

Using social network analysis to inform disease control interventions



Nelly Marquetoux^{a,*}, Mark A. Stevenson^b, Peter Wilson^c, Anne Ridler^c, Cord Heuer^a

^a EpiCentre, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand

^b Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Werribee, Victoria, Australia

^c Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand

ARTICLE INFO

Article history:

Received 30 June 2015

Received in revised form 21 January 2016

Accepted 25 January 2016

Keywords:

Network analysis

Livestock movements

Epidemiology

Infectious diseases dynamics

Basic reproduction ratio

Control strategies

ABSTRACT

Contact patterns between individuals are an important determinant for the spread of infectious diseases in populations. Social network analysis (SNA) describes contact patterns and thus indicates how infectious pathogens may be transmitted. Here we explore network characteristics that may inform the development of disease control programmes.

This study applies SNA methods to describe a livestock movement network of 180 farms in New Zealand from 2006 to 2010. We found that the number of contacts was overall consistent from year to year, while the choice of trading partners tended to vary. This livestock movement network illustrated how a small number of farms central to the network could play a potentially dominant role for the spread of infection in this population. However, fragmentation of the network could easily be achieved by “removing” a small proportion of farms serving as bridges between otherwise isolated clusters, thus decreasing the probability of large epidemics.

This is the first example of a comprehensive analysis of pastoral livestock movements in New Zealand. We conclude that, for our system, recording and exploiting livestock movements can contribute towards risk-based control strategies to prevent and monitor the introduction and the spread of infectious diseases in animal populations.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Movement of livestock between farms or markets is a strong determinant of the spread of transmissible pathogens in animal populations (Keeling and Eames 2005; House and Keeling, 2011). Detailed knowledge about movement of livestock can thus be a useful tool to inform control strategies. A classic example is provided by the foot-and-mouth disease (FMD) epidemics of 2001 in the UK. A few long-range movements, mostly via markets and dealers, spread the infection widely at an early stage (Kao 2002; Kao et al., 2006; Shirley and Rushton, 2005). Consequently, more stringent and accurate livestock tracing systems, in addition to mandatory movement restrictions, were implemented in the UK (Vernon, 2011). Analysis of animal movement can also provide a useful framework to study the spread of endemic diseases and has been extensively used for tuberculosis, both in cattle populations (Gilbert et al., 2005; Woolhouse et al., 2005) and wildlife (Corner et al., 2003; Drewe et al., 2011; Porphyre et al., 2008).

Landcorp farming limited (LC) is a state-owned enterprise, comprised of 122 farms¹ located throughout New Zealand, representing the regional variety of farm types of the country's pastoral livestock industry. New Zealand farming is characterized by all-year pastoral farming in which pasture availability drives the annual production cycle and different livestock species are often co-grazed on the same pasture. LC farms are typical of this farming system and most LC farms host multiple livestock species (cattle, sheep and/or deer—mostly red deer). LC keeps detailed records of the shipments of livestock off and onto their farms. Most movements occur between LC farms with a small proportion of movements to non-LC farming enterprises. The movement records of LC were unique in New Zealand, in that they provided a complete set of movement events over several years within a corporate group of farms representing a relatively closed population. We propose that this information could provide some insight into factors potentially influential in disease spread. A better understanding of these

* Corresponding author. Fax: +64 3505716.

E-mail address: nelly.marquetoux@gmail.com (N. Marquetoux).

¹ 122 farms as per 2012, this number is subject to annual changes and increased to 137 in 2014. The number of different LC properties involved in the network analysis is 112.

factors is a necessary first step in developing rational interventions to limit disease spread.

In this study we analysed the contact pattern arising from movements of sheep, cattle and deer to and from LC farms from 2006 to 2010. Our aim was to use social network analysis (SNA) to describe the trading pattern of LC farms and to understand how contact through trading might influence the spread of infectious diseases. A preliminary step was to assess the consistency of network characteristics over time, and a sequel was to discuss the effect of targeted control measures with respect to various aspects of disease spread in this particular network. This analysis will help to prioritize the allocation of resources to enhance biosecurity in this network of farms.

2. Material and methods

Definitions of the technical terms related to SNA used in this paper are provided in Table 1.

2.1. Movement data

The data available for analysis included all livestock movement records to and from the properties of LC for the period 1 July 2006–30 June 2010 (inclusive). LC farms primarily exchange livestock within the company, but some movements also involved properties outside LC. Livestock species involved in the movements were dairy cattle, beef cattle, sheep and deer. Two types of movements were recorded. A transfer was defined as the *permanent movement* of livestock from one property to another (equivalent to a sale). The format for a transfer was an annual summary record: for every pair of farms between which livestock transfers occurred in a given year, there was one summary record for each species moved corresponding to the sum of all the annual transfers for this species. Agistments were defined as *temporary movements* of livestock (equivalent to a lease). These occurred when animals were sent from one LC property to another – or to a property outside LC – for a limited period for grazing management, followed by another agistment movement record back to the property of origin. Detailed data for agistments were available, thus each agistment corresponded to an actual shipment of animals including the date on which the movement event occurred.

The LC movement database contained: the source and destination property; the type of movement (transfer or agistment); the date (year for transfers, calendar date for agistments); the species involved: sheep, beef, dairy or deer; and the total number of animals of each species moved in one year for transfers, or the actual batch size for agistment events.

Using these data, we constructed yearly networks consisting of all recorded livestock movements within the LC enterprise from the 01st July to the 30th of June and we described the contact pattern in terms of consistency over time, size, centrality measures and cohesion (see Table 1). We finally analyzed the characteristics of this contact pattern that were believed to be key for disease transmission.

2.2. Data analysis

The consistency of farm movement events was evaluated across successive years. All subsequent analyses were carried out using data for the year 2009–2010 only, as movement records were most complete for this year. The recorded data represented a census of the movements to/from the LC farms. However, some non-LC farms traded with LC farms, thus acting as satellites of this network. For these commercial farms, only the movements to/from LC farms were known while movements to/from other commercial farms were not included in the data. To avoid biases, all the

Table 1

Definitions of social network terms used in this study.

Parameter	Definition
General terms	
Directed path	The pathway between nodes (farms) accounting for the direction of the contacts (i.e. livestock movements). A movement from nodes A to B or a movement from nodes B to A thus defines two different pathways.
Undirected path	The pathway between nodes (farms) ignoring the direction of the contacts. A movement from nodes A to B or a movement from nodes B to A thus defines the same pathway.
Measures of centrality	
In-, out-degree	The number of contacts to or from a node (farm), respectively, during a defined period. In-degree is potentially positively correlated with the probability of introduction of infectious agents. Out-degree is potentially positively correlated with the probability of spreading infection.
Weighted in-, out-degree	The weighted in- and out-degree were defined as the total number of animals (as opposed to the total number of contacts) received or sent by a farm, respectively, during a defined period.
Betweenness	The frequency by which a node falls between pairs of other nodes on the shortest path connecting them (Freeman, 1978). Betweenness is a measure of centrality used to quantify a node's potential to 'control' the flow, or curtail paths within a network.
Measures of cohesion:	
Clustering coefficient	Clustering coefficient (CC) can be either a local or a global network attribute. In this study CC is expressed as a global measure, corresponding to the probability that any two nodes <i>j</i> and <i>k</i> are connected to a node <i>i</i> and nodes <i>j</i> and <i>k</i> are in turn connected to each other (Kiss et al., 2006). As a global measure CC quantifies 'cliquishness' within the network (Watts and Strogatz, 1998).
Strongly connected component	The section of a network where any node could be reached from any other node by following the direction of existing paths (Christley et al., 2005).
Weakly connected component	The section of a network where all nodes are linked to each other irrespective of the direction of the path (Christley et al., 2005).

movements were used to calculate farm-level network properties (such as degree or betweenness) but only measures for the LC farms were reported. Similarly, network-level statistics (degree distribution, standard deviation and average degree, degree correlations and effect of targeted control) were calculated using only the LC farms. Since the non-LC farms contributed to the overall connectivity, they were kept to calculate measures of network cohesion. The analyses were performed using the software package Pajek for SNA (Batagelj and Mrvar, 2008), and the *igraph* package (Csárdi and Nepusz, 2006) within R (R Development Core Team, 2014). Networks were plotted using Gephi (Bastian et al., 2009).

