Title

Rewiring cattle movements to limit infection spread.

Authors

Thibaut Morel-Journel[a*], Pauline Ezanno[a], Elisabeta Vergu[b]

Affiliations

a: INRAE, Oniris, BIOEPAR, 44300, Nantes, France

b: INRAE, Université Paris-Saclay, MaIAGE, 78350, Jouy-en-Josas, France

Corresponding author

*Thibaut Morel-Journel: thibaut.morel-journel@inserm.fr

Abstract

The cattle tracing databases set up over the past decades in Europe have become major resources for representing demographic processes of livestock and assessing potential risk of infections spreading by trade. The herds registered in these databases are nodes of a network of commercial movements, which can be altered to lower the risk of disease transmission. In this study, we develop an algorithm aimed at reducing the number of infected animals and herds, by rewiring specific movements responsible for trade flows from high- to low-prevalence herds. The algorithm is coupled with a generic computational model describing infection spread within and between herds, based on data extracted from the French cattle movement tracing database (BDNI). This model is used to simulate a wide array of infections, with either a recent outbreak (epidemic) or an outbreak that occurred five years earlier (endemic), on which the performances of the rewiring algorithm are explored. Results highlight the effectiveness of rewiring in containing infections to a limited number of herds for all scenarios, but especially if the outbreak is recent and if the estimation of disease prevalence is frequent. Further analysis reveal that the key parameters of the algorithm affecting infection outcome vary with the infection parameters. Allowing any animal movement from high to low-prevalence herds reduces the effectiveness of the algorithm in epidemic settings, while frequent and fine-grained prevalence assessments improve the impact of the algorithm in endemic settings. According to our results, our approach focusing on a few commercial movements is expected to lead to substantial improvements in the control of a targeted disease, although

changes in the network structure should be monitored for potential vulnerabilities to other diseases. Due to its generality, the developed rewiring algorithm could be applied to any network of controlled individual movements liable to spread disease.

Keywords

Control strategy; Epidemiology; Data-based; Network; Stochastic model

Abbreviation

BDNI: *Base de données nationale d'identification animale*

Introduction

Following bovine spongiform encephalopathy and classical swine fever epidemics in the 1990s, the European Union initiated the mandatory identification and registration of cattle in Europe (EU, 2000). This decision led to the creation of national identification databases, such as the cattle tracing system in the United Kingdom (Kao et al., 2006, Vernon, 2011), the French national bovine identification database (BDNI) (Rautureau et al., 2011, Dutta et al., 2014), the Italian national bovine database (Natale et al., 2009, Bajardi et al., 2011) and the database of the Swedish board of agriculture (Nöremark et al., 2009, 2011). These animal tracing systems have enabled the monitoring of infectious livestock diseases and the development of strategies to prevent their spread (Gilbert et al., 2005, Moslonka-Lefebvre et al., 2016, Beaunée et al., 2017), since animal trade is a major transmission pathway between herds. Indeed, commercial exchanges are not only recorded comprehensively, but also controlled by farmers, unlike animal mobility in the wild. These databases, whose reliability has increased over time since their creation (Green and Kao, 2007), are therefore powerful tools for simulating infectious diseases in cattle (Ezanno et al., 2020) and assessing the impact of livestock movements on epidemics (Ezanno et al., 2021).

The information provided by these commercial animal movements can be used as a basis for representing comprehensively the demographic processes and trades between cattle farms located in a given region, using a metapopulation framework (Liu et al., 2007, Widgren et al., 2015). To this end, disease transmission between individuals within a defined set of herds can be modelled, by combining an epidemiological model with existing data on births, deaths and movements. This type of models accounts at least for two ways of spreading the infection: by contact within a herd, or by actually moving animals between herds. This is for instance the case for paratuberculosis, a cattle disease mainly spread between herds by trade (Beaunée et al., 2015, Biemans et al., 2021). Manipulating the structure of cattle movement is expected to have a direct impact on the latter and an indirect impact on the former.

The structure of these trade movements can be understood through the prism of graph theory: herds are the vertices of a commercial exchange network, whose edges are the movements of livestock (Dubé et al., 2009). Thus, each herd can be characterised using graph metrics, e.g. its in- and out-degree, i.e. the number of herds it has respectively bought animals from and sold animals to. Network-based control strategies then aim to modify the structure of the network to reduce infection risks. Removing vertices (Rautureau et al., 2011, Büttner et al., 2013) or edges (Yang et al., 2013, Green et al., 2009) through trade ban or culling is a method used to slow down epidemics. In a context of cattle exchange however, preventing farmers from buying or selling livestock entails high economic costs. Therefore, this strategy cannot be used routinely or over extended periods of time. It is likely better suited to the management of regulated diseases, the consequences of which are also very costly and for controlling outbreaks of newly introduced diseases. Conversely, the application of such drastic methods on the longer term for endemic diseases may not be feasible.

Edge rewiring is a less radical approach able to balance the trade-off between health risks and economic costs. This method corresponds to the modification of one or both vertices that an edge connects (Gross et al., 2006, Piankoranee and Limkumnerd, 2020, Britton et al., 2016, Ball and Britton, 2020). Although most of the theoretical literature on the subject rather considers rewiring in the context of human contact networks, it has also been used to study epidemic spread in cattle movement networks (Gates and Woolhouse, 2015, Mohr et al., 2018, Ezanno et al., 2021, Biemans et al., 2022). For instance, Gates and Woolhouse (2015) present a rewiring method that creates an entirely new movement network disconnecting large buyers from large sellers, while retaining the total number of animals bought or sold by each herd. This method requires information at the network level, the criteria used being the distributions of in- and out-degrees of all herds. Global-level information is also generally required for most rewiring methods in contact networks, although Piankoranee and Limkumnerd (2020) proposed a method based on local information. In their study, rewiring is decided at the vertex level, according to its status and those of its direct neighbours. Controlling cattle movements depending on the sanitary status of their origin has been proposed in previous studies, e.g. by Hidano et al. (2016). Their study presents different scenarios regarding farmers' practices, especially their tendency to avoid buying cattle from regions with a higher incidence of bovine tuberculosis. The approach presented here is similar, albeit at a finer grain: preventing farmers from buying cattle from herds with a higher prevalence of the target disease.

This study presents a new rewiring method to reduce the spread of infections in a cattle movement network. To do this, we developed a rewiring algorithm aimed at preventing the movements of animals from higher-prevalence herds to lower-prevalence ones. It was based on an edge-level criterion: the estimated difference in prevalence between the herd of origin and the herd of destination of the movement considered. For this study, we tested the algorithm in conjunction with a computational epidemiological

model describing the spread of a nonspecific disease, whose infectiousness was parametrically defined. The impact of the algorithm was tested using a real commercial movement network, based on dataset from the French cattle tracing system (BDNI). In contrast with similar rewiring approaches developed recently to target specific diseases (Ezanno et al., 2021, Biemans et al., 2022), we propose a more generalist approach aimed at investigating the effectiveness of this type of method in a broader context. After presenting the movement network used as an example, the model and the algorithm, we consider various outputs of simulations with and without rewiring, concerning the functioning of the algorithm itself, its impact on infection propagation, and on the structure of the cattle movement network.

