by visitors or from hospitals is rare, but do not explicitly model [42] disease spread throughout a network of care homes. In [166] individual care homes and [43] the general public are independent, deterministic SEIR models, with a stochastic external [44] FOI connecting the general public to each home. This FOI depends on the prevalence [45] of COVID-19 in the general public, and the size of each home. Transmission rates in [46] homes and in the general public do not vary over time. In [167], two weakly-coupled [47] SEIR sub-models with time-dependent transmission rates define the dynamics; one sub- [48] model describes the general public and one describes all care home residents in Stockholm [49] as a single homogeneous group. Again, a single FOI acts on the residents to capture [50] infections from staff and visitors. The models [166, 167] do not differentiate between, and [51] therefore allow comparison of, the COVID-19 pathways into care homes. Bunnik et al. [168] [52] use a compartmental metapopulation model to explore the trade-offs between increasing [53] protection for a "vulnerable" population and relaxing restrictions for the "non-vulnerable" [54] after the first lockdown in Scotland. They use time-dependent transmission rates with three [55] metapopulation groups; vulnerable, shielders and general public. We extend and apply the [56] methodology of [168] in our model, investigating protection to a vulnerable group (care [57] home residents) in ways other than shielding. [58]

We construct a **SEIARD** compartmental metapopulation model to describe the first wave of COVID-19 in a health board in Scotland. The population is divided into groups of care home residents, staff, and general public. Our care home resident group are not a single homogeneous unit as in [167, 168] but are separate units, creating a refined spatial/geographic structure. These units are not independent as in [166] but are linked by a staff sharing network which, to our knowledge, is unique. We calibrate this model to data from the NHS Lothian Health Board and explore the sensitivity of the results to changes in key parameters. We investigate the importation of infections by staff from the community, visitation, staff sharing, and additionally, we shed light on the exposure of care homes at the beginning of the first wave, e.g., via hospital discharges [169]. The aim is to identify and rank the modes of COVID-19 ingress into and throughout the susceptible care home community in order to improve future pandemic responses. Our model allows for this investigation by coupling the general public and individual care homes with the explicit movement of staff and visitors between the two.

Materials and methods

Mathematical model

We develop a deterministic **SEIARD** compartmental metapopulation model with heterogeneous subpopulation sizes. Each subpopulation consists of a host human population, categorised further into six compartments of COVID-19 infection status: Susceptible (**S**), i.e., everyone who is not infected; Exposed (**E**), those exposed to the virus (and infected) but not yet infectious; Infectious and reported (**I**), infectious individuals that have been identified with a positive test; Unreported infectious (**A**), infectious individuals that have not been identified with a positive test; Recovered (**R**), those who had COVID-19 and recovered; and Dead (**D**), those who died from their illness. Symptomatic and asymptomatic individuals are not modelled explicitly; instead, asymptomatic infections contribute towards a reduction in the reporting rates. This model is illustrated in Figure 5.1 (a).
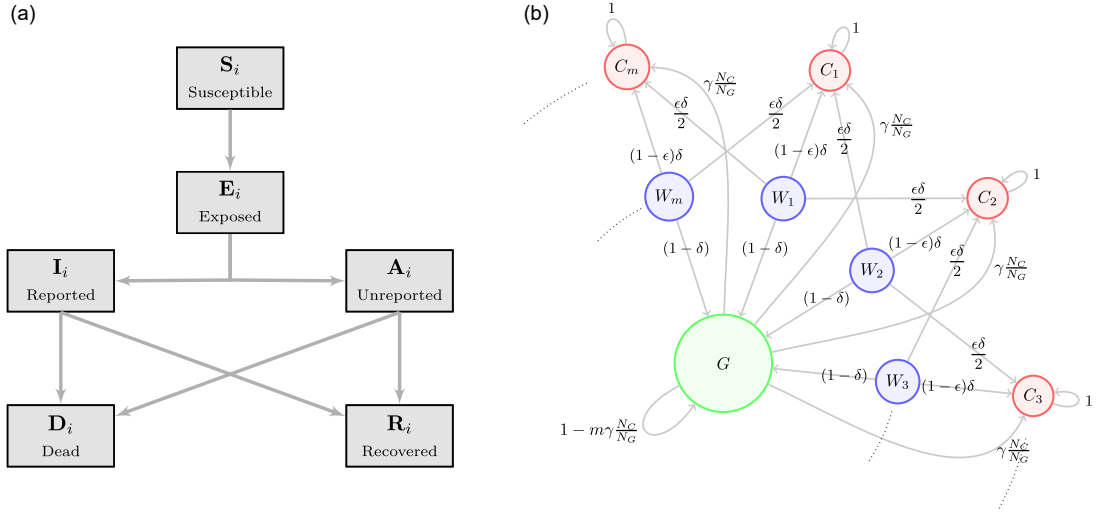
Figure 5.1. Schematics for the compartmental and metapopulation structure. (a): **SEIARD** compartmental structure of the model; **(b)**: Time-share network of interaction amongst subpopulations. Directed edge weights are t_{ik}, the proportion of people from subpopulation i who travel to mix at effective population k.

The metapopulation structure represents the population of the NHS Lothian health board in Scotland. We distinguish between care home residents, care home workers and the general population, modelling the $m = 109$ care homes for older adults in NHS Lothian [170]. The jth home has a resident subpopulation, C_j, with a corresponding care home worker subpopulation, W_j. The general population is encapsulated by the subpopulation G. Each care home includes the same number of residents, a simplifying assumption made due to lack of publicly available data on care home sizes in Lothian. We also assume the worker subpopulations are the same size as residents' [171].

Each node of the network, $i \in X := \{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$ with $|X| = n$, is described in terms of the **SEIARD** compartmental model with equations:

$$\frac{d\mathbf{S}_i}{dt} = -\mathbf{S}_i\Lambda_i$$

$$\frac{d\mathbf{E}_i}{dt} = \mathbf{S}_i\Lambda_i - \frac{\mathbf{E}_i}{\lambda}$$

$$\frac{d\mathbf{I}_i}{dt} = r_i\frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{I}_i}{\tau}$$

$$\frac{d\mathbf{A}_i}{dt} = (1 - r_i)\frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{A}_i}{\tau} \qquad (5.1)$$

$$\frac{d\mathbf{R}_i}{dt} = (1 - \mu_i)\frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau}$$

$$\frac{d\mathbf{D}_i}{dt} = \mu_i\frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau}$$

Susceptibles in subpopulation i ($\mathbf{S}_i$), are infected with a FOI Λ_i, and moved to the exposed class ($\mathbf{E}_i$). After a non-infectious latent period of λ days, they become infectious, testing positive at a reporting rate of r_i. These identified infections move to the class $\mathbf{I}_i$. Hence, any unidentified infections, $\mathbf{A}_i$, occur at rate 1 - r_i. After τ infectious days, a person either recovers or dies at the rate μ_i. For simplicity, and considering the short span of time the model is designed to describe, non-COVID related deaths are not considered. For similar reasons, we do not include a birth rate or admission of new residents to care homes from the general population.

We assume a constant reporting rate for care home residents ($r_{i\in\{C_1,...,C_m\}} = r_C$), workers ($r_{i\in\{W_1,...,W_m\}} = r_W$), and the general public (r_G). The parameters τ and λ describe the infectious period and latency period, respectively, and are assumed to be the same across all subpopulations. Mortality rates, μ_i, vary by subpopulation, reflecting the positive association of serious outcomes of COVID-19 with age [172]. As we are modelling over a period of 4 months (approx. first wave), and immunity after COVID-19 infection lasts as long as 5 months [173] [174], we do not consider a transition from Recovered to Susceptible.

We model visitation to each care home by multiplying the proportion of the population, N_C/N_G, that visit the care home, and their length of stay, $\gamma(t)$. The proportion remains constant over time while $\gamma(t)$ varies over time. Up until 13th March 2020, each

resident has one visitor per day [162]. Then $\gamma(t)$ drops to 0, reflecting the policy change to essential visitation only [170, 175]. $\gamma(t)$ is described by the function $logi(\Omega_x)$, with $\Omega_x = \{t, \omega^x_{end}, \omega^x_{rate}, \omega^x_{low}, \omega^x_{high}\}$, defined below:

$$logi(\Omega_x) = \frac{(\omega^x_{high} - \omega^x_{low})}{(1 + \exp(\omega^x_{rate}(t - \omega^x_{end})))(1 + \exp(-\omega^x_{rate}(t + 82)))} + \omega^x_{low}, \qquad (5.2)$$

with the shape of a sigmoidal logistic function. The value of 82 is used in the function so that when $t = 0$, $logi(\Omega_x) = \omega^{high}$. The function drops from ω^{high} to ω^{low} at a time controlled by ω^x_{end}, such that when $t = \omega^x_{end}$, $logi(\Omega_x) = (\omega^x_{high} + \omega^x_{low})/2$. The ω_{rate} parameter changes the gradient of the descent at $t = \omega_{end}$.

Thus, visitation rate is described by $\frac{N_C(0)}{N_G(0)}\gamma(t) = \frac{N_C(0)}{N_G(0)}logi(\Omega_\gamma)$. Given that visitation drops to 0 in the first 2 weeks of the simulation, the changes in population size over that time is negligible, hence we can keep the proportion of the population constant and control visitation by solely changing $\gamma(t)$.

A constant proportion of workers, δ, spend their time at care homes. With $\delta = 0.5$, half of a worker compartment, W_i, are at care homes, C_i, over the course of a day. The proportion of workers not at care homes, $1 - \delta$, reside in the general population, G. During the first wave of COVID-19 in Scotland, there was both intra-organizational staff sharing between homes (i.e., staff who work at multiple homes belonging to the same care provider), as well as inter-organizational staff sharing (use of bank or agency staff) [176, 177]. Therefore, a constant proportion of each homes' assigned workers, ε, are exchanged between homes. We refer to this as *staff sharing*. We have made the simplifying assumption that the staff sharing network has a topology of a circle, whereby the shared staff for home j are split evenly between homes $j - 1$ and $j + 1$. We assume care home residents do not leave their homes.

Interaction across subpopulations is heterogeneous and is described in terms of time-sharing, determining proportions of subpopulations mixing in groups with each other. In the ith subpopulation there are $N_i(t) = \mathbf{S}_i(t) + \mathbf{E}_i(t) + \mathbf{I}_i(t) + \mathbf{A}_i(t) + \mathbf{R}_i(t)$ active individuals

who can mix with others. The proportion from subpopulation i who travel to, and mix with, subpopulation k is t_{ik}. The effective population size of subpopulation k, given that others have travelled to it and some people from k have left, is $\hat{N}_k(t) = \sum_{j \in X} t_{jk} N_j(t)$. We assume these effective populations $\hat{N}_k(t)$ are well mixed, so people who travel to each population can meet all others there. There are two types of effective populations; the care homes and the general population. Care home j, comprises $\hat{N}_{C_j}(t)$ people: its residents, its working staff, staff from other care homes, and visitors. The general population consists of $\hat{N}_G(t)$ people; this includes all the staff not at work and the non-visiting general population.

Our specific time-share assumptions are represented visually as a directed, weighted network in Figure 5.1 (b). The corresponding weighted adjacency matrix, the travel/time-share matrix, is $\boldsymbol{T} \in \mathbb{R}^{n \times n}$, whose $[i,j]^{th}$ element is t_{ij}. The rows and columns of $\boldsymbol{T}$ are in the order of $\{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$. $\boldsymbol{T}$ consists of the partitions $\{T_{CC}, T_{CW}, T_{WC}, T_{WW}, T_{CG}, T_{WG}, T_{GC}, T_{GW}\}$. To clarify notation: matrix $\boldsymbol{I}_m$ indicates the identity matrix of dimension m, matrix $[a]_{m \times m}$ indicates a matrix of dimension $m \times m$ with all entries a. Hence $\boldsymbol{T}$ and the subsequent sub-matrices are as follows:

$$
\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T_{CC}} & \boldsymbol{T_{CW}} & \boldsymbol{T_{CG}} \\ \boldsymbol{T_{WC}} & \boldsymbol{T_{WW}} & \boldsymbol{T_{WG}} \\ \boldsymbol{T_{GC}} & \boldsymbol{T_{GW}} & \boldsymbol{T_{GG}} \end{bmatrix}_{n \times n}, \tag{5.3}
$$

$$
\boldsymbol{T_{CC}} = \boldsymbol{I_m}, \quad \boldsymbol{T_{CW}} = \boldsymbol{T_{WW}} = \begin{bmatrix} 0 \end{bmatrix}_{m \times m}, \quad \boldsymbol{T_{CG}} = \begin{bmatrix} 0 \end{bmatrix}_{m \times 1}, \quad \boldsymbol{T_{GW}} = \begin{bmatrix} 0 \end{bmatrix}_{1 \times m},
$$

$$
\boldsymbol{T_{WG}} = \begin{bmatrix} 1 - \delta \end{bmatrix}_{m \times 1}, \quad \boldsymbol{T_{GC}} = \begin{bmatrix} \dfrac{N_C(0)}{N_G(0)} \gamma(t) \end{bmatrix}_{1 \times m}, \quad \boldsymbol{T_{GG}} = \begin{bmatrix} 1 - m \dfrac{N_C(0)}{N_G(0)} \gamma(t) \end{bmatrix}_{1 \times 1},
$$

$$
\boldsymbol{T_{WC}} = \begin{bmatrix} (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & 0 & \ldots & 0 & \frac{\varepsilon\delta}{2} \\[2ex] \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & \ldots & 0 & 0 \\[2ex] 0 & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \ldots & 0 & 0 \\[2ex] \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\[2ex] 0 & 0 & 0 & \ldots & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} \\[2ex] \frac{\varepsilon\delta}{2} & 0 & 0 & \ldots & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta \end{bmatrix}_{m \times m} . \tag{5.4}
$$

Disease transmission in the model is assumed to be frequency-dependent. The FOI integrates which infections occur to whom, from whom and where the infection takes place, as in [178,179]. The FOI acting on subpopulation i, Λ_i (see Equation 5.5), accounts for the different groups' people from i mix in over a day and who they encounter. It is most easily understood by considering $\Lambda_i \mathbf{S}_i$:

$$
\Lambda_i \mathbf{S}_i = \sum_{k \in L_i} \frac{t_{ik} \mathbf{S}_i}{\hat{N}_k(t)} \sum_{j \in X} \beta_{ji}(t) t_{jk} (\mathbf{I}_j + \mathbf{A}_j) \tag{5.5}
$$

The set of effective populations that subpopulation i travels to is L_i, consistent with the non-zero elements in the i_{th} row of the travel matrix $\boldsymbol{T}$. At effective population k, there is $t_{ik}\mathbf{S}_i$ susceptible individuals from i. At k there will also be $t_{jk}(\mathbf{I}_j + \mathbf{A}_j)$ infectious people from j who have travelled to k. The transmission rate between subpopulation j and i is $\beta_{ji}(t)$. Therefore,

$$
\frac{t_{ik}\mathbf{S}_i}{\hat{N}_k} \beta_{ji}(t) t_{jk} (\mathbf{I}_j + \mathbf{A}_j)
$$

is the number of new daily infections in i caused by people from j at the effective population k.

The transmission rates $\beta_{ji}(t)$ allow us to represent heterogeneous interaction patterns of individuals between and within different subpopulations. They incorporate the transmission dynamics of COVID-19 changing over time and location, for example, through lockdowns or other changes in behaviour [178]. We write $\beta_{ji}(t) = \frac{R(t)_{ji}}{\tau}$, describing the transmission rate $\beta_{ji}(t)$ between subpopulations j and i, with the reproduction rate, $R(t)_{ji}$, divided by the infectious period, τ. The contact rate and infection probability between subpopulations i and j is captured by $R(t)_{ji}$. We assume only the transmission rates between and within the subpopulation types (residents C, workers W, general public G) differ. Therefore, the transmission rates are arranged in a symmetric partitioned matrix $\boldsymbol{\beta} \in \mathbb{R}^{n \times n}$ whose $[j, i]^{th}$ element is $\beta_{ji}(t)$. The rows and columns of $\boldsymbol{\beta}$ are in the order of $\{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$. $\boldsymbol{\beta}$ contains block sub-matrices;

$$
\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta_{CC}} & \boldsymbol{\beta_{WC}^T} & \boldsymbol{\beta_{GC}^T} \\ \boldsymbol{\beta_{WC}} & \boldsymbol{\beta_{WW}} & \boldsymbol{R_{GW}^T} \\ \boldsymbol{\beta_{GC}} & \boldsymbol{\beta_{GW}} & \boldsymbol{\beta_{GG}} \end{bmatrix}
$$

$$
\boldsymbol{\beta_{CC}} = \Big[\beta_C(t)\Big]_{m \times m}, \; \boldsymbol{\beta_{WC}} = \Big[\beta_C(t)\Big]_{m \times m}, \; \boldsymbol{\beta_{WW}} = \Big[\beta_W(t)\Big]_{m \times m}
$$

$$
\boldsymbol{\beta_{GC}} = \Big[\beta_C(t)\Big]_{1 \times m}, \; \boldsymbol{\beta_{GW}} = \Big[\beta_G(t)\Big]_{1 \times m}, \; \boldsymbol{\beta_{GG}} = \Big[\beta_G(t)\Big]_{1 \times 1}
$$

The matrix notation above is the same as for the travel matrix $\boldsymbol{T}$. For simplicity, we have assumed that the resident-resident, worker-resident, and general population-resident transmission rates are equal. Similarly, we assume the general population-worker and general population-general population transmission rates are the same. The transmission rates are described by:

$$\beta_C(t) = \frac{logi(t, \omega_{end}^c, \omega_{rate}^c, \omega_{low}^c, \omega_{high}^c)}{\tau},$$

$$\beta_W(t) = \frac{logi(t, \omega_{end}^c, \omega_{rate}^c, \omega_{low}^c, (\omega_{high}^c + \omega_{high}^G)/2)}{\tau}$$

$$\beta_G(t) = \frac{logi(t, \omega_{end}^G, \omega_{rate}^G, \omega_{low}^G, \omega_{low}^G, \omega_{high}^G)}{\tau}$$

Where the logi function, Equation 5.2, models the reproduction rate. To simplify and to reduce the number of parameters, we relate the reproduction rate for workers in terms of the residents and general population. As care home workers balance their time between care homes and the general population, we assume the workers pre-lockown maximum reproduction rate is the average of the care homes and general populations, $\omega_{high}^W = (\omega_{high}^C + \omega_{high}^G)/2$. We assume the reproduction rate for workers and residents drops at the same time, and to the same value.

Model calibration process

We used data from the network of care homes in NHS Lothian [170] complemented by Public Health Scotland Open Data, breaking down COVID-19 cases and deaths per health board [180,181], to inform and calibrate our model. Parameters were found using a mixture of methods, as indicated in Table 5.1, including literature search, sensitivity of results, and rigorous fit based on minimising the sum of squares of residuals.

Table 5.1. Parameter definitions, alongside their base case values and source.

Parameter	Description	Value	Source
ε	Staff sharing	0.4	Data fit
δ	Proportion of workers at care home	0.5	Assumption[a]
$r_{i\in\{C_1,\dots,C_m\}}=r_C$	Reporting rate for residents	0.53	Estimated [170, 182–184]
$r_{i\in\{W_1,\dots,W_m\}}=r_W$	Reporting rate for workers	0.52	Estimated [170, 180, 184–186]
r_G	Reporting rate for general public	0.077	Estimated [180, 185, 186]
$\mu_{i\in\{C_1,\dots,C_m\}}=\mu_C$	Death rate for residents	0.25	Estimated [170, 182–184]
$\mu_{i\in\{G,W_1,\dots,W_m\}}=\mu_G$	Death rate for general public (and workers)	0.017	Estimated [180, 181, 185, 186]
τ	Infectious period	7 days	[187]
λ	Latent period	5.8 days	[188]
m	Number of care homes	109	[170]
$N(0)=\sum_i N_i(0)$	Total initial population	907,580	[186]
$N_{C_i}(0)$	Initial resident subpopulation size	48	[170]
$N_{W_i}(0)$	Initial worker subpopulation size	48	[171]
$N_G(0)$	Initial general public subpopulation size	897,116	Estimated [170, 171, 186]
$\omega^{\mathbf{C}}_{\mathbf{end}}$	Timing of R_t descent for residents and workers	42 days	Manually fitted [b]
$\omega^{\mathbf{C}}_{\mathbf{rate}}$	Rate of descent of R_t for residents and workers	0.5	Assumption[c]
$\omega^{\mathbf{C}}_{\mathbf{low}}$	Post-descent R_t for residents and workers	0.6	Assumption[d]
$\omega^{\mathbf{C}}_{\mathbf{high}}$	Pre-descent R_t for residents	4.7	Data fit
$\omega^{\mathbf{G}}_{\mathbf{end}}$	Timing of R_t descent for general population	22 days	Manually fitted [b]
$\omega^{\mathbf{G}}_{\mathbf{rate}}$	Rate of descent of R_t for general population	0.5	Assumption[c]
$\omega^{\mathbf{G}}_{\mathbf{low}}$	Post-descent R_t for general population	0.6	[189]
$\omega^{\mathbf{G}}_{\mathbf{high}}$	Pre-descent R_t for the general population	4.1	Data fit
$\omega^{\gamma}_{\mathbf{end}}$	Timing of descent for visitation	10 days	[170, 175]
$\omega^{\gamma}_{\mathbf{rate}}$	Rate of descent of visitation	3	[170, 175]
$\omega^{\gamma}_{\mathbf{low}}$	Post-descent value for visitation	0	Assumption[e]
$\omega^{\gamma}_{\mathbf{high}}$	Pre-descent R_t for visitation	0.083	Data fit
H_{seeded}	Number of homes seeded	4	Data fit
$E_G(0)$	Initial general population infections	120	Data fit

[a] We assume workers spend half day at work, other half mixing in general population. Alternatively, workers do 12hr shifts. Units are given where appropriate in the Value column.

[b] Manually set by matching the model output to the infection peak dates in the NHS Lothian data [170, 180].

[c] Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [170, 180]. The value of 0.5 corresponds to a descent of $\sim$ 2 weeks.

[d] We assume the reproductive rate for every sub-population drops to the Scottish government's [189] estimated R_t after lockdown (so $\omega^C_{low} = \omega^G_{low}$).

[e] Equals 0 to reflect the policy change to essential visitation only [170, 175], and to avoid the complication of modelling end-of-life visitation.

Data

NHS Lothian is the second-largest health board in Scotland [190], providing public health services to an estimated 907,580 people (2019 mid-year population estimate [186]). The daily confirmed positive tests of COVID-19 cases reported across the entire health board were taken from the Public Health Scotland Open Data [180]. This data does not delineate which cases occurred in care homes, and thus, we retrieved the subset of cases in care homes from Burton et al. [170], which reports a 7-day average of confirmed cases in care home residents. Weekly COVID-19 deaths at the NHS Lothian health board level come from National Records Scotland [181]. Care home resident deaths are a subset of these and are published in [170]. Both death data are weekly counts of registered deaths where COVID- 19 is mentioned on the death certificate (either as the underlying cause or as a contributory factor) [181].

Parameters set using evidence and assumptions

In this section, we describe our assumptions on parameters, parameters estimated from the literature, and parameters manually calibrated to the data.

A Scottish population study between 10th April to 15th June [185] estimated a combined adjusted seroprevalence across their study period (first wave = 10th April to 15th June) of 4.3% (95% CI 4.2%-4.5%). As of the week beginning 15th June 2020, there had been 18,077 positive tests [180], which as a percentage of Scotland's population (2019 census [186]) is $\sim 0.33\%$. We use this information to assume a constant reporting rate in the first wave for the general public of $r_G = 0.33/4.3 \sim 0.077$.

In Scotland, the policy from the start of March to 16th April 2020 was to test only the first few symptomatic care home residents, and afterwards, was to test all symptomatic residents [170]. Assuming when there is an outbreak in a home, 40% of the residents end up infected (40% incidence) [182, 183]. Given 48 residents per care home, until 16th of April we assume a reporting rate of (a few tested)/(total infected) = $3/(0.4 \times 48) = 5/32$. After 16th April, we assume all the symptomatic cases are reported, giving a reporting rate of 4/5 (an estimated symptomatic proportion of COVID-19 cases in long term aged care

is 80% [184]). Between the start of our simulation (6th March 2020) and 16th April 2020

is a time difference of 42 days, and between 17th April 2020 and the end of our simulation

period (15th June 2020) is a time difference of 60 days. Therefore, for 42 days, we assume

a reporting rate of 5/32, and for 60 days, it is 4/5. The weighted average and constant CH

reporting rate over the simulation period is $r_C = (5/32)(42/102) + (4/5)(60/102) \sim 0.53$.

Until the 17th of April, we assume the staff reporting rate was the same as the general

public (0.077). From then on, we assume the care home testing policy change (on the 17th

of April) extended to their staff [170], and the reported percentage of cases was 83% (the

symptomatic proportion [184]). Our weighted average and constant staff reporting rate

over the simulation period is $r_S = 0.077 \times (42/102) + 0.8360/102 \sim 0.52$.

There are two constant death rates in our model: a resident death rate (μ_C) and a

general population death rate (μ_G). We assume care home staff have the same death

rate as the general population. There were ~ 899 positive tests and 423 deaths in NHS

Lothian care home residents over the study period. Using our resident reporting rate, we

estimate there were $899.1/0.53 \sim 1697$ total residents infected with COVID-19 over the

study period. Therefore, we estimate a resident death rate of $\mu_C = 423/1697 \sim 0.25$.

Similarly, there were 3123 total positive tests and 709 deaths over the study period in NHS

Lothian overall. Using our general reporting rate, r_G, we estimate a general population

death rate of $\mu_G = 709/(3123/0.077) \sim 0.017$.

Under our parameterisation, the timing of the drop in reproductive rates for care home

residents (ω_{end}^C) and for the general population (ω_{end}^G) control the timing of peak infections

in each respective population, independent of all other parameter values. This is linked to

the reproductive rate function ($logi$) at an inflection point at $t = \omega_{end}^x$ (where $logi(\Omega_x)$ takes

the value of ($\omega_{high}^x + \omega_{low}^x$)/2). Therefore, we manually set these parameters (ω_{end}^C, ω_{end}^G) by

matching the model output to the infection peak dates in the NHS Lothian data [170, 180].

The ω_{rate}^{G} and ω_{rate}^{C} parameters were assumed to be 0.5. Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [170, 180]. From sensitivity analysis, we found that changing the values of these parameters does not affect the disease dynamics (in terms of total infections/deaths). The ω_{rate} parameter controls the steepness of the descent from ω_{high} to ω_{low}, however the timing of the start of the descent changes to almost cancel out the effect of changing the steepness of this drop. The value of $\omega_{rate}^{G} = \omega_{rate}^{C} = 0.5$, corresponds to a descent of ~ 2 weeks. For ω_{rate}^{γ} we chose the value of 3 to follow the rapid visitation policy changes in care homes [170, 175].

The ω_{low}^{G} and ω_{low}^{C} values were set to 0.6, the estimated R_t after the first wave in Scotland [189]. Due to the uncertainty in the timing of the drop and the R_t peak value, we did not use this source for the ω_{high}^{G} and ω_{end}^{G} parameters.

We have made the simplifying assumption of $\omega_{low}^{\gamma} = 0$, to avoid the complications of modelling end-of-life visitation in care homes.

Closed environments are conducive to COVID-19 transmission and superspreading events [191], therefore we assume pre-lockdown transmission rates within care homes are not less than the general populations, $\omega_{high}^{C} \geqslant \omega_{high}^{G}$.

Care homes operate with staff under differing working hour schedules. This includes care homes having a day and night shift ($\delta = 0.5$), three 8 hour shifts ($\delta = 0.33$) or an uneven distribution of staff spread throughout the day. For simplicity, we assume that all homes operate under two 12 hour shifts per day, i.e., $\delta = 0.5$. Other shifts are explored in the sensitivity analysis.

We make a number of assumptions about the population initially infected. Workers were not initially infected in the model. In the general population, we assume an equal amount of exposed and infected individuals (with and without symptoms), i.e., $E_G(0) = I_G(0) + A_G(0)$. $H_{seeded} = |\{C_j \in \{C_1, \ldots, C_m\} : E_{C_j}(0) > 0\}|$ care homes were seeded with infections, representing introductions such as hospital discharges. To account for the delay in infections at the start of the pandemic in care homes compared to the general population,

as seen in the data Figure 5.2, we assume for all $j \in \{1, \dots, m\}$, $I_{C_j}(0) = A_{C_j}(0) = 0$. We seeded the homes so that initially infected homes lay equally spaced on the circle sharing structure (see Figure 5.1 (b)). If a home is seeded then we assume $E_{C_j}(0) = 1$, and if not, $E_{C_j}(0) = 0$.

Data fit

While some parameter values can be found based on the external data and literature, as shown in the previous section, other parameters were estimated using a formal fit to the cases and deaths data for the Lothian NHS health board (Table 5.1). These parameters were free to vary subject to constraints based on a combination of assumptions and information from the literature. We used the method of least-squares, aggregating the error of model output against the four data sets for NHS Lothian cases and deaths and choosing the parameter set which minimised this error. The data for NHS Lothian population cases and care home cases were in the form of daily and seven day averages respectively. The death data for both the NHS Lothian population and care homes were in weekly counts. To make the fitting consistent, we transformed the daily and seven day average data into weekly data for conformity (Figure 5.2). The constraints on the parameters in our model, described in the previous section, left 6 free parameters for formal fitting. Their ranges used for the data fit are shown in Table 5.2.

Table 5.2. Parameters used for the data fit and the sets of values simulated over.

Parameter	Values considered
ω_{high}^C	$\{3.3, 3.4, \dots, 5\}$
ω_{high}^G	$\{3.3, 3.4, \dots, 4.5\}$
ε	$\{0.1, 0.2, \dots, 0.5\}$
$E_G(0)$	$\{100, 110, \dots, 180\}$
ω_{high}^γ	$\{0.042, 0.083, 0.17\}$
H_{seeded}	$\{1, 2, \dots 10\}$

To investigate the question of how many care homes were exposed at the start of the pandemic, we ran the fitting separately for H_{seeded} fixed at 1 through 10. We simulated the model over 21,060 combinations of the remaining parameters to calculate the least-squares, for each value of H_{seeded}. We investigate the distribution of the parameters in Table 5.2 in the top ten best fitting scenarios, for each value of H_{seeded}.

Sensitivity analysis

After identifying the parameter set that minimises the least-squares, the base case (Table 5.2), we performed a sensitivity analysis. We measured the change in each population's deaths when shifting individual parameters in Table 5.3 from the base case. This allowed us to assess the relative impact of individual parameters on each population. We also measured the change in deaths in each population when changing pairs of the time-share parameters (ω^{γ}_{high}, δ, ε), keeping all other parameters at the base case. The results were stored in a 50×50 grid and visualised using heat-maps to determine the key modes of COVID-19 ingress into and throughout care homes.

216

Table 5.3. Parameters involved in the sensitivity analysis. Sensitivity shift is the unit of change used for each parameter from its base case. These values were chosen to measure the change in each population's deaths to small perturbations of individual parameters from it's base case.

Parameter	Sensitivity shift
ω_{end}^{C}	1
ω_{high}^{C}	0.1
ω_{low}^{C}	0.1
ω_{rate}^{C}	0.1
ω_{end}^{G}	1
ω_{high}^{G}	0.1
ω_{low}^{G}	0.1
ω_{rate}^{G}	0.1
δ	0.1
ε	0.05
ω_{high}^{γ}	0.0167
$E_G(0) = I_G(0) + A_G(0)$	10
λ	0.3
τ	0.4

Results

In this section we first show how the model captures the NHS Lothian data for cases and deaths in the period from March to June 2020, and subsequently show how sensitive the results are to changes in key parameters.

Data fit

The model captures the key features of the COVID-19 related cases and deaths in both care home and general populations, Figure 5.2. The minimum aggregate least-squares error was 33,042, with our model predicting 3,165 total cases and 817 total deaths compared to the total 3,123 cases and 709 deaths in the data. The average difference between data and predictions was 3.5 cases/deaths per week. In care homes we predicted 871 cases and 411 deaths compared to 903 cases and 423 deaths in the data. Our model does not predict

the initial jump in deaths in care homes due to our assumption that infection reporting 320
is constant. Further, our model overestimates the number of deaths for all populations 321
despite a good fit for the cases, as the calculation of death rates is tied to the reporting 322
rates. 323
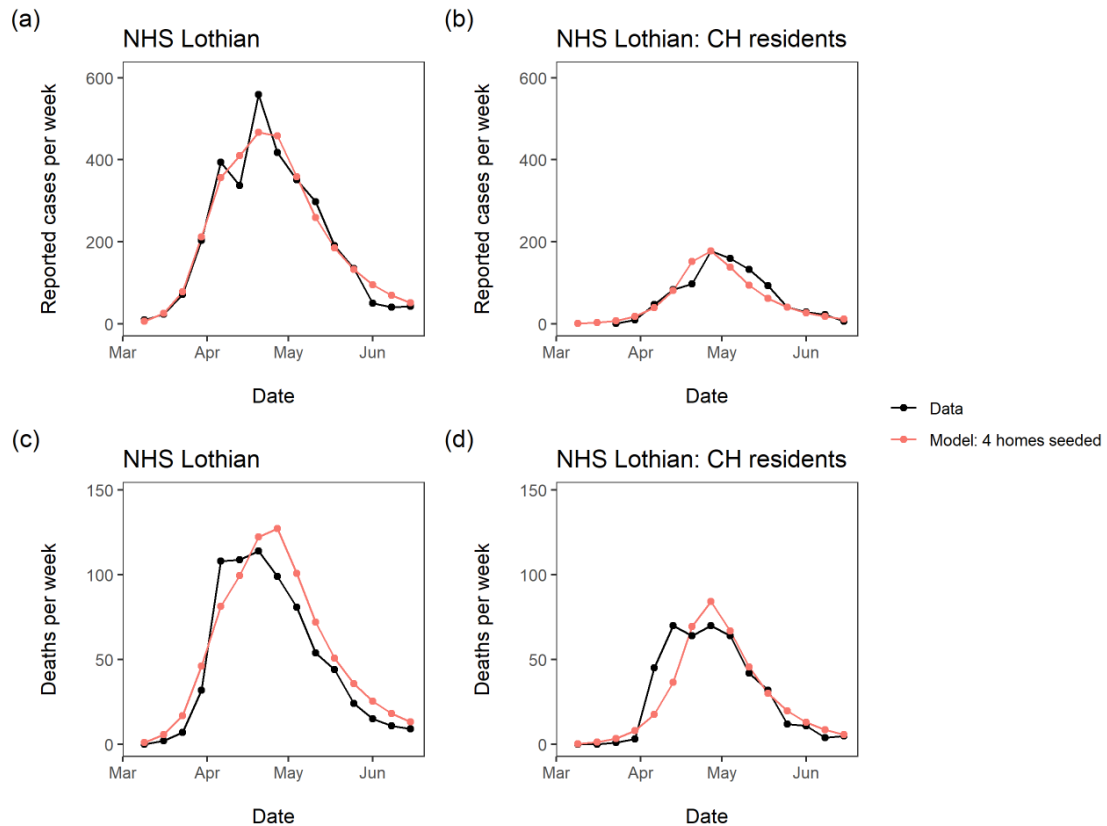
Figure 5.2. Surveillance data and fitted model. Data used for fitting are black lines, and model solution with parameter values in Table 5.1 are red lines. (a) reported cases per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (b) reported cases per week in NHS Lothian care home residents; (c) deaths per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (d) deaths per week in NHS Lothian care home residents.