2.2.1. Consistency of the contact pattern over time

Networks are dynamic structures, therefore evaluating the consistency (or the lack of it) of global and individual network properties can reveal important evolutions in the network topology (Kossinets and Watts, 2006; Robinson et al., 2007). Year to year

consistency was first assessed in terms of the number of contacts of each type (transfers and agistments). In addition, we analysed the consistency of the pair-wise relationships between farms. To do this, we identified each unique pair of farms (two farms with at least one movement occurring between them) and how many years this link existed between 2006 and 2010.

To assess the consistency over the years of the identity of the more central farms in the network, according to different centrality measures, farms were ranked by their in- or out-degree (weighted or not) and their betweenness for each of the four years. The top 10% farms (or 'hubs') according to each centrality measure were then identified for each year (i.e. the 12 top farms in each year). We then calculated the proportion of farms among those 48 top farms that were present at least three years out of four. This proportion was called a similarity index and was used to compare how similar across years the hub farms were, according to different centrality measures. In addition, we computed the probability of the estimated similarity index if the rank of the farms followed a random order using a Fisher test (corresponding to a p -value).

2.2.2. Description of the degree distribution

A common feature of many real-world networks is the heterogeneity in patterns of contact (Barabási and Albert, 1999; Albert and Barabási, 2002), whereupon small numbers of nodes are highly connected whereas most other nodes in the network have relatively few connections. Heterogeneity in the number of contacts can have a strong influence on the spread and persistence of an infectious disease (Hethcote, 1978) and infection can spread more readily via hub nodes. It is therefore important to examine the distribution of the number of contacts per node to determine where the heterogeneity lies, between the assumption of homogenous mixing and that of pure scale-free connectivity, when the degree distribution follows a power law (Bansal et al., 2007).

We used standard statistical methods combining maximum-likelihood (ML) and goodness-of-fit (gof) tests to fit a parametric probability distribution to the degree distribution data (number of contacts per farm) for the 107 LC farms involved in the network in 2009–10.

We assessed the observed data for consistency with the power law distribution. The guidelines described by Clauset et al. (2009) provided the framework, while the implementation was carried using the functions `plfit` and `plpva`². The `plfit` function computes a ML estimate for the power-law exponent and determines the best cut-off value for a power-law tail in the observed distribution with a Kolmogorov–Smirnov test. The `plpva` function then computes a p -value to test the plausibility of a power-law fit for the observed data, by comparing the observed gof test to that obtained with n random samples from a true power-law distribution. Hence a high p -value indicates that a large proportion of the power-law samples fluctuates “further away” from the model than the observed data, which is unlikely to occur by chance and indicates that the power-law model fit is plausible for this data, see (Clauset et al., 2009). The `plpva` function allowed adjustment for finite-size bias (small sample size) and the p -value for the power law tail was computed over 1000 non-parametric repetitions of the fitting procedure.

When attempting to fit a power-law to a distribution, it is good practice to check if the data could also be fitted with other skewed distributions (Clauset et al., 2009). The package `fitdistrplus` implemented within R (Delignette-Muller et al., 2010) was used to assess the best fit from the following empirical distributions: gamma, Weibull, exponential and log normal.

2.2.3. Small-world properties of the network and overall connectivity

Many networks found in nature, unlike random networks, are highly structured and display strong clustering (or clique behavior), meaning that two connected nodes are likely to share social ties with a common third node, forming a triangle of inter-connected nodes. This property is referred to as ‘small-world’ (Watts and Strogatz, 1998). The tendency to form social (or regional) groups with more contacts within than between tends to enhance transmission within clusters, making it more difficult for pathogens to overcome local structures and spread globally. This can result in smaller effective reproduction rates and smaller final epidemic (Cross et al., 2005; Keeling, 1999). However the presence of a few long-range connections via hub-nodes, acting like shortcuts between communities, can result in a fast dissemination of pathogens through all parts of the network and increase the likelihood of rapid disease spread (Keeling and Eames, 2005).

We assessed the small-world properties of the network of LC farms in 2009–2010 by calculating the average path length between connected LC farms and the clustering coefficient and by comparing these values with those obtained from 10000 simulated random Erdos–Renyi network of the same size (Watts and Strogatz, 1998). We assessed the overall connectivity of the network by determining the size of the largest or ‘giant’ weakly connected component (GWCC) and of the largest, giant strongly connected component (GSCC).

2.2.4. Effect of network properties on the basic reproduction number (R_0)

At the farm level, the basic reproduction rate (R_0) is the expected number of farms infected by a first infected farm during its infectious period in a fully susceptible, homogenous population at equilibrium (Anderson and May, 1979; Diekmann et al., 1990). In homogenous mixing populations, $R_0 > 1$ classically represents the threshold for pathogen invasion (Anderson and May, 1991; Diekmann et al., 1990). Heterogeneity in the contact pattern may enhance the transmission of infection throughout a network (Pastor-Satorras and Vespignani, 2001); however this partly depends on the correlation between in- and out-degree (Kiss et al., 2006; Woolhouse et al., 2005). In the absence of correlations between the in-degree and the out-degree in a directed network, the heterogeneous structure of the contact pattern (that is, the presence of hubs) is unlikely to have an effect on the value of R_0 (Woolhouse et al., 2005). A positive correlation between in-degree and out-degree increases the value of R_0 at a given infectious rate of a pathogen whereas a negative correlation decreases it (Woolhouse et al., 2005; Bansal et al., 2007). These network properties should be accounted for in calculations of R_0 .

We used the method of Volkova et al. (2010) to estimate the basic reproduction number for the network, taking into account the variance and covariance in contact rates, as follows:

$$R_{0(\text{network})} = \rho_0 \times \sqrt{(m_{\text{in}} \times m_{\text{out}}) + (SD_{\text{in}} \times SD_{\text{out}} \times r_{\text{in-out}})} \quad (1)$$

In Eq. (1) ρ_0 is an unknown constant depending on pathogen specific virulence characteristics such as the probability of infection given contact or the duration of the infectious period; m_{in} (m_{out}) is the average in-degree (resp. out-degree), SD_{in} (SD_{out}) is the standard deviation of the in-degree (resp. out-degree) distribution and $r_{\text{in-out}}$ is the Pearson correlation coefficient between the in- and the out-degree. The in- and out-degree were calculated using all the movements since they represented a census of LC movements; however the mean, standard deviation and correlations were calculated after removing the non-LC farms for which degree calculations were biased (but keeping the edges between those and the LC farms). By using the Pearson correlation coefficient in this equation, the assumption was that only the linear part of the

² available on the webpage <http://tuvalu.santafe.edu/~aaronc/powerlaws/>.

correlations had an impact on disease spread. In a homogenous-mixing population of farms, there would be no variability in the contact rate between farms; hence the standard deviations would be zero. In this case, according to Eq. (1), the expression for R_0 reduces to:

$$R_{0(\text{homogenous})} = \rho_0 \times \sqrt{(m_{\text{in}} \times m_{\text{out}})} \quad (2)$$

It is noteworthy that by aggregating farm contacts in a static yearly network, our approach would more adequately address chronic diseases, while overestimating the transmission of infections such as FMD. Nonetheless the principle of this analysis was not specific to any given pathogen, hence we were not interested in the absolute value of R_0 *per se*, rather in the relative increase that could be attributed to the presence of highly connected farms. Thus we assumed that ρ_0 was constant and calculated a relative value defined as $R_{0(\text{network})}/R_{0(\text{homogenous})}$, to evaluate the multiplicative impact of heterogeneity and correlations in the contact pattern in 2009–2010 on the value of R_0 (Volkova et al., 2010).