Data and methods

Cattle movement network

In order to test the algorithm on a actual network of commercial bovine movements, we use an extraction from the French national bovine identification database (BDNI). It includes all cattle herds in Brittany (a French region) that sold or bought at least one animal during the year 2014. This set of 21,548 herds is referred to as the 'metapopulation' thereafter. Every animal in the dataset is included regardless of breed or age, in order to have a larger number of movements per herd over this period of time. Three types of commercial exchanges are considered: (i) 'internal movements' have an origin and a destination among the herds in the dataset, (ii) 'imports' have only a destination in the dataset and (iii) 'exports' have only an origin in the dataset. They represent respectively 64%, 16% and 20% of the commercial exchanges involving at least one herd of the dataset. Each commercial exchange of animals is assumed to take place directly from one herd to another, neglecting intermediaries. This means that markets and sorting centres are not considered for this study. They differ from herds in that they tend to concentrate a large number of animals, but for a limited period of time (less than a day for markets, a few days for sorting centres). In addition, the dataset also includes information about the demographic events in the herd, which are considered as a special type of movements: (iv) births have only a destination, corresponding to the herd where the animal is born, and (v) deaths have only an origin, corresponding to the last herd recorded for the animal.

The dataset is represented as a network with herds and internal movements corresponding to the vertices and edges, respectively. This network is (i) dynamic, i.e. movements are characterised by the date at which they occur, (ii) weighted, i.e. a single edge represents the set of all movements from herd A to herd B, with a weight corresponding to the number of movements, and (iii) directed, i.e. movements from herd A to herd B are accounted for separately from movements from herd B to herd A. The network therefore includes 21,548 vertices and 100,088 edges. The total number of internal movements over 2014 is 206,640, thus the average edge weight is 2.06.

Epidemiological model: within and between-herd dynamics and infection settings

The model developed aims to simulate pathogen transmission within herds, and infection spread between herds through cattle movements. A full description of the model is available in Supplementary material 1. The model is stochastic in discrete time – each time-step corresponding to a day of 2014 – and in discrete space – by integrating the network of herds and movements described above. Commercial exchanges and demography are data-based: movement m is characterised by its origin O_m, its destination D_m, its date according to the dataset T_m^* and the date at which it is simulated T_m. By default, movements are simulated according to the dataset, i.e. $T_m = T_m^*$. Within-herd dynamics are based on a SIRS model with three parameters: the infection rate β, the recovery rate γ – therefore the average infection duration is $1/\gamma$ – and the rate of return to susceptibility δ. At each time-step t, herd h is characterised by its number of susceptible, infected and recovered individuals, noted respectively $S_h(t)$, $I_h(t)$ and $R_h(t)$. The total herd size $N_h(t)$ is defined as the sum of these three values and infection prevalence as $P_h(t) = I_h(t)/N_h(t)$.

Each simulated infection begins with an initial outbreak in a metapopulation without infection, i.e. with only susceptible individuals. At $t = t_I$, the date of the outbreak, 10% of all herds in the metapopulation are infected, by replacing 1 susceptible individual with 1 infected individual in each of the herds. The probability of a herd being part of this 10% is proportional to the number of imports in the herd according to the 2014 dataset. The rationale is that herds receiving the most individuals from herds outside of the metapopulation are the most likely to introduce a new infection.

Two types of infections are considered for the study: epidemic and endemic. An infection is defined as 'epidemic' if it starts at the outbreak, i.e. if $t_0 = t_I$. The initial state of the infection is then as described above. An infection is defined as 'endemic' if its start date is five years after the outbreak, i.e. $t_0 = t_I + 1825$ days. The initial state of infection is then the result of a five-year infection, simulated using the same epidemiological model and an extraction from the BDNI over Brittany between 01/01/2009 and 31/12/2013. Endemic simulations for which the infection goes extinct before t_0 are discarded, so that only initial states that are not disease-free are considered.

Prevalence status of the herds

The algorithm developed aims at identifying and preventing movements of cattle 'at risk', i.e. those from higher-prevalence herds to lower-prevalence herds. The differences in prevalence are based on prevalence classes, numbered from 1 to c. Class i corresponds to prevalence values between b_i and b_{i+1}, with the lowest boundary $b_1 = 0$ and the highest boundary $b_{c+1} = 1$. The prevalence status of herd h at time t, noted $V_h^r(t)$, is then the class including its prevalence, i.e. $V_h^r(t) = i$ if $P_h(t) \in [b_i; b_{i+1}[$, and $V_h^r(t) = 1$ if $P_h(t) = 1$. Yet, this 'real' prevalence status is not the one used by the algorithm. Rather, it uses an 'observed' prevalence status, noted $V_h^o(t)$, which is recorded at t_{obs} and then remains the same for q time-steps., i.e. $V_h^o(t) = V_h^r(t_{obs}) \ \forall \ t_{obs} \in [t; t+q[$. No additional error on the observed status (e.g.

because of imperfect test specificity or sensitivity) is assumed, so that it always corresponds to the real prevalence status at t_{obs}. Movements are considered 'at risk' if the observed prevalence status of their origin is strictly greater than that of their destination.

Sequential rewiring

The algorithm works by permuting the origins of pairs of movements, one of which is at risk, so that neither of them is at risk after the rewiring. The pairs of movements are created such that $1 \leq c_{ON} \leq c_{DR} < c_{OR} \leq c_{DN} \leq c$, with c_{OR} and c_{DR} the observed status of the origin and destination of the movement at risk and c_{ON}, c_{DN} those of the origin and destination of the other movement. By permuting the origins, the algorithm creates a movement with an origin of status c_{ON} and a destination of status c_{DR}, and another movement with an origin of status c_{OR} and a destination of status c_{DN}. Then neither of the two movements is at risk, since $c_{ON} \leq c_{DR}$ and $c_{OR} \leq c_{DN}$.

For all movements to occur at a given time-step, the algorithm performs these permutations in a specific order to ensure that no potential rewiring is missed. Supplementary material 2 describes this functioning of the algorithm over a single time-step in pseudo-code. Firstly, it defines all possible quadruplets of prevalence classes $\{c_{OR}, c_{DR}, c_{ON}, c_{DN}\}$. These quadruplets are arranged primarily in ascending order of c_{DR}, secondarily in descending order of c_{OR}, thirdly in ascending order of c_{ON} and fourthly in descending order of c_{DN}. This order ensures that no potential permutation is missed by the algorithm. For each quadruplet, the algorithm then permutes the origins of k pairs of movements, with k the minimum between the number of movements at risk and the number of other movements considered.

Once all possible permutations are performed, there might be remaining movements at risk set to be performed on this time-step. Firstly, these movements are postponed to the next day, to be potentially rewired with another set of movements. The postponed movements are then prioritised for rewiring on the following day. Yet, postponing commercial movement represents a constrain for farmers. Therefore, a maximal delay during which a movement can be postponed Δ_{MAX} is fixed for the algorithm. Thus, remaining movement m is postponed to the next day only if it was not already postponed Δ_{MAX} days, i.e. if $T_m - T_m^* < \Delta_{MAX}$. If the algorithm prohibits any movement at risk, the remaining movements that cannot be postponed (called 'problematic' movements) are replaced by one export with the origin of the problematic movement as origin and one import with the destination of the problematic movement as destination. Otherwise, the problematic movement is conserved as such. Overall, the algorithm therefore depends on four parameters: the number of prevalence classes c, the period at which observed status is updated q, the maximum delay Δ_{MAX} and whether movements at risk are prohibited.