The reproduction rates change rapidly over the period of April - May 2020, Figure 5.3 324
reflecting the delayed effect of the lockdown. The care home resident population's fall in 325
reproduction rate is delayed by ~ 3 weeks compared to the fall in the general population. 326
This delay is informed by the data, due to the ω_{end} parameters controlling the timing of 327
the peaks in Figure 5.2 - we must be careful when attempting to interpret this delay. 328
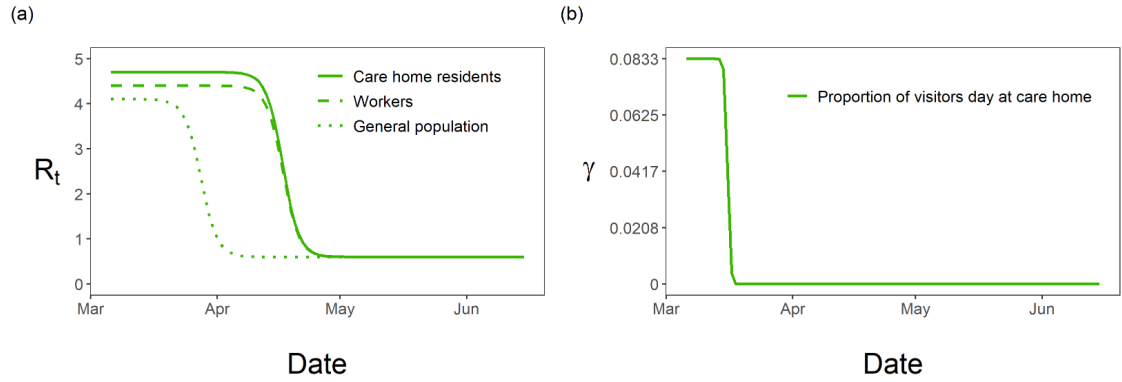
Figure 5.3. Fitted time-dependent parameters. (a) Fitted reproductive numbers over time for care home residents, $R_C(t)$, workers, $R_W(t)$, and general population, $R_G(t)$; (b) fitted visitation, γ, over time with drop highlighting the change in policy.

In order to assess the initial level of care home exposure to virus, we consider the quality   329
of fit as a function of H_{seeded}. The minimum sum of squares of residuals takes the shape of   330
a parabola, with a minimum at $H_{seeded} = 4$, see Figure 5.4. This suggests that a relatively   331
small number of homes were initially exposed to COVID-19.   332
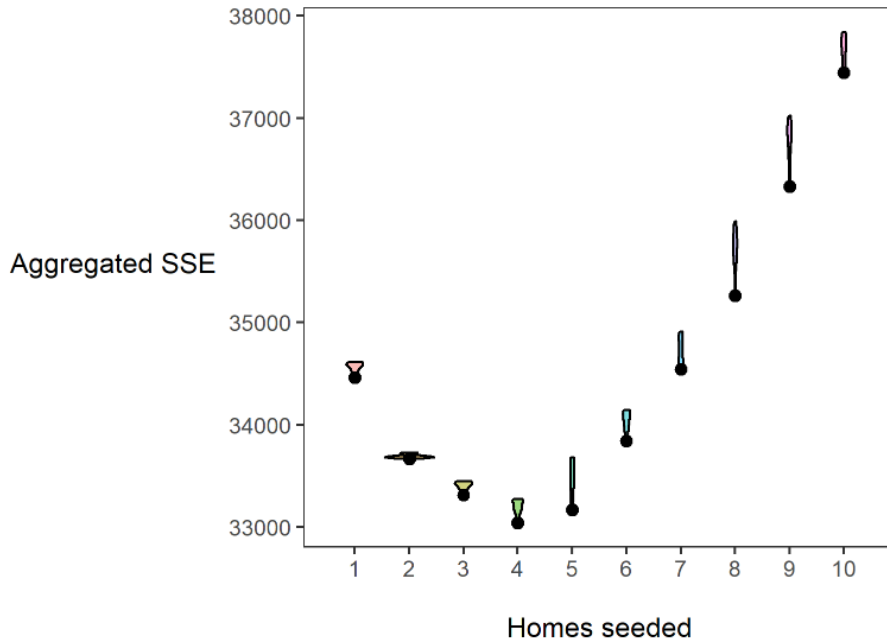
Figure 5.4. Quality of fit as a function of homes seeded. Each violin is the distribution of aggregated sum of squared errors (SSE) in the top ten best-fitting parameter sets, for a number of homes seeded. Black dots indicate the minimum aggregated SSE achieved for each home seeded.

The optimal choice (in terms of the least-squares criterion) for the parameters as used in the data fit (Table 5.2) is relatively stable with respect to changes in H_{seeded}, Figure 5.5. The pre-lockdown reproduction rate in care homes, ω_{high}^C, appears stable in the range of 4.5 to 4.7, changing for 10 homes seeded with the optimal value lowering to 3.9. This highlights the clear link between the reproduction rate and the exposure of care homes at the beginning of the pandemic. For the pre-lockdown reproduction rate in the general population, ω_{high}^G, we see a stable optimal value in the range of 3.9 to 4.2. We see a lot of uncertainty in the proportion of staff shared ε, for H_{seeded} outside of the range of 4 to 8. The distribution for ε changes for $H_{seeded} = 10$, with the optimal value going back up to 0.5. This coincides with the rise in $E_G(0)$ and the substantial fall in ω_{high}^C. This points to the correlation between these 3 parameters. Similarly, there is a lot of uncertainty in the value of pre-lockdown visitation, ω_{high}^γ, with an optimal choice for every value considered in our fitting as we vary H_{seeded}. This uncertainty highlights that the parameters in our model are highly correlated. As seen in Figure 5.6 and Figure 5.7, Figure 5.5 can also be seen to hint at how effective these parameters are at affecting the outcome of the model. For example, the variability in the chosen value of ε and ω_{high}^γ can also be attributed to the relatively small affect they have on the model outcome.
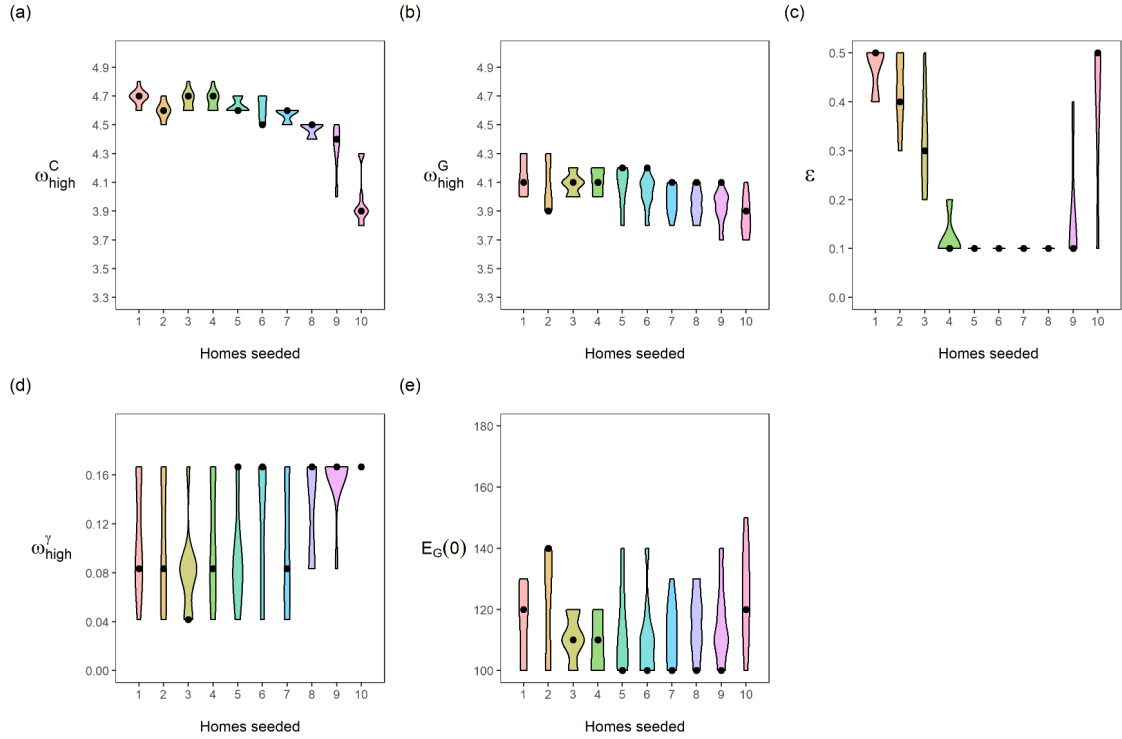
Figure 5.5. Distribution of fitted parameters as a function of homes seeded. Each panel is a different calibrated parameter. Each violin in a panel is the distribution of individual parameters in the top ten best fitting parameter sets, for a number of homes seeded. (a) pre-lockdown care home resident R_t, ω_{high}^C; (b) pre-lockdown general public R_t, ω_{high}^G; (c) staff sharing, ε; (d) visitation pre-lockdown, ω_{high}^γ; (e) general public seeded cases, $E_G(0) = I_G(0) + A_G(0)$. Black dots indicate the parameter value giving the lowest aggregated least square error, for each number of homes seeded.

Sensitivity analysis

Figure 5.6 indicates the sensitivity of the predicted deaths to the parameters in Table 5.3 for care home residents, workers and general population. Predictions are most sensitive to the infectious period, τ, and latency period, λ. The parameters $\{\omega_{end}^C, \omega_{high}^C, \omega_{low}^C\}$ significantly influence care home and worker deaths, without affecting predicted deaths in the general population. Interestingly, a change in one day from when the care home reproduction rate drops, results in almost 10% change in predicted resident and worker deaths. The parameters, $\{\omega_{end}^G, \omega_{high}^G, \omega_{low}^G\}$, controlling the timing of the reproduction rate and it's value before and after lockdown for the general population, significantly affect

predicted deaths in the general population. Interestingly, changing the value of δ by 0.1 ₃₅₉ (20%) results in very little effect to the residents ($< 5\%$) and general population ($< 1\%$) ₃₆₀ but a 15% change in predicted worker deaths. ₃₆₁

predicted deaths in the general population. Interestingly, changing the value of δ by 0.1 [359] (20%) results in very little effect to the residents ($< 5\%$) and general population ($< 1\%$) [360] but a 15% change in predicted worker deaths. [361]
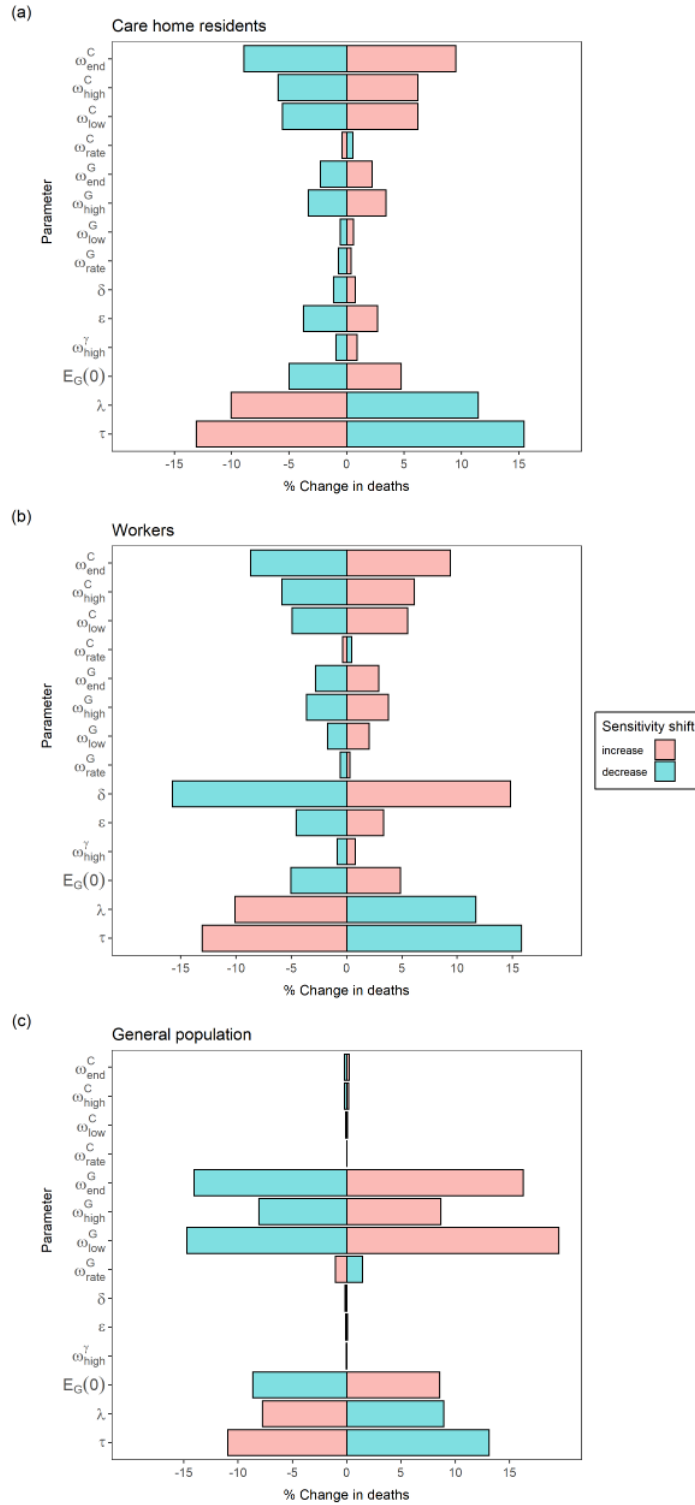
Figure 5.6. Sensitivity of the final deaths in each population to perturbations in model parameters. Each bar shows the % change in final deaths in a population caused by shifting an individual parameter from the base case, keeping all other parameters fixed at the base case (Table 5.1). Each parameter is increased or decreased from its base case value by the corresponding 'sensitivity shift' value in Table 5.3.

In our model, the path of staff catching infections from the community is controlled by δ; staff spreading infections between homes through staff sharing by ε; and visitors bringing infections into homes from the outside community by ω^{γ}_{high}. We investigate the relative risk of these COVID-19 pathways into care homes using Figure 5.7. This shows the combined impact of varying pairs of these parameters on the total resident deaths. Changing staff sharing, ε, and staff shift patterns, δ, the final number of predicted resident deaths do not change significantly, apart from when $\delta \sim 1$ (Figure 5.7). When considering $\delta = 1$, we should also restrict $\varepsilon = 0$, as staff living in care homes would not be shared across them. Varying ε and ω^{γ}_{high} does not significantly effect predicted deaths while holding all other parameters constant at their respective values in Table 5.1. Changing staff shift patterns (δ), has the largest impact on the predicted deaths, especially in the extremes. Reducing ε and ω^{γ}_{high} alone are weaker, but ε has a comparatively larger effect than ω^{γ}_{high}. Reducing ε to 0 and increasing δ to 1 together creates an even higher impact on the predicted deaths Figure 5.7, showing an outbreak severity reduction of $\sim 75\%$ compared to the observed values in Lothian during the first wave. Reducing pre-lockdown visitation from 2 visiting hours per resident to 0 hours per resident, would reduce our predicted first-wave deaths by 20 to 40, about 10% of the death count in the first wave in NHS Lothian.
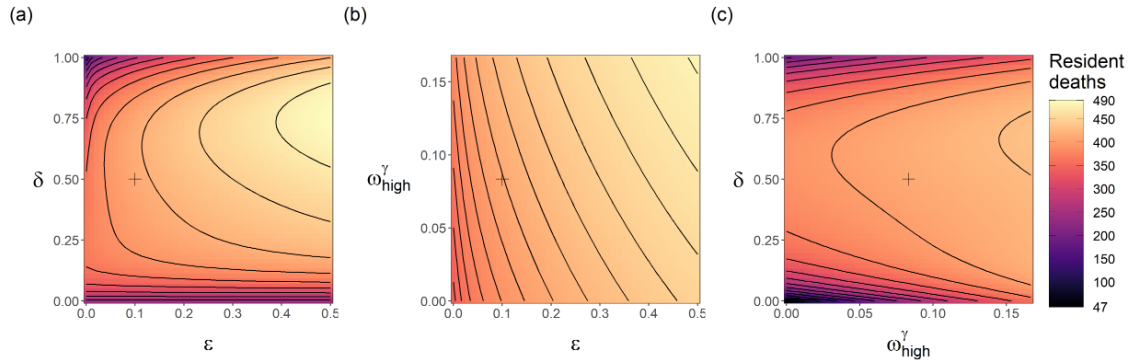
Figure 5.7. Sensitivity of the final resident deaths to the time-share/mixing parameters $(\delta, \varepsilon, \omega^{\gamma}_{high})$. Proportion of CH staff at work is δ, proportion of staff shared between homes is ε, and pre-lockdown visitation is ω^{γ}_{high}. Each panel shows the combined impact of varying two of the time-share/mixing parameters, with all other model parameters fixed as the base case (Table 5.2). The black lines in each panel are isoclines. The cross in each panel indicates the base case value for each parameter.

Discussion

To identify and rank the key modes of ingress into care homes, we used a combination of
modelling, data fit and simulations. We find that homes in our model are more at risk
to outbreaks through staff infections from the general population, relative to visitation or
staff sharing. We also find that outbreaks in our model were not significantly driven by
hospital discharges. These findings coincide with the results from Rosello et al., who used
a stochastic compartmental model on single care homes in England [165]. We additionally
find that in our model the drop in within-home reproduction rate was 3 weeks behind that
of the general population.

Changing worker shift patterns in care homes (δ) only weakly affects the model outcome
for most "reasonable" values e.g. a 2-shift pattern ($\delta = 0.5$) or 3-shift pattern ($\delta = 0.33$). It
is only in the extremes where substantial differences are seen. δ close to 1, alongside low pre-
lockdown visitation, greatly reduces the final outbreak size in care homes. Thus, our results
point to a strategy of staff living-in with residents, in conjunction with timely lowering of
visitation, as an effective pandemic response. This happened in France, where care home
outbreak size was reduced significantly in care homes where staff self-confined [192]. If
living within the care home is not possible, this result of very high levels of δ may imply
that the strategy of segmenting the staff away from both care home residents and the
general population whilst they are not at work would be effective, e.g., organisation of
accommodation for care home workers [159]. From our model, we see the most effective
solution to keeping care homes safe from infection is to focus on the pathway from general
population to workers to residents.

Eliminating staff sharing did not eliminate outbreaks in our model simulations, sug-
gesting that this was not the primary route of infection entering homes. Supporting the
literature [159, 176, 193], reducing staff sharing does reduce the outbreak severity; in our
model, this impact is low compared to other routes. This conclusion is limited due to
our assumption of the circle sharing contact structure, which in turn reflects limited data
availability regarding the contact structure of the care home industry in Lothian (due to
commercial sensitivity). A different contact structure could result in staff sharing lead-

ing to more/less exportation of infection from homes with outbreaks. A more thorough examination on the contact-structure of this system and how that impacts disease spread dynamics would be an important contribution to the literature. One way to achieve this would be to consider an addition of highly connected hubs [194]. We expect that including highly connected worker populations will increase the effect of staff sharing. In our simulations, staff sharing has an effect when there exists a non-uniform distribution of infections in worker sub-populations. Worker sub-populations acting as hubs would acquire disease quicker and skew the distribution of infection amongst worker sub-populations. However, the general population strongly connects all nodes in the network, and dominates the impact of the staff sharing network on disease spread. This is because we assume a single general population with full mixing. At the geographical scale we model (a health board) this is an appropriate assumption, although it would not hold for larger scales of heterogeneity, e.g., the national level (Scotland).

A reduction in visitation reduces predicted resident deaths, as speculated in [159]. Our model predictions support findings that visitation to care homes was not the driving cause of infection in care homes [195]. Since visitation was banned, the evidence for visitation causing outbreaks is limited. Investigation with constant visitation would be necessary to see how the outcome would be different if visitation did not change at all; this was not the focus of our investigation.

Our parameter estimations suggest that after the nation-wide lock down, R_t within care homes dropped three weeks after it did in the wider population. Several possible explanations exist, including differences in testing availability or testing strategies and the difficulty in controlling care home resident interactions to lower disease transmission. Dropping care homes' reproduction rate 1 day earlier, results in a 10% reduction in resident and worker deaths. Therefore a more thorough investigation into the delay between the general populations' and care homes' reproduction rate is necessary to aid closing this gap in future outbreaks.

From the data fit, there were a low number of care homes infected at the beginning of the first wave ($H_{seeded} = 4$). These initial infections could represent hospital discharges or any other pathway. The result supports the claim from a report on English care homes that resident discharges from hospitals were not the primary cause of care home outbreaks [196]. The hospital discharges are only included as initial care home infections, H_{seeded}. In reality, discharges continued during the first wave [170] and more detailed data will be needed to address this problem.

We made a number of simplifying assumptions. Our model does not explicitly account for the variation in susceptibility with age [197]. It is only implicitly addressed by considering different values of β within and outwith of care homes, while keeping the staff and general population homogeneous. Due to the unavailability of data regarding care home worker infections, we expressed worker transmission rates in terms of transmission rates for care home residents and the general population. We assumed the resident-resident and resident-worker transmission rates were equal. However, contacts between care giving staff and residents are likely more frequent and closer than between residents. On the other hand there may be more adherence or better knowledge of how to use PPE among staff. Also, contact between residents could be reduced more easily during the pandemic [159].

The data fit was achieved by minimising the aggregated sum of squared error for each of the four time-series. This method requires the errors to be independent, follow a normal distribution and for the variance to be constant. With the data, we could not estimate the variance over time. To mitigate the effect of the differing variance of the data sets, we shifted the four time-series to the same scale, this being weekly cases/deaths.

The size of individual care homes is believed to be the main factor that influences the likelihood of a care home outbreak [170, 176, 198]. Larger homes typically have more staff and therefore a higher chance of experiencing an outbreak before the smaller ones. In general, we expect larger care homes to receive an increased force of infection from all sources, proportional to its increased size, and therefore an increased outbreak risk. This in turn could increase risk for smaller homes directly connected to the larger ones through staff sharing and visitations, and the overall outbreak risk. However, this effect could

227

balanced by a lowered risk associated with small care homes, with the total population 464
size kept constant. In our model, we assume a uniform home size in order to keep the model 465
generic. As a result (and since the model is deterministic) the risk of staff and visitors 466
bringing in infections is the same for all care homes, which may result in underestimation 467
of the initial rate of spread. An obvious extension of this chapter would be to consider 468
various sources of heterogeneity, including size. 469

The National Records Scotland death data used were the dates of death registration, not 470
the date of death. This is limiting, as we are an average of 3 days behind in the prediction of 471
deaths [181]. The data for care home resident deaths includes deaths in hospital; including 472
nosocomial infections, which we do not take into account into our model. We expect this 473
not to limit the interpretation of our results, considering hospital deaths of care home 474
residents make up approximately 5% of the total care home resident deaths [181]. 475

We do not distinguish explicitly between symptomatic and asymptomatic individuals, 476
as in (**A**) there are people with symptoms that would have been missed by testing. However, 477
asymptomatic infections implicitly affect this model's reporting rates. We do not explicitly 478
model self-isolation or any behavioural change after infection, nor delays or changes in 479
reporting. For simplicity, reported infected individuals are reported immediately. Since we 480
are not explicitly modelling behaviour change once infected, we do not expect incorporating 481
reporting delays to significantly affect our results on the infection pathways. Reporting 482
differed over time, especially in the early weeks of the pandemic when testing was scarce; in 483
care homes, the national policy was to test only the first few symptomatic residents [170]. 484
With our constant reporting rate, we overestimate the number of positive cases prior to 485
the policy change, and underestimate the cases afterwards. A time-dependent reporting 486
rate would impact our death rate and reproductive rate parameters and therefore requires 487
further study. 488

Data regarding care home outbreaks were limited due to the commercial nature of care 489
home organisations in Scotland. Making this data available would allow for additional 490
modelling approaches. Adding in further heterogeneity into the system by including a dis- 491
tribution of home sizes and types would further improve the modelling approach. Includ- 492

ing a stochastic component to this model could lead to more insight into "super-spreader" events in care homes [191] and their effect on epidemic response. Finally, this work focuses on disease dynamics over the course of the first wave, it would be interesting to model this system for future waves and to incorporate the implementation of vaccinations. Extending this system to the rest of Scotland or even the entirety of the UK would be a clear extension of this work.

Conclusion

In our model, the primary route of introduction to the care home sector is from the staff via the general population. Our results also agree with other findings in the literature which suggest that visitation was not the major cause of COVID-19 establishment in care homes [195] and that hospital discharges accounted for a small number of care home outbreaks early on in the pandemic [196]. Prioritising the protection and monitoring of staff, second to timely reductions in visitation and staff sharing, are what our model suggests for an effective reduction in outbreak size. Our findings also show difficulties in halting the spread of COVID-19 within the homes in tandem with the general population. This highlights the need for more planning and support for care homes and their staff in organising quick and effective responses to emerging pandemics. The lessons to be learned from the first wave of COVID-19 will be crucial to minimise the damage from future outbreaks.

Acknowledgements

Chapter 6

Conclusions

This thesis aimed to understand the structure of UK plant trade in relation to the spread of pests and diseases. Data is hard to obtain, often of a commercially sensitive nature and therefore very little is known on the structure of plant trade in the UK at the national level. Even less is known about how pests and diseases spread throughout this network. In response to this, we have formulated a method of trade network construction, which can be improved from additional data. Using the unique access to data that we had, we constructed a trade network model and characterised nurseries that trade on a spectrum as opposed to the simplified roles of wholesaler, grower and retailer as seen in the literature. The network we constructed is also novel in the inclusion of customers in the network and our application of edge weights which results in heterogeneous trade.

Our social network analysis highlighted the importance of including edge-weights to network models as different conclusions can be made when analysing a network with or without edge-weights. We also found that customers do not contribute significantly to network structure, supporting the network analysis of previous research. From our analysis of the hubs, authorities and betweenness centrality measures we concluded that retailers are most at risk of acquiring infection, nurseries which specialise sales to retailers cause the largest outbreaks and nurseries which specialise sales to other nurseries are good candidates for trade restriction measures. Additionally, we found that this is robust for small changes in market structure. This is an important result, given that there is interest regarding how changes to plant trade in the UK may affect disease spread (in response to the UK's

exit from the European Union for example). However, a change in UK connections to the international trade network is something our research does not address.

To understand how to control the spread of disease from trade, we extended our network model into a compartmental, metapopulation framework. From our analysis on this model, we found that a disease seeded in nurseries that sell mostly to other nurseries greatly increased the rate of spread throughout the network and raised the probability of an outbreak when compared to seedings in other nurseries. Changes in the rate of transmission suggested a requirement of infection to more than double every month for a large epidemic to occur. However, this requirement may be lower for infected plants that stay in nurseries and retailers for longer than one month. We found that the utility of inspecting consignments to control disease decreased as the frequency of inspecting all stock on site increased. Depending on price and efficacy, more frequent inspections could be cost-effective for plant nurseries. Our model gave estimations for the inspection price thresholds where this change occurs. The most effective control strategy we found from this analysis was to inspect all plants in nurseries each month with at least 50% efficacy. Therefore efforts should be made to make this feasible, such as reducing inspection costs and/or diverting resources from inspecting consignments. The results concerning inspection frequency are particularly relevant to the continued development of the Plant Healthy scheme [154].

The COVID-19 pandemic presented an opportunity to use compartmental metapopulation models, similar to Chapter 4, to understand COVID-19 transmission into the care home community. Our analysis showed that the primary driver of COVID-19 into care homes was the infection that care home staff acquired from the general population, providing theoretical support for staff living in with care home residents during a pandemic scenario. Our results also showed that visitation from the general population was not the major cause of the establishment of COVID-19 into care homes. The data fitting suggested that the number of care homes initially seeded with infection was low. This chapter was an additional example of the understanding that networked metapopulation models can give to the way diseases spread throughout isolated populations via the movements of individuals.

The plant trade models in this thesis have been constructed with many unknown parameters, the analysis has thus been limited and many assumptions have been made. Future work with a focus on a specific pest/disease, with access to more data to provide parameter estimations can overcome these limitations. This thesis acts as an initial exploration in using sales data to inform trade networks and to understand how disease spreads throughout the UK plant trade network and the vulnerabilities that lie within.

Future work

Our plant trade network model captured trader behaviour and the variability in the volume of plants traded across the network. We chose not to embed our plant trade network model onto a grid-based geography of the UK, as in *Harwood et al.*, because the scope of our research was focused solely on modelling the structure of the network [46]. As a result, our research is restricted to the introduction of currently absent pests and diseases to the UK. A clear extension to this trade model would be to superimpose the network trade model onto a spatial model of the UK, allowing the modelling of currently present pests/diseases.

In our analysis for chapters 2 and 3, we investigated structural changes in the market, represented by a change in the distribution of nursery subcategories. We found that our results were robust to small changes in market structure. This investigation could be extended by looking at the effects of changes in the definition of each nursery subcategory. For example, we could alter the network model to coincide more closely to the wholesaler, grower and retailer paradigm, by assuming that N_{Nur} nurseries have an in-degree of 0, i.e., they only sell. This would have significant impacts on network structure, with one nursery type being cut off from receiving any infected plants from the rest of the network.

Scheduled inspections in Chapter 3 were applied uniformly to nurseries and retailers in the model. An interesting further investigation would be to look into scheduled inspections to different subsets of nurseries and retailers. For instance, only applying scheduled in-

spections to one nursery subcategory would have enabled a comparison of the results we obtained from our social network analysis in Chapter 2. It would also be interesting to see the effect of commercial customers applying their own trade inspections, as this could lower the burden of disease control on nurseries.

Bibliography

1. DEFRA, "Plant Biosecurity Strategy for Great Britain," 2023. [Online]. Available: https://www.gov.uk/government/publications/plant-biosecurity-strategy-for-great-britain-2023-to-2028/plant-biosecurity-strategy-for-great-britain-2023-to-2028

2. S. McKelvey, F. Koch, and B. Smith, "Predicting Movement of Nursery Hosts Using a Linear Network Model," *Proceedings of the Sudden Oak Death Third Science Symposium Predicting*, pp. 249–256, 2007.

3. N. D. Wolfe, C. P. Dunavan, and J. Diamond, "Origins of major human infectious diseases," *Nature*, vol. 447, no. 7142, pp. 279–283, 2007.

4. A. Santini, A. Liebhold, D. Migliorini, and S. Woodward, "Tracing the role of human civilization in the globalization of plant pathogens," *ISME Journal*, vol. 12, no. 3, pp. 647–652, 2018.

5. F. E. Cox, "History of human parasitology," *Clinical Microbiology Reviews*, vol. 15, no. 4, pp. 595–612, 2002.

6. C.A. Janeway, P. Travers, M. Walport, M.J Shlomchik, *Immunobiology: The Immune System in Health and Disease.* Garland Science, 2001, ch. 10.

7. A. Homei and M. Worboys, *Fungal Disease in Britain and the United States 1850-2000: Mycoses and Modernity.* Palgrave Macmillan, 2013.

8. J. R. Porter, "Antony Van Leeuwenhoek. Tercentenary of his discovery of bacteria," *Bacteriological Reviews*, vol. 40, no. 2, pp. 260–269, 1976.

9. J. Bové and H. Lecoq, "Discovery of the first virus, Tobacco mosaic virus: 1892 or 1898 ?" *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, vol. 324, no. 10, pp. 929–933, 2001.

10. C. Boyles, P. G. Koehler, "All about pests," 2013. [Online]. Available: https://ipm.ifas.ufl.edu/pdfs/Cir543-All_About_Pests.pdf

11. F. M. Fishel, "Pest Management and Pesticides: A Historical Perspective," *Edis*, vol. 2010, no. 1, pp. 1–6, 2010.

12. J. K. Brown, "A cost of disease resistance: Paradigm or peculiarity?" *Trends in Genetics*, vol. 19, no. 12, pp. 667–671, 2003.

13. I. Saccheri and I. Hanski, "Natural selection and population dynamics," *Trends in Ecology and Evolution*, vol. 21, no. 6, pp. 341–347, 2006.