2.2.5. Efficacy of targeted control strategies to disrupt the transmission via livestock movements

The farms potentially playing a key role in the transmission of infectious agents via livestock movements could be of two kinds:

- farms with high degree: high degree farms are more likely to exchange livestock with a number of different trading partners. They therefore could be more at risk of becoming infected and/or to pass the infection on;
- farms with high betweenness; high betweenness farms act as links between inter-connected cliques, acting as short-cuts and resulting in a 'small-world', highly navigable network. This can favor epidemics of large size involving most nodes in the network, due to percolation behavior in connectivity.

We first wanted to explore whether a highly connected farm was also more likely to control the flow between clusters of inter-connected farms. We therefore examined the correlation between total degree and betweenness for the LC farms.

Next, we simulated a targeted ban of movements to/from the farms within the LC network. The list of all the 107 LC farm identifiers involved in the network in 2009–10 was sorted in descending order of total degree. One-hundred-and-six simulations (n-1) were then carried out wherein for each simulation the LC farm with the highest total degree was removed from the network, simulating a movement ban to/from this farm. The effect of farm removal was then quantified in terms of: (1) the size of the remaining largest GSCC (which provided an estimate of the lower bound of potential final epidemic size); and (2) $R_{0(\text{reduced network})}/R_{0(\text{full network})}$, where the reduced network was the result of a targeted farm removal. Farm removal was cumulative: once removed, a farm was not eligible anymore and the next highest-degree farm was determined in the next step. The value of R_0 at each step, called $R_{0(\text{reduced network})}$, was calculated using Eq. (1) (with average in- and out-degree being calculated after excluding the non-LC farms for which the complete set of movements was unknown, to avoid bias). $R_{0(\text{reduced network})}/R_{0(\text{full network})}$ represented the proportion of the contribution of the movement pattern to R_0 after the movement ban for a given number of target farms. A second set of simulations were then carried out where farm removal was on the basis of betweenness (instead of total degree).

To evaluate the specific effect of a targeted control of movements towards key-farms, we compared these results with those obtained with a simulated ban of movements to/from randomly chosen farms. To do this we simulated the removal of farms in 100 randomly generated orders and computed the size of the largest GSCC and $R_{0(\text{reduced network})}/R_{0(\text{full network})}$ on each occasion.

■ pairs only connected between 2006 and 2009
■ pairs also connected in 2009–10

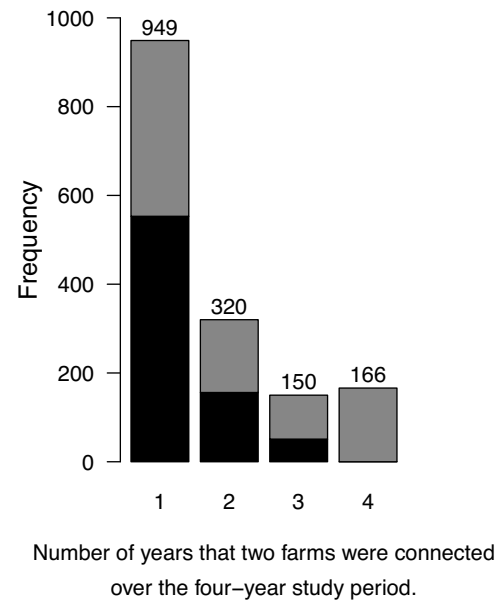


Fig. 1. Frequency of pairs of farms being connected in only one year or repeatedly in multiple years (2006–2010).

The mean and 95% quantiles for both quantities were then calculated, representing the effect of a random ban of movements on the potential of disease spread throughout the network.

3. Results

During the four-year study period there were 3,531 movement events, involving a total of 180 farms (112 LC farms and 68 non-LC farms in total). Most movements occurred between LC farms; 312 (9%) occurred between a LC farm and a farm outside LC.

3.1. Consistency of the contact pattern over time

The frequency of movements of each type (transfers and agistments) or each species was consistent over the first 3 years of the study period. In year 4 (2009–2010) the total number of movements increased by a factor of 1.8 from the value recorded for 2008–2009 (see Table 2). The change was due to an increase in the number of recorded agistments (temporary movements), particularly for dairy cows for which it was common practice to graze and mate heifers outside the property of origin before the first calving. This apparent increase in the frequency of agistments in the final year was due to a change in LC's recording policy: unlike transfers, most agistments were not accurately recorded before 2009. The pattern of transfers was consistent throughout the observation period. The data for 2009–2010 were therefore considered to be the most complete and robust of the LC network dataset. During 2009–2010, the average distance traveled by livestock in this network was 150 kilometers (median: 71 km, range: 0.8–1091 km).

The frequency with which a pair of farms was found to be connected throughout the study period (from one year only, to all four years) is shown in Fig. 1. Only 20% of the directed movements connecting two farms were repeated in at least three years out of four between 2006 and 2010, while 60% of the pairs occurred only once. This low frequency of pairs consistently trading with each other over the years was not due to the lack of reporting of some

Table 2
Total number of livestock movements between farms and number of animals of each species moved (in parentheses) in four years.

	Agistments	Transfers	Total contacts	Sheep	Beef	Deer	Dairy
2006/7	62	708	770	266 (214,295)	258 (23,414)	49 (15,076)	197 (17,147)
2007/8	45	719	764	282 (232,422)	232 (24,052)	52 (17,767)	198 (19,860)
2008/9	30	682	712	234 (201,555)	216 (24,125)	55 (13,724)	207 (24,159)
2009/10	586	699	1285	310 (216,689)	261 (25,544)	62 (15,457)	652 (63,660)
Total years	723	2808	3531	1092 (864,961)	967 (97,135)	218 (62,024)	1254 (124,826)

Table 3
Similarity between years 1– 4 for the top 10% ranked LC farms according to one measure of the degree.

Outcome	Similarity % ^a	p-value
In-degree	48% (23/48)	p < 0.0001
Out-degree	65% (31/48)	p < 0.0001
In-degree (weighted)	71% (34/48)	p < 0.0001
Out-degree (weighted)	63% (30/48)	p < 0.0001
Betweenness	40% (19/48)	p < 0.01

^a Percentage of LC farms that are present at least 3 years out of the four-year study period among the 10% highest ranked, according to different centrality measures.

movements before 2009, as indicated by the stratification by 2006–2009 and 2009–2010 (Fig. 1). This shows that trading partners tended to vary from one year to the next in the network of LC farms, based on the 4 years of data available for analysis.

Farms with the highest centrality scores in a given year tended to have similar scores in subsequent years (Table 3), with 48–71% of farms being among the top 10% highest degree farms in at least three years during the study period. For betweenness centrality however, only 40% of the farms were consistent hubs.

3.2. Description of the degree distribution

The methodology proposed by Clauset et al. (2009) was used to assess the presence of a power law tail for the degree distribution (Fig. 2). For the total degree distribution the best cut-off for a power law tail in the empirical distribution was more than 19 contacts per farm, and the exponent of the power law was $\gamma = 3.09$. The gof test for the tail of the distribution (>19 contacts per farm) indicated that the power law was a plausible fit for these data (p -value = 0.854). With only 50 observations above 19 contacts/farm, the power to reject the assumption of a power-law fit was limited; nevertheless, the high p -value indicated that there was no evidence for lack of fit. Similarly, a power law fit could not be ruled out for the tail of the distribution of in-degree ($x_{\min} = 8$, $\gamma = 2.62$, $n = 47$ observations, p -value = 0.22) or for out-degree ($x_{\min} = 11$, $\gamma = 2.87$, $n = 39$ observations, p -value = 0.4).

Considering other skewed distributions, the lognormal distribution provided the best fit for both the in-degree and total degree and the Weibull distribution for the out-degree distribution, based on the lowest Akaike Information Criterion (among lognormal, gamma, Weibull, or exponential distributions). This did not indicate that these distributions appropriately fitted the observed distribution, but only that the other distributions provided a worse fit. Different gof statistics were performed to test the plausibility of the fit. The less sensitive of the four tests (Kolmogorov–Smirnov test) could not rule out the null hypothesis for the following distributions: lognormal fit for the total degree and in-degree, Weibull fit for the out-degree. However, the authors of the package indicated that that “this approximate test may be too conservative” (Delignette-Muller et al., 2010).