Simulations

Simulations are performed on the dataset between 01/01/2014 (defined as $t = 0$) and 01/01/2015 ($t = 365$). Different epidemiological settings are explored by manipulating the SIRS model parameters (β, γ and δ) and infection type (epidemic or endemic). Two clustering analyses are performed on the preliminary simulations to define six epidemiological settings (Supplementary material 3): weak, moderate and strong epidemic settings and weak, moderate and strong endemic settings (Fig. S2).

The effectiveness of the algorithm is tested by running simulations with $3 \times 3 \times 3 \times 2$ combinations of the algorithm parameters, respectively (i) the number of prevalence classes c (2, 3 or 4 classes), (ii) the update period q (1, 28 or 91 days), the maximum delay Δ_{MAX} (1, 3 or 7 days) and (iv) the prohibition of movements at risk (yes or no). Each combination, as well as a control without rewiring, are simulated 100 times for each of the six epidemiological settings.

Preliminary simulations are also carried out for each epidemiological setting between 01/01/2009 ($t = -1825$) and 31/12/2013 ($t = -1$), with an initial outbreak at $t_I = -1825$. On the one hand, the number of susceptible, infected and recovered individuals of each herd at $t = -1$ are used as the starting numbers for the endemic simulations (starting at $t = 0$). On the other hand, the boundaries of the prevalence classes b_i used by the algorithm are set as quantiles of the distribution of prevalence values. These boundaries ensure that the number of herds of each class is roughly the same at the start of the simulation. If fewer than $1/c$ herds have a null prevalence, b_i is the $((i-1)/c)^{th}$ quantile of the distribution. If it is greater than $1/c$, $b_1 = b_2 = 0$ and b_i is the $((i-2)/(c-1))^{th}$ quantile of the distribution.

Outcomes and analyses of numerical explorations

The simulations outcomes are listed in Table 1. They are related either to (i) the functioning of the algorithm, (ii) the infection or (iii) the network of internal movements modified by the algorithm.

The algorithm-related outcomes $n_{rew}(t)$, $n_{del}(t)$ and $n_{prob}(t)$ are computed each time-step after rewiring, while $n_{risk}(t)$ and $n_{err}(t)$ are computed before. These latter outcomes are computed by using the real prevalence status of the herds, rather than the observed ones. A movement m is included in $n_{risk}(t)$ if $V^r_{O_m}(t) > V^r_{D_m}(t)$, and also included in $n_{err}(t)$ if $V^o_{O_m}(t) \leq V^o_{D_m}(t)$ at the same time. The proportion of undetected movements at risk is computed on a weekly basis, to account for intra-week variability in the number of livestock movements. Over week w, this proportion $p_{err}(w)$ is:

$$p_{err}(w) = \frac{\sum_{t=7(w-1)+1}^{7w} n_{err}(t)}{\sum_{t=7(w-1)+1}^{7w} n_{risk}(t)}.$$

The Spearman's correlation coefficient ρ between $p_{err}(w)$ and the number of weeks since last update (from 1 to 4 weeks if $q = 28$ days, from 1 to 13 weeks if $q = 91$ days) is also computed to assess the

Outcomes related to	Notation	Description
Algorithm	$n_{rew}(t)$	Number of movements rewired at time t
	$n_{del}(t)$	Number of delayed movements at time t
	$n_{prob}(t)$	Number of problematic movements at time t
	$n_{risk}(t)$	Number of movements at risk at time t
	$n_{err}(t)$	Number of movements undetected as at risk at time t
Infection	n_{inf}	Number of herd infections
	n_{ext}	Number of herds in which the infection goes extinct
	a_{dur}	Average duration of infection
	$c_{inc}(t)$	Cumulative incidence at time t
	$n_{herd}(t)$	Number of infected herds at time t
	$n_{ind}(t)$	Number of infected individuals in the metapopulation at time t
	$a_{prev}(t)$	Average prevalence in the infected herds at time t
Network	n_{SCC}	Number of strongly connected components
	max_{SCC}	Size of the largest strongly connected component
	ind_h	In-degree of herd h
	$outd_h$	Out-degree of herd h

Table 1: List of the outcomes computed from the simulations. The infection-related outcomes were computed for each simulation separately. The algorithm and network-related ones were computed for each simulation with the algorithm.

relationship between errors in herd prevalence status and time. The Spearman's coefficient is preferred because it does not assume any particular distribution of the involved variables.

The impact of the algorithm on the infection dynamic is estimated through $c_{inc}(t)$, i.e. the cumulative number of herds newly infected over the simulation. The variations in $n_{herd}(t)$ and $n_{ind}(t)$ over time are also presented in Supplementary material 4. Besides, the overall impact of the algorithm on the infection is assessed using a global multivariate sensitivity analysis, following Lamboni et al. (2011) and using the *multisensi* package of the *R* software (Bidot et al., 2018), which is used to perform sensitivity analyses on a multivariate output. For this analysis, twelve variables are derived from the infection-related outcomes. The three outcomes computed once per simulation n_{inf}, n_{ext} and a_{dur} are used as such. In addition, the maximum, minimum and final values over the whole period simulated (respectively noted $max(u(t))$, $min(u(t))$ and $u(365)$ for outcome $u(t)$) of $n_{herd}(t)$, $n_{ind}(t)$ and $a_{prev}(t)$ are also computed. The analysis includes a principal component analysis (PCA) on the scaled variables, which are used as the multivariate output for the sensitivity analysis. Two generalised sensitivity indices (GSI), which are weighted means of the sensitivity indices over all the dimensions of the PCA, are computed for each algorithm parameter: the total index (tGSI) including interactions with other parameters, and the first-order index (mGSI), not including them. The first principal component of the PCA is also used to assess the distribution of the simulations depending on the algorithm parameters.

The network-related outcomes are based on an static view of the network aggregating all the internal movements performed during the simulation, from $t = 0$ to $t = 365$. Therefore, they take into account the rewiring performed by the algorithm, and the potential removal of problematic movements if movements at risks are completely prohibited. The outcomes recorded for the modified networks are compared to the same metrics for the original network defined by the 2014 dataset. The strongly connected components –

from which n_{SCC} and max_{SCC} are computed – correspond to groups of vertices linked to each other by a directed path. The percentiles of the distributions of ind_h and $outd_h$ of all herds in the static network are used to assess the in-degree and out-degree distributions, respectively.

Results

Outcomes related to the algorithm

Our results show that number of movements rewired varies greatly depending on the date of the outbreak. It is negligible in the epidemic settings, with 80% of simulations with a total of rewired movements between 192 (fewer than 0.1% of all movements) and 2250 (1.1%). However, it is larger in the endemic settings, with 80% of simulations with between 17,344 (8.4% of all movements) and 33,640 (16.3%) movements rewired. Besides, increasing the value of Δ_{MAX} logically increases the number of delayed movements (which is 0 by definition for $\Delta_{MAX} = 0$) and decreases the number of problematic movements. In the endemic settings, the problematic movements represent a small proportion of the movements detected as high risk (median: 5.4%, 9th decile: 17.4%). In the epidemic settings however, they represent a larger part (median: 14.3%, 9th decile: 59.7%), although their absolute numbers remain low (median: 129, 9th decile: 651). Because of the overwhelming number of initially non-infected herds in these simulations, the movements at risk are likely more difficult to rewire, and thus more likely to be tagged as problematic by the algorithm.