14. J. Rimer, I. R. Cohen, and N. Friedman, "Do all creatures possess an acquired immune system of some sort?" *BioEssays*, vol. 36, no. 3, pp. 273–281, 2014.

15. G. G. Dimijian, "Evolving together: the biology of symbiosis, part 1," *Baylor University Medical Center Proceedings*, vol. 75234, pp. 217–226, 2000.

16. C. M. Brasier, "The biosecurity threat to the UK and global environment from international trade in plants," *Plant Pathology*, vol. 57, no. 5, pp. 792–808, 2008.

17. R. Epanchin-Niell, C. McAusland, A. Liebhold, P. Mwebaze, and M. R. Springborn, "Biological invasions and international trade: Managing a moving target," *Review of Environmental Economics and Policy*, vol. 15, no. 1, pp. 180–190, 2021.

18. C. Potter and J. Urquhart, "Tree disease and pest epidemics in the Anthropocene: A review of the drivers, impacts and policy responses in the UK," *Forest Policy and Economics*, vol. 79, pp. 61–68, 2017.

19. N. Nunn and N. Qian, "The Columbian exchange: A history of disease, food, and ideas," *Journal of Economic Perspectives*, vol. 24, no. 2, pp. 163–188, 2010.

20. T. L. O'Halloran, B. E. Law, M. L. Goulden, Z. Wang, J. G. Barr, C. Schaaf, M. Brown, J. D. Fuentes, M. Göckede, A. Black, and V. Engel, "Radiative forcing of natural forest disturbances," *Global Change Biology*, vol. 18, no. 2, pp. 555–565, 2012.

21. S. Y. Padmanabhan, "The Great Bengal Famine," *Annual Review of Phytopathology*, vol. 11, no. 1, pp. 11–24, 1973.

22. K. Yoshida, V. J. Schuenemann, L. M. Cano, M. Pais, B. Mishra, R. Sharma, C. Lanz, F. N. Martin, S. Kamoun, J. Krause, M. Thines, D. Weigel, and H. A. Burbano, "The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine," *eLife*, vol. 2013, no. 2, pp. 1–25, 2013.

23. J. Flood and R. Day, "Managing risks from pests in global commodity networks – policy perspectives," *Food Security*, vol. 8, no. 1, pp. 89–101, 2016.

24. N. J. Grünwald, M. Garbelotto, E. M. Goss, K. Heungens, and S. Prospero, "Emergence of the sudden Oak death pathogen *Phytophthora ramorum*," *Trends in Microbiology*, vol. 20, no. 3, pp. 131–138, 2012.

25. DEFRA, "*Phytophthora ramorum* and *kernoviae*." [Online]. Available: https://planthealthportal.defra.gov.uk/pests-and-diseases/high-profile-pests-and-diseases/phytophthora/

26. M. Pautasso, T. F. Döring, M. Garbelotto, L. Pellis, and M. J. Jeger, "Impacts of climate change on plant diseases-opinions and trends," *European Journal of Plant Pathology*, vol. 133, no. 1, pp. 295–313, 2012.

27. L. Hill, G. Jones, N. Atkinson, A. Hector, G. Hemery, and N. Brown, "The £15 billion cost of Ash dieback in Britain," *Current Biology*, vol. 29, no. 9, pp. R315–R316, 2019.

28. P.A. Arneson, "Coffee rust," 2000. [Online]. Available: https://www.apsnet.org/edcenter/disandpath/fungalbasidio/pdlessons/Pages/CoffeeRust.aspx

29. C. J. Rodrigues, A. J. Bettencourt, and L. Rijo, "Races of the Pathogen and Resistance to Coffee Rust," *Annual Review of Phytopathology*, vol. 13, no. 1, pp. 49–70, 1975.

30. J. B. Ristaino, M. L. Gumpertz, "New frontiers in the study of dispersal and spatial analysis of epidemics caused by species in the genus phytophthora," *Annual Review of Phytopathology*, vol. 38, no. 1, pp. 541–576, 2000, pMID: 11701854.

31. A. MacLeod, M. Pautasso, M. J. Jeger, and R. Haines-Young, "Evolution of the international regulation of plant pests and challenges for future plant health," *Food Security*, vol. 2, no. 1, pp. 49–70, 2010.

32. DEFRA, "*Xylella fastidiosa* - UK Plant Health Information Portal," 2017. [Online]. Available: https://planthealthportal.defra.gov.uk/pests-and-diseases/high-profile -pests-and-diseases/xylella/

33. Woodland Trust, "Oak Processionary Moth." [Online]. Available: https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/tree-pests-and-diseases/key-tree-pests-and-diseases/oak-processionary-moth/

34. DEFRA, Forestry Commission, "Public urged to report sightings of the tree pest Oak processionary moth," 2022. [Online]. Available: https://www.gov.uk/government/news/public-urged-to-report-sightings-of-the-tree-pest-oak-processionary-moth

35. Forestry Commission, "Oak processionary moth (*Thaumetopoea processionea*)," 2019. [Online]. Available: https://www.forestresearch.gov.uk/tools-and-resources/pest-and-disease-resources/oak-processionary-moth-thaumetopoea-processionea/

36. DEFRA, "UK Risk Register Details for *Thaumetopoea processionea*," 2022. [Online]. Available: https://planthealthportal.defra.gov.uk/pests-and-diseases/uk -plant-health-risk-register/viewPestRisks.cfm?cslref=7319

37. R. H. Baker, H. Anderson, S. Bishop, A. Macleod, N. Parkinson, and M. G. Tuffen, "The UK Plant Health Risk Register: A tool for prioritizing actions," *EPPO Bulletin*, vol. 44, no. 2, pp. 187–194, 2014.

38. Public Health England, "Health effects of exposure to setae of Oak processionary moth larvae Systematic review About Public Health England," *Public Health England*, 2015. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/432003/Oak_Processionary_Moth_FINAL___2_.pdf

39. M. Bergman, "The Potato Blight in The Netherlands," *International Review of Social History*, vol. 12, no. 3, pp. 390–431, 1967.

40. K. Ivors, M. Garbelotto, I. D. Vries, C. Ruyter-Spira, B. T. Hekkert, N. Rosenzweig, and P. Bonants, "Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations," *Molecular Ecology*, vol. 15, no. 6, pp. 1493–1505, 2006.

41. E. M. Goss, M. Larsen, G. A. Chastagner, D. R. Givens, and N. J. Grünwald, "Population genetic analysis infers migration pathways of *Phytophthora ramorum* in US nurseries," *PLoS Pathogens*, vol. 5, no. 9, 2009.

42. E. M. Goss, M. Larsen, A. Vercauteren, S. Werres, K. Heungens, and N. J. Grünwald, "*Phytophthora ramorum* in Canada: Evidence for migration within North America and from Europe," *Phytopathology*, vol. 101, no. 1, pp. 166–171, 2011.

43. S. Prospero, E. M. Hansen, N. J. Grünwald, and L. M. Winton, "Population dynamics of the sudden Oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004," *Molecular Ecology*, vol. 16, no. 14, pp. 2958–2973, 2007.

44. S. Prospero, N. J. Grünwald, L. M. Winton, and E. M. Hansen, "Migration patterns of the emerging plant pathogen *Phytophthora ramorum* on the West Coast of the United States of America," *Phytopathology*, vol. 99, no. 6, pp. 739–749, 2009.

45. V. Chadfield and M. Pautasso, "*Phytophthora ramorum* in england and wales: which environmental variables predict county disease incidence?" *Forest Pathology*, vol. 42, no. 2, pp. 150–159, 2012.

46. T. D. Harwood, X. Xu, M. Pautasso, M. J. Jeger, and M. W. Shaw, "Epidemiological risk assessment using linked network and grid based modelling: *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK," *Ecological Modelling*, vol. 220, no. 23, pp. 3353–3361, 2009.

47. E. Schieber and G. Zentmyer, "Coffee Rust in the Western Hemisphere," *Plant Disease*, vol. 68, pp. 89–93, 1984.

48. Oxford Economics, "The economic impact of ornomental horticultural and landscaping in the UK," Oxford Economics, Technical Report, 2018. [Online]. Available: https://www.oxfordeconomics.com/resource/3b5ce883-cc72-4cf9-910e -be267fe93f46/

49. I. Tomlinson and C. Potter, "'Too little, too late'? Science, policy and Dutch Elm Disease in the UK," *Journal of Historical Geography*, vol. 36, no. 2, pp. 121–131, 2010.

50. D. Chapman, B. V. Purse, H. E. Roy, and J. M. Bullock, "Global trade networks determine the distribution of invasive non-native species," *Global Ecology and Biogeography*, vol. 26, no. 8, pp. 907–917, 2017.

51. Forest research, "Xylella (Xylella distribution)." [Online]. Available: https: //www.forestresearch.gov.uk/tools-and-resources/fthr/pest-and-disease-resources/ xylella-xylella-fastidiosa/

52. C. Bragard, K. Dehnen-Schmutz, F. Di Serio, P. Gonthier, M. Jacques, J. A. J. Miret, A. F. Justesen, A. MacLeod, C. S. Magnusson, P. Milonas, J. A Navas-Cortés, R. Potting, P. L. Reignault, H. Thulke, W. van der Werf, A. V. Civera,J. Yuen, L. Zappalà, D. Boscia, ..., and Stephen Parn, "Update of the scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the eu territory," *EFSA Journal*, vol. 17, no. 5, 2019.

53. M. Jeger, D. Caffier, T. Candresse, E. Chatzivassiliou, K. Dehnen-Schmutz, G. Gilioli, J. C. Grégoire, J. A. J. Miret, A. MacLeod, M. N. Navarro, B. Niere, S. Parnell, R. Potting, T. Rafoss, V. Rossi, G. Urek, A. Van Bruggen, W. Van der Werf, J. West, ... , and C. Bragard, "Updated pest categorisation of *Xylella fastidiosa*," *EFSA Journal*, vol. 16, no. 7, 2018.

54. K. P. Tumber, J. M. Alston, and K. B. Fuller, "Pierce's disease costs California $104 million per year," *California Agriculture*, vol. 68, no. 1-2, pp. 20–29, 2014.

55. DEFRA, "UK Risk Register Details for *Xylella fastidiosa*," 2022. [Online]. Available: https://planthealthportal.defra.gov.uk/pests-and-diseases/uk-plant-health-risk-register/viewPestRisks.cfm?cslref=12570

56. S. H. Reichard and P. White, "Horticulture as a pathway of invasive plant introductions in the United States," *BioScience*, vol. 51, no. 2, pp. 103–113, 2001.

57. J. M. Davidson, A. C. Wickland, H. A. Patterson, K. R. Falk, and D. M. Rizzo, "Transmission of *Phytophthora ramorum* in mixed-evergreen forest in California," *Phytopathology*, vol. 95, no. 5, pp. 587–596, 2005.

58. M. Javal, A. Roques, J. Haran, F. Hérard, M. Keena, and G. Roux, "Complex invasion history of the Asian long-horned beetle: fifteen years after first detection in Europe," *Journal of Pest Science*, vol. 92, no. 1, pp. 173–187, 2019.

59. K. Dehnen-Schmutz, J. Touza, C. Perrings, and M. Williamson, "The horticultural trade and ornamental plant invasions in Britain," *Conservation Biology*, vol. 21, no. 1, pp. 224–231, 2007.

60. N. E. Lehan, J. R. Murphy, L. P. Thorburn, and B. A. Bradley, "Accidental introductions are an important source of invasive plants in the continental united states," *American Journal of Botany*, vol. 100, no. 7, pp. 1287–1293, 2013. [Online]. Available: https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1300061

61. A. M. Liebhold, E. G. Brockerhoff, L. J. Garrett, J. L. Parke, and K. O. Britton, "Live plant imports: The major pathway for forest insect and pathogen invasions of the US," *Frontiers in Ecology and the Environment*, vol. 10, no. 3, pp. 135–143, 2012.

62. R. M. Smith, R. H. Baker, C. P. Malumphy, S. Hockland, R. P. Hammon, J. C. Ostojá-Starzewski, and D. W. Collins, "Recent non-native invertebrate plant pest establishments in Great Britain: Origins, pathways, and trends," *Agricultural and Forest Entomology*, vol. 9, no. 4, pp. 307–326, 2007.

63. J. R. Newhouse, "Chestnut Blight," *Scientific american*, vol. 263, no. July, pp. 106–111, 1990.

64. P. H. Freer-Smith and J. F. Webber, "Tree pests and diseases: the threat to biodiversity and the delivery of ecosystem services," *Biodiversity and Conservation*, vol. 26, no. 13, pp. 3167–3181, 2015.

65. M. J. MacLachlan, A. M. Liebhold, T. Yamanaka, and M. R. Springborn, "Hidden patterns of insect establishment risk revealed from two centuries of alien species discoveries," *Science Advances*, vol. 7, no. 44, pp. 1–10, 2021.

66. G. Gordh and S. McKirdy, *The Handbook of Plant Biosecurity*. Springer, 2014.

67. A. M. Liebhold, B. Ludek, E. G. Brockerhoff, R. S. Epanchin-Niell, A. Hastings, D. A. Herms, J. M. Kean, D. G. McCullough, D. M. Suckling, P. C. Tobin, and T. Yamanaka, "Eradication of Invading Insect Populations: From Concepts to Applications," *Annual Review of Entomology*, vol. 61, pp. 335–352, 2016.

68. J. L. Lockwood, P. Cassey, and T. Blackburn, "The role of propagule pressure in explaining species invasions," *Trends in Ecology & Evolution*, vol. 20, no. 5, pp. 223–228, 2005, special issue: Invasions, guest edited by Michael E. Hochberg and Nicholas J. Gotelli.

69. K. Dehnen-Schmutz, J. Touza, C. Perrings, and M. Williamson, "A century of the ornamental plant trade and its impact on invasion success," *Diversity and Distributions*, vol. 13, no. 5, pp. 527–534, 2007.

70. M. Pautasso, X. Xu, M. J. Jeger, T. D. Harwood, M. Moslonka-Lefebvre, and L. Pellis, "Disease spread in small-size directed trade networks: The role of hierarchical categories," *Journal of Applied Ecology*, vol. 47, no. 6, pp. 1300–1309, 2010.

71. M. Newman, *Networks: An Introduction.* Oxford University Press, 2010.

72. S. Rahman, *Basic Graph Theory.* Springer, 2017.

73. A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network Analysis.* Academic Press, 2016, ch. 4, pp. 115–136.

74. R. A. Stein, "Super-spreaders in infectious diseases," *International Journal of Infectious Diseases*, vol. 15, no. 8, pp. e510–e513, 2011.

75. W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London*, vol. 115, no. 772, pp. 700–721, 1927.

76. A. M. Bate, G. Jones, A. Kleczkowski, A. MacLeod, R. Naylor, J. Timmis, J. Touza, and P. C. White, "Modelling the impact and control of an infectious disease in a plant nursery with infected plant material inputs," *Ecological Modelling*, vol. 334, pp. 27–43, 2016.

77. S. M. White, J. A. Navas-Cortés, J. M. Bullock, D. Boscia, and D. S. Chapman, "Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives," *Plant Pathology*, vol. 69, no. 8, pp. 1403–1413, 2020.

78. M. T. Signes-Pont, J. J. Cortés-Plana, H. Mora, and R. Mollá-Sirvent, "An epidemic model to address the spread of plant pests. The case of *Xylella fastidiosa* in almond trees," *Kybernetes*, vol. 50, no. 10, pp. 2943–2955, 2021.

79. O. Martin, Y. Fernandez-Diclo, J. Coville, and S. Soubeyrand, "Equilibrium and sensitivity analysis of a spatio-temporal host-vector epidemic model," *Nonlinear Analysis: Real World Applications*, vol. 57, p. 103194, 2021.

80. P. Magal and S. Ruan, *Structured Population Models in Biology and Epidemiology.* Springer Berlin, Heidelberg, 2008, vol. 1.

81. L. Wang and X. Li, "Spatial epidemiology of networked metapopulation: an overview," *Chinese Science Bulletin*, vol. 59, no. 28, pp. 3511–3522, 2014.

82. A.-L. Barabási and M. Pósfai, *Network science.* Cambridge University Press, 2016.

83. M. Moslonka-lefebvre, A. Finley, I. Dorigatti, K. Dehnen-schmutz, T. Harwood, M. J. Jeger, X. Xu, O. Holdenrieder, and M. Pautasso, "Networks in Plant Epidemiology: From Genes to Landscapes, Countries, and Continents," *Phytopathology*, vol. 101, pp. 392–403, 2011.

84. R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

85. R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, 2002.

86. M. J. Jeger, M. Pautasso, O. Holdenrieder, and M. W. Shaw, "Modelling disease spread and control in networks: Implications for plant sciences," *New Phytologist*, vol. 174, no. 2, pp. 279–297, 2007.

87. D. Watts ,S. Strogatz , "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

88. M. Pautasso, M. J. Jeger, "Epidemic threshold and network structure: The interplay of probability of transmission and of persistence in small-size directed networks," *Ecological Complexity*, vol. 5, no. 1, pp. 1–8, 2008.

89. M. Moslonka-Lefebvre, M. Pautasso, and M. J. Jeger, "Disease spread in small-size directed networks: Epidemic threshold, correlation between links to and from nodes, and clustering," *Journal of Theoretical Biology*, vol. 260, no. 3, pp. 402–411, 2009.

90. D. Karemera, V. I. Oguledo, and B. Davis, "A gravity model analysis of international migration to north america," *Applied Economics*, vol. 32, no. 13, pp. 1745–1755, 2000.

91. L. D. Benedictis and D. Taglioni, *The Gravity Model in International Trade.* Springer Berlin Heidelberg, 2011, pp. 55–89.

92. R. Jandarov, M. Haran, O. Bjørnstad, and B. Grenfell, "Emulating a Gravity Model to Infer the Spatiotemporal Dynamics of an Infectious Disease," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 63, no. 3, pp. 423–444, 11 2013.

93. M. Stefanouli and S. Polyzos, "Gravity vs radiation model: two approaches on commuting in Greece," *Transportation Research Procedia*, vol. 24, pp. 65–72, 2017.

94. T.-D. Hua, A.-T. Nguyen-Thi, and T.-A. H. Nguyen, "Link prediction in weighted network based on reliable routes by machine learning approach," *2017 4th NAFOS-TED Conference on Information and Computer Science*, pp. 236–241, 2017.

95. C. Fu, M. Zhao, L. Fan, X. Chen, J. Chen, Z. Wu, Y. Xia, and Q. Xuan, "Link weight prediction using supervised learning methods and its application to yelp layered network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 8, pp. 1507–1518, 2018.

96. I. Z. Kiss, J. C. Miller, and P. L. Simon, *Introduction to networks and diseases.* Springer, 2017, vol. 46.

97. V. Colizza and A. Vespignani, "Invasion threshold in heterogeneous metapopulation networks," *Physical Review Letters*, vol. 99, no. 14, pp. 1–4, 2007.

98. M. Pautasso and M. J. Jeger, "Network epidemiology and plant trade networks," *AoB PLANTS*, vol. 6, pp. 1–14, 2014.

99. M. Morris and M. Kretzschmar, "Concurrent partnerships and transmission dynamics in networks," *Social Networks*, vol. 17, no. 3-4, pp. 299–318, 1995.

100. R. M. Christley and N. P. French, "Small-world topology of UK racing: The potential for rapid spread of infectious agents," *Equine Veterinary Journal*, vol. 35, no. 6, pp. 586–589, 2003.

101. M. D. Shirley and S. P. Rushton, "The impacts of network topology on disease spread," *Ecological Complexity*, vol. 2, no. 3, pp. 287–299, 2005.

102. S. Bansal, B. T. Grenfell, and L. A. Meyers, "When individual behaviour matters: Homogeneous and network models in epidemiology," *Journal of the Royal Society Interface*, vol. 4, no. 16, pp. 879–891, 2007.

103. J. A. Drewe, K. T. Eames, J. R. Madden, and G. P. Pearce, "Integrating contact network structure into tuberculosis epidemiology in meerkats in South Africa: Implications for control," *Preventive Veterinary Medicine*, vol. 101, no. 1-2, pp. 113–120, 2011.

104. D. Smilkov and L. Kocarev, "Influence of the network topology on epidemic spreading," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 1, pp. 1–10, 2012.

105. N. Schwartz, R. Cohen, D. Ben-Avraham, A. L. Barabási, and S. Havlin, "Percolation in directed scale-free networks," *Physical Review E*, vol. 66, no. 1, pp. 1–4, 2002.

106. M. F. Nelson and C. E. Bone, "Effectiveness of dynamic quarantines against pathogen spread in models of the horticultural trade network," *Ecological Complexity*, vol. 24, pp. 14–28, 2015.

107. G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006. [Online]. Available: http://igraph.sf.net

108. A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

109. A. E. Jones, L. A. Munro, D. M. Green, K. L. Morgan, A. G. Murray, R. Norman, D. Ryder, N. K. Salama, N. G. Taylor, M. A. Thrush, I. S. Wallace, and K. J. Sharkey, "The contact structure of Great Britain's salmon and trout aquaculture industry," *Epidemics*, vol. 28, p. 100342, 2019.

110. H. H. Lentz, A. Koher, P. Hövel, J. Gethmann, C. Sauter-Louis, T. Selhorst, and F. J. Conraths, "Disease spread through animal movements: A static and temporal network analysis of pig trade in Germany," *PLoS ONE*, vol. 11, no. 5, pp. 1–32, 2016.

111. A. S. Ruget, G. Rossi, P. T. Pepler, G. Beaunée, C. J. Banks, J. Enright, and R. R. Kao, "Multi-species temporal network of livestock movements for disease spread," *Applied Network Science*, vol. 6, no. 1, 2021.

112. J. Enright and R. R. Kao, "A descriptive analysis of the growth of unrecorded interactions amongst cattle-raising premises in Scotland and their implications for disease spread," *BMC Veterinary Research*, vol. 12, no. 1, pp. 1–6, 2016.

113. C. Hautefeuille, G. Dauphin, and M. Peyre, "Knowledge and remaining gaps on the role of animal and human movements in the poultry production and trade networks in the global spread of avian influenza viruses - A scoping review," *PLoS ONE*, vol. 15, pp. 1–21, 2020.

114. J. M. Radin, R. A. Shaffer, S. P. Lindsay, M. R. G. Araneta, R. Raman, and J. H. Fowler, "International chicken trade and increased risk for introducing or reintroducing highly pathogenic avian influenza A (H5N1) to uninfected countries," *Infectious Disease Modelling*, vol. 2, no. 4, pp. 412–418, 2017.

115. G. Fournié, A. Tripodi, T. T. T. Nguyen, V. T. Nguyen, T. T. Tran, A. Bisson, D. U. Pfeiffer, and S. H. Newman, "Investigating poultry trade patterns to guide avian influenza surveillance and control: A case study in Vietnam," *Scientific Reports*, vol. 6, pp. 1–10, 2016.

116. G. Fournie, J. Guitian, S. Desvaux, V. C. Cuong, D. H. Dung, D. U. Pfeiffer, P. Mangtani, and A. C. Ghani, "Interventions for avian influenza A (H5N1) risk management in live bird market networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 22, pp. 9177–9182, 2013.

117. A. Brioudes and B. Gummow, "Understanding Pig and Poultry Trade Networks and Farming Practices Within the Pacific Islands as a Basis for Surveillance," *Transboundary and Emerging Diseases*, vol. 64, no. 1, pp. 284–299, 2017.

118. V. Martin, X. Zhou, E. Marshall, B. Jia, G. Fusheng, M. A. FrancoDixon, N. de Haan, D. U. Pfeiffer, R. J. Soares Magalhães, and M. Gilbert, "Risk-based surveillance for avian influenza control along poultry market chains in South China: The value of social network analysis," *Preventive Veterinary Medicine*, vol. 102, no. 3, pp. 196–205, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.prevetmed.2011.07.007

119. R. J. Soares Magalhães, X. Zhou, B. Jia, F. Guo, D. U. Pfeiffer, and V. Martin, "Live Poultry Trade in Southern China Provinces and HPAIV H5N1 Infection in Humans and Poultry: The Role of Chinese New Year Festivities," *PLoS ONE*, vol. 7, no. 11, 2012.

120. R. J. Soares Magalhães, A. Ortiz-Pelaez, K. L. Thi, Q. H. Dinh, J. Otte, and D. U. Pfeiffer, "Associations between attributes of live poultry trade and HPAI H5N1 outbreaks: A descriptive and network analysis study in northern Vietnam," *BMC Veterinary Research*, vol. 6, pp. 6–8, 2010.

121. X. Zhou, Y. Li, Y. Wang, J. Edwards, F. Guo, A. C. Clements, B. Huang, and R. J. Soares Magalhaes, "The role of live poultry movement and live bird market biosecurity in the epidemiology of influenza A (H7N9): A cross-sectional observational study in four eastern China provinces," *Journal of Infection*, vol. 71, no. 4, pp. 470–479, 2015.

122. M. E. Arnold, R. M. Irvine, O. Tearne, D. Rae, A. J. Cook, and A. C. Breed, "Investigation into sampling strategies in response to potential outbreaks of low pathogenicity notifiable avian influenza initiated in commercial duck holdings in Great Britain," *Epidemiology and Infection*, vol. 141, no. 4, pp. 751–762, 2013.

123. G. Fournié, F. J. Guitian, P. Mangtani, and A. C. Ghani, "Impact of the implementation of rest days in live bird markets on the dynamics of H5N1 highly pathogenic avian influenza," *Journal of the Royal Society Interface*, vol. 8, no. 61, pp. 1079–1089, 2011.

124. J. E. Dent, I. Z. Kiss, R. R. Kao, and M. Arnold, "The potential spread of highly pathogenic avian influenza virus via dynamic contacts between poultry premises in Great Britain," *BMC Veterinary Research*, vol. 7, 2011.

125. C. Guinat, J. Artois, A. Bronner, J. L. Guérin, M. Gilbert, and M. C. Paul, "Duck production systems and highly pathogenic avian influenza H5N8 in France, 2016–2017," *Scientific Reports*, vol. 9, no. 1, pp. 2016–2017, 2019.

126. T. J. Hagenaars, G. J. Boender, R. H. Bergevoet, and H. J. van Roermund, "Risk of poultry compartments for transmission of highly pathogenic avian influenza," *PLoS ONE*, vol. 13, no. 11, pp. 1–18, 2018.

127. S. Nickbakhsh, L. Matthews, S. W. Reid, and R. R. Kao, "A metapopulation model for highly pathogenic avian influenza: Implications for compartmentalization as a control measure," *Epidemiology and Infection*, vol. 142, no. 9, pp. 1813–1825, 2014.

128. K. A. Patyk, J. Helm, M. K. Martin, K. N. Forde-Folle, F. J. Olea-Popelka, J. E. Hokanson, T. Fingerlin, and A. Reeves, "An epidemiologic simulation model of the spread and control of highly pathogenic avian influenza (H5N1) among commercial and backyard poultry flocks in South Carolina, United States," *Preventive Veterinary Medicine*, vol. 110, no. 3-4, pp. 510–524, 2013.

129. K. J. Sharkey, R. G. Bowers, K. L. Morgan, S. E. Robinson, and R. M. Christley, "Epidemiological consequences of an incursion of highly pathogenic H5N1 avian influenza into the British poultry flock," *Proceedings of the Royal Society B*, vol. 275, no. 1630, pp. 19–28, 2008.

130. J. Truscott, T. Garske, I. Chis-Ster, J. Guitian, D. Pfeiffer, L. Snow, J. Wilesmith, N. M. Ferguson, and A. C. Ghani, "Control of a highly pathogenic H5N1 avian influenza outbreak in the GB poultry flock," *Proceedings of the Royal Society*, vol. 274, no. 1623, pp. 2287–2295, 2007.

131. A. Kasyanov, L. Kirkland, and M. T. Mihaela, "A Spatial SIRS Boolean Network Model for the Spread of H5N1 Avian Influenza Virus among Poultry Farms," *Proceedings of the 5th International Workshop Computational Systems Biology*, pp. 73–76, 2008.

132. A. Le Menach, E. Vergu, R. F. Grais, D. L. Smith, and A. Flahault, "Key strategies for reducing spread of avian influenza among commercial poultry holdings: Lessons for transmission to humans," *Proceedings of the Royal Society B*, vol. 273, no. 1600, pp. 2467–2475, 2006.

133. 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, no. 3, pp. 73–85, 2009.

134. M. Andraud and N. Rose, "Modelling infectious viral diseases in swine populations: A state of the art," *Porcine Health Management*, vol. 6, no. 1, pp. 1–12, 2020.

135. E. A. Kukielka, B. Martínez-López, and D. Beltrán-Alcrudo, "Modeling the live-pig trade network in Georgia: Implications for disease prevention and control," *PLoS ONE*, vol. 12, no. 6, pp. 1–15, 2017.

136. M. A. Morgane Salines and N. Rose, "Combining network analysis with epidemiological data to inform risk-based surveillance: Application to hepatitis E virus (HEV) in pigs," *Preventive Veterinary Medicine*, vol. 149, no. 2018, pp. 125–131, 2018.

137. M. Salines, M. Andraud, N. Rose, and S. Widgren, "A between-herd data-driven stochastic model to explore the spatio-temporal spread of hepatitis e virus in the French pig production network," *PLoS ONE*, vol. 15, no. 7, pp. 1–18, 2020.

138. A. G. Murray and N. K. Salama, "Modelling disease in aquatic systems that are spread by processes operating at different temporal and spatial scales: Examples from salmon aquaculture," *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, vol. 12, no. 32, pp. 1–11, 2017.

139. D. M. Green, A. Gregory, and L. A. Munro, "Small- and large-scale network structure of live fish movements in Scotland," *Preventive Veterinary Medicine*, vol. 91, no. 2-4, pp. 261–269, 2009.

140. D. M. Green, M. Werkman, L. A. Munro, R. R. Kao, I. Z. Kiss, and L. Danon, "Tools to study trends in community structure: Application to fish and livestock trading networks," *Preventive Veterinary Medicine*, vol. 99, no. 2-4, pp. 225–228, 2011.

141. L. A. Munro and A. Gregory, "Application of network analysis to farmed salmonid movement data from Scotland," *Journal of Fish Diseases*, vol. 32, no. 7, pp. 641–644, 2009.

142. T. Adams, K. Black, C. MacIntyre, I. MacIntyre, and R. Dean, "Connectivity modelling and network analysis of sea lice infection in Loch Fyne, west coast of Scotland," *Aquaculture Environment Interactions*, vol. 3, no. 1, pp. 51–63, 2012.

143. M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

144. B. Martínez-López, T. Alexandrov, L. Mur, F. Sánchez-Vizcaíno, and J. M. Sánchez-Vizcaíno, "Evaluation of the spatial patterns and risk factors, including backyard pigs, for classical swine fever occurrence in Bulgaria using a Bayesian model," *Geospatial Health*, vol. 8, no. 2, pp. 489–501, 2014.

145. G. Robins and D. Lusher, *What Are Exponential Random Graph Models?* Cambridge University Press, 2012, p. 9–15.

146. K. Kokott and H. Hartmann, "Herkunftssicherungs- und Informationssystem für Tiere," 2015. [Online]. Available: https://www.hi-tier.de/

147. C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks," *Journal of Complex Networks*, vol. 3, no. 2, pp. 221–248, 06 2014.

148. Y. Dodge, *The Concise Encyclopedia of Statistics.* Springer New York, 2008, ch. Negative Binomial Distribution, pp. 369–370.

149. ——, *The Concise Encyclopedia of Statistics.* Springer New York, 2008, ch. Log-normal Distribution, pp. 321–322.

150. Pye tait consulting, "2019 Horticulture Sector Skills Survey – Sub-Sector Report : Arboriculture A report for the Ornamental Horticulture Roundtable Group," Royal Horticultural Society, Technical Report, 2019. [Online]. Available: https://www.rhs.org.uk/science/pdf/horticulture-skills-report/horticulture-sector-skills-survey-report.pdf

151. DEFRA, "Information Note: Xylella high risk hosts," 2018. [Online]. Available: https://planthealthportal.defra.gov.uk/assets/factsheets/Xylella-host-info-notev 8final.pdf

152. G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society. Series B*, vol. 26, no. 2, pp. 211–252, 1964. [Online]. Available: http://www.jstor.org/stable/2984418

153. J. A. Rice, *Mathematical Statistics and Data Analysis.* Duxbury, 2007, vol. 31, no. 3.

154. Plant Healthy, "Certification Scheme Manual Plant Health Management Standard," 2022. [Online]. Available: https://planthealthy.org.uk/assets/images/Plant-Heal thy-Certification-Scheme-Manual-V1.2-1.pdf

155. D. Eddelbuettel and R. François, *Rcpp: Seamless R and C++ integration*, 2011, vol. 40, no. 8.

156. D. Eddelbuettel, R. Francois, D. Bates, B. Ni, and C. Sanderson, *Package RcppArmadillo*, 2023. [Online]. Available: https://cran.r-project.org/web/package s/RcppArmadillo/RcppArmadillo.pdf

157. World Health Organisation, "WHO Coronavirus (COVID-19) Dashboard With Vaccination Data," 2021. [Online]. Available: https://covid19.who.int/

158. D. Bell, A. Comas-Herrera, D. Henderson, S. Jones, E. Lemmon, M. Moro, S. Murphy, D. O'Reilly, and P. Patrignani, "COVID-19 mortality and long-term care: a UK comparison," LTCcovid, Technical Report August, 2020. [Online]. Available: https://ltccovid.org/wp-content/uploads/2020/08/COVID-19-mortality-in-long-term-care-final-Sat-29-1.pdf

159. UK Government Social Care Working Group, "Commission : What are the appropriate layers of mitigation to deploy for care homes in the context of post vaccination risk landscape?" Technical Report, 2021. [Online]. Available: https://www.gov.uk/government/publications/scwg-what-are-the-appropriate-mitigations-to-deploy-in-care-homes-in-the-context-of-the-post-vaccination-risk-landscape-26-may-2021

160. L. K. Nguyen, I. Megiddo, and S. Howick, "Challenges of infection prevention and control in Scottish long-term care facilities," *Infection control and hospital epidemiology*, vol. 41, no. 8, pp. 943–945, 8 2020.