Table 4
Impact of the correlation between in- and out-degree in the LC farm network on R_0 , the estimated potential initial spread of epidemics in year 2009–2010.

	Un-weighted degree (number of contacts)	Weighted degree (number of animals moved)
m_{in}	10.8	2642.3
m_{out}	11.3	2903.2
SD_{in}	9.4	5062.8
SD_{out}	9.2	3585.5
$r_{\text{in-out}}$ [95% CI]	0.56 [0.42–0.68]	0.01 [−0.18–0.20]
$R_0(\text{network})/R_0(\text{homogenous})$ [95% CI]	1.18 [1.14–1.22]	1.02 [0.76–1.22]

3.3. Small-world properties of the network and overall connectivity

The network of LC farms during the commercial year 2009–2010 had a diameter of nine. This means that there existed a directed pathway between any two pairs of farms of the network, with no more than nine steps to reach one farm from another. The average path length between any two pairs of farms was of 3.9 steps and the clustering coefficient was 0.34. By way of comparison, the 95% quantiles of the distribution of clustering coefficients obtained with 10,000 simulated Erdos–Renyi random networks comprising 164 nodes and 1285 ties were between 0.088 and 0.1. Hence, the observed clustering in the LC network was about 4-fold stronger than that of a random network. The short path length between farms observed in the network of LC farms, associated with high clustering, is characteristic of a small-world network. Graphical representation of the network (Fig. 4) clearly shows the clustering of farms into two groups corresponding to North Island and South Island farms. Livestock movements occurred preferentially within the same island (97% of movements), but the North and South Island farms were interconnected by small numbers of trans-island movement events (3%) that acted as short-cuts between the two main clusters. Thus, despite high clustering and even topographic separation into two islands, the network of LC farms showed a high level of overall cohesiveness, as assessed by the size of the interconnected components of the network (Table 5). In 2009–2010, 100% of the total number of farms was weakly interconnected, meaning that an (undirected) pathway existed between all pairs of farms, while 79% of farms were strongly connected (Table 5).

3.4. Effect of network properties on the basic reproduction number (R_0)

Estimates of the increase in the magnitude of R_0 that could be attributed to the heterogeneity of the contact network are shown in Table 4.

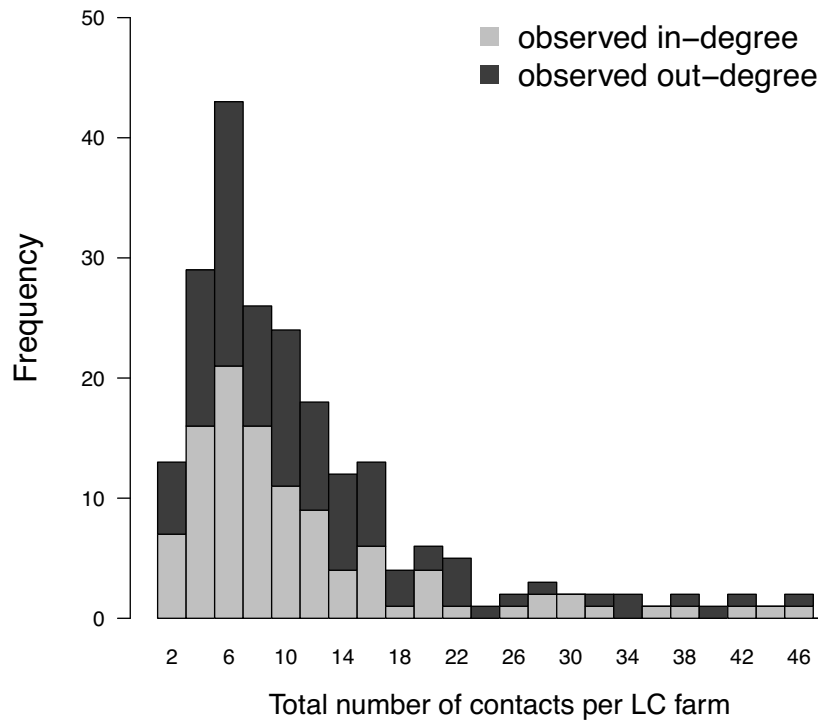


Fig. 2. Un-weighted degree distribution for the 107 LC farms involved in the network in 2009–10.

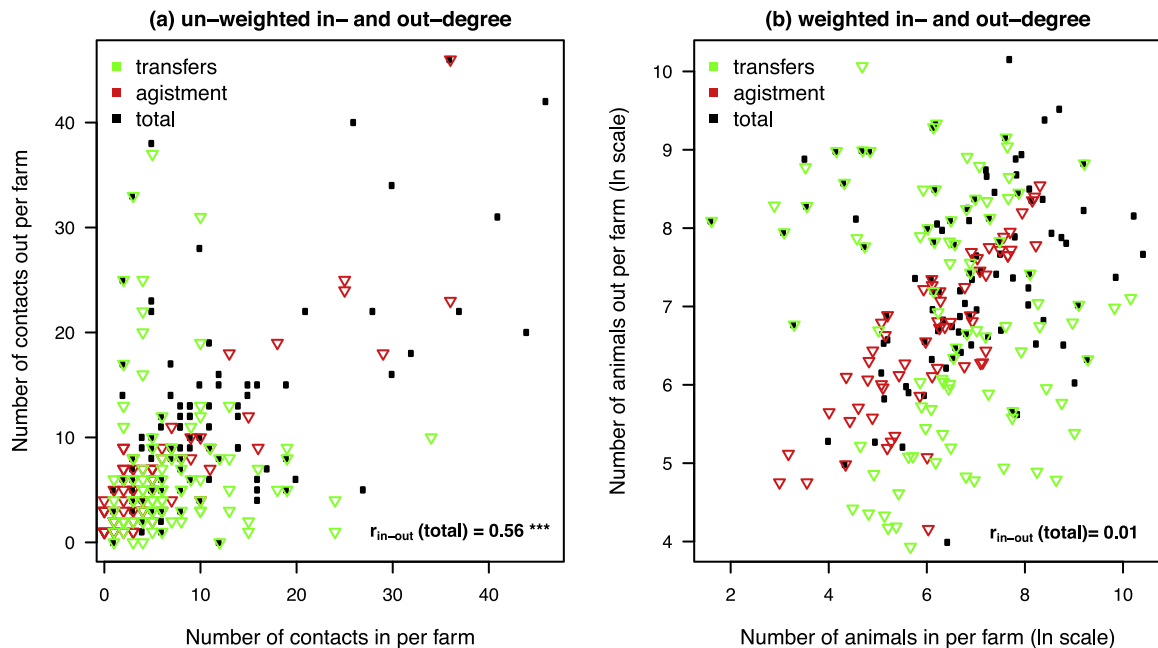


Fig. 3. Correlations observed in the LC network in 2009–10 between the number of contacts in and out per farm (un-weighted degree, a) or the number of animals in and out per farm (weighted degree, b), respectively for transfers (green), agistments (red) and the total of both movements (black). r_{in-out} corresponds to the Pearson correlation coefficient between the total in- and out-degree. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

Size of the giant strong component (GSC) and giant weak component (GWC) in the network, during the fourth year of the study period.

	2009–10
Total number of farms	164
Size of GSC	129
% Network	79%
Size of GWC	164
% Network	100%

Pearson's correlation coefficient for the un-weighted in- and out-degree was strongly positive (Table 4), indicating a linear correlation between the numbers of contacts in and out of farms in 2009–2010. As can be seen in Fig. 3, this correlation was mostly due to the agistments. This type of movement corresponded to a temporary lease of unproductive livestock (typically dairy heifers or replacement lamb ewes) to a farm in a different region than the farm of origin, according to seasonal and regional variations of pasture availability, followed by a return in the farm of origin.