Increasing the herd status update period q is not associated with a decrease in the number of rewiring events (Fig. 1A, 1B). The value of q is even rather positively correlated with the number of rewiring events in epidemic settings. This suggests that the algorithm performs more erroneous rewiring as q increases. This is confirmed by the distributions of Spearman's correlation coefficient between $p_{err}(w)$ and the number of weeks since last update ρ with $q = 91$ days (Fig. 1D), in epidemic settings (80% of values of ρ between -0.01 and 0.50) and in endemic settings (80% of values of ρ between 0.39 and 0.75). This is also somewhat the case with $q = 28$ days (Fig. 1C), although the correlations are weaker, in endemic (80% of values of values between -0.09 and 0.79) as well as in epidemic settings (80% of values of values between -0.05 and 0.34).

The average proportions of undetected movements at risk $p_{err}(w)$ all tend to increase with the number of weeks since the last update w (Fig. 1E, 1F). This increase is systematically greater for the largest value of q, up to $p_{err}(w) = 0.3$. However, they also appear to have reach a plateau after 10 weeks. This suggests that a further increase in the update period q would not strongly increase the proportion of undetected movements at risk. As for Spearman's correlation coefficient ρ, the increase is greater in endemic settings than in epidemic settings.

Outcomes related to the infection

Comparison of the results with and without rewiring shows the overall effectiveness of the algorithm in containing the infection (Fig. 2). Regardless of the epidemiological setting and the combination of

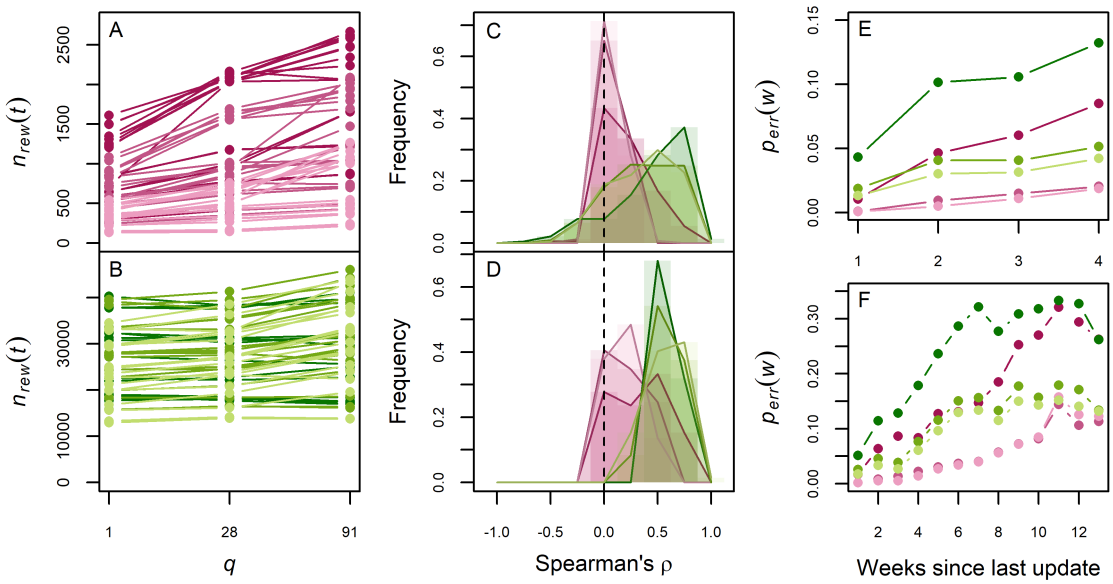

Figure 1: Impact of the update period q on the undetected movements at risk, in epidemic (magenta) or endemic settings (green), weak (light), moderate (medium) or strong (dark). First column: total number of rewiring events as a function of the update frequency q, averaged over all simulations for a same algorithm parameter combination, in epidemic (A) and endemic settings (B). Second column: distribution of Spearman's correlation coefficients (ρ), with $q = 28$ days (C) and $q = 91$ days (D). Third column: average proportion of undetected movements at risk $p_{err}(w)$ as a function of the number of weeks since the last update, with $q = 28$ days (E) and $q = 91$ days (F).

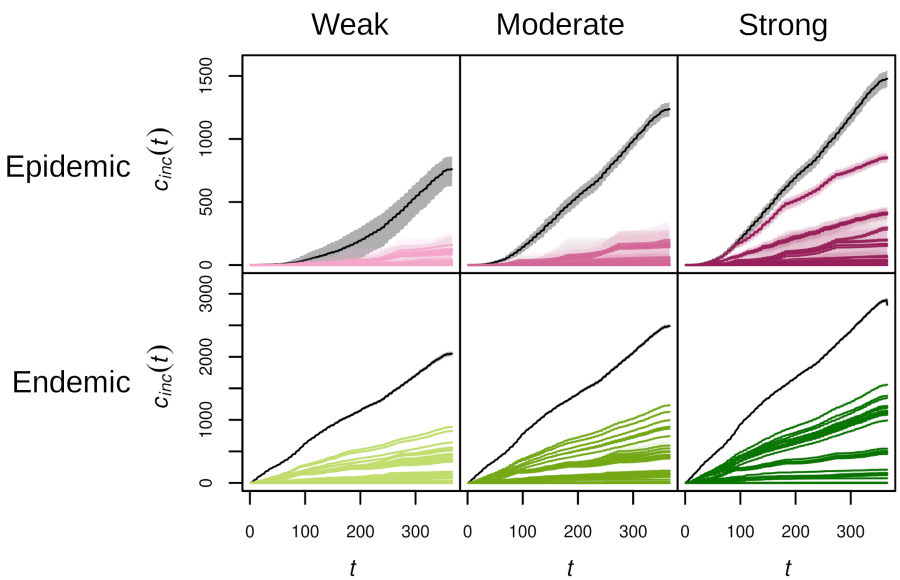

Figure 2: Cumulative incidence $c_{inc}(t)$, in number of herd infections, as a function of time (t, in days), for simulations with (colour) or without rewiring (black), in epidemic (1st row, magenta) or endemic settings (2nd row, green), weak (1st column, light), moderate (2nd column, medium) and strong (3nd column, dark). Each combination of algorithm parameters is represented by its mean over the repetitions (solid line) and an interval of 80% of simulations (envelope).