161. D. R. M. Smith, A. Duval, K. B. Pouwels, D. Guillemot, J. Fernandes, B.-T. Huynh, L. Temime, and L. Opatowski, "Optimizing COVID-19 surveillance in long-term care facilities: a modelling study," *BMC Medicine 2020 18:1*, vol. 18, no. 1, pp. 1–16, 12 2020.

162. L. K. N. Nguyen, S. Howick, D. McLafferty, G. H. Anderson, S. J. Pravinkumar, R. Van Der Meer, and I. Megiddo, "Evaluating intervention strategies in controlling COVID-19 spread in care homes: An agent-based model," *Infection Control and Hospital Epidemiology*, vol. 2019, pp. 1–11, 2020.

163. L. K. N. Nguyen, I. Megiddo, and P. S. Howick, "Report 3 : Impact of various vaccination coverages on the spread of COVID-19 and deaths in care homes," Department of Management Science, University of Strathclyde, Technical Report January, 2021.

164. L. K. N. Nguyen, S. Howick, D. McLafferty, G. H. Anderson, S. J. Pravinkumar, R. Van Der Meer, and I. Megiddo, "Impact of visitation and cohorting policies to shield residents from COVID-19 spread in care homes: an agent-based model," *AJIC: American Journal of Infection Control*, vol. 49, 2021.

165. A. Roselló, R. C. Barnard, D. R. M Smith, S. Evans, F. Grimm, N. G. Davies, Centre for Mathematical Modelling of Infectious Diseases COVID-19 modelling working group, S. R. Deeny, G. M. Knight, and W. John Edmunds, "Impact of non-pharmaceutical interventions on SARS-CoV-2 outbreaks in English care homes: a modelling study," *BMC Infectious Diseases*, vol. 22, 2022.

166. C. E. Overton, H. B. Stage, S. Ahmad, J. Curran-Sebastian, P. Dark, R. Das, E. Fearon, T. Felton, M. Fyles, N. Gent, I. Hall, T. House, H. Lewkowicz, X. Pang, L. Pellis, R. Sawko, A. Ustianowski, B. Vekaria, and L. Webb, "Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example," *Infectious Disease Modelling*, vol. 5, pp. 409–441, 2020.

167. J. Oberhammer, "Social-distancing effectiveness tracking of the COVID-19 hotspot Stockholm," *medRxiv [Preprint]*, pp. 853–853, 2020.

168. B. A. D. v. Bunnik, A. L. K. Morgan, P. R. Bessell, G. Calder-Gerver, F. Zhang, S. Haynes, J. Ashworth, S. Zhao, R. N. R. Cave, M. R. Perry, H. C. Lepper, L. Lu, P. Kellam, A. Sheikh, G. F. Medley, and M. E. J. Woolhouse, "Segmentation and shielding of the most vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown," *Philosophical Transactions of the Royal Society B*, vol. 376, no. 1829, 7 2021.

169. Public Health Scotland, "Discharges from NHS Scotland Hospitals to Care Homes between 1 March and 31 May," Public Health Scotland, Technical Report May, 2020. [Online]. Available: https://publichealthscotland.scot/publications/dischar ges-from-nhsscotland-hospitals-to-care-homes/discharges-from-nhsscotland-hospi tals-to-care-homes-between-1-march-and-31-may-2020/

170. J. K. Burton, G. Bayne, C. Evans, F. Garbe, D. Gorman, N. Honhold, D. Mc-Cormick, R. Othieno, J. E. Stevenson, S. Swietlik, K. E. Templeton, M. Tranter, L. Willocks, and B. Guthrie, "Evolution and effects of COVID-19 outbreaks in care homes: a population analysis in 189 care homes in one geographical region of the UK," *The Lancet Healthy Longevity*, vol. 1, no. 1, pp. e21–e31, 2020.

171. E. S. Knock, L. K. Whittles, J. A. Lees, P. N. Perez-Guzman, R. Verity, R. G. FitzJohn, K. A. Gaythorpe, N. Imai, and et al, "Report 41 - The 2020 SARS-CoV-2 epidemic in England: key epidemiological drivers and impact of interventions," Imperial College London, Technical Report December, 2020. [Online]. Available: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-41-rtm/

172. P. M. McKeigue and H. M. Colhoun, "Evaluation of "stratify and shield" as a policy option for ending the COVID-19 lockdown in the UK," *medRxiv [Preprint]*, p. 2020.04.25.20079913, 2020.

173. T. J. Ripperger, J. L. Uhrlaub, M. Watanabe, R. Wong, Y. Castaneda, H. A. Pizzato, M. R. Thompson, C. Bradshaw, C. C. Weinkauf, C. Bime, H. L. Erickson, K. Knox, B. Bixby, S. Parthasarathy, S. Chaudhary, B. Natt, E. Cristan, T. El Aini, F. Rischard, ... , and D. Bhattacharya, "Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity," *Immunity*, vol. 53, no. 5, pp. 925–933, 2020.

174. J. M. Dan, J. Mateus, Y. Kato, K. M. Hastie, E. D. Yu, C. E. Faliti, A. Grifoni, S. I. Ramirez, S. Haupt, A. Frazier, C. Nakao, V. Rayaprolu, S. A. Rawlings, B. Peters, F. Krammer, V. Simon, E. O. Saphire, D. M. Smith, D. Weiskopf, A. Sette, and S. Crotty, "Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection," *Science*, vol. 371, no. 6529, pp. 1–23, 2021.

175. The Scottish Government, "Coronavirus (COVID-19) social care response - 13 March 2020," pp. 1–5, 2020. [Online]. Available: https://www.careinspectorate.com/images/COVID-19_-_Letter_from_Cabinet_Secretary_for_Health_and_Sport_-_Social_care_guidance_-_13_March_2020.pdf

176. J. Reilly, D. Crawford, and D. O. Boyle, "Care home review: A rapid review of factors relevant to the management of COVID-19 in the care home environment in Scotland," Scottish Government: Cabinet Secretary for Health and Sport, Edinburgh, Technical Report, 2020. [Online]. Available: https://www.gov.scot/publications/root-cause-analysis-care-home-outbreaks/

177. Office For National Statistics, "Impact of coronavirus in care homes in England: 26 May to 19 June 2020," Office For National Statistics, Technical Report July, 2020. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommun ity/healthandsocialcare/conditionsanddiseases/articles/impactofcoronavirusincare homesinenglandvivaldi/26mayto19june2020

178. A. Rădulescu, C. Williams, and K. Cavanagh, "Management strategies in a SEIR-type model of COVID 19 community spread," *Scientific Reports*, vol. 10, no. 1, pp. 1–16, 2020.

179. D. Calvetti, A. P. Hoover, J. Rose, and E. Somersalo, "Metapopulation Network Models for Understanding, Predicting, and Managing the Coronavirus Disease COVID-19," *Frontiers in Physics*, vol. 8, no. June, pp. 1–16, 2020.

180. Public Health Scotland, "Daily COVID-19 Cases in Scotland - Daily Case Trends By Health Board - Scottish Health and Social Care Open Data," 2020. [Online]. Available: https://www.opendata.nhs.scot/dataset/covid-19-in-scotland/resource/ 2dd8534b-0a6f-4744-9253-9565d62f96c2

181. National Records of Scotland, "Deaths involving coronavirus (COVID-19) in Scotland - National Records of Scotland," 2020. [Online]. Available: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme /vital-events/general-publications/weekly-and-monthly-data-on-births-and-death s/deaths-involving-coronavirus-covid-19-in-scotland

182. S. N. Ladhani, J. Y. Chow, R. Janarthanan, J. Fok, E. Crawley-Boevey, A. Vusirikala, E. Fernandez, M. S. Perez, S. Tang, K. Dun-Campbell, E. W. Evans, A. Bell, B. Patel, Z. Amin-Chowdhury, F. Aiano, K. Paranthaman, T. Ma, M.

Saavedra-Campos, R. Myers, ..., and Zambon, Maria, "Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020," *EClinicalMedicine*, vol. 26, p. 100533, 2020.

183. N. Graham, C. Junghans, R. Downes, C. Sendall, H. Lai, A. McKirdy, P. Elliott, R. Howard, D. Wingfield, M. Priestman, M. Ciechonska, L. Cameron, M. Storch, M. Crone, P. Freemont, P. Randell, R. McLaren, N. Lang, S. Ladhani, F. Sanderson, and D. Sharp, "Sars-cov-2 infection, clinical features and outcome of covid-19 in united kingdom nursing homes," *Journal of Infection*, vol. 81, no. 3, pp. 411–419, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0163445320303480

184. O. Byambasuren, M. Cardona, K. Bell, J. Clark, M. L. McLaws, and P. Glasziou, "Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis," *Journal of the Association of Medical Microbiology and Infectious Disease Canada*, vol. 5, no. 4, pp. 223–234, 2020.

185. E. Dickson, N. Palmateer, J. Murray, C. Robertson, C. Waugh, L. Wallace, L. Mathie, K. Heatlie, S. Mavin, P. Gousias, B. Von Wissman, D. Goldberg, and A. McAuley, "Enhanced surveillance of covid-19 in scotland: population-based sero-prevalence surveillance for sars-cov-2 during the first wave of the epidemic," *Public Health*, vol. 190, pp. 132–134, 2021.

186. National Records of Scotland, "Population Estimates Time Series Data," 2019. [Online]. Available: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data

187. X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, ... , and G. M. Leung , "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nature Medicine*, vol. 26, no. 5, pp. 672–675, 2020.

188. C. McAloon, A. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, and S. J. More, "Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research," *BMJ Open*, vol. 10, no. 8, pp. 1–9, 2020.

189. The Scottish Government, "Coronavirus (COVID-19): modelling the epidemic in Scotland (Issue No. 25)," Scottish Government, Technical Report 29, 2020. [Online]. Available: https://www.gov.scot/publications/coronavirus-covid-19-mod elling-epidemic-issue-no-29/

190. NHS Lothian, "NHS Lothian," 2021. [Online]. Available: https://www.nhslothian .scot/Pages/default.aspx

191. D. Majra, J. Benson, J. Pitts, and J. Stebbing, "SARS-CoV-2 (COVID-19) super-spreader events," *Journal of Infection*, vol. 82, no. 1, pp. 36–40, 2021.

192. J. Belmin, N. Um-Din, C. Donadio, M. Magri, Q. D. Nghiem, B. Oquendo, S. Pariel, and C. Lafuente-Lafuente, "Coronavirus Disease 2019 Outcomes in French Nursing Homes That Implemented Staff Confinement with Residents," *JAMA Network Open*, vol. 3, no. 8, pp. 1–9, 2020.

193. T. M. McMichael, D. W. Currie, S. Clark, S. Pogosjans, M. Kay, N. G. Schwartz, J. Lewis, A. Baer, V. Kawakami, M. D. Lukoff, J. Ferro, C. Brostrom-Smith, T. D. Rea, M. R. Sayre, F. X. Riedo, D. Russell, B. Hiatt, P. Montgomery, A. K. Rao, ... , and J. S. Duchin, "Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington," *N Engl J Med*, vol. 382, no. 21, pp. 2005–2011, 2020.

194. A.-L. Barabási and E. Bonabeau, "Scale-Free Networks," *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003.

195. A. Comas-Herrera, "Rapid review of the evidence on impacts of visiting policies in care homes during the COVID-19 pandemic," *Internal long-term care policy network*, 2020.

196. P. H. England, "A data linkage approach to assessing the contribution of hospital-associated SARS-CoV-2 infection to care home outbreaks in England, 30 January to 12 October 2020," Public Health England, Technical Report October, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/983349/Data_linkage_approach_to_assessing_the_contribution_of_hospital-associated_SARS-CoV-2_infection_to_care_home_outbreaks_in_England.pdf

197. N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. Pearson, B. J. Quilty, A. J. Kucharski, H. Gibbs, S. Clifford, A. Gimma, K. van Zandvoort, J. D. Munday, C. Diamond, W. J. Edmunds, R. M. Houben, J. Hellewell, T. W. Russell, S. Abbott, ..., and R. M. Eggo, "Age-dependent effects in the transmission and control of COVID-19 epidemics," *Nature Medicine*, vol. 26, no. 8, pp. 1205–1211, 2020.

198. J. K. Burton, M. McMinn, J. E. Vaughan, J. Fleuriot, and B. Guthrie, "Care-home outbreaks of COVID-19 in Scotland March to May 2020: National linked data cohort analysis," *Age and Ageing*, vol. 26, no. May 2020, pp. 1–11, 2021.

Chapter 7

Appendices

Appendix A

Customer demographics and consignment size distributions for Xylella hosts

This section shows the customer demographics and consignment size distributions using the nursery sales data of key Xylella hosts (Lavender, Olive, Prunus and Rosemary). These results coincide with the results on Oak sales in Chapter 3.
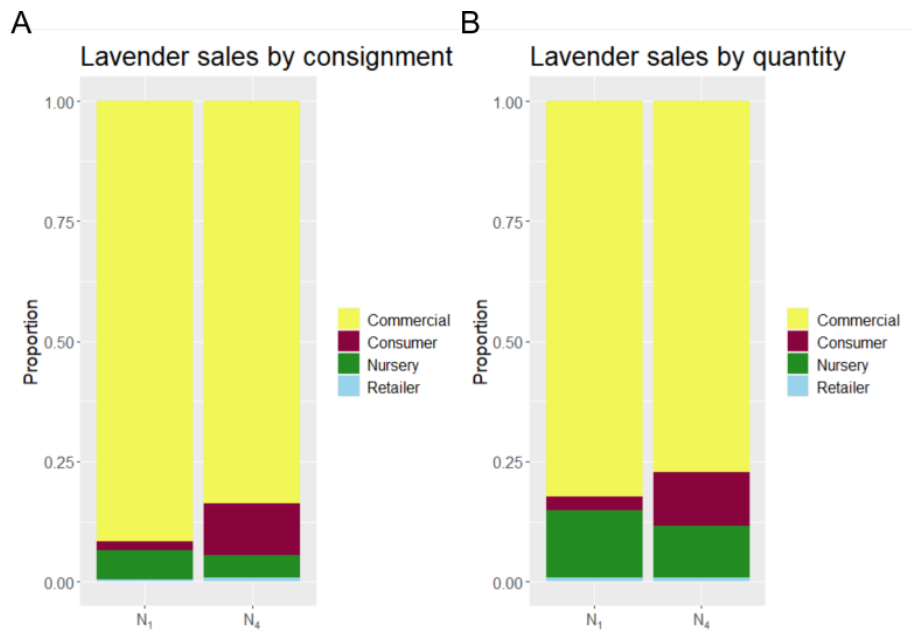
Figure A.1. Customer demographics displayed of the two nurseries in our data that trade in Lavender (N_1, N_4) measured by: **A**.) sales by consignment (number of orders), **B**.) sales by quantity (total number of plants sold).
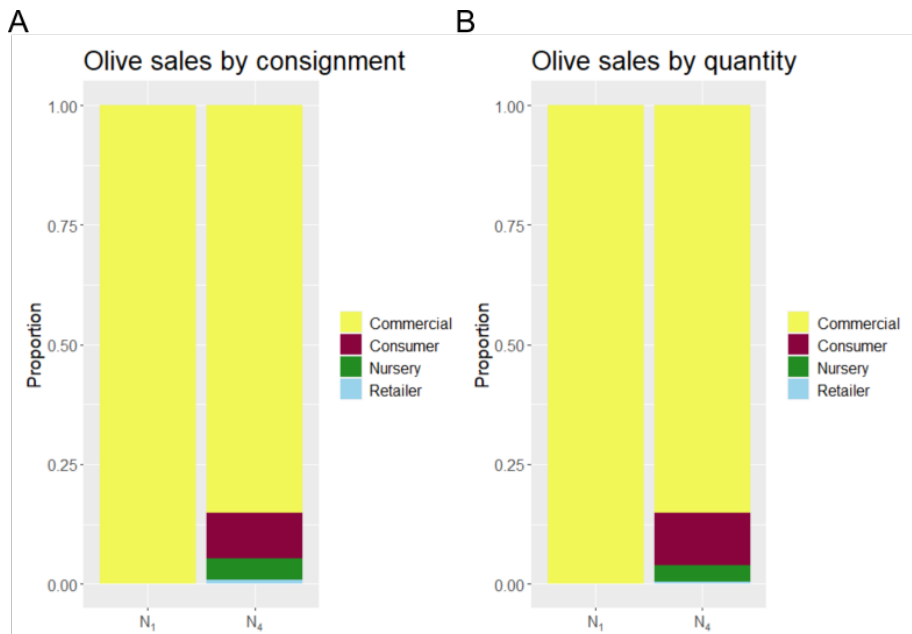
Figure A.2. Customer demographics displayed of the two nurseries in our data that trade in Olive (N_1, N_4) measured by: **A**.) sales by consignment (number of orders), **B**.) sales by quantity (total number of plants sold).
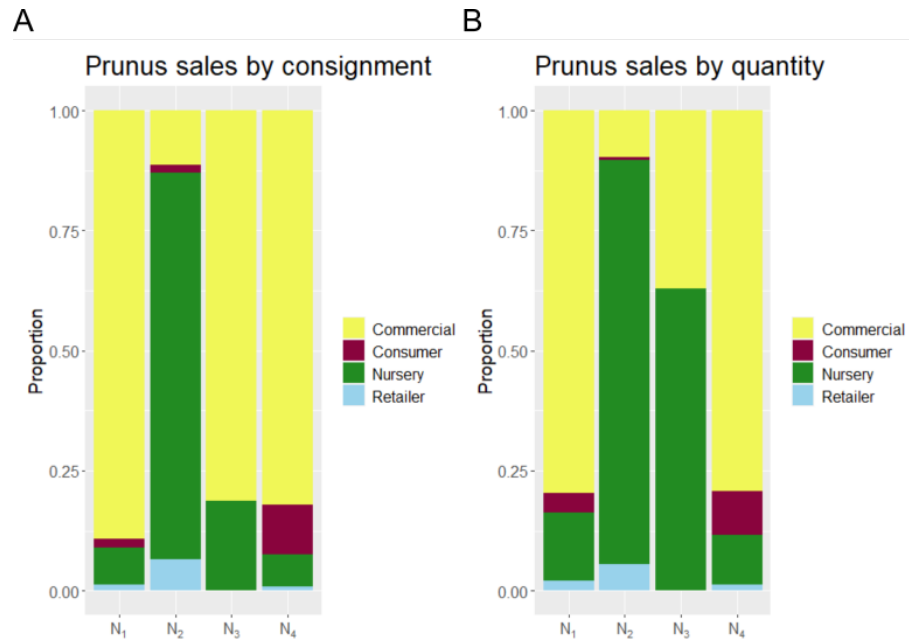
Figure A.3. Customer demographics displayed for Prunus sales of the four nurseries from our data (N_1, N_2, N_3, N_4) measured by: **A**.) sales by consignment (number of orders), **B**.) sales by quantity (total number of plants sold).
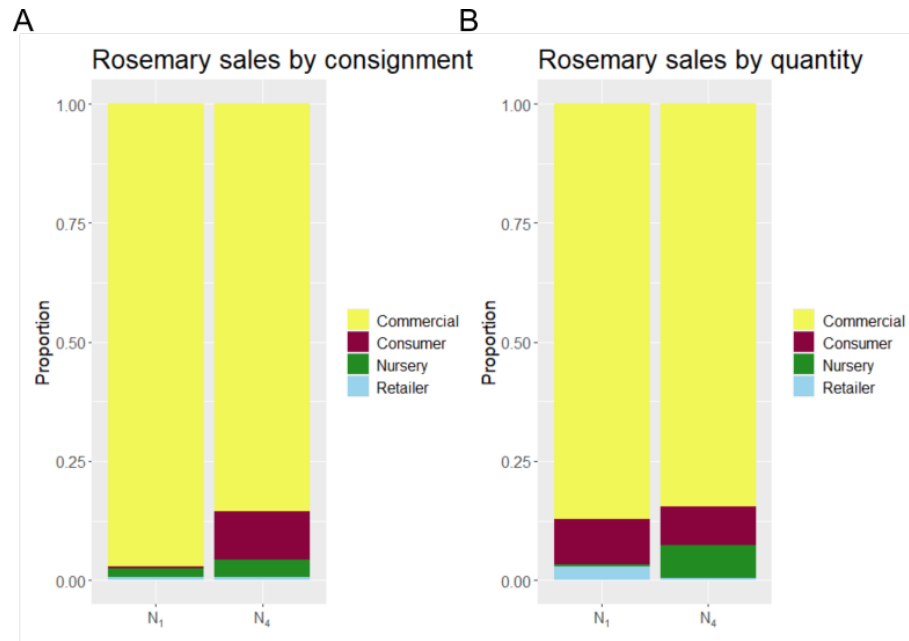
Figure A.4. Customer demographics displayed of the two nurseries in our data that trade in Rosemary (N_1, N_4) measured by: **A**.) sales by consignment (number of orders), **B**.) sales by quantity (total number of plants sold).
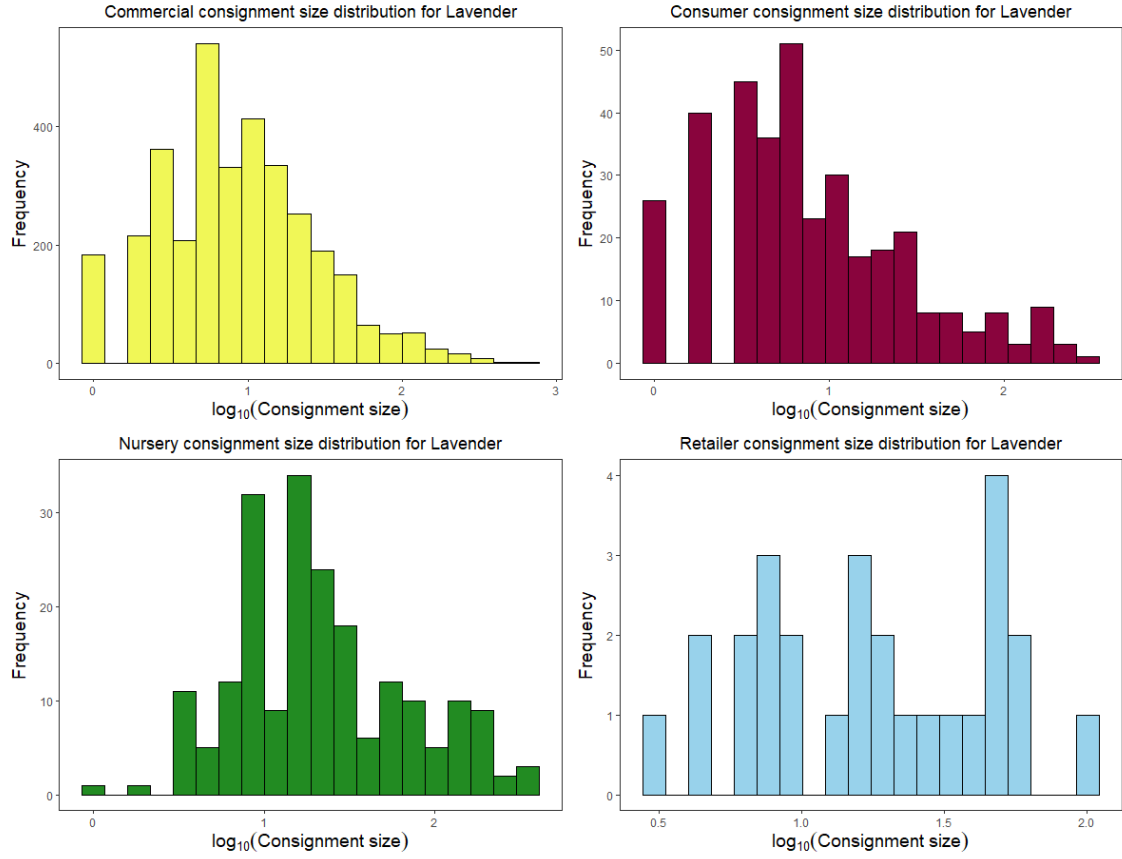
Figure A.5. The distribution of consignment sizes for sales in Lavender, separating for each customer group. We use the sales from all nurseries in our data that trade in Lavender $\{N_1, N_4\}$.
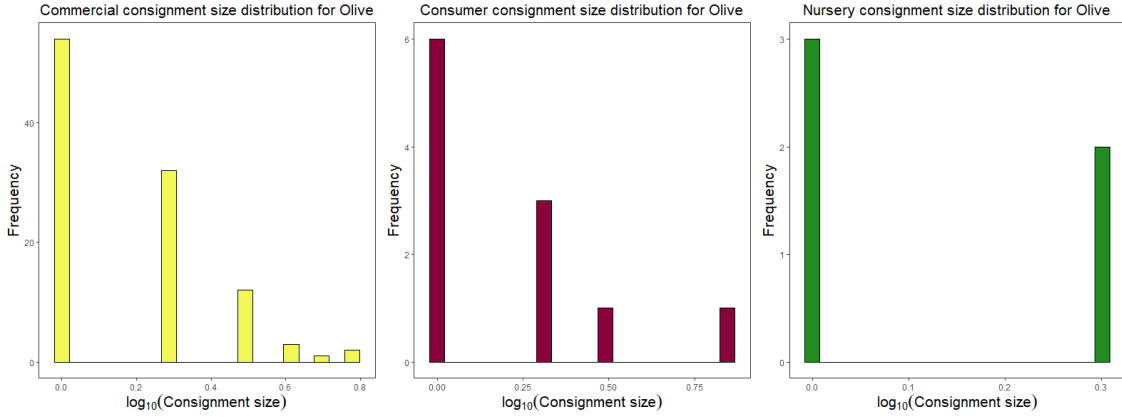
Figure A.6. The distribution of consignment sizes for sales in Olive, separating for each customer group. We use the sales from all nurseries in our data that trade in Olive $\{N_1, N_4\}$.
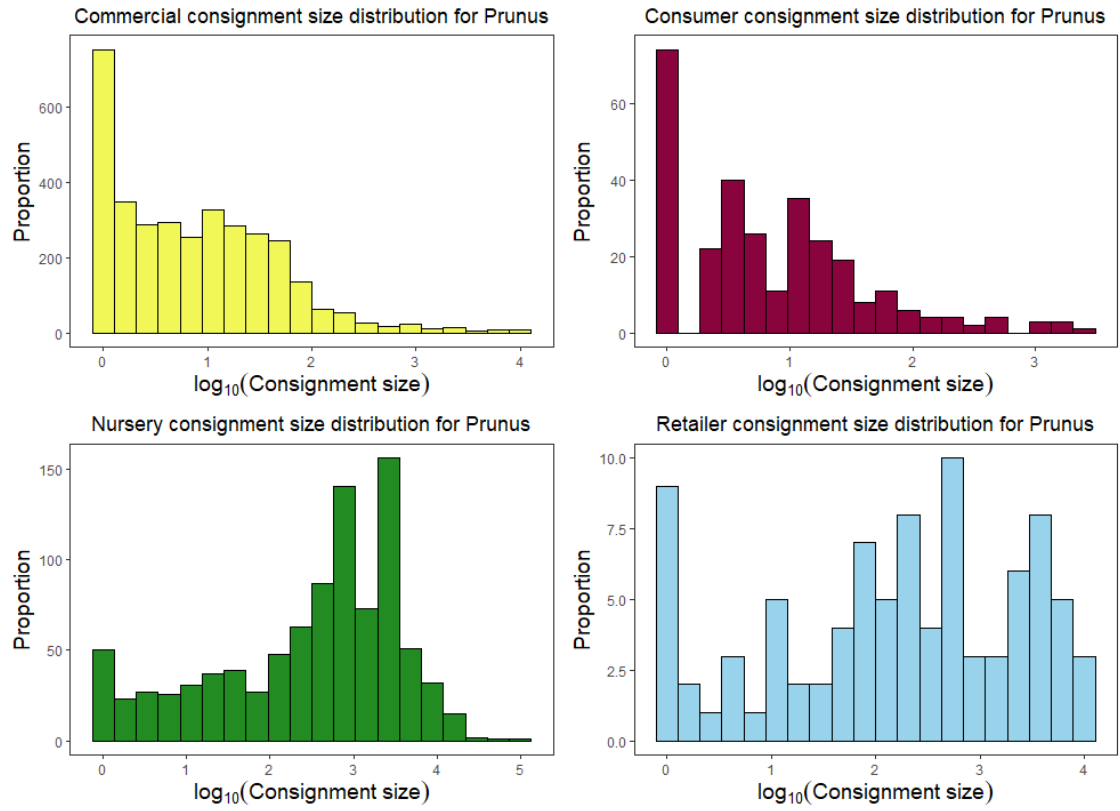
Figure A.7. The distribution of consignment sizes for sales in Prunus, separating for each customer group. We use the sales from all nurseries in our data that trade in Prunus $\{N_1, N_2, N_3, N_4\}$.
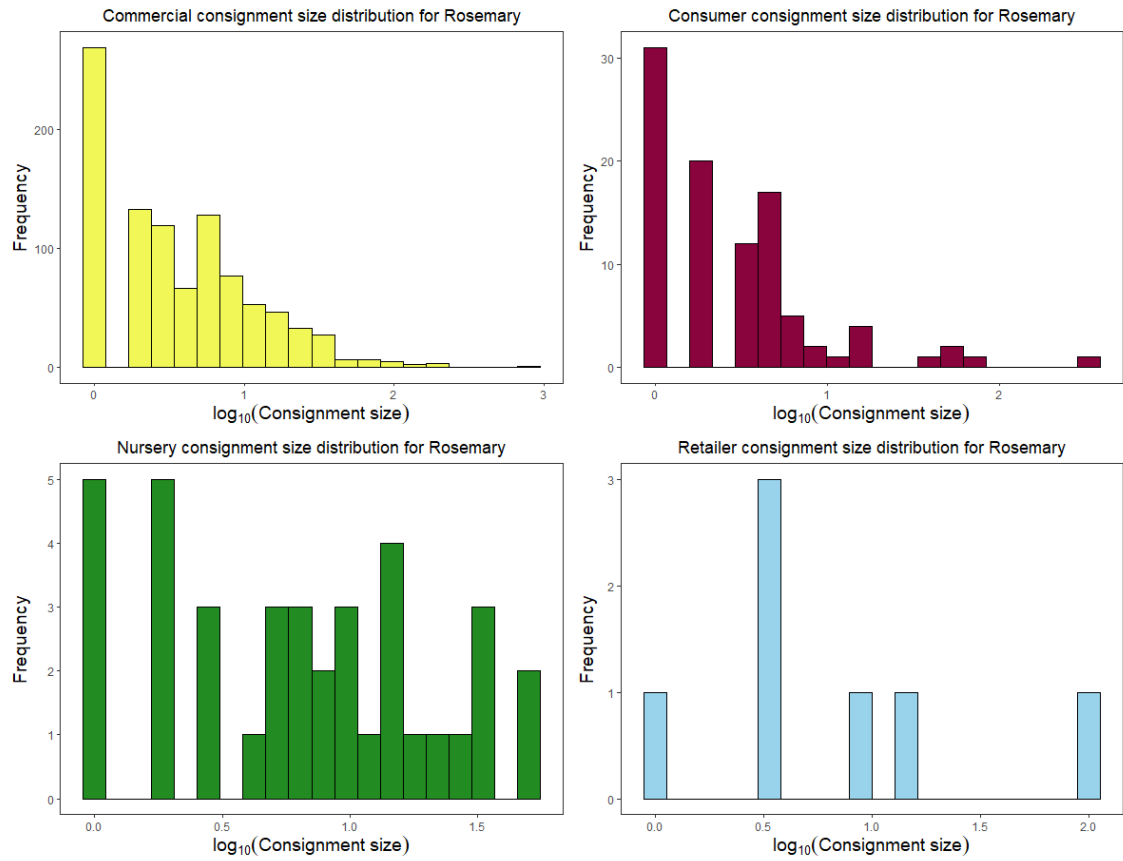
Figure A.8. The distribution of consignment sizes for sales in Rosemary, separating for each customer group. We use the sales from all nurseries in our data that trade in Rosemary $\{N_1, N_4\}$.