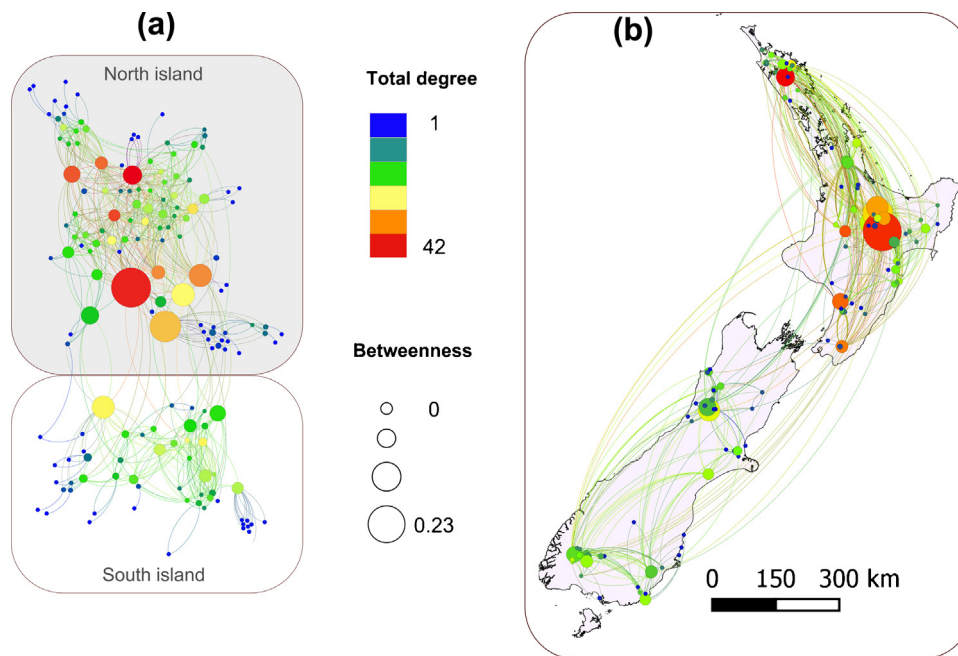


Fig. 4. Representation of the network of LC farms in 2009–2010, showing the importance of high-degree farms (in red) and high-betweenness farms (big size) in the contact structure. Each farm is represented as a node, the links between nodes are livestock movements. (a) projection according to the contact pattern only, not the geographical space, although the two main clusters correspond to farms of the North Island versus farms of the South Island, (b) contact pattern projected on a map of New Zealand according to the spatial coordinates of farms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Given this correlation between the in- and the out-degree and the variance in the degree distribution of the observed contact pattern, the initial spread was estimated to be 118% of the spread that would occur in a homogenous mixing population with similar average degree (Table 4). Our inference is that the presence of highly connected hubs is likely to enhance the potential for pathogen spread by farms that both send and receive large number of contacts, phenomenon amplified by the inherent features of agistments that increased correlations. For the animal-weighted degree the Pearson coefficient of 0.01 was not significant, hence there was no overall linear association between the number of animals received or sent per farm, despite an apparent high correlation when considering agistments only (Fig. 3). Therefore, the high variance observed in the contact pattern (see standard deviation in Table 4) did not significantly contribute to increase the value of R_0 when movements were expressed in number of individual animals moved, according to Eq. (1) (Table 4).

3.5. Efficacy of targeted control strategies to disrupt the transmission via livestock movements

The total degree and the betweenness of LC farms were highly positively correlated (Spearman's rank correlation coefficient 0.7, $p < 0.0001$). Thus farms with a number of contacts well above the average also tended to control the flow by being on the path in between many pairs of farms (see Fig. 4). These central farms tended to link different communities of farms (results not shown), thus forming one giant interconnected component as can be seen in Fig. 4. In particular, high degree/betweenness farms often received/sent livestock across islands, thus linking together the cluster of the North Island farms and the cluster of the South Island farms.

As expected from the high correlation between degree and betweenness, a ban on the movements to and from both high-degree or high-betweenness farms would be an effective measure for decreasing the potential of disease spread throughout the network (Fig. 5). More specifically, the removal of farms with the

highest betweenness scores would be the most efficient way to decrease the size of the largest GSCC of the network. After the removal of just 10% of farms with the highest betweenness scores (16 of 164 farms in 2009–2010), the size of the largest GSCC was decreased from 129 farms to just 17 farms whereas the number of different strong components in the network increased from 27 to 71. Thus, the targeted removal of bridge-farms efficiently disrupted network connectivity by isolating clusters of farms. On the other hand, the proportion of the contribution of the movement pattern to the initial spread of infection ($R_{0(\text{reduced network})}/R_{0(\text{full network})}$) was most efficiently reduced by targeting the farms with the highest total degree. Removal of 10% of the highest-degree-farms resulted in a drop to 40% of the full network spreading potential.

These results show how by removing high betweenness farms the network is transformed into a set of unconnected, smaller components (that is, a highly fragmented network). Similar effects were obtained by targeting high degree farms, which could be expected considering those farms with high degree were also those with high betweenness (Fig. 4). In contrast, random removal of farms provided a relatively inefficient means to reduce the network connectivity and the spreading potential of infectious agents.

4. Discussion

In New Zealand, the National Animal Identification and Tracing Scheme (NAIT) was launched in July 2012. Until then, little data were available to conduct detailed studies of farm-to-farm movement of livestock in New Zealand. Prior to NAIT, studies of farm-to-farm movement of livestock, individuals and animal product were based on cross sectional questionnaires (Lockhart et al., 2010). Sanson (2005) conducted a prospective study of movement of livestock off farms in the North Island of New Zealand to establish realistic parameters to inform a simulation model of FMD. That study addressed all types of off-farm movements occurring during three-week time slots, including people, material and manure; less than 4% of the total number of reported movements were actually livestock.