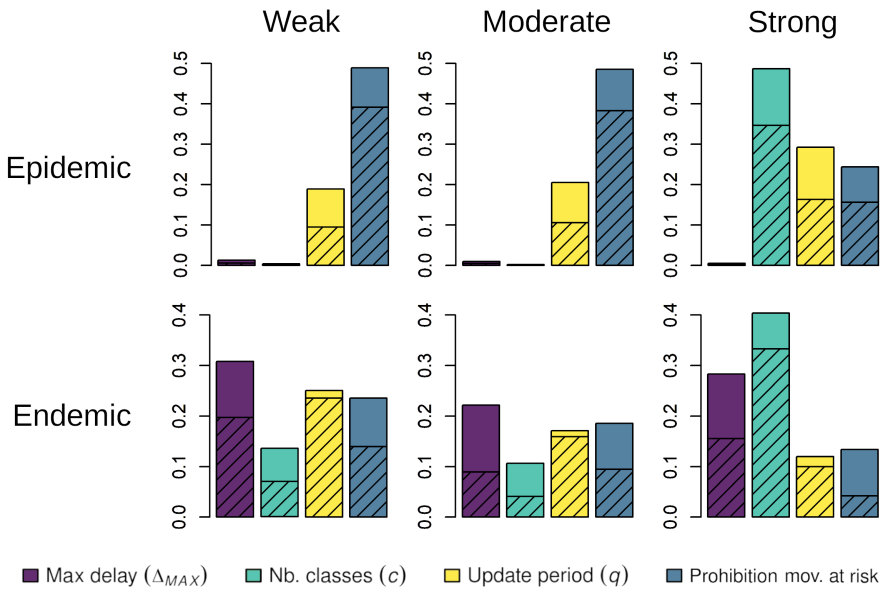

Figure 3: Generalised sensitivity indices (GSI) of the maximum delay Δ_{MAX} (purple) the number of prevalence classes c (cyan), the update period q (yellow) and the prohibition of movements at risk (blue), in epidemic (1st row) or endemic settings (2nd row), weak (1st column), moderate (2nd column) and strong (3rd column). The total indices (tGSI) are in solid colour and the first-order indices (mGSI) are hatched.

parameters considered, the cumulative number of herds newly infected $c_{inc}(t)$ remains systematically lower after rewiring. The algorithm is particularly effective in weak and moderate epidemic settings, where very few herds are infected during the year. In other epidemiological settings, the impact of the algorithm varies more strongly depending on the scenario considered. Results for $n_{herd}(t)$ and $n_{ind}(t)$ are presented in Supplementary material 4. In epidemic settings, variations in $n_{herd}(t)$ logically follow closely those of $c_{inc}(t)$. Hence, the algorithm also reduces the increase in the total number of infected herds. It also reduces the total number of infected individuals, although the impact is not as strong as for herds. In endemic settings, the value of $n_{herd}(t)$ remains similar during the whole simulation without rewiring (Fig. S3), despite new infections according to variations in $c_{inc}(t)$. This indicates a turnover in the infection at the metapopulation level, with populations losing the infection through the acquisition of resistance or the culling and trade of infected animals. By reducing the number of new infections, the algorithm actually therefore reduces the total number of infected herds over time. However, its impact is smaller on the total number of infected individuals (Fig. S4).

The sensitivity analysis shows differences in the relative importance of the algorithm parameters on the reduction of the infection (Fig. 3). Three different patterns of sensitivity to the algorithm parameters are observed. Firstly, simulations in weak and moderate epidemic settings exhibit an overwhelming sensitivity to the prohibition of movements at risk. Secondly, those in strong epidemic or endemic settings exhibit a strong sensitivity to the number of prevalence classes c. Finally, those in weak and moderate endemic settings exhibit a more balanced sensitivity to all parameters, with a substantial difference between total

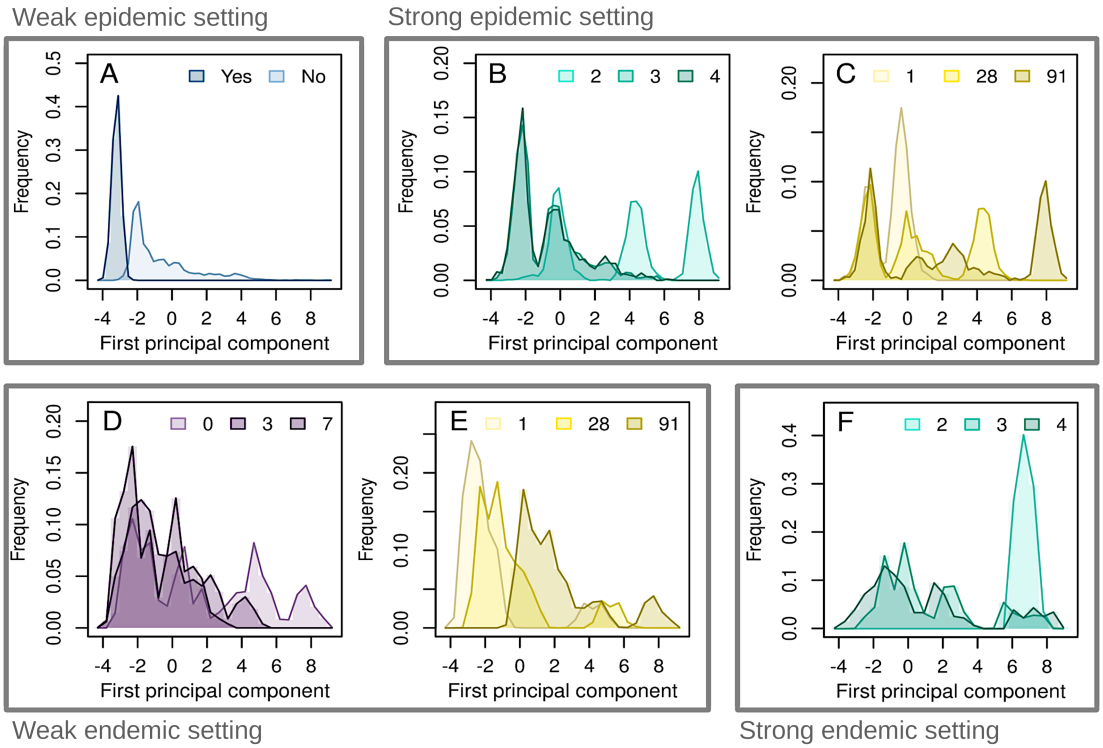

Figure 4: Distribution of the simulations on the first component of the PCA performed as a first step of the sensitivity analysis, in the weak epidemic setting (A), the strong epidemic setting (B, C), the weak endemic setting (D, E) and the strong endemic setting (F). The outputs are divided by maximum delay (purple, D), management of problematic movements (blue, A), number of prevalence classes (cyan, B and F) and herd status update period (yellow, C and E).

and first-order indices for the maximum delay Δ_{MAX}, the number of classes and the prohibition of movements at risk. These differences suggest an interaction between the three algorithm parameters. Besides, simulations for every epidemiological setting are somewhat sensitive to the update period q.