Appendix B

Full network and network subset comparison figures

This section shows the comparisons for the centrality measures we considered in Chapter 3. For each centrality measure, we compared the scores in each node subcategory with the same measures computed on the network with the commercial and consumer nodes removed.
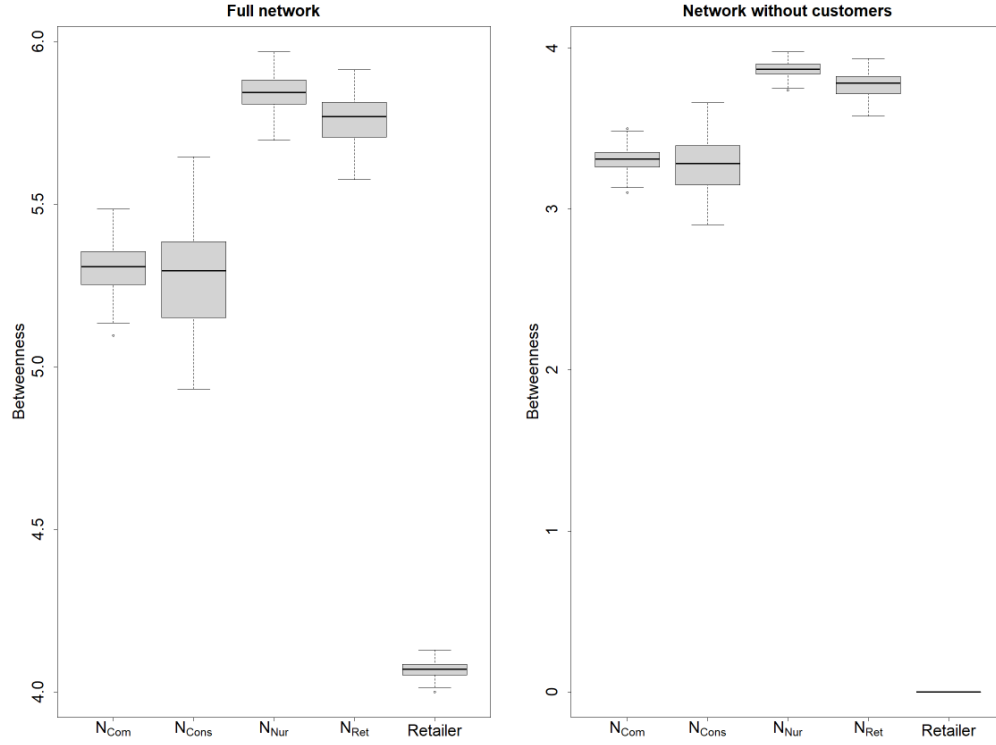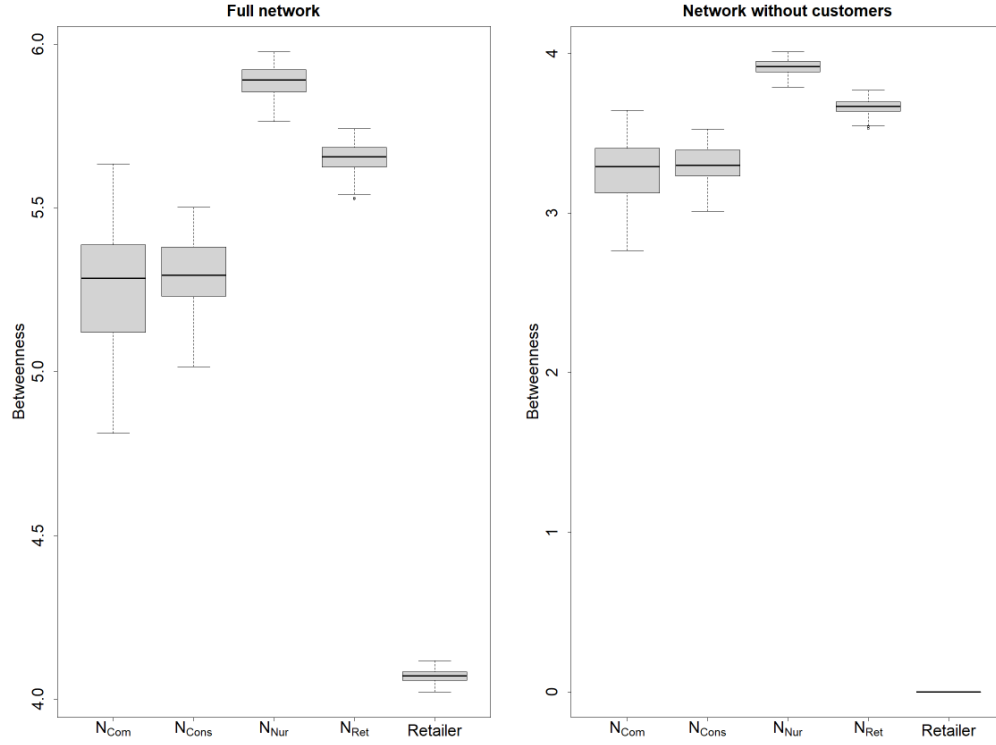
Figure B.1. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
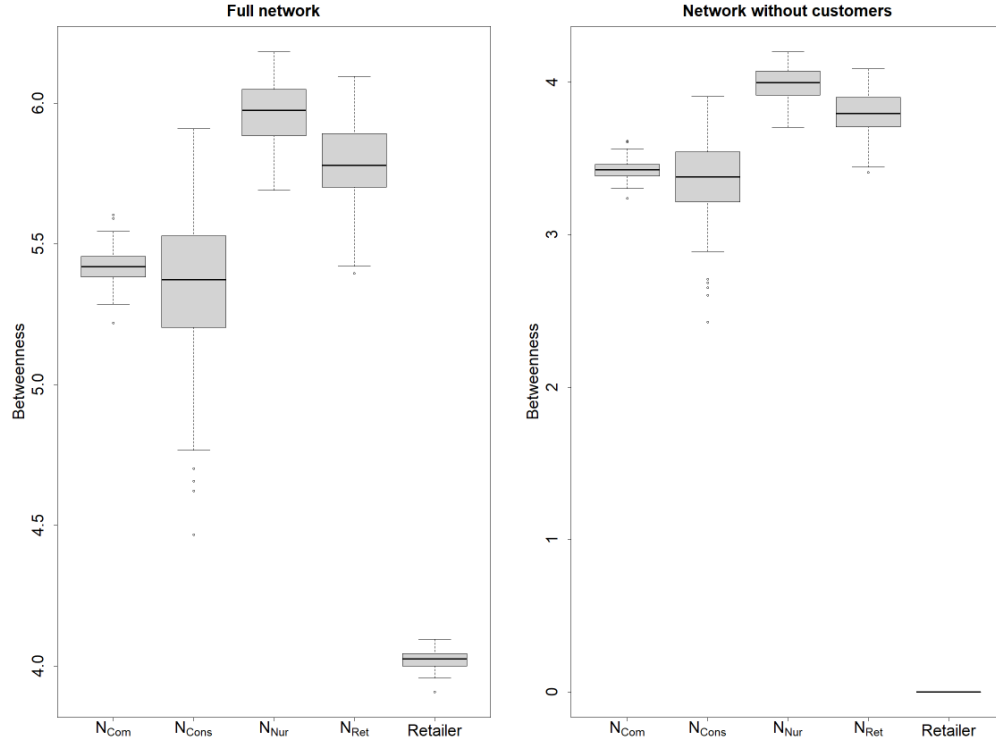
Figure B.2. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
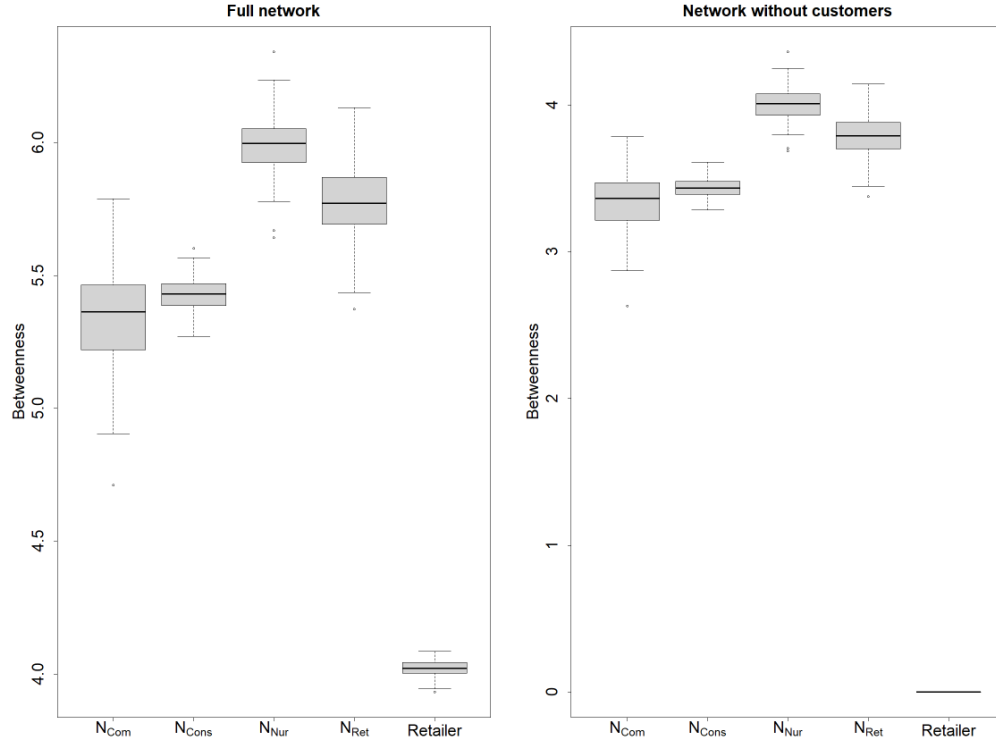
Figure B.3. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
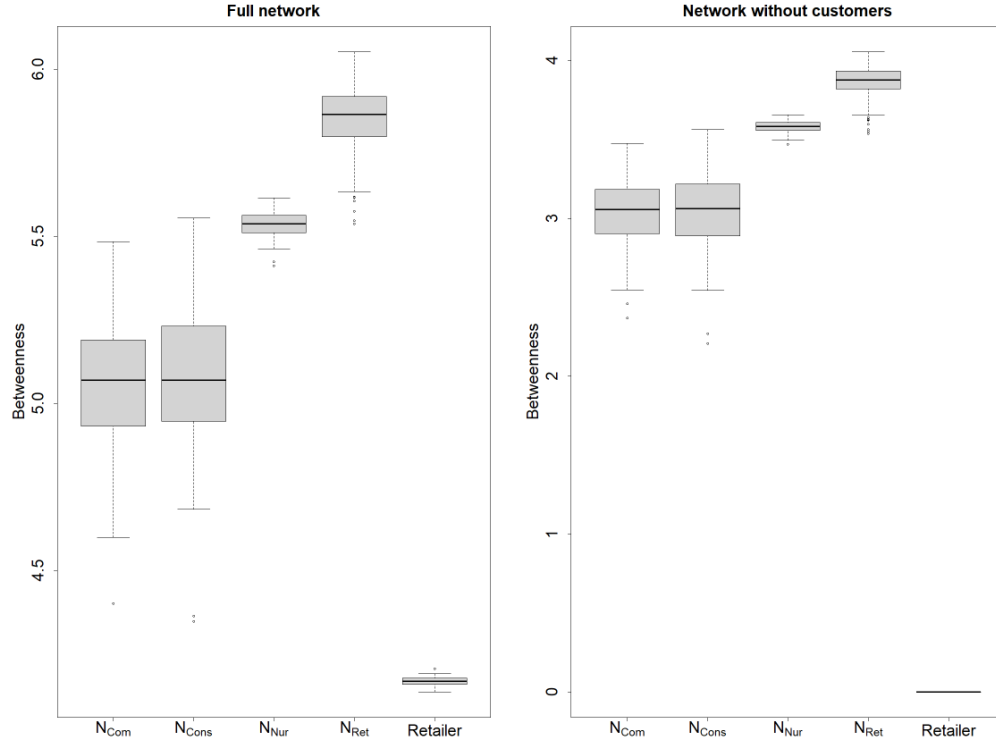
Figure B.4. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
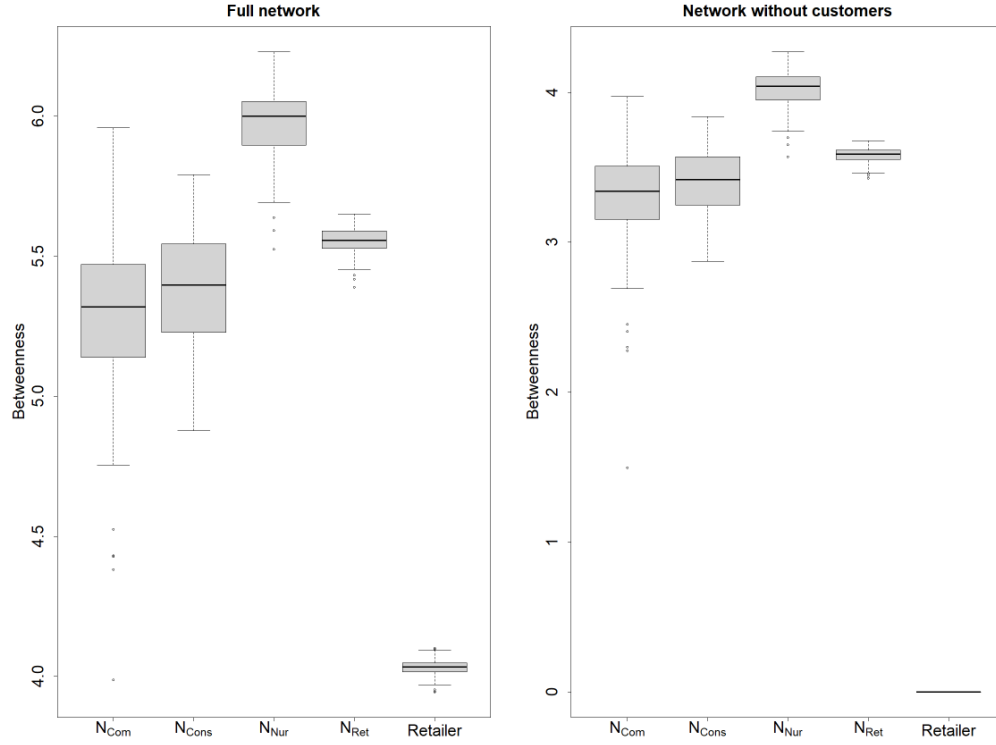
Figure B.5. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
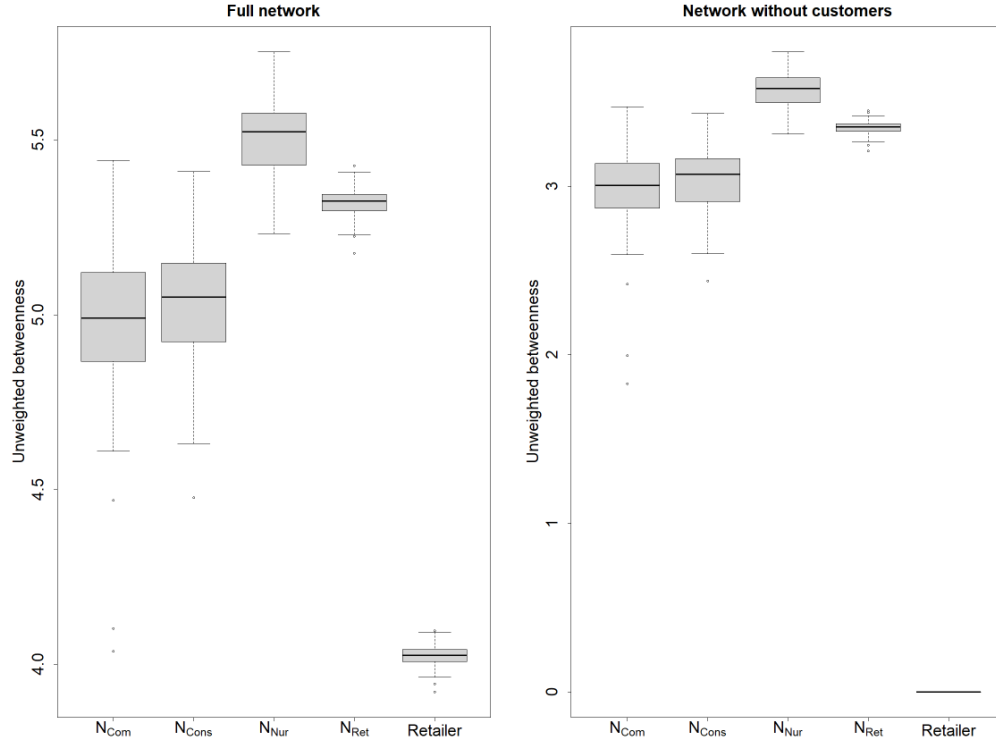
Figure B.6. Box plots of the average betweenness score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
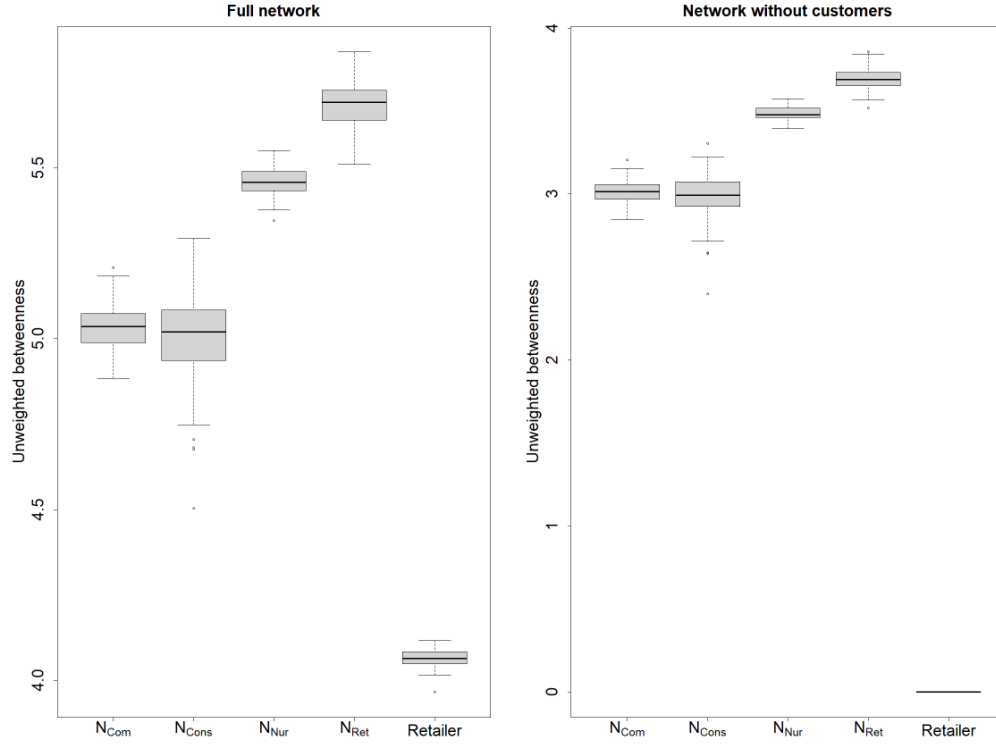
Figure B.7. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
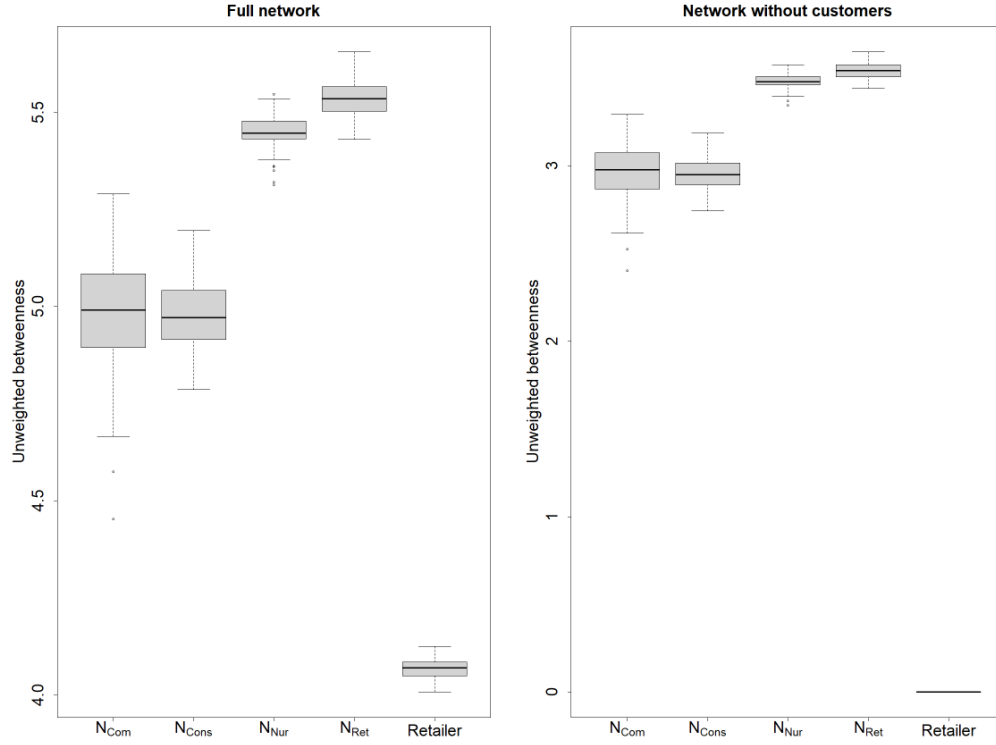
Figure B.8. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
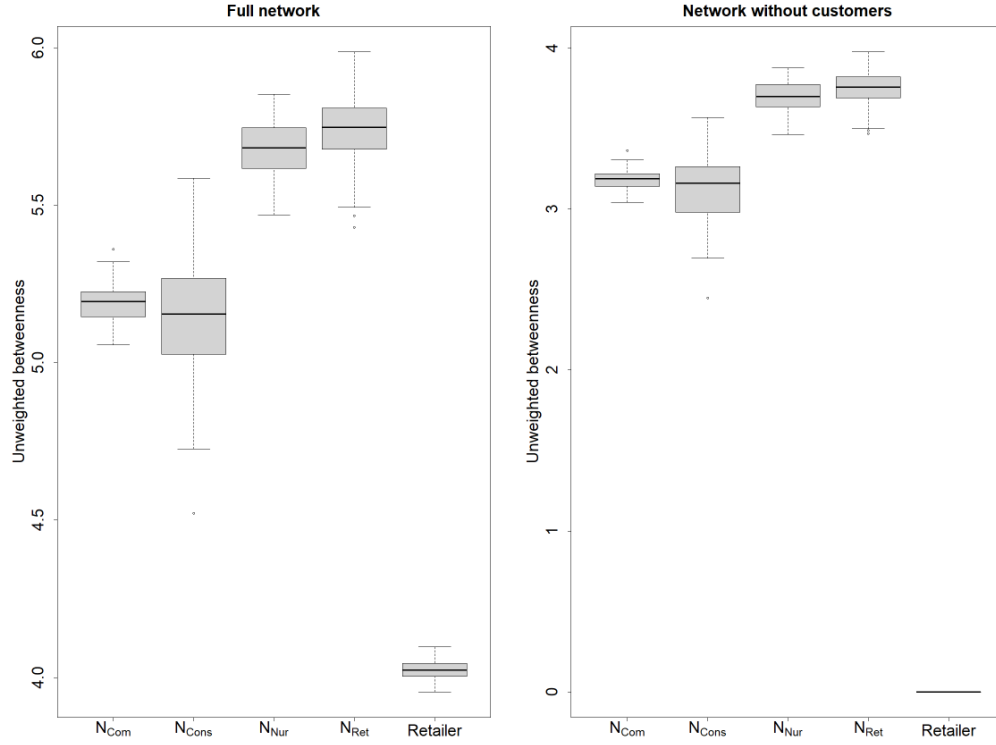
Figure B.9. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
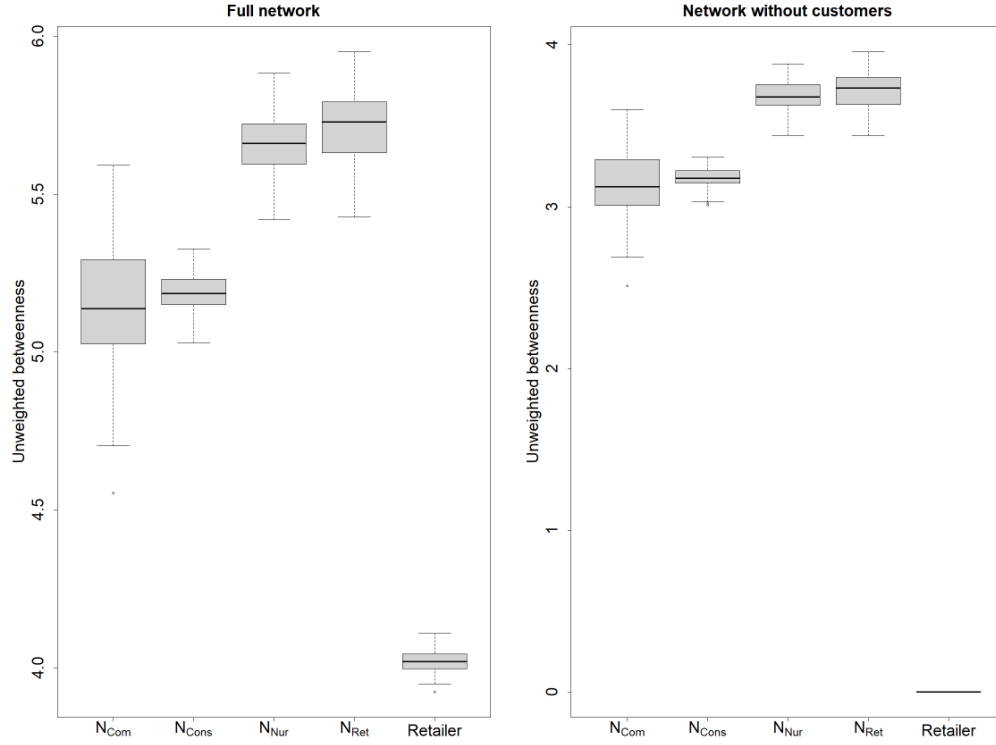
Figure B.10. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
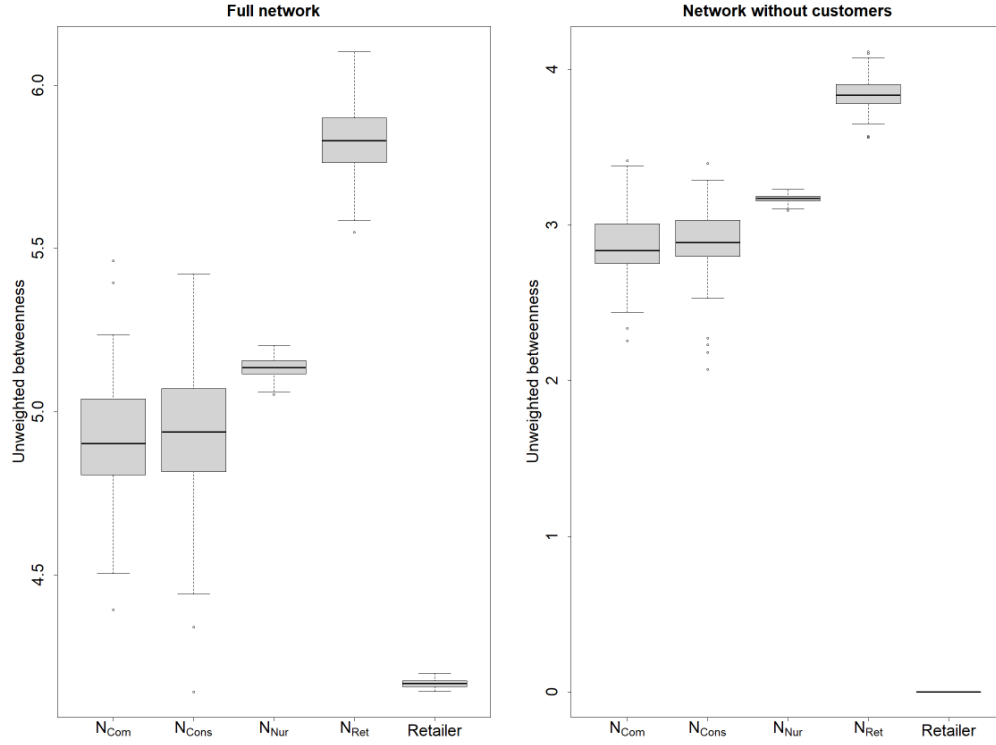
Figure B.11. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
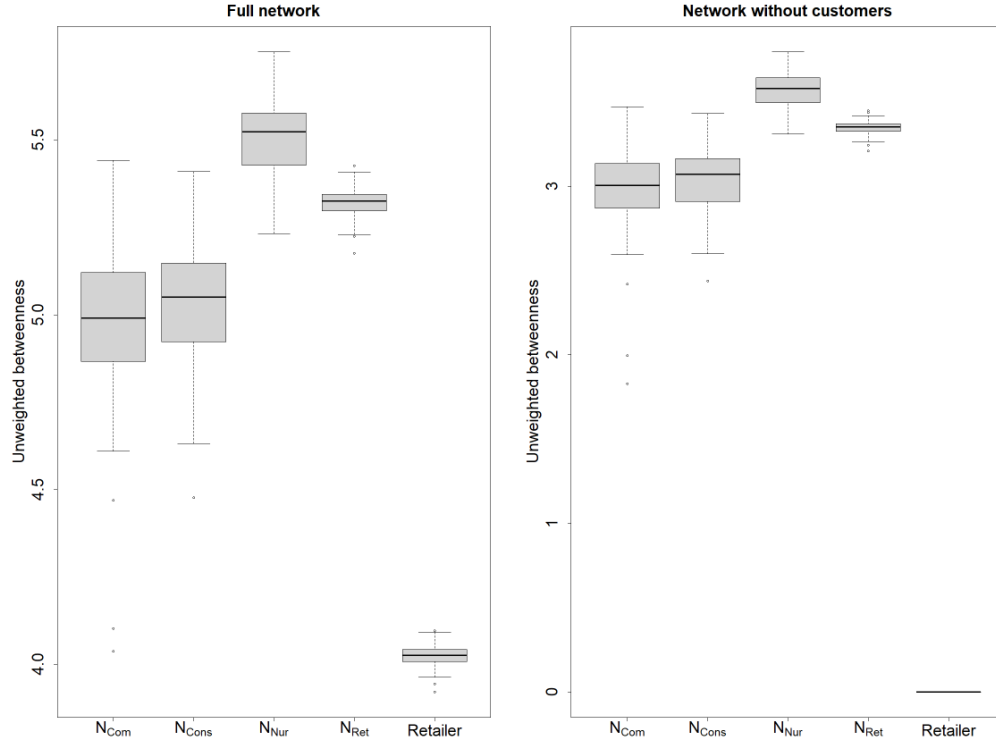
Figure B.12. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
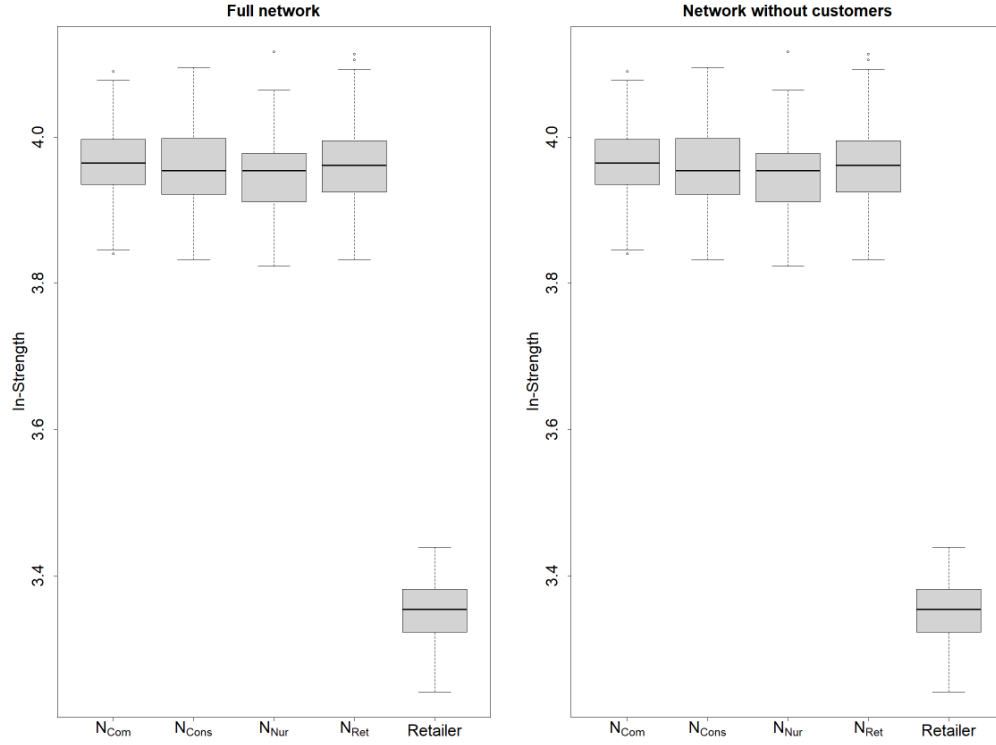
Figure B.13. Box plots of the average unweighted betweenness score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
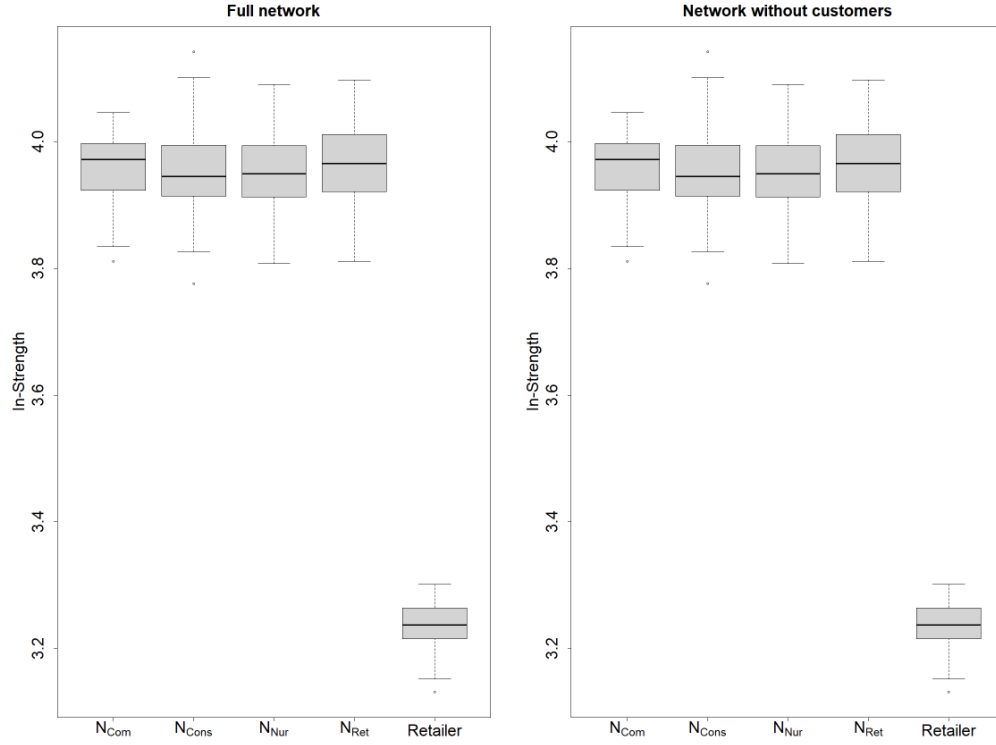
Figure B.14. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
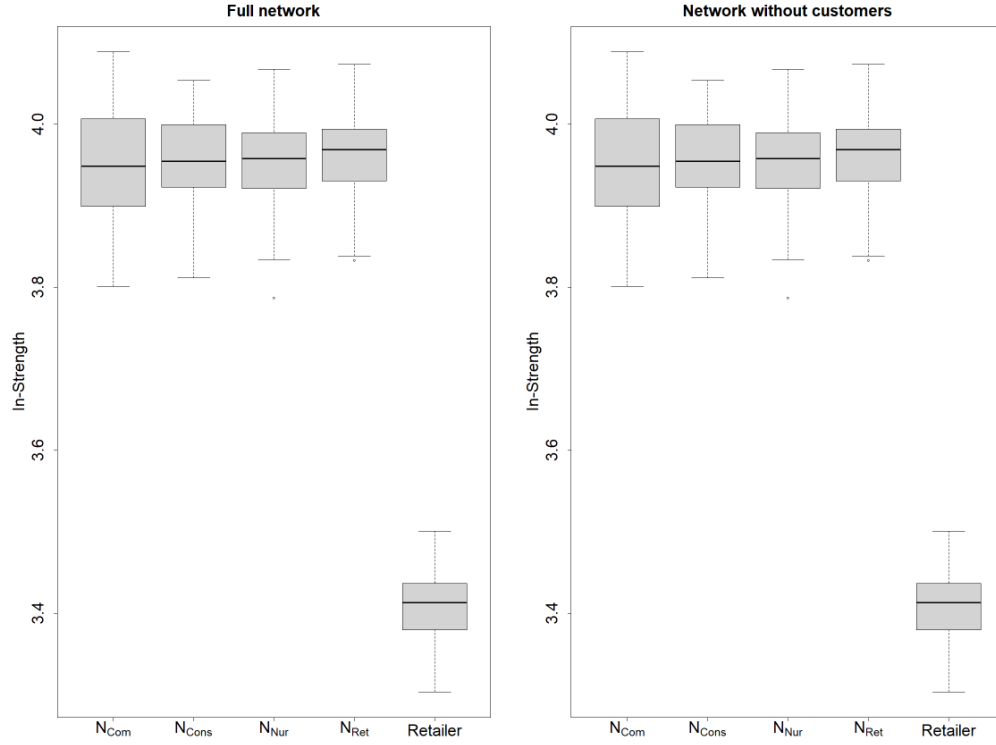
Figure B.15. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
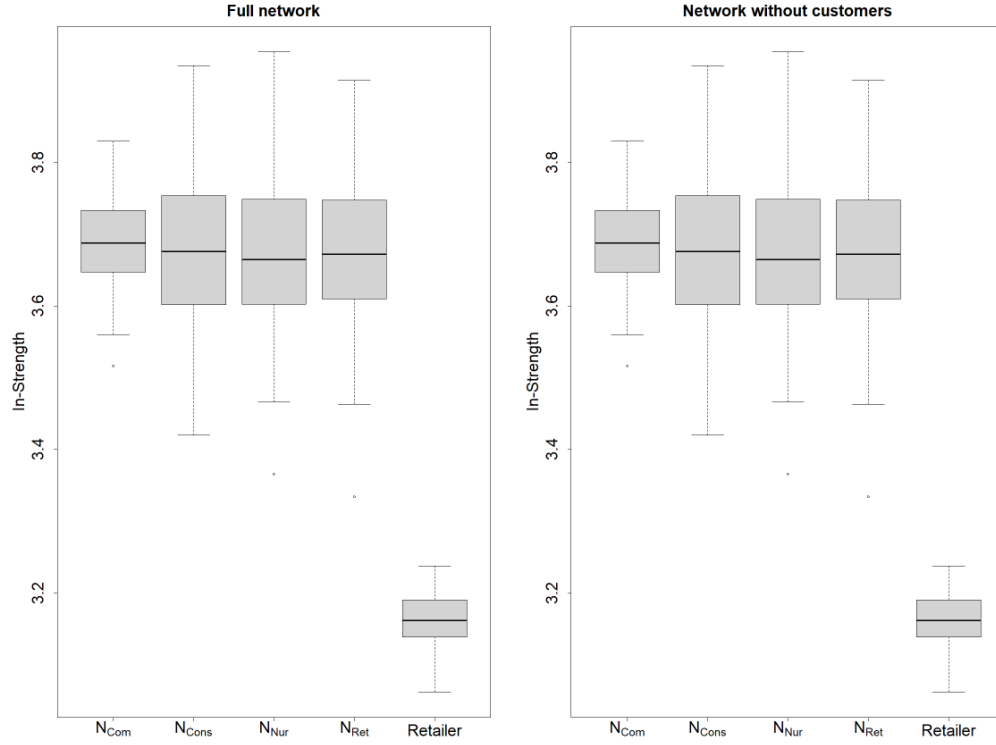
Figure B.16. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
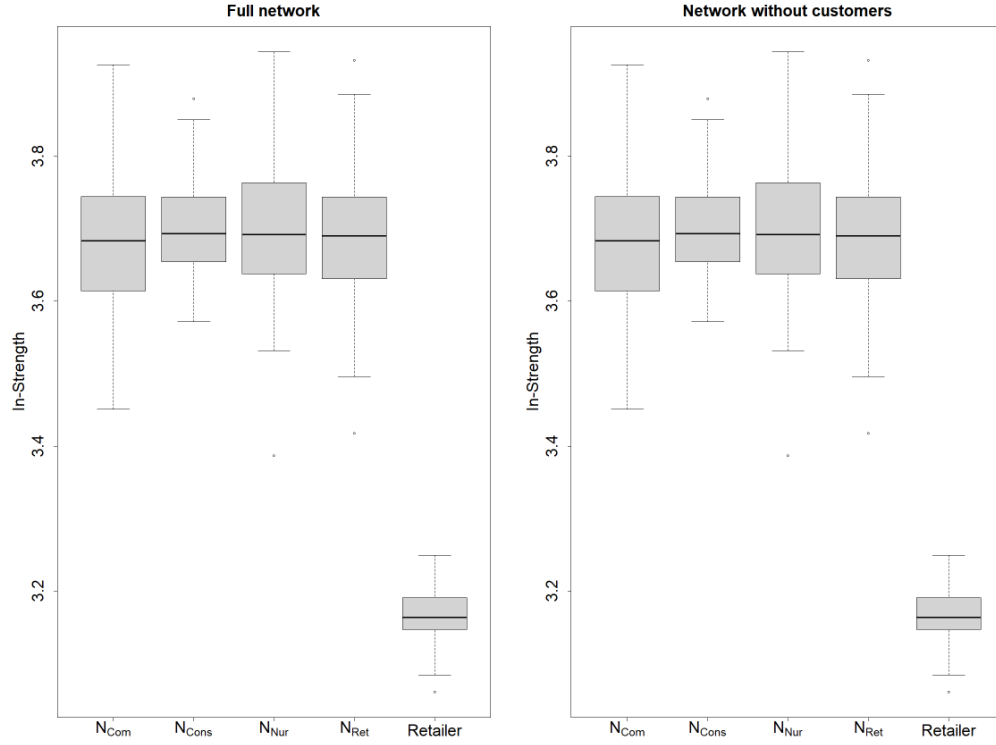
Figure B.17. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
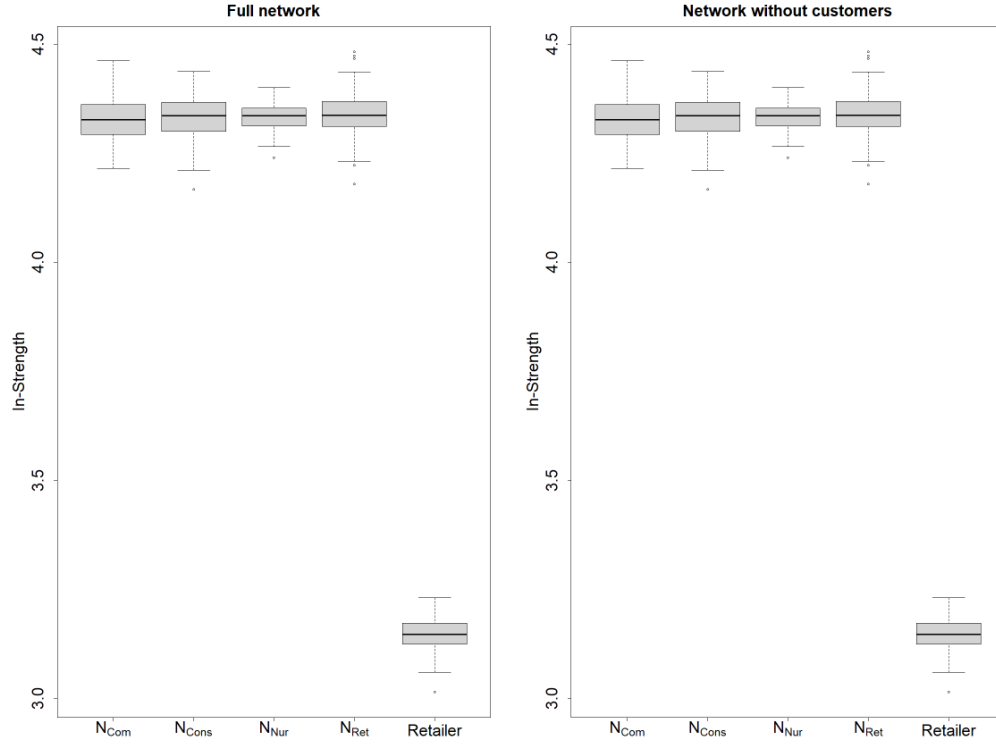
Figure B.18. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
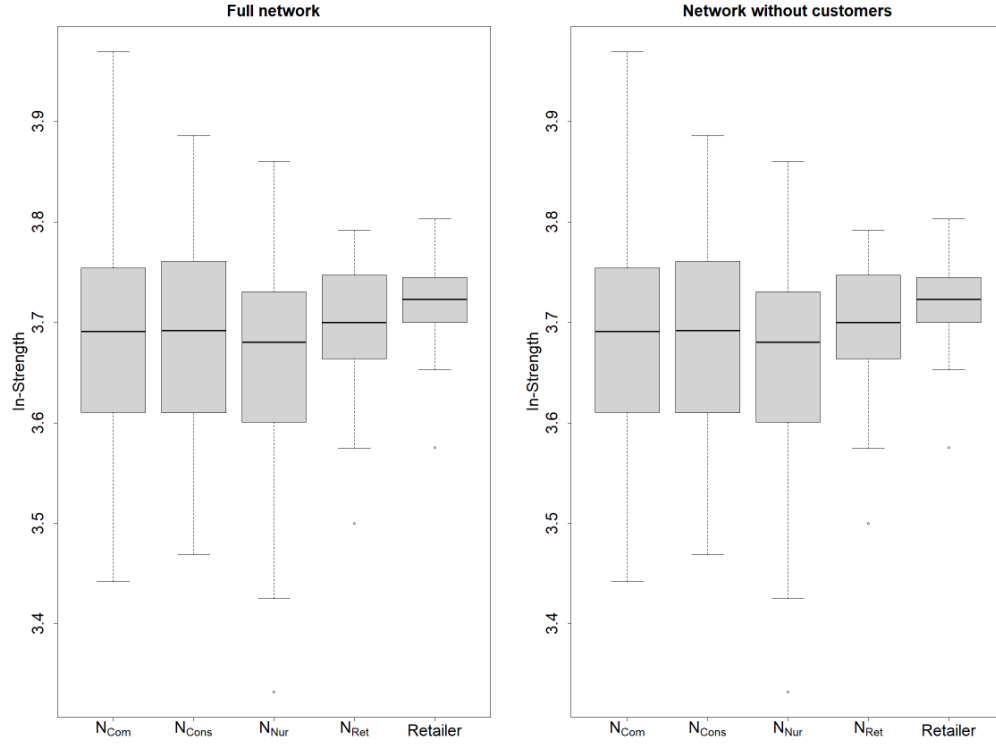
Figure B.19. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
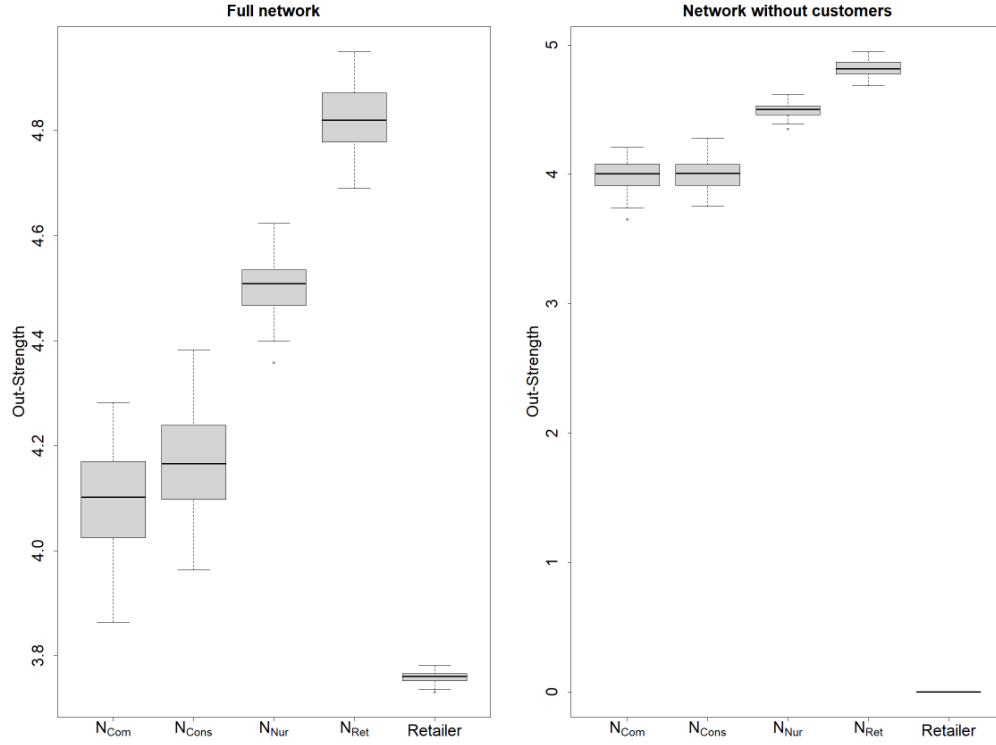
Figure B.20. Box plots of the average in-strength score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
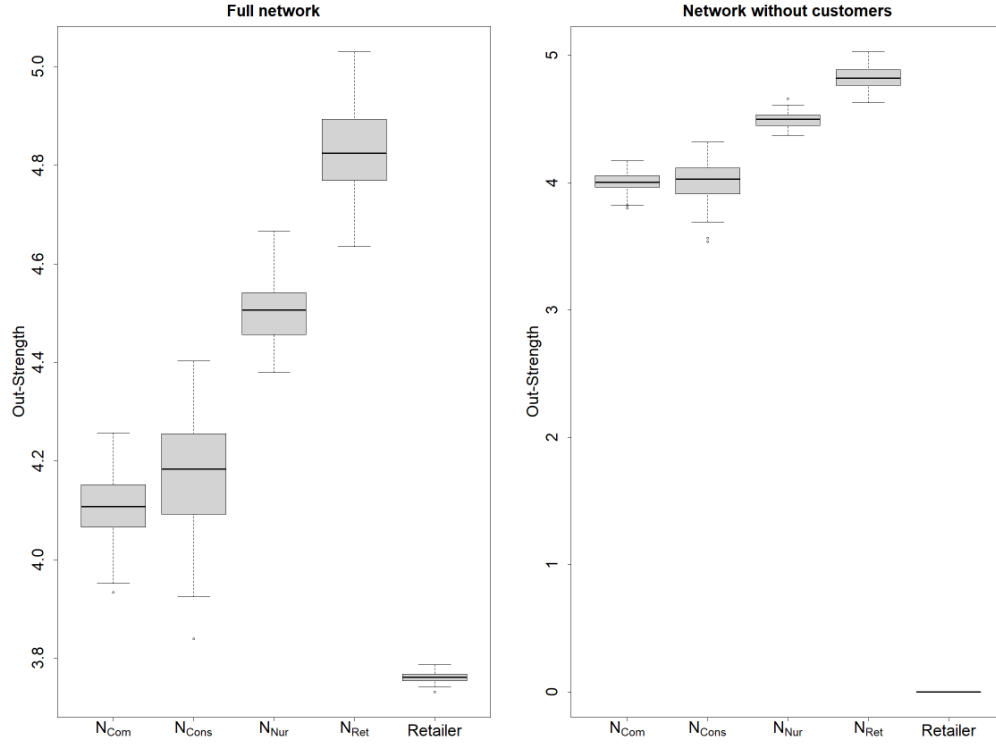
Figure B.21. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
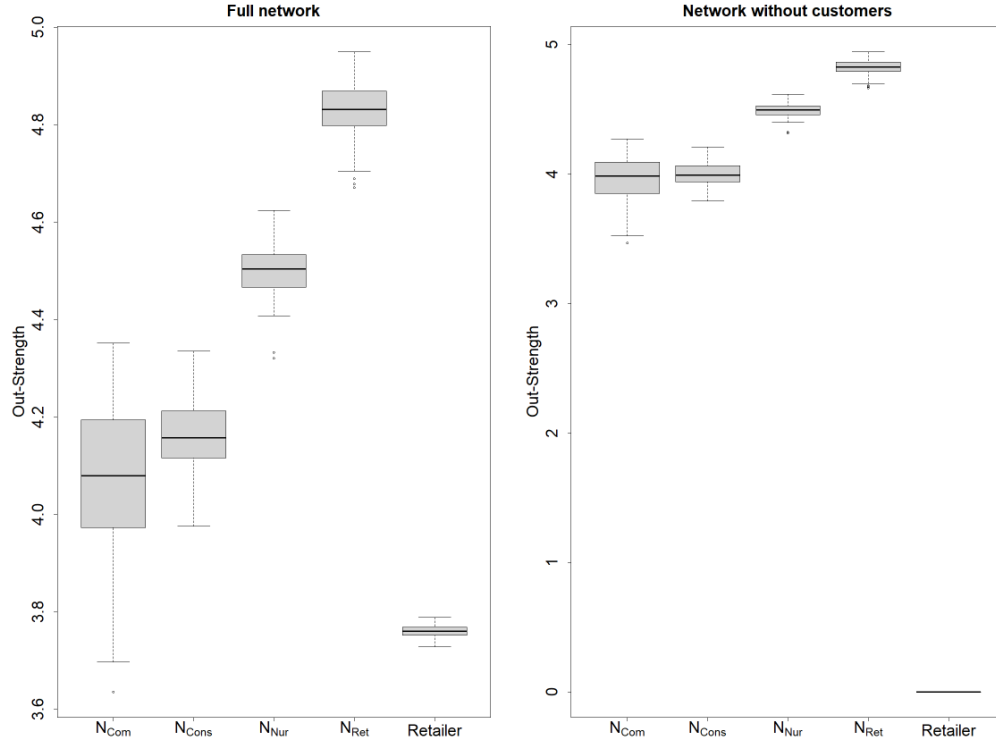
Figure B.22. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.