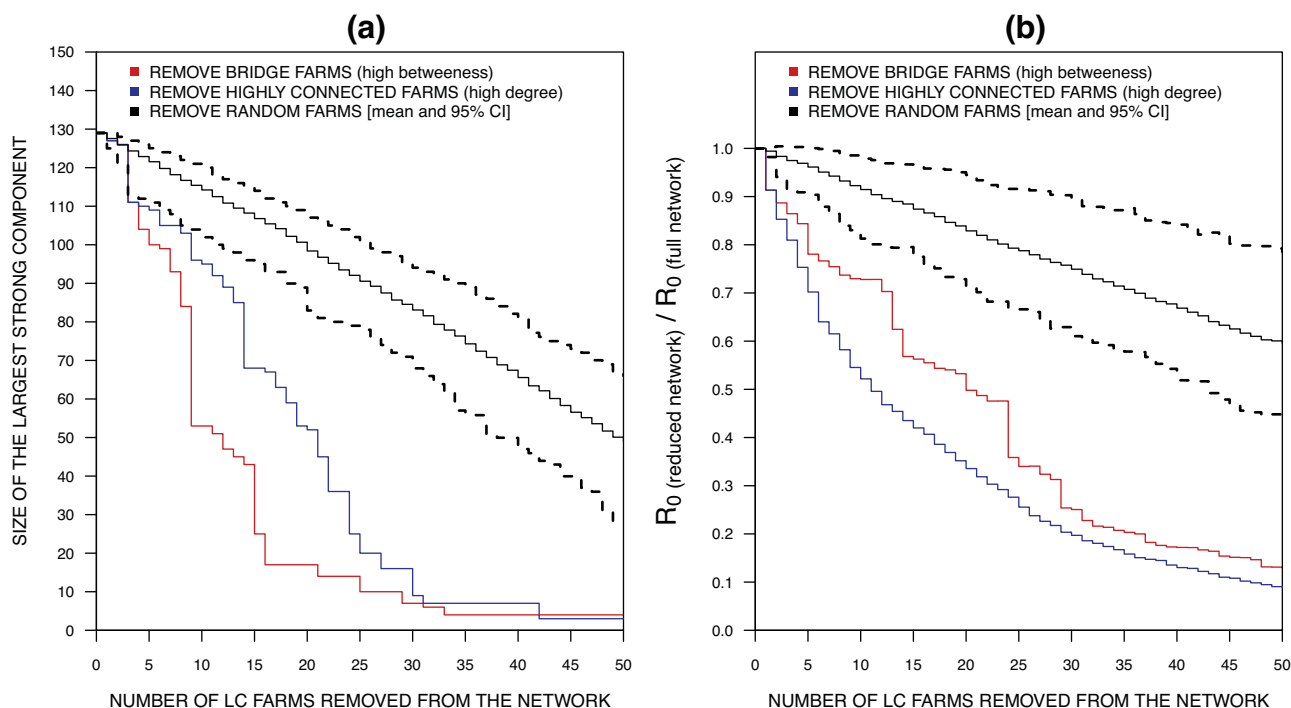


Fig. 5. Line plots showing the effect of targeted (in red and blue) or random (in black with 95% credible limits) removal of farms on two measures of potential spread: (a) showing the size of the largest giant strongly connected component as a function of the number of farms removed, (b): showing the value of $R_0(\text{reduced network})/R_0(\text{full network})$ as a function of the number of farms removed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the absence of a national livestock tracing scheme, the study of (Sanson, 2005) represented the only realistic way of quantifying farm movement patterns in New Zealand, but with key drawbacks. Firstly, it provided little insight into contact patterns for all livestock premises throughout the country. Secondly, it was a one-off cross-sectional study from which the consistency of the documented contact patterns over time could not be evaluated. Accurate recording of livestock movement data could only arise, in the absence of a mandatory system, from recording the source and destination of all animal shipments moved onto and off a given farm enterprise, the number of animals moved and movement date, similar to the system implemented in this study by LC. The recorded data represented a census of the movements to/from the LC farms. However, LC farms also traded with commercial farms outside LC; these were not central but rather behaved like satellites of this network. The presence of movements to farms outside LC meant that the LC network would be embedded into a larger network of NZ farms rather than isolated. The focus of the study was limited to the LC farms but the effect of links to commercial farms outside LC notably on the spread of infectious diseases should not be ignored from a biosecurity perspective. An overall livestock network was considered in this study, encompassing movement of cattle, sheep and deer. This approach is suitable for infections shared by all Ruminant species. It is particularly relevant in the New Zealand context where different species are often co-grazed on the same pastures, enhancing the possibility of inter-species pathogen transmission.

The contact pattern presented in this study is unlikely to be representative of all livestock enterprises in New Zealand. However, these analyses provide an indication of how a large corporate farming enterprise can derive value from routinely recorded movement data to identify individual farms at risk for contagion

4.1. Consistency of the contact pattern over time

We analysed the consistency of the pairs of farms exchanging livestock together in successive years. Most pairs of farms (60%) had been exchanging livestock for only one of the four years of the study period. This effect was not affected by the sub-optimal reporting of agistments observed before 2009 since the same pattern was observed when only data for the first three years of the study were analysed (Fig. 1). We conclude that a characteristic feature of this network was that farms tended to interact with different partners (other farms) in subsequent years. It is possible that over a longer period of time (more than four years) consistency of farm-to-farm trading relationships would, if they were present, become more apparent.

While overall network properties in the LC network (such as scale-free, small-world properties) remained similar from one year to the next (results not shown), the properties of individual nodes (farms) could still vary from year to year (Kossinets and Watts, 2006). When evaluating the consistency of the hub's identity year after year, we found that 40 to 71% of the top 10% farms with the highest centrality measure were similar every year. Considering this level of similarity between years, it can be assumed that decisions about targeted control strategies based upon the contact pattern observed in previous years would still be effective in future years, although sub-optimal. Based on a one-year time frame to measure contacts, (Volkova et al., 2010) analysed movements of sheep in Scotland and showed that the top 20% farms contributing the most to the spread of infectious diseases varied from year to year, similar to the LC network. Hence the effect of targeting the 'hubs' for control strategies based on the previous year contact network instead of real-time data consistently showed a reduced efficacy. Regular up-dates of farm-to-farm contact patterns are recommended to inform biosecurity measures, since we observed in the LC network significant variations in the efficacy of control strategies for small changes in the identity of the targeted farms

(results not shown). When working at the national level the high volumes of data and the computational power can be a limitation in this respect (Martínez-López et al., 2009).

Our analyses were on a one-year basis rather than aggregated for the whole study period. This represented a compromise in terms of time-scale aggregation, since livestock movements occur daily, but not all movement dates were available for analysis. As commercial ties between farms were not constant over the entire study period, aggregation of network data could result in an artificial co-existence of links that did not actually co-exist temporally. For this reason it was postulated that static networks appear to be constantly inaccurate for making predictions through epidemic simulations, as they fail to capture the intermittence of the connections between farms, thus artificially increasing the density and clustering of contacts (Vernon and Keeling, 2009). In a study of cattle movement patterns in the UK (Vernon and Keeling, 2009), a fully dynamic network was the only type to render appropriate temporal correlations in cattle movement patterns. The inaccuracy of predictions about final epidemic size arising from the representation of evolving networks as static would however, be minimal for diseases with low transmission probability and long infectious periods (Cross et al., 2005; Vernon and Keeling 2009), such as mycobacterial infections.

4.2. Description of the degree distribution

In veterinary epidemiology, most social network studies claim 'scale-free properties' based on the fact that degree distributions are skewed (Aznar et al., 2011; Kiss et al., 2006; Lockhart et al., 2010; Porphyre et al., 2008). A visual method to 'test' this hypothesis consists in plotting the cumulative distribution of the number of contacts per node at the log-log scale. A tail falling on a straight line is a necessary condition for a degree distribution following a power law and is therefore commonly used as a criterion to assess the scale-free nature of a network (Aznar et al., 2011; Kiss et al., 2006; Porphyre et al., 2008). However, such conclusions can be 'substantially incorrect' since a variety of skewed distributions other than power law could also be fitted by a linear regression at a double-logarithmic scale (Clauset et al., 2009), and the assumptions underlying ordinary least square regression are violated (Jones and Handcock 2003). This visual evaluation of contact heterogeneity in networks is nevertheless sufficient for highlighting the preponderant role of 'hubs' in the network and for assessing the effect of risk-based movement control strategies as is often of interest in veterinary science (Woolhouse et al., 1997; Woolhouse et al., 2005).

However it can also be useful to unravel the mathematical properties of the observed contact pattern when attempting to model the spread of infectious disease. The theoretical law that best fits the empirical set of contacts can be incorporated as an approximation of the underlying heterogeneity of contacts in epidemic models (Keeling and Eames 2005; Bansal et al., 2007). Robust mathematical methods to assess the scale-free nature of networks are often lacking in the graph literature, casting doubts as to whether scale-free properties indeed always arise from real-world networks (Clauset et al., 2009; Li et al., 2005). Statistical methods to evaluate a power-law fit, combining maximum-likelihood fit and goodness-of-fit tests (Clauset et al., 2009), were presented here as an example; these methods should be preferred to specifically assess the fit with a power law. There is at least one previous example in the field of veterinary epidemiology where the above framework was implemented (Dorjee et al., 2013). When applied to the LC network data, this approach indicated that a power law tail was highly plausible for the degree distribution for the period 2009–2010. However, even robust methods such as this present an inherent lack of power for small datasets such as the LC network.

4.3. Small-world properties and connectivity

In most real-world networks, scale-free properties are associated with social clustering that also impacts pathogen spread. Similarly, the network of LC farms exhibited small-world properties, with a clustering coefficient approximately 4 times greater than expected in a random network of same size. We observed a specific and strong geographical clustering of contacts within islands (either North or South Island of New Zealand), with few trans-island contacts (less than 3% of the total volume of movements). In such structured networks, the navigability for pathogens can be limited to localised neighborhoods (Eguíluz and Klemm, 2002). Clustering or even fragmentation of the network can therefore prevent the occurrence of large epidemics despite values of R_0 greater than one (Cross et al., 2005), whereupon a group-level R^* describing pathogen invasion through structured groups could be a much better predictor for pandemics. In the network of livestock movements in Great Britain in 2003–2004, Kao et al. (2006) assessed the risk of spread of an incursion of FMD; they showed that clustering of contact protected against epidemic spread, with a large epidemic of FMD only possible when R_0 was greater than 4. Hence $R_0 > 1$ is a necessary, albeit not sufficient, condition for infection spread in a structured network such as the network of LC farms.