The PCA performed as a first step of the sensitivity analysis is used to explore further the way algorithm parameters impact the infection-related outputs. Supplementary material 5 shows that the first principal component of the PCA is globally positively correlated with outputs describing the extent of the infection. The distributions of simulations along this first principal component therefore provides information about the way algorithm parameter values affects the extent of the infection. Supplementary material 6 presents these distributions for every epidemiological setting and every algorithm parameter, while Fig. 4 displays some of the most relevant distributions. Fig. 4A shows that, in the weak epidemic setting, simulations in which movements at risk are prohibited almost always score lower on the first principal component than those in which they are not. The distribution is similar in the moderate epidemic setting (Fig. S6), which has similar sensitivity indices (Fig. 3). Interestingly, distributions of simulations in strong epidemic or endemic settings show that those with $c = 2$ score higher on their respective first component, while those with $c = 3$ and $c = 4$ are not different (Fig. 4B, 4F). A similar pattern is observed with the maximum delay in the weak endemic setting: only simulations with $\Delta_{MAX} =$

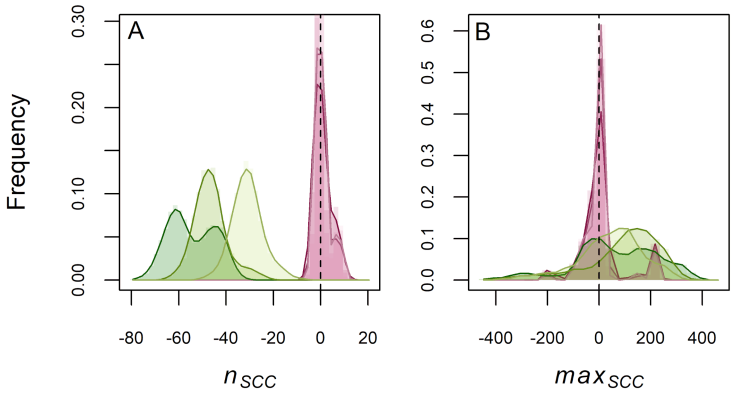

Figure 5: Distributions of the differences in number of strongly connected components (A, n_{SCC}) and in size of the largest strongly connected component B, max_{SCC}) between rewired networks and the original one from the dataset, for epidemic (magenta) and endemic (green) settings, weak(light), moderate (medium) and strong (dark).

0 score higher on the first principal component (Fig. 4D). In the strong epidemic setting, the two high-scoring peaks in the distribution according to c (Fig. 4B) correspond to the simulations with $q = 28$ and $q = 91$ (Fig. 4C), highlighting an interplay between the number of classes c and the update period q. No interplay between Δ_{MAX} and q is visible in the weak endemic setting, although Fig. 4E show that the score of simulations on the first principal component is positively correlated with q. Distributions in the moderate endemic setting are similar to those in the weak endemic setting (Fig. S6).

Outcomes related to the movement network

In endemic settings, rewiring movements increase the in- and out-degrees of the herds, i.e. the number of different herds they are connected to (see Supplementary material 7). The increase is small but systematic, for every algorithm parameter value (Fig. S7). In addition, the algorithm also affects the strongly connected components of the network in endemic settings. On the one hand, the algorithm reduces their number, all the more that the infection was strong (Fig. 5). On the other hand, the size of the largest strongly connected component is increased in most, but not all simulations (64%, 67% and 80% of simulations in low, moderate and high endemic settings, respectively). It should be noted that the lesser impact of the algorithm on the network in epidemic settings can be explained by a number of rewiring events 25 times smaller on average than in endemic settings.

Discussion

The rewiring algorithm we developed for this study is able to reduce the extent of infections, in the absence of any other restriction measure and for a large panel of disease parameters (infection rate β, recovery rate γ or rate of return to susceptibility δ). However, the extent of the reduction varies between the different epidemiological settings considered. Indeed, infections are almost completely prevented with

weak or moderate epidemic settings, while they still develop or persist for other settings, although not as much as without any rewiring. However, the decrease in the number of infected herds is not necessarily coupled with a decrease in the number of infected individuals. This result highlights the tendency of the algorithm to concentrate infected individuals in the already infected herds. The algorithm therefore performs a trade-off that is beneficial to the metapopulation as a whole – with fewer infected herds – but detrimental to the smaller number of already infected herds, in such a situation where movement rewiring is not combined with complementary on-farm measures to reduce within-herd infection prevalence. This is the case for the infections in an epidemic setting, in which the prevalence in the infected herds increases over the year. This is also the case for infections in endemic settings, in which new sensitive individuals could still be born or imported.

The sensitivity analysis on the infection-related outcomes reveals that the impact of the parameters of the algorithm is highly dependent on the epidemiological setting. Prohibiting the movements at risk, i.e., removing the movements that cannot be rewired and are delayed as much as possible, is mostly significant if the infection is not too strong and is just beginning. Only in these cases can the infection be fully contained by the prohibition. Increasing the maximal delay improves the performance of the algorithm in an endemic setting, for which the number of movements rewired is much larger than in epidemic settings. In those, delaying the movement to the next day increases substantially the opportunities for rewiring. The other two parameters are both related to the definition of the prevalence statuses used by the algorithm. A greater number of prevalence classes, which mainly impacts rewiring during strong infections, improves the separation of disease-free herds from the rest. Indeed, considering more prevalence classes lowers the upper boundary of the lowest one, which included only herds with very few or no infected animals, thus allowing the algorithm to effectively protecting disease-free herds. A longer update period between updates of the prevalence status makes the rewiring algorithm more error-prone, with a proportion of undetected movements at risk increasing with the time since the last update, at least up to ten weeks. This result is visible for any epidemiological setting, suggesting that any increase in the frequency of update to the status of the herds should improve the effectiveness of the algorithm. Conversely, the results indicate that increasing the number of prevalence classes to more than two, or having a maximum delay greater than zero, improves the efficiency of the algorithm much more than further increases.

As expected, the impact of rewiring on the commercial movements network structure is limited, as it targeted a few movements only: less than 20% of the movements for endemic infections and less than 2% of them for epidemic infections. Nevertheless, rewiring tends to increase the overall connectedness of the herds during endemic infections. Indeed, the increase in degree and in size of the largest strong component indicates that the algorithm has connected herds that were originally not so. These metrics are generally correlated with higher expected epidemic risks (Kiss et al., 2006, Dubé et al., 2009). The use

of such a rewiring method to manage actual bovine movements should take into account this potential increase in the risk of spreading other diseases. The algorithm could be extended to assess multiple diseases at once, but the additional constraints on rewiring would likely reduce its effectiveness.

The main hurdle to implementing this rewiring method in a real-life setting is its reliance on accurate and frequent prevalence data from a large number of farms. Firstly, a lack of specificity or sensitivity in the tests used might lead to an overestimation or underestimation of the prevalence in the herds, depending on the disease considered. Although we show that the algorithm remains efficient even though the observed prevalence differed from the real ones, this additional error could add up with the one observed in our study. However, the impact of such errors is also expected to be mitigated by the use of prevalence classes, so that small differences do not necessarily change the prevalence status of the herds.

Secondly, obtaining frequent prevalence data for a large number of farms remains challenging. Bulk milk-based sampling systems could be used for some diseases in cattle (e.g. Garoussi et al., 2008, Humphry et al., 2012, with bovine viral diarrhoea), which would facilitate prevalence estimation for multiple herds at once, thus reducing the associated costs. It would also be possible to reduce the sampling effort by focusing on a subset of herds to monitor. Firstly, this sampling effort should take into account additional information available thanks to measures already in place. For instance, the status of some herds could be approximated through health accreditation schemes (e.g. Ezanno et al., 2021), with herds already identified as disease-free could be automatically assigned to the lowest prevalence status for a given duration. Secondly, herds to monitor could be selected based on their role in disease spread, notably through network metrics. Indeed, central herds in the movement network, i.e. those through which a large proportion of animal movements pass, are expected to play a larger role in the spread of infection (Rautureau et al., 2011, Natale et al., 2011). Hoscheit et al. (2021) reviewed centrality measures taking into account the dynamic nature of the movement network, based on the BDNI. They found that the TempoRank index would for example be a good candidate for selecting a subset of herds to be specifically monitored and taken into account by the algorithm.