Figure B.23. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
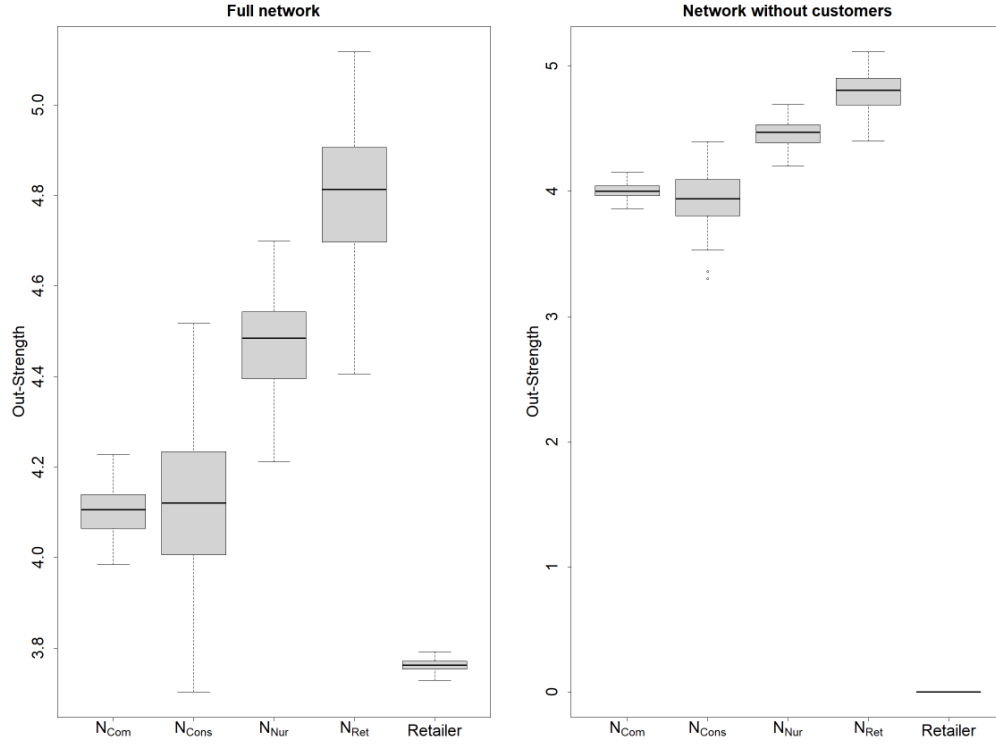
Figure B.24. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
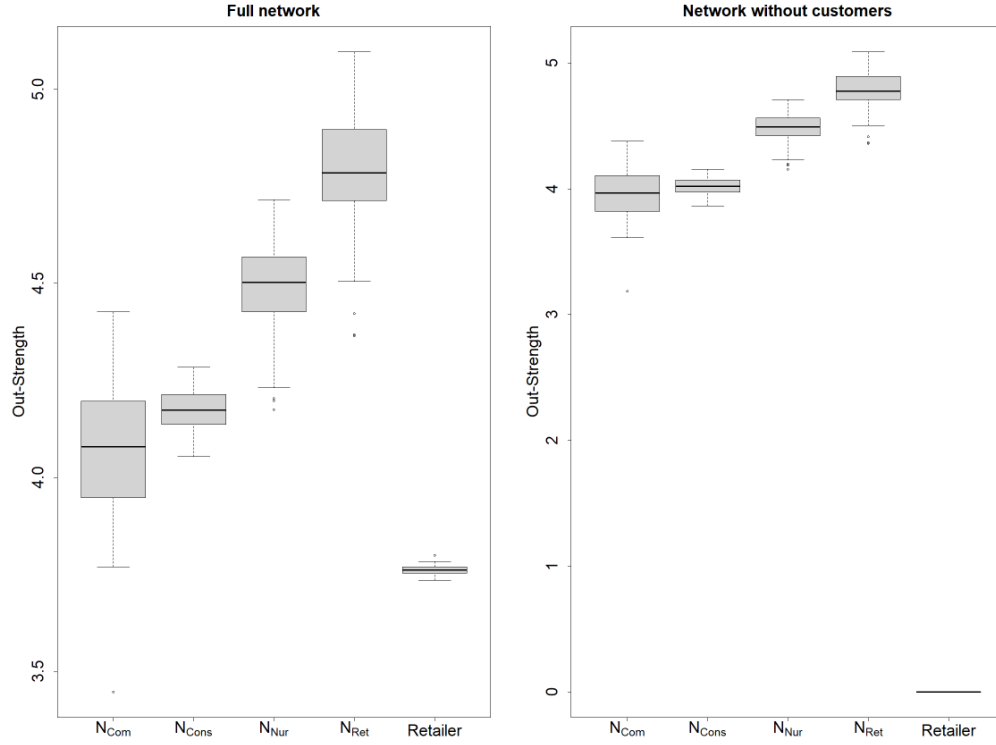
Figure B.25. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
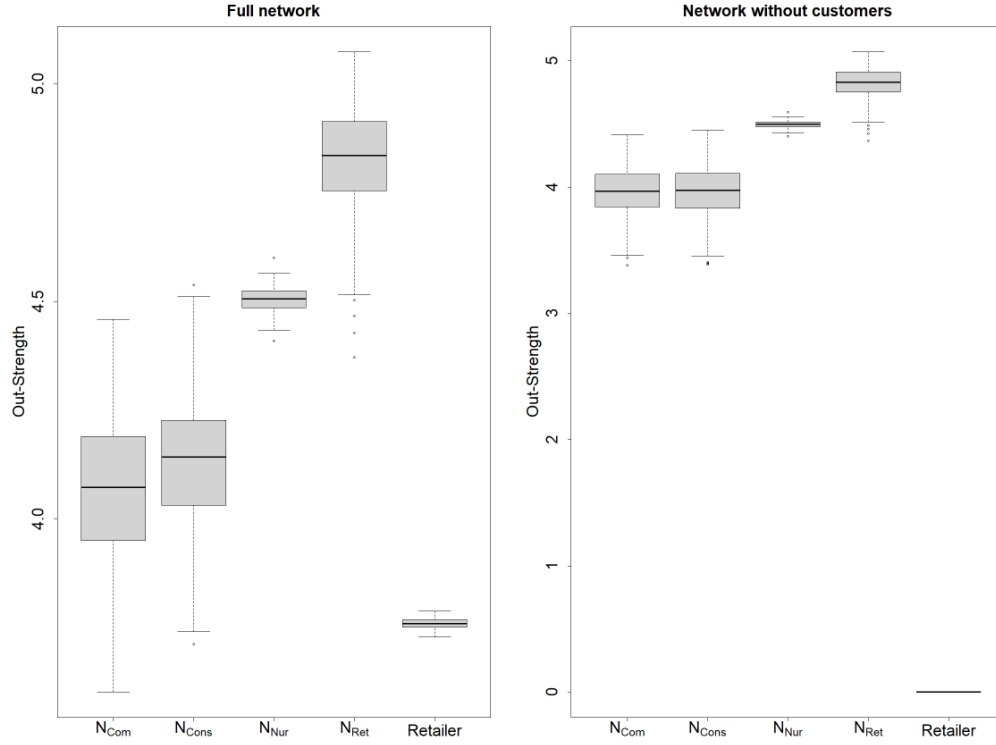
Figure B.26. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
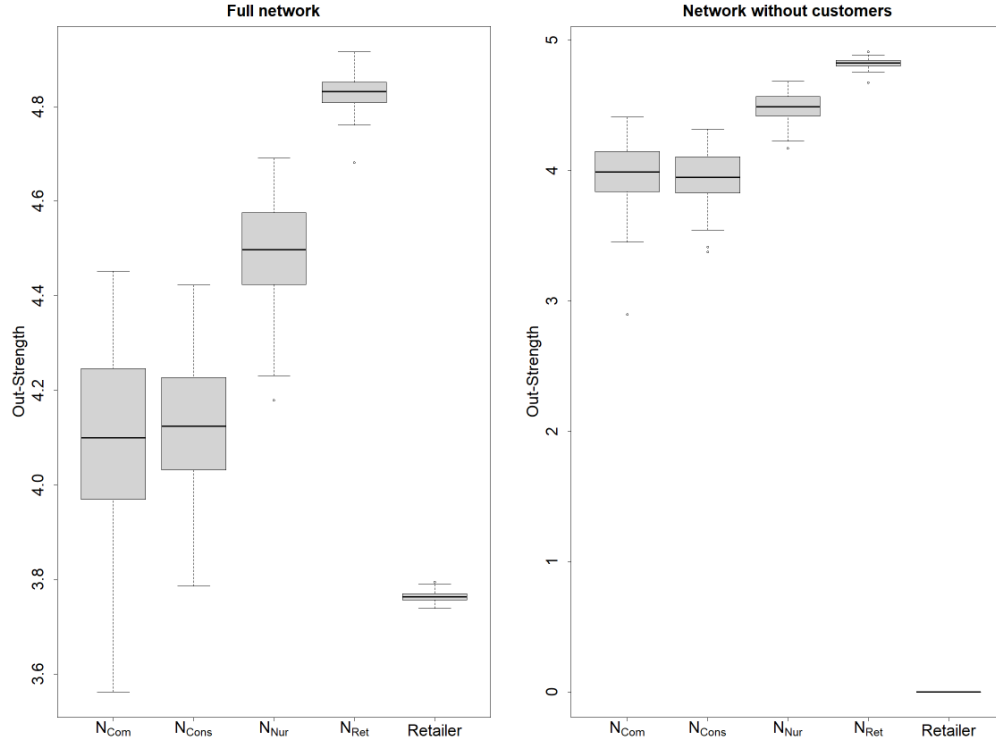
Figure B.27. Box plots of the average out-strength score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
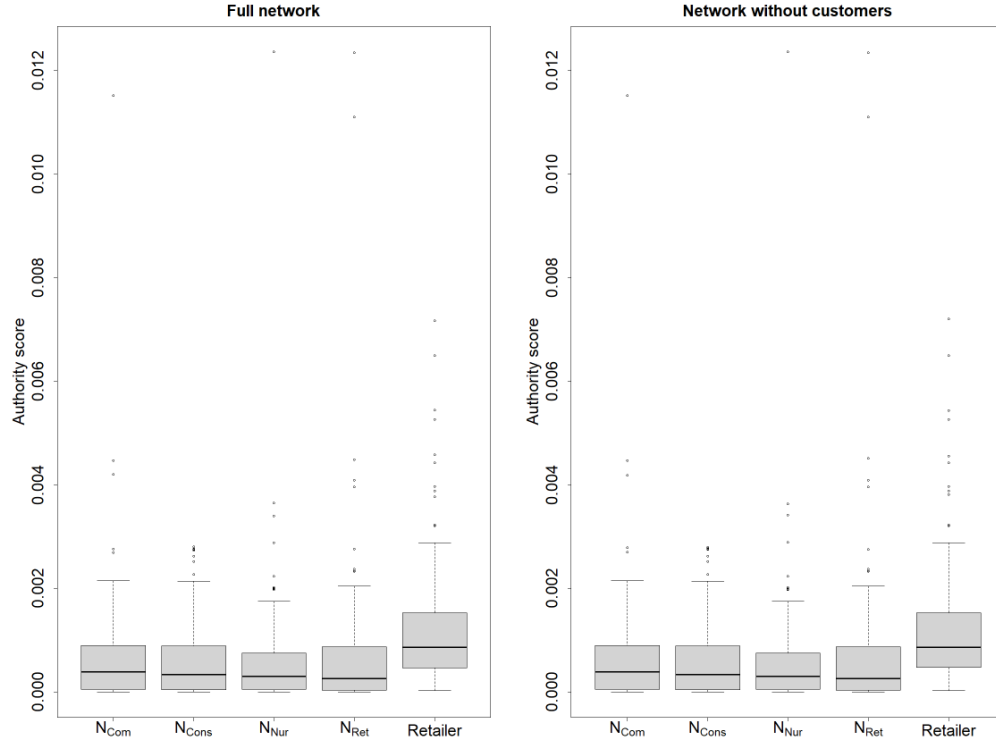
Figure B.28. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
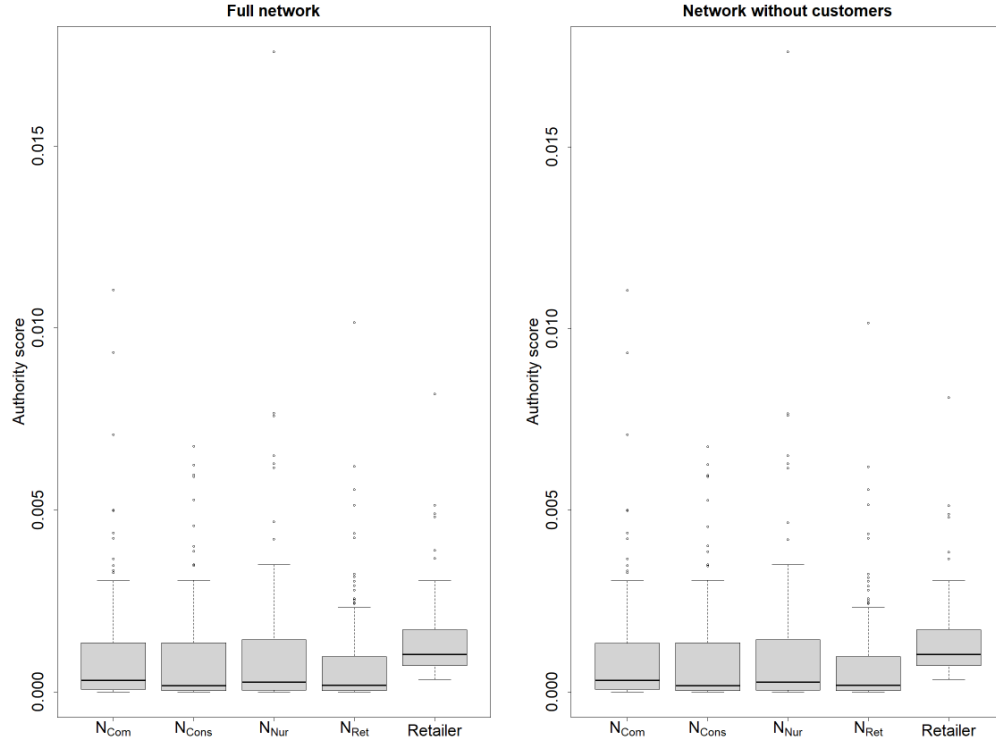
Figure B.29. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
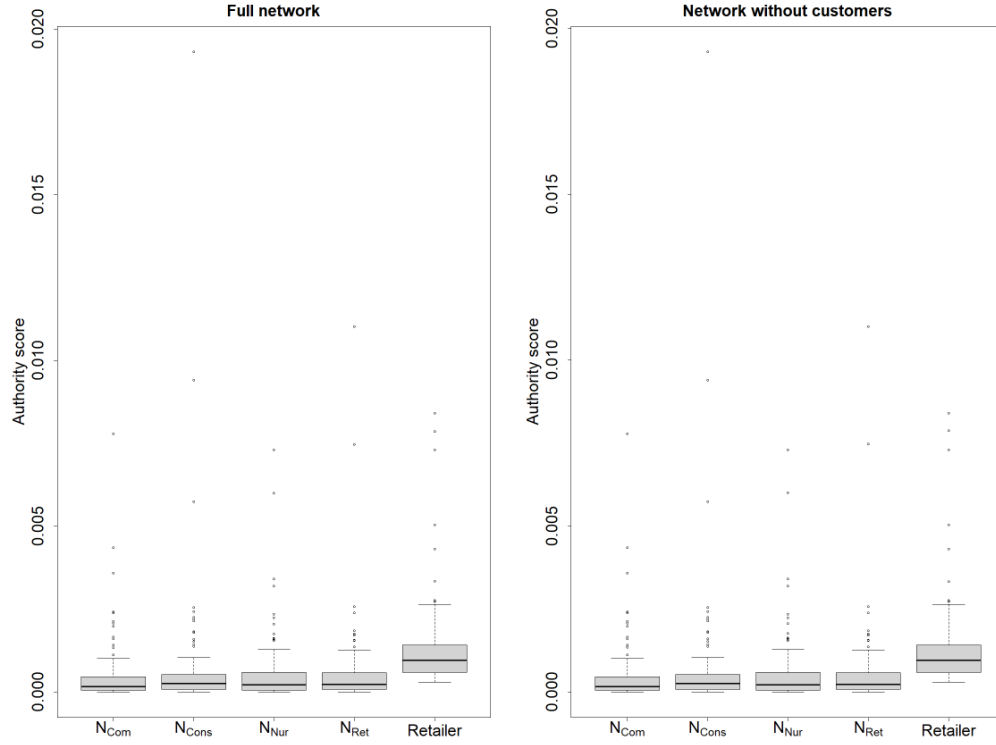
Figure B.30. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
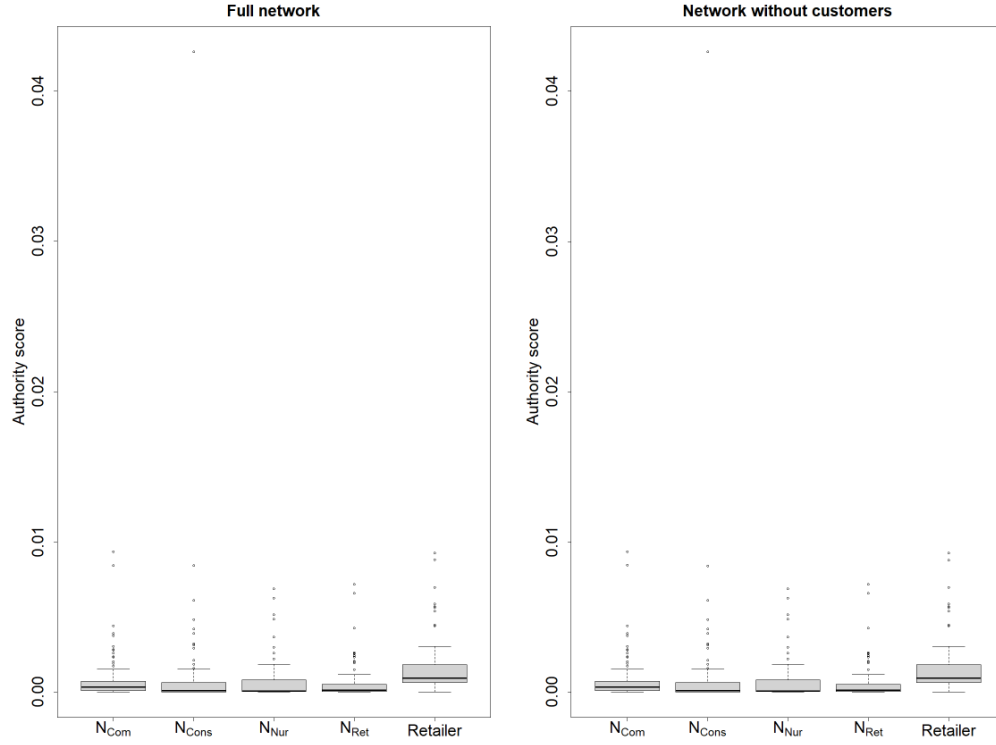
Figure B.31. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
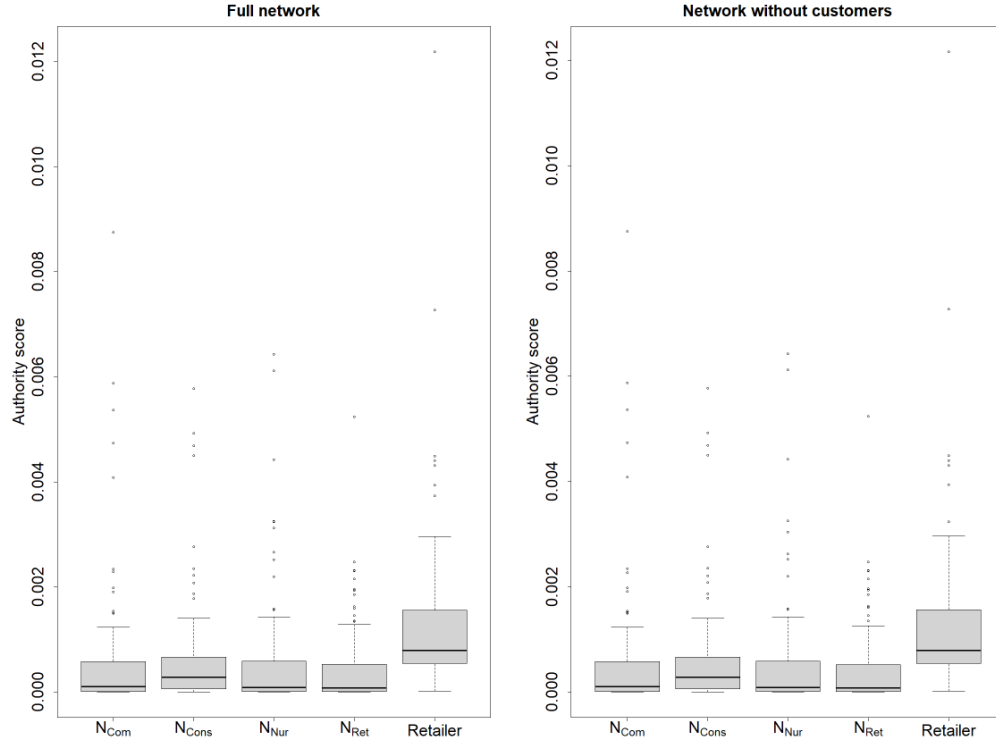
Figure B.32. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
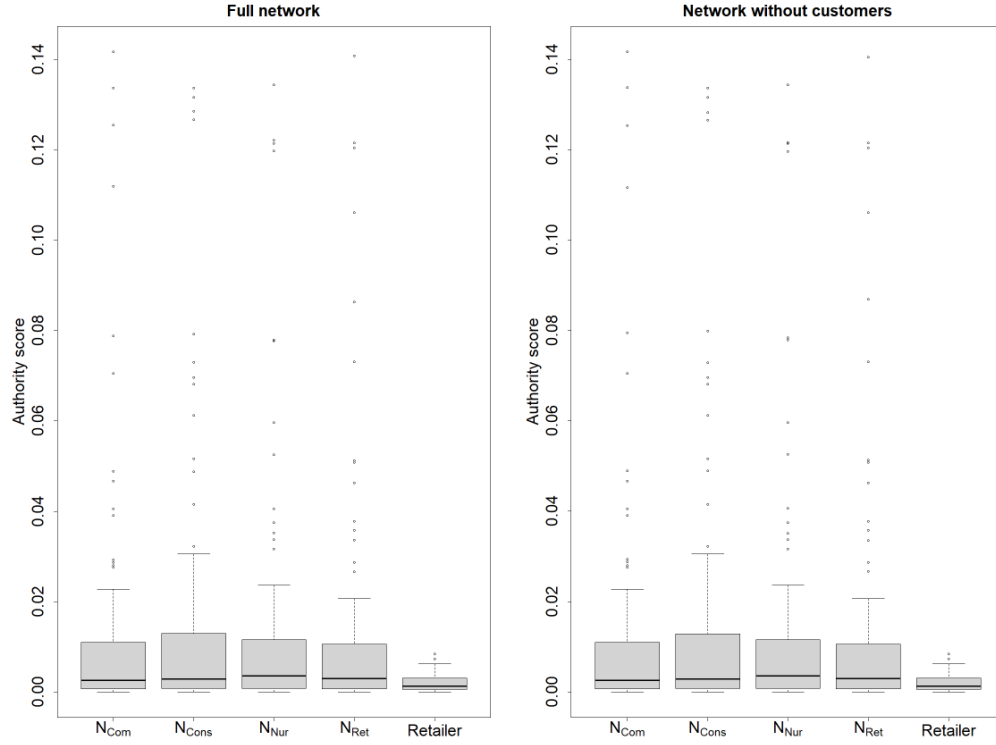
Figure B.33. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
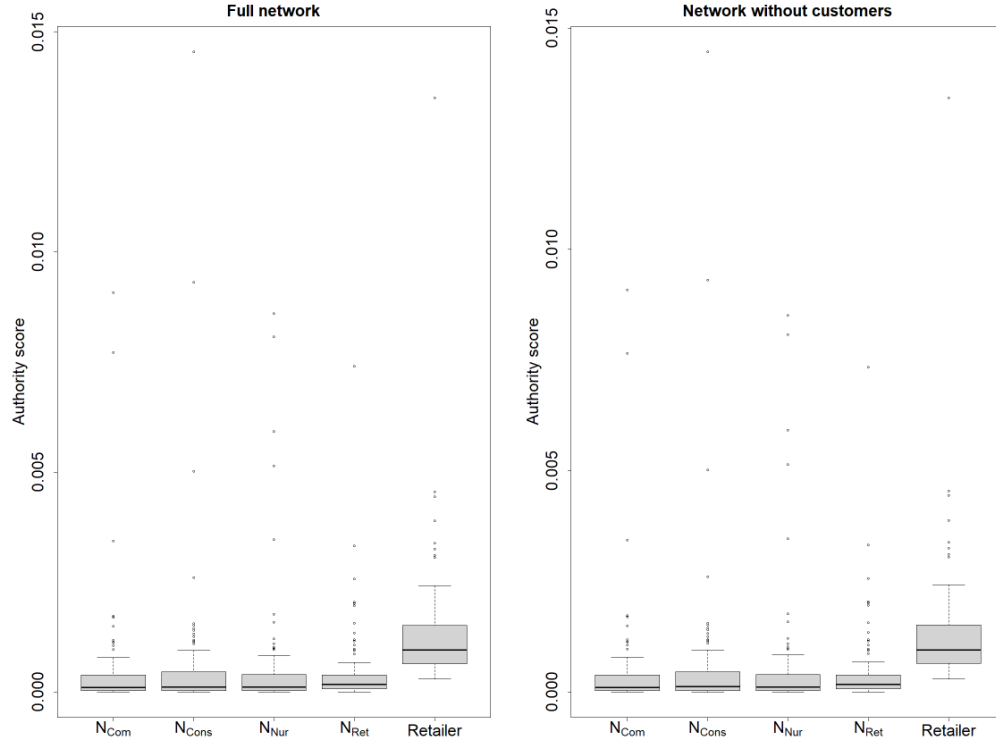
Figure B.34. Box plots of the average authority score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
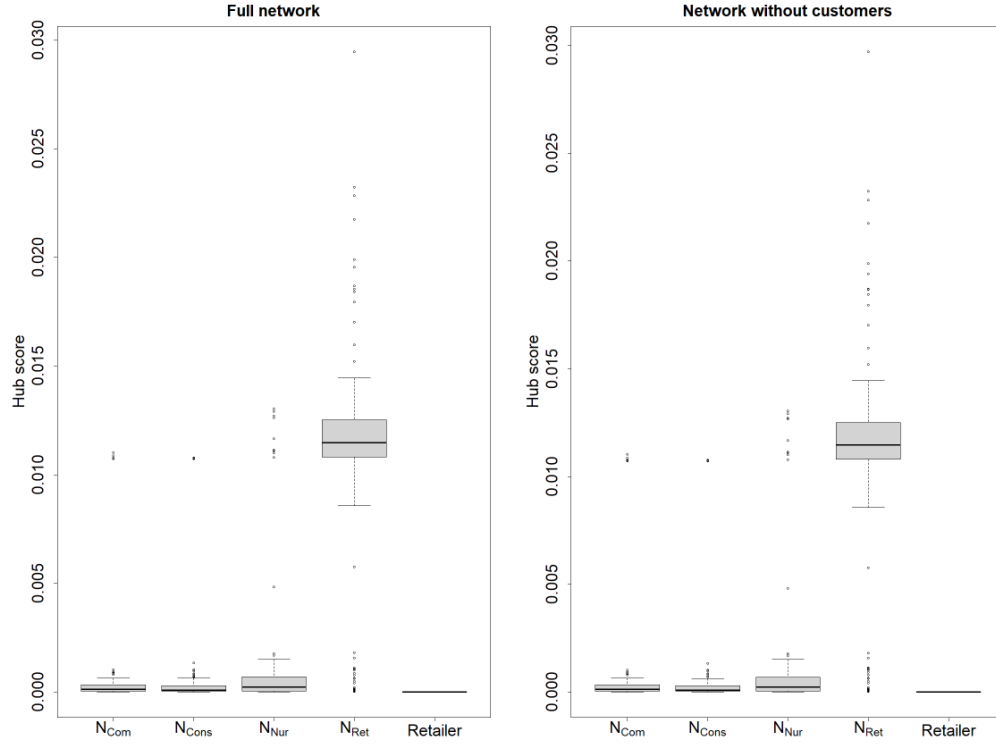
Figure B.35. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
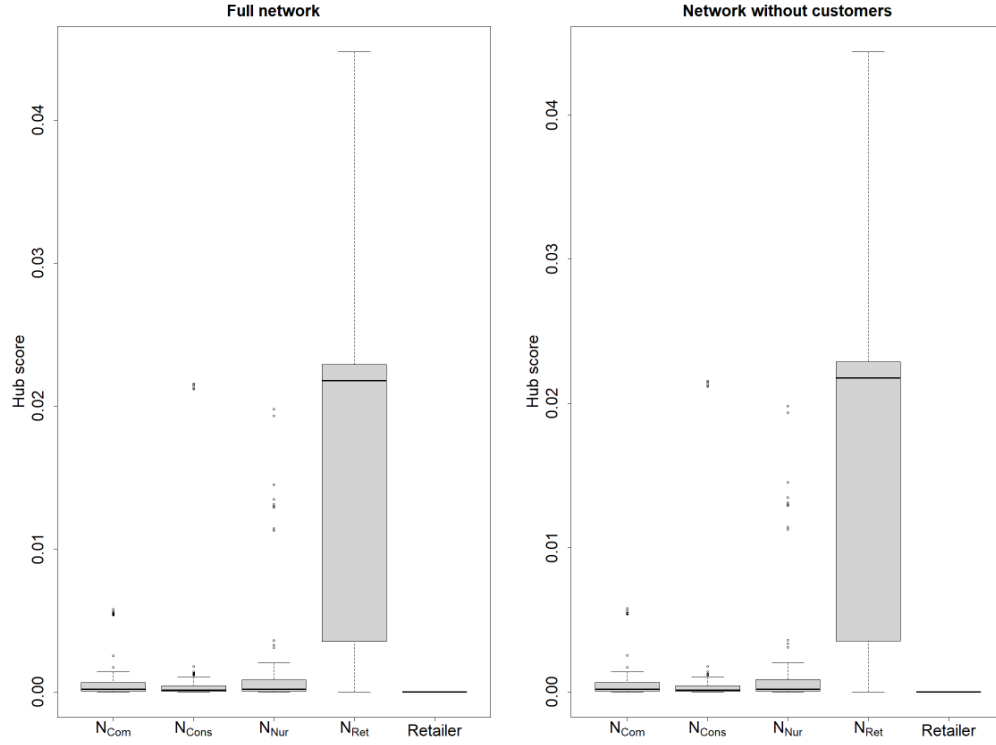
Figure B.36. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
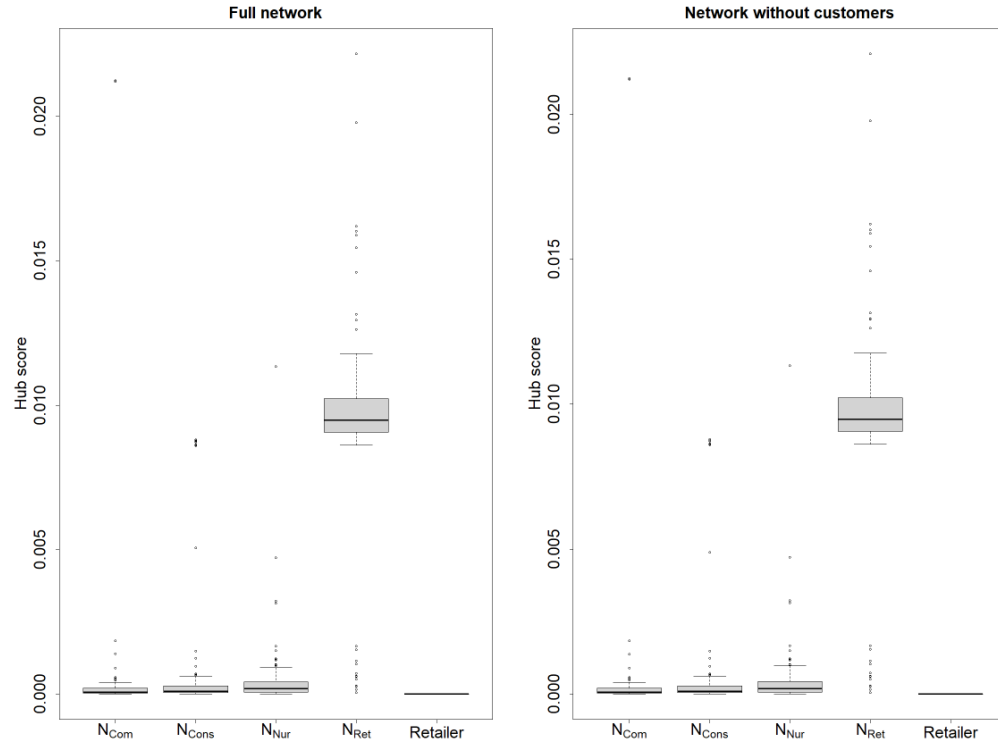
Figure B.37. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
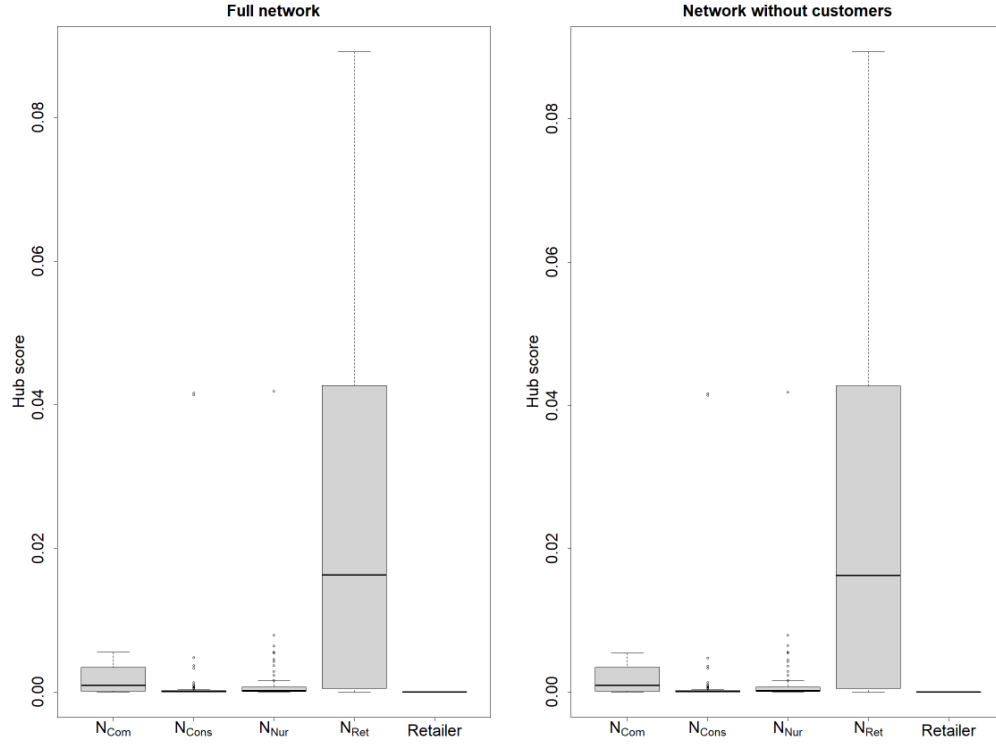
Figure B.38. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
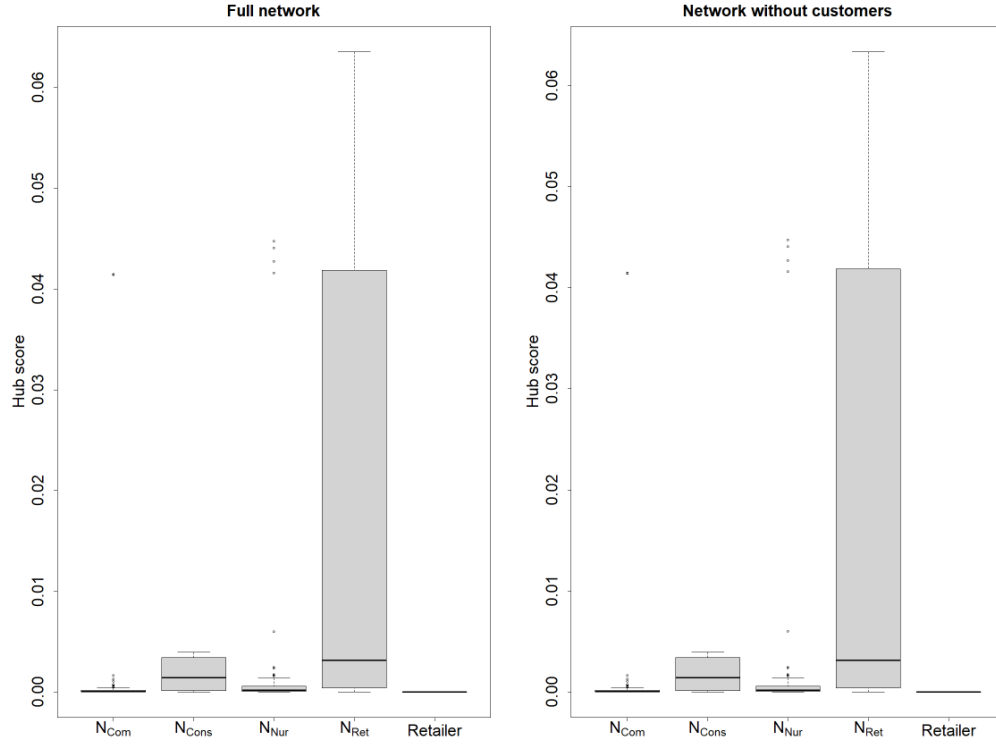
Figure B.39. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
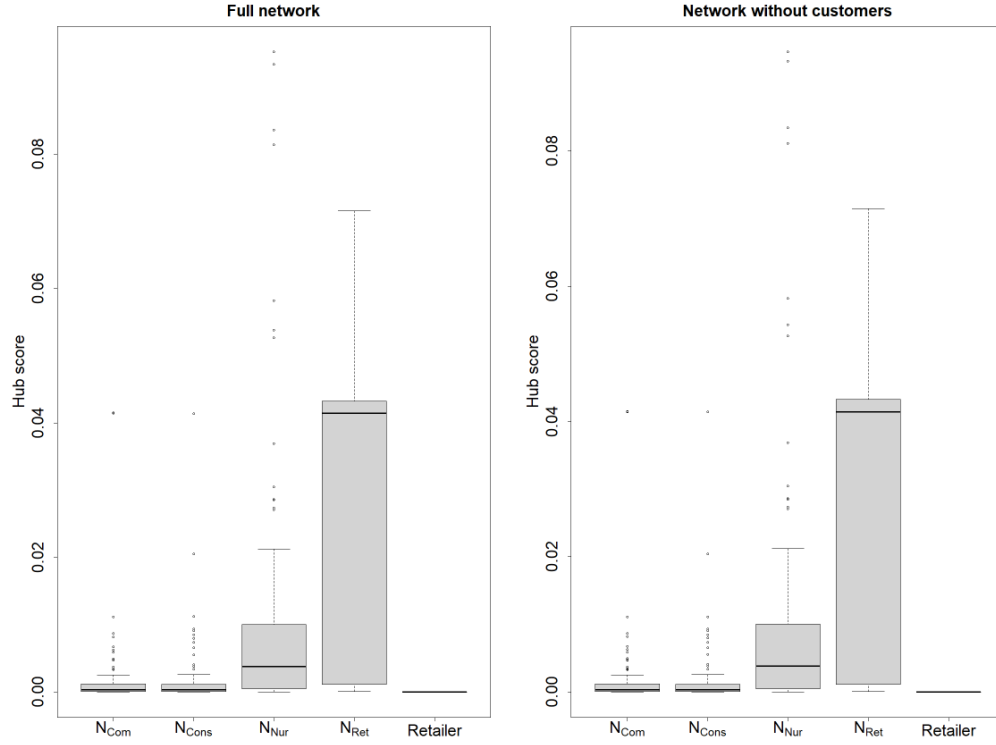
Figure B.40. Box plots of the average hub score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
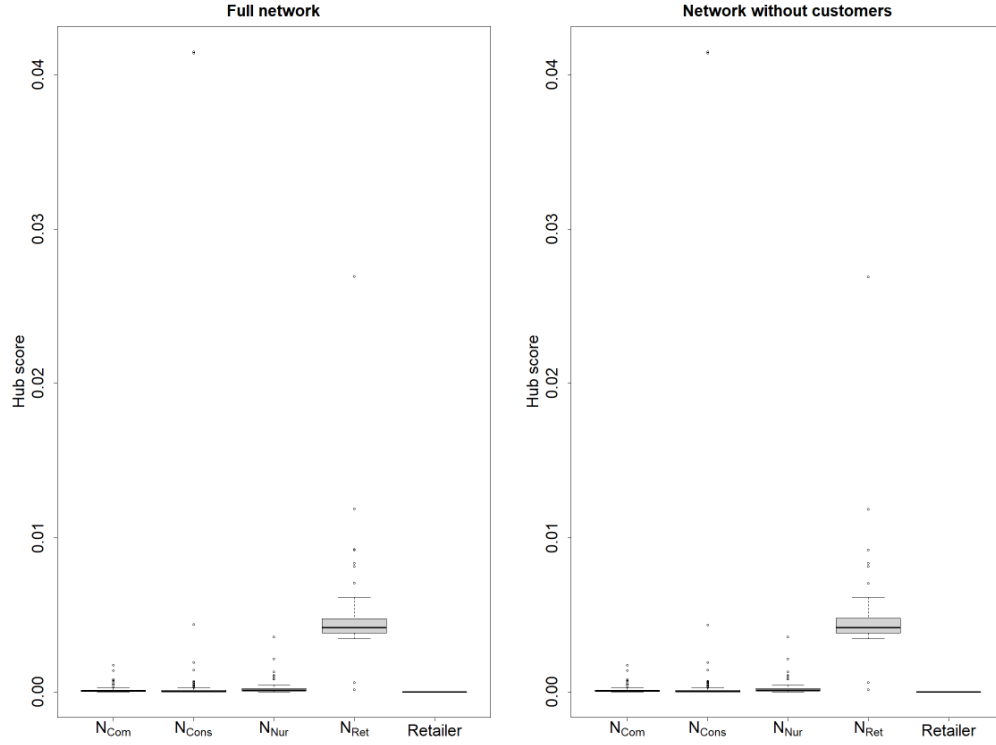
Figure B.41. Box plots of the average hub score score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
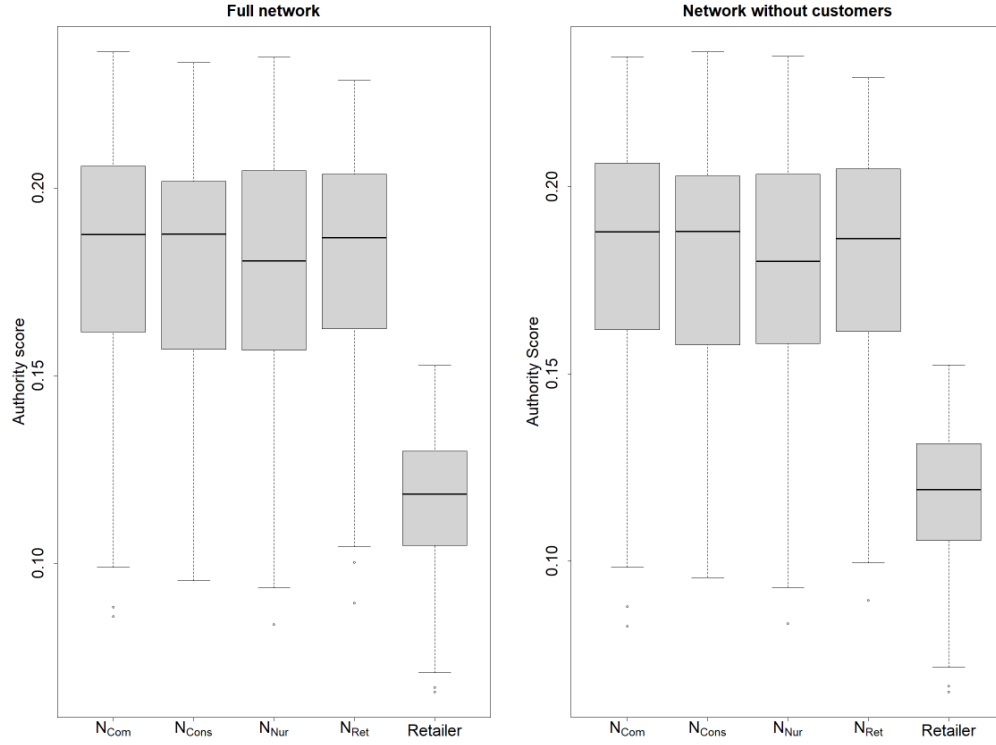
Figure B.42. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
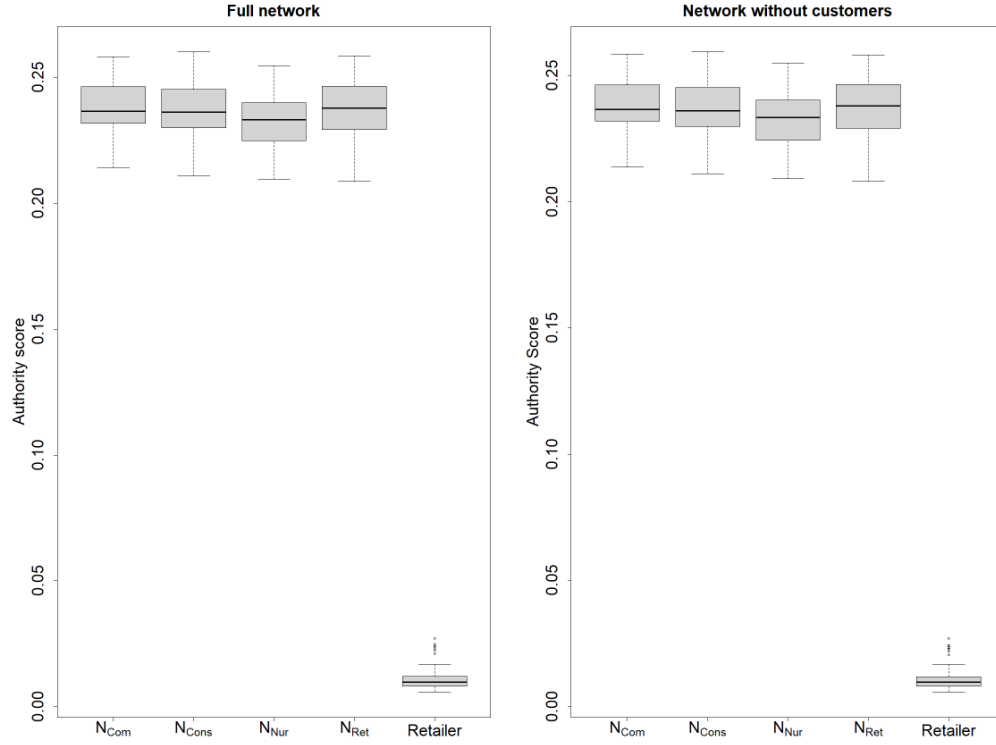
Figure B.43. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
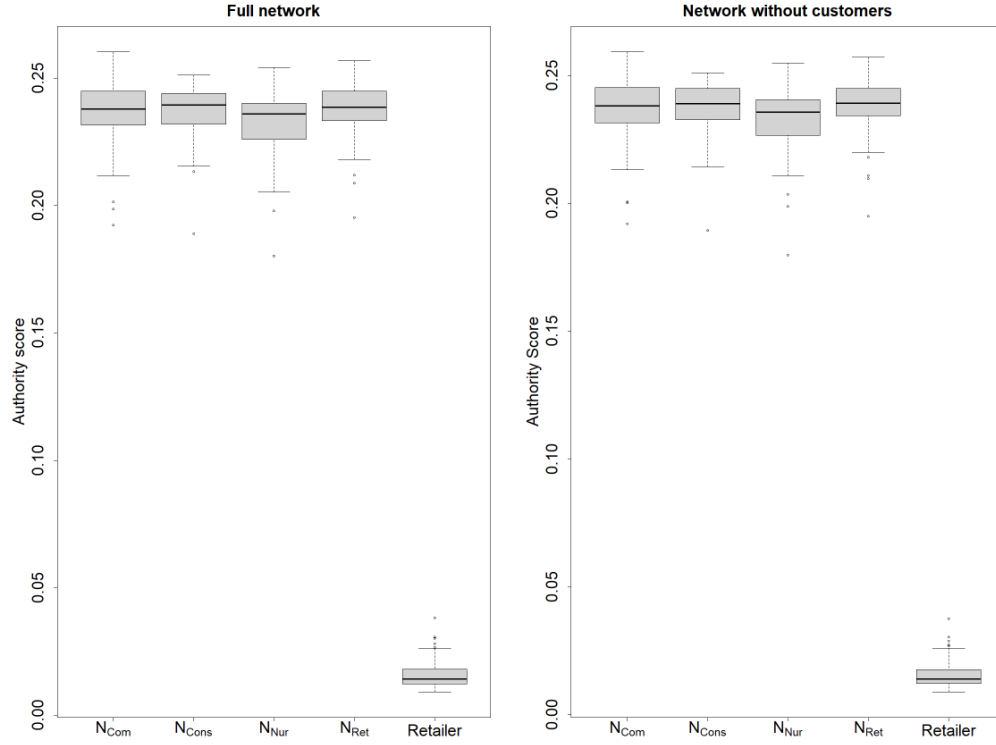
Figure B.44. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.