Complementary to knowledge about R_0 , the size of the largest GSCC in the network represents an estimate of the total number of farms that a pathogen could reach if introduced into the network (Christley et al., 2005; Kao et al., 2006; Kiss et al., 2006; Robinson et al., 2007). Along these lines, we looked at the size of the largest GSCC or GWCC to assess the overall connectivity. Depending on the year, the largest GSCC in the LC network comprised 79%–95% of the total number of farms in the network. This number represented the lower bound for the potential final epidemic size. Additionally, infection could spread to all the sink nodes receiving animals from (but not sending animals to) farms in the GSCC, thus reaching virtually all the farms in the LC network. Hence, despite the presence of community structures, the density of contacts as well as the presence of bridge-farms linking communities allowed for a very high overall connectivity in this network. Previous research additionally shows that large epidemics are more likely to occur in such a structured meta-population for chronic diseases than acute infections (Cross et al., 2005).

4.4. Effect of network properties on the basic reproduction number (R_0)

In homogenous-mixing populations, the value of R_0 depends primarily on the average number of contacts per node, so called *first order moment* of the network (May and Lloyd 2001). In skewed networks the heterogeneity in the number of contacts induces a second order relationship between nodes (May and Lloyd 2001; Volkova et al., 2010). This corresponds to a non-null variance and covariance in contact rates and modifies the threshold value of infectious rates above which pathogens are propagated (Bansal et al., 2007; Volkova et al., 2010). In a network in which farms with a high in-degree tend to have few off-farm movements (for example finishing farms), those farms –once infected– would be unlikely to spread infection to other farms. By contrast, we observed a strong correlation between in-degree and out-degree (un-weighted) for the LC network in 2009–2010. In that year, 46% of movements were 'agistments' which correspond to temporary exchanges of young stock, particularly dairy cattle, leased to graze in areas of the country with better pasture availability and later returned to their farm of origin. This practice, well exemplified by the LC corporation movement pattern, is typical of the pastoral system in New Zealand. It contributes greatly to the circulation of livestock in all parts of the

country; moreover, the inherent back-and forth nature of this type of movements increases the correlation between in and out degree, as observed for the LC network in our study. This in turn could favor the transmission of infectious diseases.

Methods to assess the epidemic invasion in a network include a range of simulation modeling procedures. Other, more tractable methods have also been described to estimate values of R_0 that take into account the variance and covariance in contact rates, so that a negative correlation in the degree would lead to a decrease in the value of R_0 (May and Lloyd 2001; Volkova et al., 2010). We used the approach of (Volkova et al., 2010) to evaluate the impact of the second order moment on the basic reproductive number (R_0), compared to what it would be in a network of same size with homogenous mixing (that is, a network in which the variance of the degree distribution was 0). In the paper by (Volkova et al., 2010) in- and out- degree correlations were examined for Scottish sheep farms during the period 2003–2007. The correlation between the number of contacts in and out was close to zero, whereas it appeared much stronger when taking into account the numbers of animals per batch (still weak on an absolute scale), ranging from 0.18 to 0.36. For the LC network this relationship was reversed. The number of contacts in and out (in-degree and out-degree) were strongly positively correlated (Pearson's $\rho=0.56$ [0.42–0.68]), whereas correlations between the number of animals imported and exported per farm (weighted in-degree and out-degree) were not correlated. According to (Volkova et al., 2010), different scales used to define contacts could apply to different diseases, with unweighted contact more appropriate to study the spread of highly contagious diseases (such as FMD) and weighted contact for diseases with low intra-herd prevalence (such as mycobacterial infections). For the LC network, the presence of few highly connected farms could contribute to enhance the initial spread of a pathogen by 20% (un-weighted model), which could be a cause of concern in the case of an incursion, such as a FMD epidemic. Our analyses confirm that the way the contacts were weighted had a strong influence on the inferences.

We calculated relative values of R_0 . Absolute values for R_0 could not be computed and would be meaningless, since they depend upon pathogen properties such as the duration of the infectious period and the probability of infection given contact.

Finally, it should be noted that even higher order relationships ignored in this study, such as assortative or dis-assortative mixing also have an impact on disease spread in networks (Kiss et al., 2006).

4.5. Efficacy of targeted control strategies to disrupt the transmission via livestock movements

While scale-free properties are not easily defined, the vulnerability to targeted removal of hub-nodes represents the most prominent feature of scale-free networks (Li et al., 2005). This phenomenon is known in statistics as the Pareto principle. Applied to infectious diseases it means that “20% of the host population contributes to at least 80% of the net transmission potential, as measured by the basic reproduction number R_0 ” (Woolhouse et al., 1997).

We explored the impact of removing farms from the LC network that ranked highly for either their total degree (total number of contacts per farm) or their betweenness centrality. Farms with high betweenness had a strong tendency to also display a high total degree, thus the removal of farms according to their betweenness or to their degree both had a strong effect on the potential spread of infection. By contrast, the removal of randomly chosen farms was always significantly less effective (Fig. 5). Removing high degree farms was the most efficient way to reduce R_0 (Fig. 5(b)), while targeting high-betweenness farms was most effective to fragment the network into smaller, unconnected components (Fig. 5(a)).

We observed percolation-type transitions (Kiss et al., 2006): as important bridge-farms acting as short-cuts between socially or geographically remote clusters were removed, the size of the largest GSCC suddenly dropped, thus the network connectivity fell apart. Commonly, livestock markets act as easily identifiable ‘bridges’ in a network of livestock farms and are often primary targets for livestock movement control. However some farms (trader holdings) can also play a role as a bridge between cliques (Lockhart et al., 2010) or links between markets (Kao et al., 2006). Although these farms tend to be less readily identifiable than livestock markets (Ortiz-Pelaez et al., 2006), they might represent a very high risk to spread infectious agents (Dubé et al., 2010). Betweenness centrality is a measure of how each node lies ‘in between’ pairs of other nodes, therefore providing an indication of farms that control connectivity in the network (Ortiz-Pelaez et al., 2006). In the case of the LC network, the livestock operations displaying high betweenness were farms that sent/received livestock trans-island.

This feature could be exploited to decrease the risk of large epidemics, if the most influential farms for the overall connectivity are identified and targeted for control measures, such as movement bans, quarantines or other biosecurity measures. An obvious example in the case of the LC network is related to the topography of the country: if only trans-island movements were banned, the network would be disconnected into two independent clusters of farms; a pathogen introduced in one island could not spread to LC farms of the other island, no matter how contagious. This principle of isolation between the two main islands is already exploited by the Ministry of Primary Industries to implement legal biosecurity measures for pest control in New Zealand. In particular, the South Island is declared at the time of writing “controlled area” regarding the unwanted *Didyma* algae to prevent the spread to waterways of the North Island³; similarly, a movement control of bee colonies and other risk goods from the North to the South Island was enforced until *Varroa destructor* finally became established in the South Island⁴.

Our results, similar to previous work (Dubé et al., 2009), suggest that risk-based control strategies for biosecurity regulations could be efficient to reduce farm-to-farm spread of infection after an incursion. Risk-based surveillance targeting farms importing livestock from a large number of trading partners could also be implemented routinely for incursion detection. In an attempt to assess the possibility of spread throughout a contact network, considerations both on the magnitude of R_0 and on the presence of a GSCC in the network should be made as they inform two different aspects of the dynamics of an epidemic.

5. Conclusions

This study represents the first example of an analysis of a comprehensive contact network of livestock farms in New Zealand.

As one of the largest corporation of farms in the country, LC is well suited for implementing routine records of livestock movements. Although the results presented here are unlikely to be representative of all livestock enterprises in New Zealand, they show the value of such data for epidemiological studies. They also illustrate the contribution of a very common grazing practice in the New Zealand pastoral farming industry, which consists of dispatching young stock to lease farms in different areas depending on pasture availability and scarcity for a temporary period; this contributes to increase correlations in the contact pattern which could favor the transmission of pathogens. Similar insights would likely be obtained if the methodologies applied in this study were applied

³ <http://www.biosecurity.govt.nz/didymo>.