In this study, we use a network corresponding to commercial movements between every farm in Brittany (an administrative region of France) over a year to test the efficiency of the algorithm. The choice to limit the size of the network is notably motivated by computational limitations. Indeed, simulating a stochastic spread of the disease on a national scale over six years - five for the preliminary simulations and one for the main simulations - would have been considerably more costly, thus limiting the exploration of variations in the parameters of the SIRS model and the algorithm. Yet, this choice had additional implications that should be underlined.

Firstly, a substantial proportion of the movements involve herds outside of Brittany and are therefore not concerned by the rewiring. Indeed, 20% of all movements whose destination was in the metapopulation had an origin outside of it. In our simulations, these imports are assumed to not be movement at risk,

i.e. that the prevalence status of their origin is never higher than that of their destination. This is not trivial, as it presumes that imports do not create greater infection risks than internal movements. In a real-life context, applying this rewiring method in a single region would therefore require an additional management of the risk associated with imports. Yet, extending its use nationally should mitigate this problem, as the proportion of imports is expected to be much lower at this scale. Secondly, every commercial movement between farms is considered to test the algorithm, regardless of breed or age, in order to have a large enough set of movements. Indeed, additional criteria, concerning for instance the breed of the animals, could be added easily by providing the algorithm with movements for individuals in each category separately. However, such criterion would reduce the rewiring possibilities of the algorithm and therefore its effectiveness. Again, the network of commercial movements at the national scale could be large enough to separate the movements by breed or consider only movements of specific breeds.

Although the algorithm is tested on historical data from the BDNI for this study, it could also be used prospectively as part of decision-making tools, barring the limitations presented above. Indeed, the rewiring method could work without any simulation of infection, if herd statuses were provided otherwise. Given these statuses and the potential movements to occur, the algorithm would also suggest necessary changes to prevent movements at risk. In this context, the implementation of these changes would also depend on the actual decision of the informed farmers. Unless rewiring is enforced, it is expected that constraints other than sanitary ones would affect movements, which would impact the effectiveness of the algorithm. Coupling it with a decision-making model could provide additional insight on this impact. In order to make it easier to use as part of such decision-making tools, the algorithm has been specifically designed to be able to include additional, different constraints.

Besides, the rewiring method presented is not limited to cattle, but applicable to a much wider range of networks in animal and plant populations, e.g. among seed exchange networks, which face similar infection risks (Jeger et al., 2007, Pautasso et al., 2010). While the need for controlled movements makes this method more relevant to agricultural systems, the spatial and temporal scales considered can also be adapted depending on the context. Indeed, the daily time-step and the region level were used here as they correspond to the BDNI data structure, but are not necessary for the algorithm to work. The usefulness of our rewiring method could therefore extend beyond cattle concerns, even though the effectiveness of the algorithm in other contexts remains to be tested.

This study demonstrates the effectiveness of a rewiring method targeting specific movements to reduce infection risks. Our approach thus differs radically from that presented by Gates and Woolhouse (2015), as it also aims at generating minimal changes in the structure of the movement network. However, this study builds upon the results from Ezanno et al. (2021), by confirming the effectiveness of this method beyond the specific case of bovine paratuberculosis. Indeed, the algorithm presented by Ezanno et al. (2021) and later by Biemans et al. (2022), was developed specifically to address the control of bovine

paratuberculosis, notably characterised by an endemic status and a low detection rate. To do so, they used a specific age-structured epidemiological model (Camanes et al., 2018) and an algorithm calibrated to target the disease. This was also the case for instance of Mohr et al. (2018), which specifically targeted foot-and-mouth disease. Conversely, the present study aims at assessing more comprehensively the effectiveness of the algorithm. It is tested for different epidemiological settings – both endemic and epidemic – using a non-specific epidemiological model, and for broad range of parameter values. This study is therefore complementary to the previous ones, by bringing a broader perspective on the impact of rewiring in animal movement network on infectious diseases in general.

Funding

This work was supported by the French National Agency for Research (*Agence nationale de la recherche*), project CADENCE [grant number ANR-368 16-CE32-0007].

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, scripts, code, and supplementary information availability

The code for the algorithm, as well as additional scripts for formatting the data or running preliminary simulations and dummy test data, are freely available at https://sourcesup.renater.fr/projects/pub-rewir-algo/. The dataset used in the study is an extraction from the French national bovine identification database (BDNI), which is confidential, and therefore cannot be provided publicly.

References

P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza. Dynamical patterns of cattle trade movements. *PLoS One*, 6(5):e19869, 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0019869.

F. Ball and T. Britton. Epidemics on networks with preventive rewiring. *arXiv*, 2020.

G. Beaunée, E. Vergu, and P. Ezanno. Modelling of paratuberculosis spread between dairy cattle farms at a regional scale. *Vet Res*, 46(1):1–13, 2015. ISSN 1297-9716. doi: 10.1186/s13567-015-0247-3.

G. Beaunée, E. Vergu, A. Joly, and P. Ezanno. Controlling bovine paratuberculosis at a regional scale: Towards a decision modelling tool. *J Theor Biol*, 435:157–183, 2017. doi: 10.1016/j.jtbi.2017.09.012.

C. Bidot, M. Lamboni, and H. Monod. Multisensi: Multivariate Sensitivity Analysis, 2018.

F. Biemans, R. B. Romdhane, P. Gontier, C. Fourichon, G. Ramsbottom, SJ. More, and P. Ezanno. Modelling transmission and control of Mycobacterium avium subspecies paratuberculosis within Irish dairy herds with compact spring calving. *Prev Vet Med*, 186:105228, 2021. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2020.105228.

F. Biemans, S. Arnoux, S. J. More, JA. Tratalos, L. Gavey, and P. Ezanno. The effect of risk-based trading and within-herd measures on Mycobacterium avium subspecies paratuberculosis spread within and between Irish dairy herds. *Prev Vet Med*, 209:105779, 2022. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2022.105779.

T. Britton, D. Juher, and J. Saldaña. A network epidemic model with preventive rewiring: Comparative analysis of the initial phase. *Bull Math Biol*, 78(12):2427–2454, 2016. doi: 10.1007/s11538-016-0227-4.

K. Büttner, J. Krieter, A. Traulsen, and I. Traulsen. Static network analysis of a pork supply chain in Northern Germany—Characterisation of the potential spread of infectious diseases via animal movements. *Prev Vet Med*, 110(3-4):418–428, 2013. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2013.01.008.

G. Camanes, A. Joly, C. Fourichon, R. Ben Romdhane, and P. Ezanno. Control measures to prevent the increase of paratuberculosis prevalence in dairy cattle herds: An individual-based modelling approach. *Vet Rec*, 49(1):1–13, 2018. doi: 10.1186/s13567-018-0557-3.

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. *Transbound Emerg Dis*, 56 (3):73–85, 2009. ISSN 1865-1674. doi: 10.1111/j.1865-1682.2008.01064.x.

B. L. Dutta, P. Ezanno, and E. Vergu. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Prev Vet Med*, 117(1):79–94, 2014. doi: 10.1016/j.prevetmed.2014.09.005.

EU. Regulation (EC) No 1760/2000 establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing Council Regulation (EC) No 820/97, 2000.