Figure B.45. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
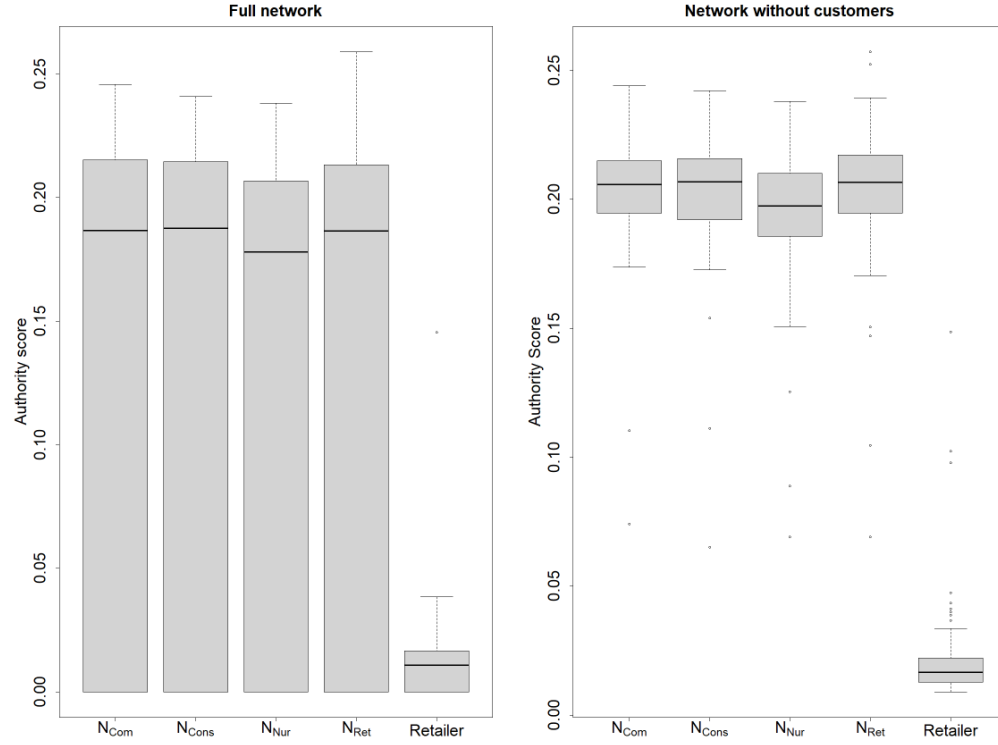
Figure B.46. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
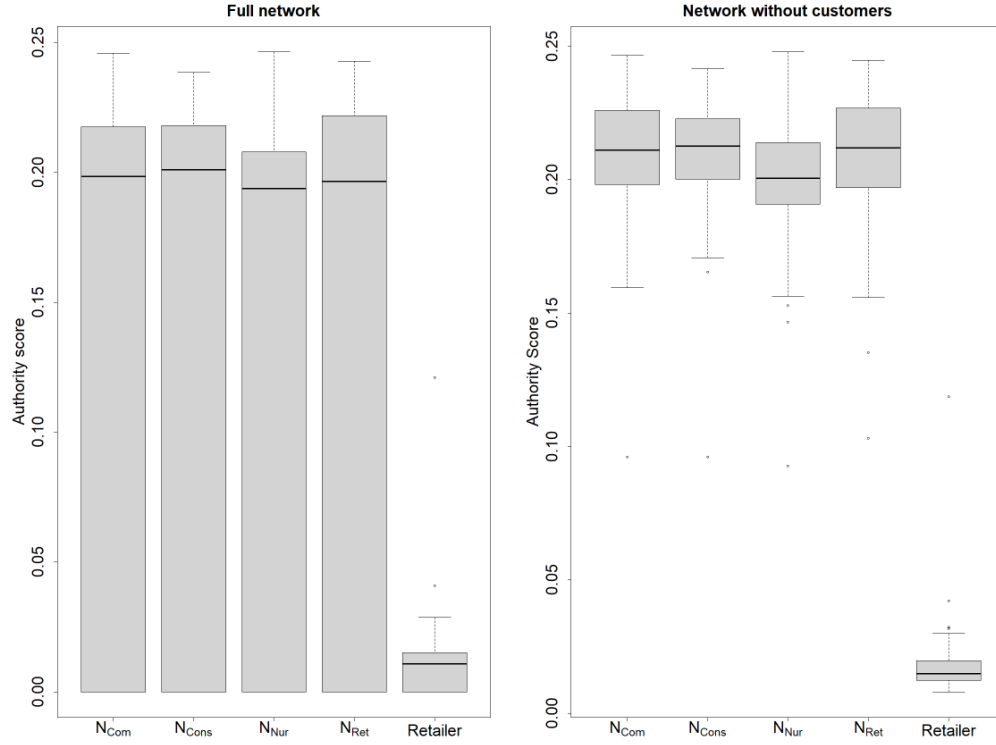
Figure B.47. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
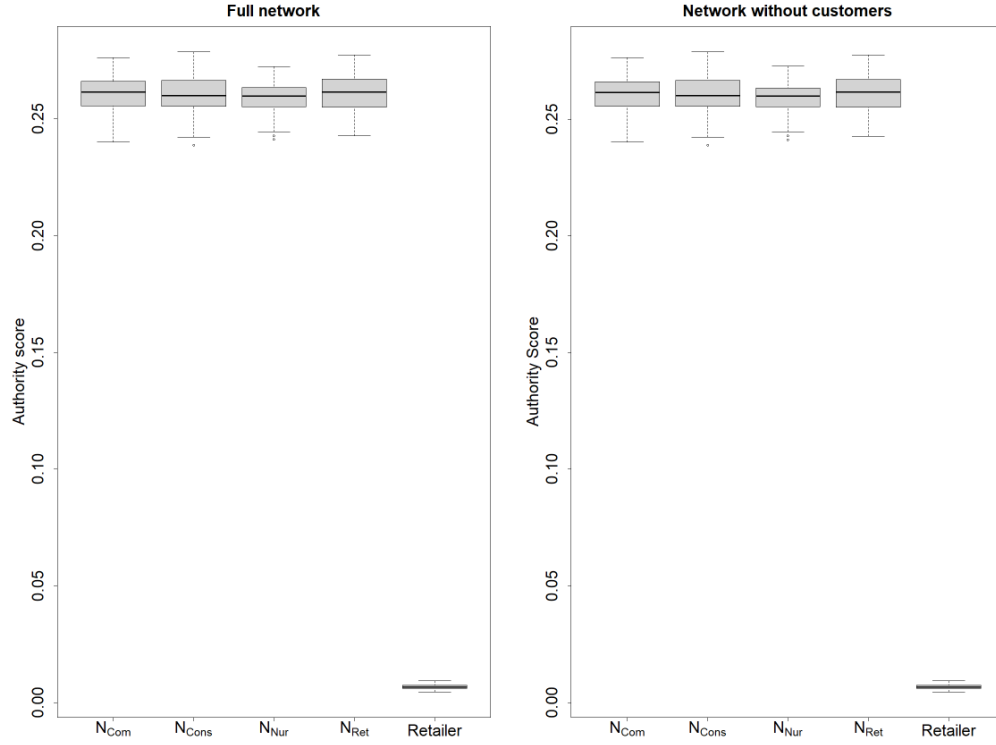
Figure B.48. Box plots of the average unweighted authority score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
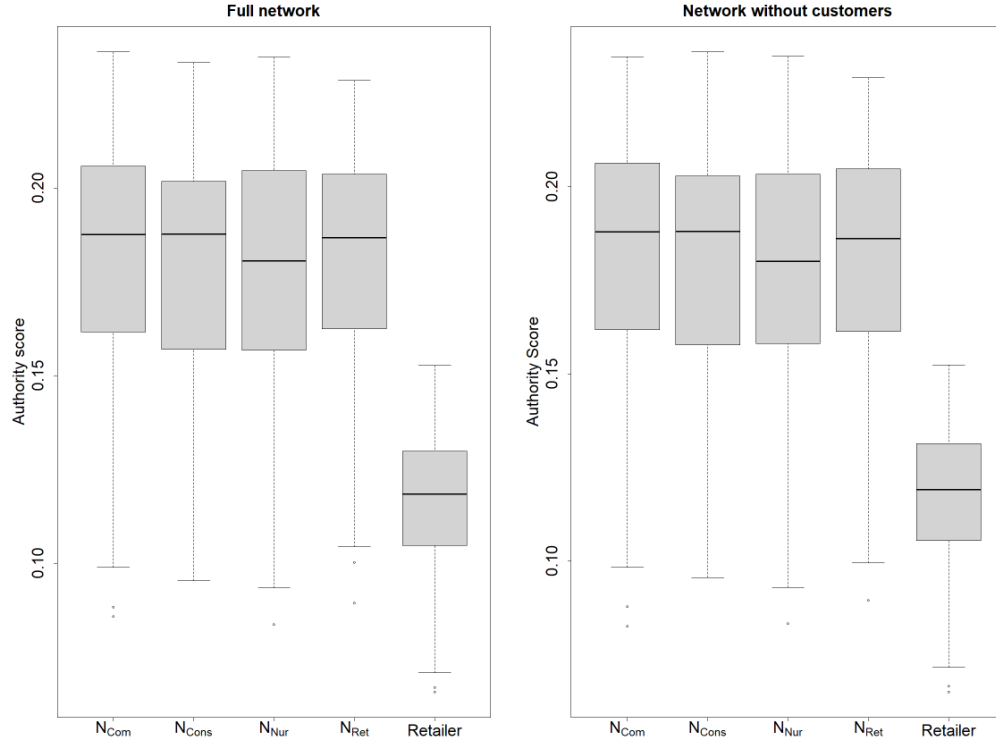
Figure B.49. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 1 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (40, 40, 40, 40)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
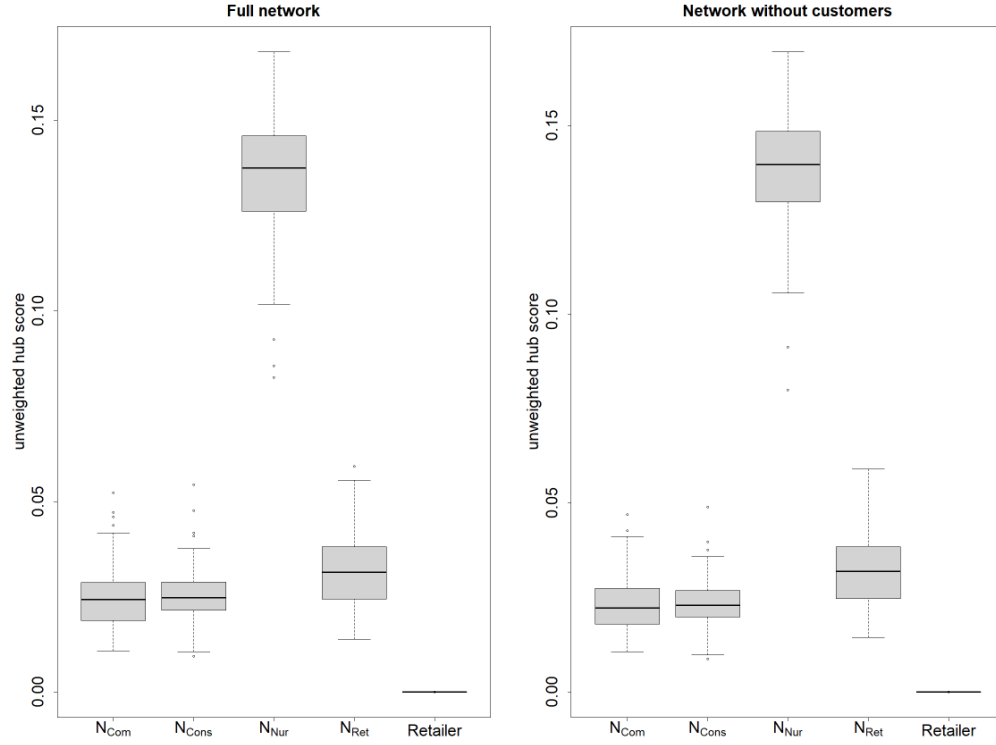
Figure B.50. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 2 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (80, 20, 40, 20)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
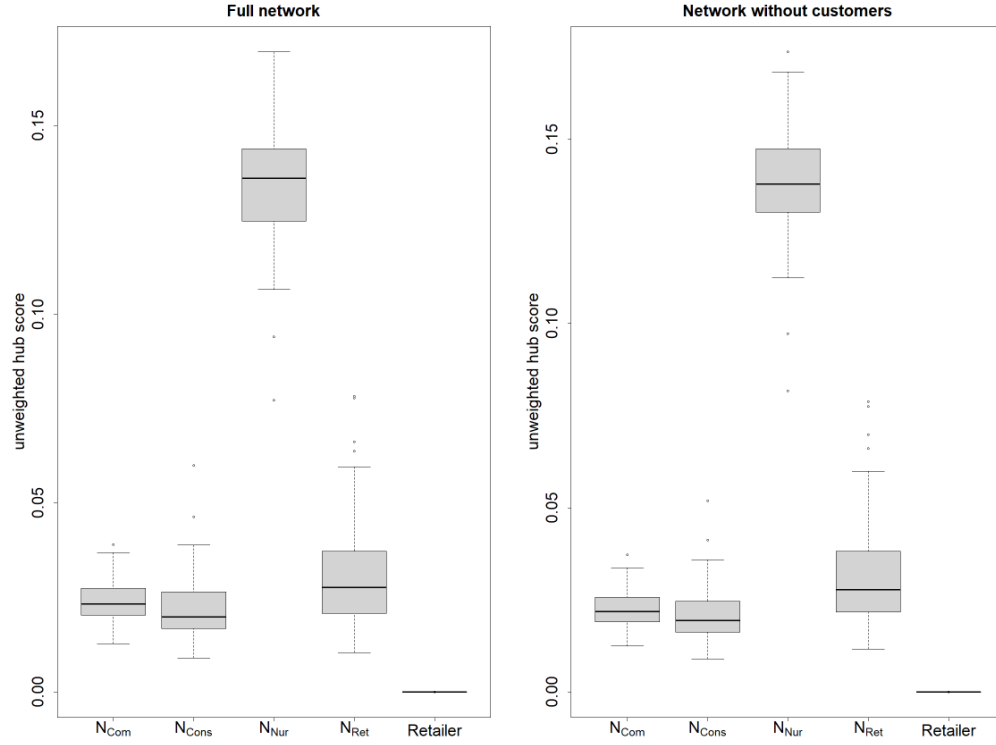
Figure B.51. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 3 ($(|N_{\mathrm{Com}}|, |N_{\mathrm{Cons}}|, |N_{\mathrm{Nur}}|, |N_{\mathrm{Ret}}|) = (20, 50, 40, 50)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \mathrm{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
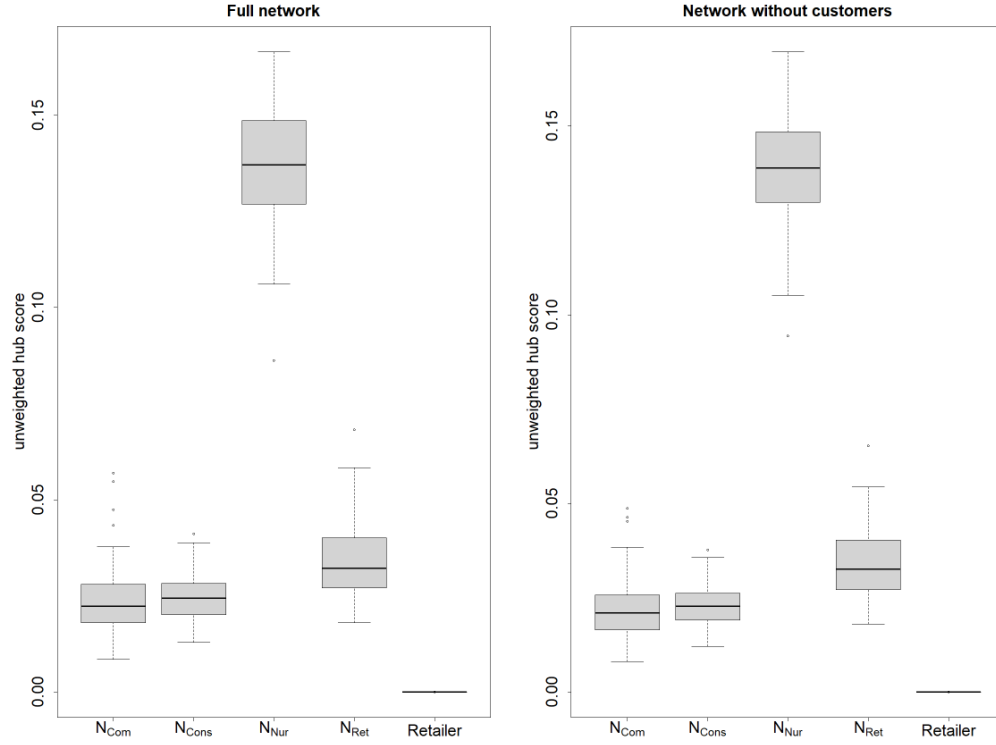
Figure B.52. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 4 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (130, 10, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
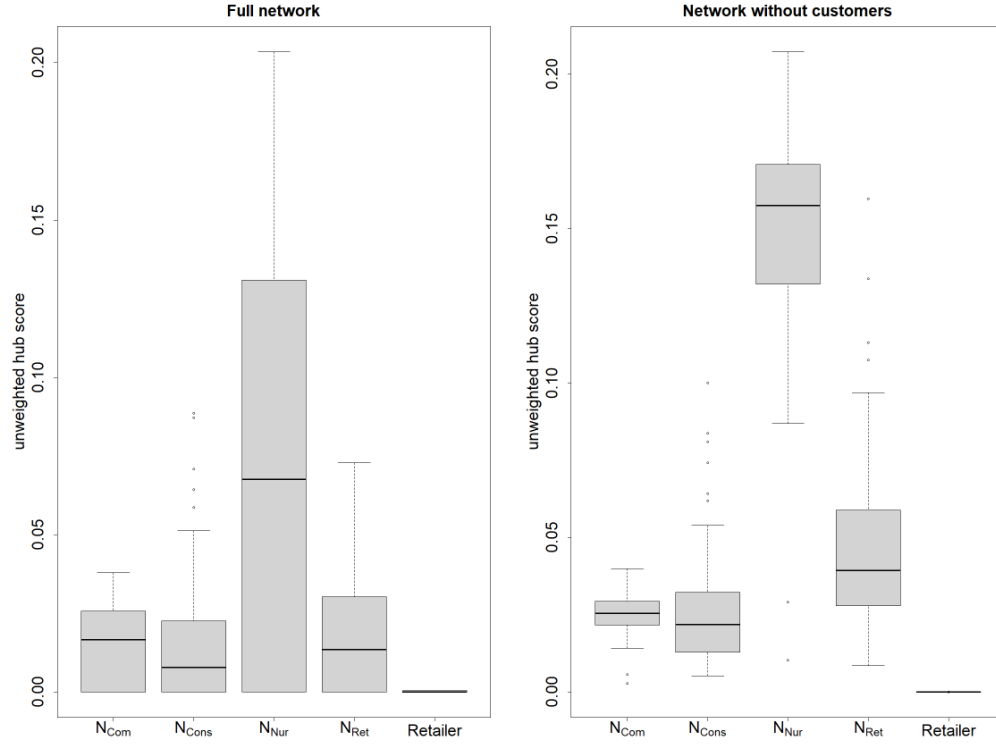
Figure B.53. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 5 $((|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 130, 10, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
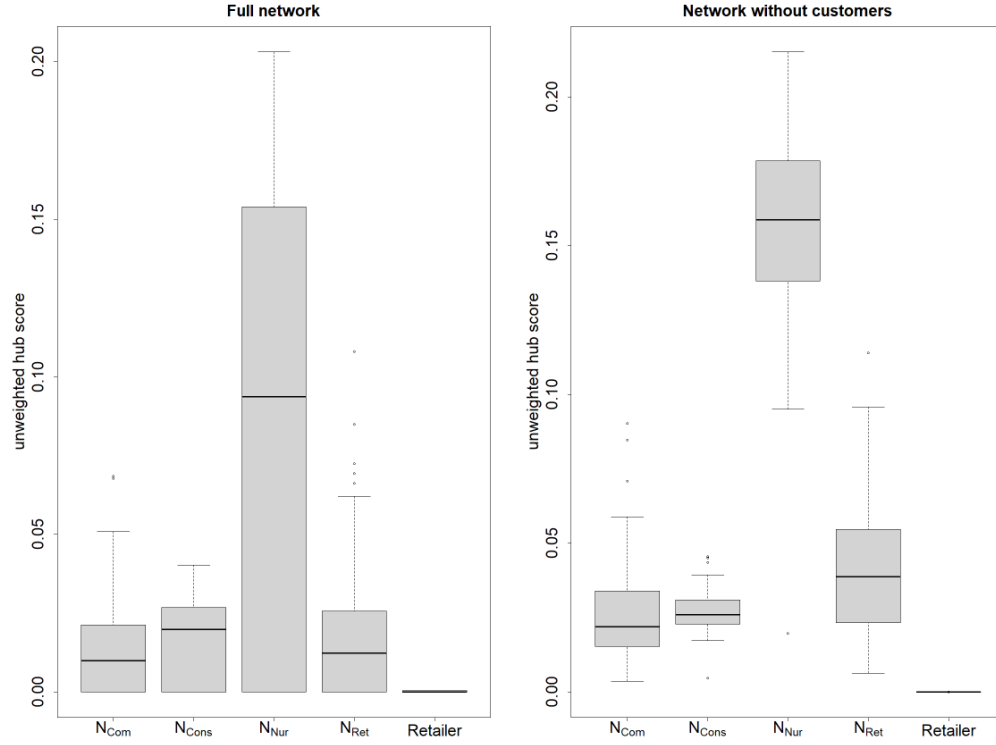
Figure B.54. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 6 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 130, 10)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.
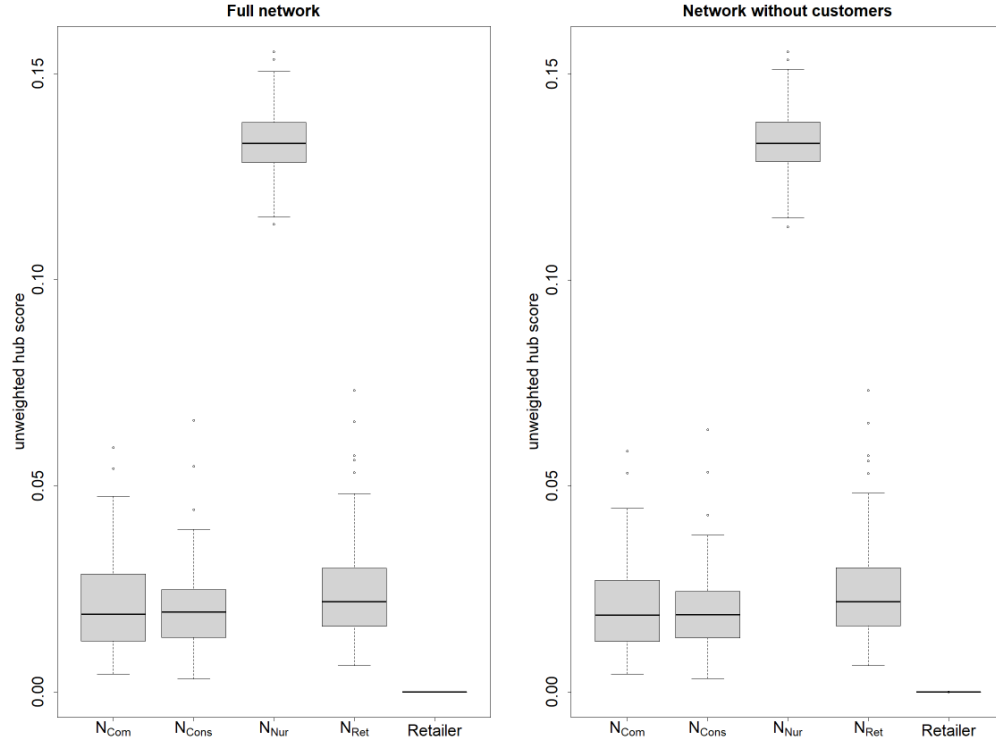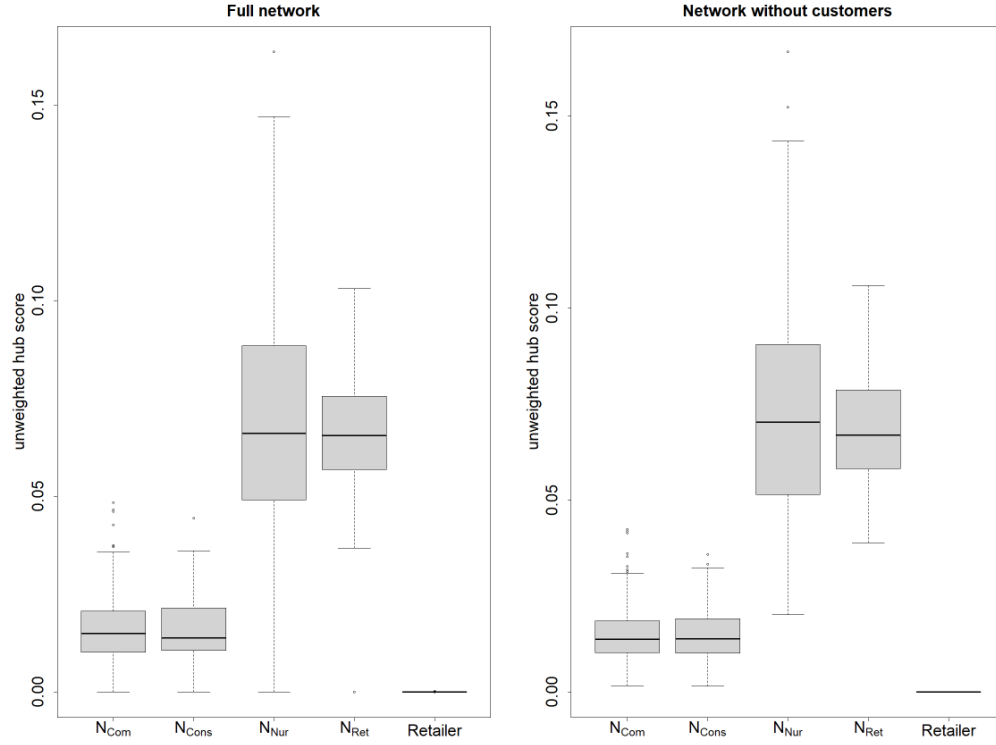
Figure B.55. Box plots of the average unweighted hub score per node subcategory for 100 networks with nursery distribution scenario 7 ($(|N_{\text{Com}}|, |N_{\text{Cons}}|, |N_{\text{Nur}}|, |N_{\text{Ret}}|) = (10, 10, 10, 130)$). All other parameters values used are shown in Table 1.2. Scores are shown on a $\log_{10}(1 + \text{data})$ scale. The plot on the left shows out-degree scores calculated on the entire network, the plot on the right shows out-degree scores calculated on the subset of the network with only nurseries and retailers. We note the differences in the scales of the y-axes.