⁴ <http://www.biosecurity.govt.nz/pests/varroa>.

to movement data collected from a wider, more representative cross-section of the population of livestock farms in New Zealand. Notably, the network of LC farms presents very similar structural properties as most other published animal movements networks, despite the fact that farming systems are different between countries and production types.

Our findings emphasize the value of control measures targeted on smaller numbers of 'at risk' farms to control contagious disease spread, as opposed to blanket policies applied to the general population of farms. Selective interventions do not only include movement bans; the key-player farms identified in this study could also be subject to enhanced biosecurity measures and could be prioritised for active surveillance of infectious diseases since they would be at higher risk of becoming infected and transmitting infection. We thus propose that there is a value in documenting and analysing movement events, either at the national or, as demonstrated here, at corporate farming level. In the future, more comprehensive and detailed network analysis, encompassing all livestock movements in New Zealand, should become possible as the National Animal Identification and Tracing system comes into full operation.

Acknowledgements

This study was partially funded by the New Zealand Johne's Disease Research Consortium and stipends were granted by Massey University and New Zealand International Doctoral Research Scholarship. The main author is particularly grateful to Landcorp farming limited for providing us the data for this analysis and Gordon Williams for invaluable help in managing these data and understanding the movement pattern in relation with Landcorp policies and production needs. We are also especially grateful to Carolyn Gates for her insightful comments concerning the analysis and the manuscript.

References

- Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74.
- Anderson, R., May, R., 1991. *Infectious Diseases of Humans*. Oxford University Press, Oxford.
- Anderson, R.M., May, R., 1979. Population biology of infectious diseases: part I. *Nature* 280 (5721), 361–367.
- Aznar, M.N., Stevenson, M.A., et al., 2011. Analysis of cattle movements in Argentina 2005. *Prev. Vet. Med.* (2–3), 119–127.
- Bansal, S., Grenfell, B.T., et al., 2007. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* 4 (16), 879–891.
- Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286 (5439), 509–512.
- Bastian, M., Heymann, S., et al., 2009. Gephi: an open source software for exploring and manipulating networks. *Int. AAAI Conf. Weblogs Soc. Media*.
- Batagelj, V., Mrvar Pajek, A., Program for large Network Analysis, Home page: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>. Last accessed February 2016.
- R.M. Christley, S.E. Robinson, et al. Network analysis of cattle movement in Great Britain. Society for Veterinary Epidemiology and Preventive Medicine Proceedings of a Meeting Held at Nairn, Inverness Scotland, 30th March–1st April 2005.
- Clauset, Aaron, Shalizi, Cosma, Rohilla, Newman, M.E.J., 2009. 'Power-Law distributions in empirical data'. *SIAM Rev.* 51 (4), 661–703.
- Corner, L.A., Pfeiffer, D.U., et al., 2003. Social-network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). *Prevent. Vet. Med.* 59 (3), 147–167.
- Cross, C. Paul, et al., 2005. 'Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations'. *Ecol. Lett.* 8 (6), 587–595.
- Csárdi, G., Nepusz, T., 2006. The igraph software package for complex network research. *Int. J. Complex Syst.*, 1695.
- Delignette-Muller, M. L., Pouillot, R., et al., 2010. fitdistrplus: help to fit of a parametric distribution to non-censored or censored data. R package version 0.1-3. URL <http://CRAN.R-project.org/package=fitdistrplus>. Last accessed February 2016.
- Diekmann, O., Heesterbeek, J.A., et al., 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* 28 (4), 365–382.
- Dorjee, S., Revie, C.W., et al., 2013. Network analysis of swine shipments in Ontario, Canada, to support disease spread modelling and risk-based disease management. *Prevent. Vet. Med.* 112 (1–2), 118–127.
- Eames, K.T.D., et al., 2011. Integrating contact network structure into tuberculosis epidemiology in meerkats in South Africa: implications for control. *Prevent. Vet. Med.* 101 (1–2), 113–120.
- Dubé, C., Ribble, C., et al., 2010. An analysis of the movement of dairy cattle through 2 large livestock markets in the province of Ontario, Canada. *Can. Vet. J.* 51, 1254–1260.
- Dubé, C., Ribble, C., et al., 2009. A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound. Emerg. Dis.* 56 (3), 73–85.
- Eguíluz, V.M., Klemm, K., 2002. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* 89 (10), 108701.
- Gilbert, M., Mitchell, A., et al., 2005. Cattle movements and bovine tuberculosis in Great Britain. *Nature* 435 (7041), 491–496.
- Hethcote, H.W., 1978. An immunization model for a heterogeneous population. *Theor. Popul. Biol.* 14 (3), 338–349.
- House, T., Keeling, M.J., 2011. Insights from unifying modern approximations to infections on networks. *J. R. Soc. Interface* 8 (54), 67–73.
- Jones, J.H., Handcock, M.S., 2003. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. R. Soc. B: Biol. Sci.* 270 (1520), 1123–1128.
- Kao, R.R., 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends Microbiol.* 10 (6), 279–286.
- Kao, R.R., Danon, L., et al., 2006. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc. R. Soc. B: Biol. Sci.* 273 (1597), 1999–2007.
- Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. *Proc. Biol. Sci.* 266 (1421), 859–867.
- Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. *J. R. Soc. Interface* 2, 295–307.
- Kiss, I.Z., Green, D.M., Kao, R.R., 2006. 'The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread'. *J. R. Soc. Interface* 3 (10), 669–677.
- Kossinets, G., Watts, D.J., 2006. Empirical analysis of an evolving social network. *Science* 311 (5757), 88–90.
- Li, L., Alderson, D., et al., 2005. Towards a theory of scale-free graphs: definition, properties, and implications. *Internet Math.* 2 (4), 431–523.
- Lockhart, C.Y., Stevenson, M.A., et al., 2010. Patterns of contact within the New Zealand poultry industry. *Prevent. Vet. Med.* 95 (3–4), 258–266.
- Martínez-López, B., Perez, A.M., et al., 2009. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound. Emerg. Dis.* 56 (4), 109–120.
- May, R.M., Lloyd, A.L., 2001. Infection dynamics on scale-free networks. *Phys. Rev. E* 64 (6), 066112.
- Ortiz-Pelaez, A., Pfeiffer, D.U., et al., 2006. Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK. *Prevent. Vet. Med.* 76 (1–2), 40–55.
- Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86 (14), 3200–3203.
- Porphyre, T., Stevenson, M., et al., 2008. Influence of contact heterogeneity on TB reproduction ratio R_0 in a free-living brushtail possum *Trichosurus vulpecula* population. *Vet. Res.* 39 (3).
- R Development Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. Home page: <http://www.R-project.org>.
- Robinson, S.E., Everett, M.G., et al., 2007. Recent network evolution increases the potential for large epidemics in the British cattle population. *J. R. Soc. Interface* 4 (15), 669–674.
- Sanson, R.L., 2005. A survey to investigate movements off sheep and cattle farms in New Zealand, with reference to the potential transmission of foot-and-mouth disease. *N. Z. Vet. J.* 53 (4), 223–233.
- Shirley, M.D.F., Rushton, S.P., 2005. 'Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic'. *Epidemiol. Infect.* 133 (6), 1023–1032.
- Vernon, M.C., 2011. Demographics of cattle movements in the United Kingdom. *BMC Vet. Res.* 7, 31.
- Vernon, M.C., Keeling, M.J., 2009. Representing the UK's cattle herd as static and dynamic networks. *Proc. R. Soc. B: Biol. Sci.* 276 (1656), 469–476.
- Volkova, V.V., Howey, R., et al., 2010. Sheep movement networks and the transmission of infectious diseases. *PLoS One* 5 (6), e11185.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (6684), 440–442.
- Woolhouse, M.E.J., Dye, C., et al., 1997. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci.* 94, 338–342.
- Woolhouse, M.E.J., Shaw, D.J., et al., 2005. Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* 1 (3), 350–352.