P. Ezanno, M. Andraud, G. Beaunée, T. Hoch, S. Krebs, A. Rault, S. Touzeau, E. Vergu, and S. Widgren. How mechanistic modelling supports decision making for the control of enzootic infectious diseases. *Epidemics*, 32:100398, 2020. doi: 10.1016/j.epidem.2020.100398.

P. Ezanno, S. Arnoux, A. Joly, and R. Vermesse. Rewiring cattle trade movements helps to control bovine paratuberculosis at a regional scale. *Prev Vet Med*, page 105529, 2021. doi: 10.1016/j.prevetmed.2021.105529.

M. T. Garoussi, A. Haghparast, and H. Estajee. Prevalence of bovine viral diarrhoea virus antibodies in bulk tank milk of industrial dairy cattle herds in suburb of Mashhad-Iran. *Prev Vet Med*, 84(1-2): 171–176, 2008. doi: 10.1016/j.prevetmed.2007.12.016.

M. C. Gates and M. E. Woolhouse. Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics*, 12:11–19, 2015. ISSN 1755-4365. doi: 10.1016/j.epidem.2015.02.008.

M. Gilbert, A. Mitchell, D. Bourn, J. Mawdsley, R. Clifton-Hadley, and W. Wint. Cattle movements and bovine tuberculosis in Great Britain. *Nature*, 435(7041):491–496, 2005. doi: 10.1038/nature03548.

D. M. Green, A. Gregory, and L. A. Munro. Small-and large-scale network structure of live fish movements in Scotland. *Prev Vet Med*, 91(2-4):261–269, 2009. doi: 10.1016/j.prevetmed.2009.05.031.

DM. Green and RR. Kao. Data quality of the cattle tracing system in Great Britain. *Vet Rec*, 161(13): 439–443, 2007. ISSN 0042-4900. doi: 10.1136/vr.161.13.439.

T. Gross, C. J. D. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Phys Rev Lett*, 96(20):208701, 2006. doi: 10.1103/PhysRevLett.96.208701.

A. Hidano, T. E. Carpenter, M. A. Stevenson, and M. C. Gates. Evaluating the efficacy of regionalisation in limiting high-risk livestock trade movements. *Prev Vet Med*, 133:31–41, 2016. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2016.09.015.

P. Hoscheit, É. Anthony, and E. Vergu. Dynamic centrality measures for cattle trade networks. *Appl Netw Sci*, 6(1):1–17, 2021. ISSN 2364-8228. doi: 10.1007/s41109-021-00368-5.

RW. Humphry, F. Brülisauer, IJ. McKendrick, PF. Nettleton, and GJ. Gunn. Prevalence of antibodies to bovine viral diarrhoea virus in bulk tank milk and associated risk factors in Scottish dairy herds. *Vet Rec*, 171(18):445–445, 2012. doi: 10.1136/vr.100542.

M. J. Jeger, M. Pautasso, O. Holdenrieder, and M. W. Shaw. Modelling disease spread and control in networks: Implications for plant sciences. *New Phytol*, 174(2):279–297, 2007. ISSN 0028-646X. doi: doi:10.1111/j.1469-8137.2007.02028.x.

R. R. Kao, L. Danon, D. M. Green, and I. Z. Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc R Soc B*, 273(1597):1999–2007, 2006. doi: 10.1098/rsif.2006.0129.

I. Z. Kiss, D. M. Green, and R. R. Kao. 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, 2006. ISSN 1742-5689. doi: doi:10.1098/rsif.2006.0129.

M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliab Eng Syst Saf*, 96(4):450–459, 2011. doi: 10.1016/j.ress.2010.12.002.

W.-c. Liu, L. Matthews, M. Chase-Topping, N. J. Savill, D. J. Shaw, and M. E. Woolhouse. Metapopulation dynamics of Escherichia coli O157 in cattle: An exploratory model. *J R Soc Interface*, 4(16): 917–924, 2007. ISSN 1742-5689.

S. Mohr, M. Deason, M. Churakov, T. Doherty, and R. R. Kao. Manipulation of contact network structure and the impact on foot-and-mouth disease transmission. *Prev Vet Med*, 157:8–18, 2018. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2018.05.006.

M. Moslonka-Lefebvre, C. A. Gilligan, H. Monod, C. Belloc, P. Ezanno, J. A. Filipe, and E. Vergu. Market analyses of livestock trade networks to inform the prevention of joint economic and epidemiological risks. *J R Soc Interface*, 13(116):20151099, 2016. ISSN 1742-5689. doi: 10.1098/rsif.2015.1099.

F. Natale, A. Giovannini, L. Savini, D. Palma, L. Possenti, G. Fiore, and P. Calistri. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Prev Vet Med*, 92 (4):341–350, 2009. doi: 10.1016/j.prevetmed.2009.08.026.

F. Natale, L. Savini, A. Giovannini, P. Calistri, L. Candeloro, and G. Fiore. Evaluation of risk and vulnerability using a Disease Flow Centrality measure in dynamic cattle trade networks. *Prev Vet Med*, 98(2-3):111–118, 2011. doi: 10.1016/j.prevetmed.2010.11.013.

M. Nöremark, N. Håkansson, T. Lindström, U. Wennergren, and S. S. Lewerin. Spatial and temporal investigations of reported movements, births and deaths of cattle and pigs in Sweden. *Acta Vet*, 51(1): 1–15, 2009. doi: 10.1186/1751-0147-51-37.

M. Nöremark, N. Håkansson, S. S. Lewerin, A. Lindberg, and A. Jonsson. Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance. *Prev Vet Med*, 99(2-4):78–90, 2011. doi: 10.1016/j.prevetmed.2010.12.009.

M. Pautasso, M. Moslonka-Lefebvre, and M. J. Jeger. The number of links to and from the starting node as a predictor of epidemic size in small-size directed networks. *Ecol Complex*, 7(4):424–432, 2010. ISSN 1476-945X. doi: doi:10.1016/j.ecocom.2009.10.003.

S. Piankoranee and S. Limkumnerd. Effect of local rewiring in adaptive epidemic networks. *Phys Lett A*, 384(15):126308, 2020. ISSN 0375-9601. doi: 10.1016/j.physleta.2020.126308.

S. Rautureau, B. Dufour, and B. Durand. Vulnerability of animal trade networks to the spread of infectious diseases: A methodological approach applied to evaluation and emergency control strategies in

cattle, France, 2005. *Transbound Emerg Dis*, 58(2):110–120, 2011. ISSN 1865-1674. doi: 10.1111/j.1865-1682.2010.01187.x.

M. C. Vernon. Demographics of cattle movements in the United Kingdom. *BMC Vet Res*, 7(1):1–16, 2011. doi: 10.1186/1746-6148-7-31.

S. Widgren, R. Söderlund, E. Eriksson, C. Fasth, A. Aspan, U. Emanuelson, S. Alenius, and A. Lindberg. Longitudinal observational study over 38 months of verotoxigenic Escherichia coli O157: H7 status in 126 cattle herds. *Prev Vet Med*, 121(3-4):343–352, 2015. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2015.08.010.

H.-X. Yang, Z.-X. Wu, and B.-H. Wang. Suppressing traffic-driven epidemic spreading by edge-removal strategies. *Phys Rev E*, 87(6):064801, 2013. doi: 10.1103/PhysRevE.87.064801.