Appendix C

COVID-19 model supplementary material

This section is relevant to Chapter 6 and includes the table of author contributions to clearly identify what each author contributed to the work. This section also includes the figure for our manual fitting of the parameters which control the timing of the fall in the reproduction rate due to lockdown for care homes and the general population (ω_{end}^C, ω_{end}^G). This figure shows the values we considered as baseline fall under a local minima in terms of squared error.

Table C.1. Description of the contribution roles of each author: Adam Kleczkowski (AK), Matthew Basiter (MB), Ewan McTaggart (EMT), Paul McMenemy (PM), Itamar Megiddo (IM).

| Contributor Role	Role Definition	Name
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	AK
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	MB, EMT
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	MB, EMT
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	AK
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	MB, EMT
Methodology	Development or design of methodology; creation of models	AK, PM, MB, EMT, IM
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AK, PM, IM, MB, EMT
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	AK, PM, MB, EMT
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	AK, PM, MB, EMT
Supervision	Oversight and leadership responsibility for the research activity planning and, execution including mentorship external to the core team.	AK, PM, IM
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	MB, EMT
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	MB, EMT
Writing- Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	MB, EMT
Writing- Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	AK, PM, IM, MB, EMT
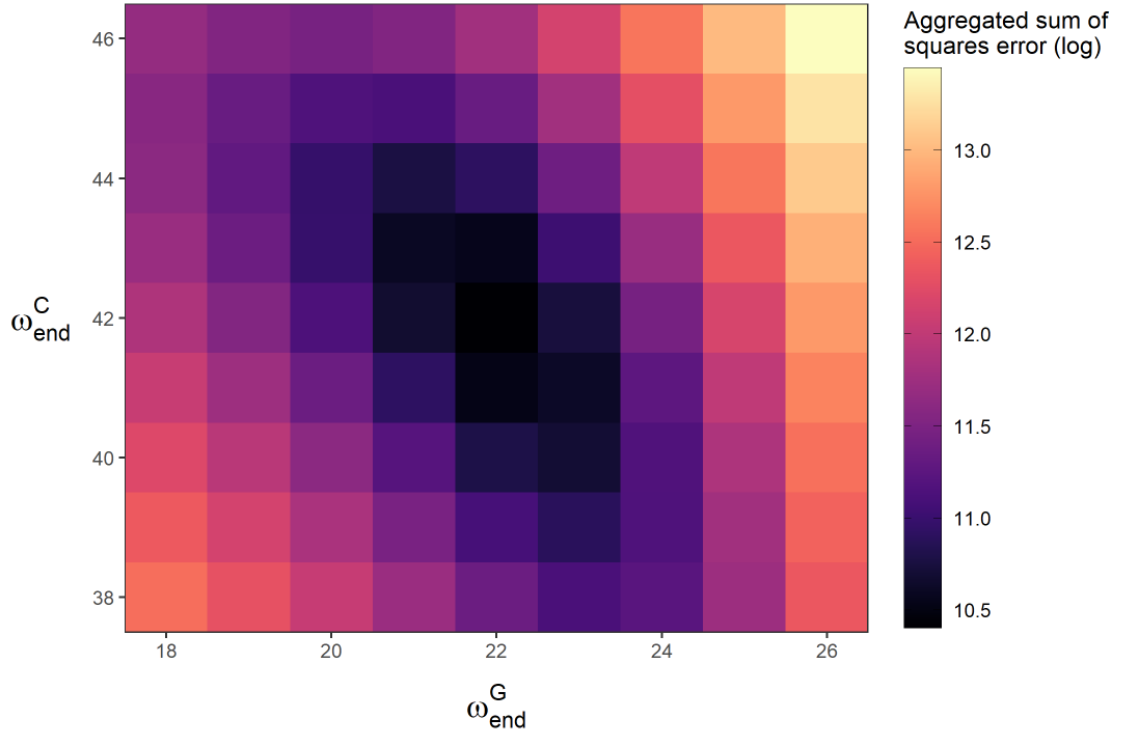
Figure C.1. Natural log of aggregated sum of squared error in a $\omega^C_{end} - \omega^G_{end}$
parameter space. ω^C_{end} controls the timing of the drop in resident R_t, and ω^G_{end} controls the
timing of the drop in general population R_t. These parameters were fitted manually, achieving a
minimum for the values shown in Table 5.1. In the plot, all other parameters are held at the base
case (Table 5.1).