Edinburgh Research Explorer

Multi-species temporal network of livestock movements for disease spread

RESEARCH

Multi-species temporal network of livestock movements for disease spread

Anne-Sophie Ruget[1] , Gianluigi Rossi[1] , P. Theo Pepler[2] , Gaël Beaunée[1,3], Christopher J. Banks[1], Jessica Enright[4] and Rowland R. Kao[1,5]*

*Correspondence:
rowland.kao@ed.ac.uk
[1] The Roslin Institute,
University of Edinburgh,
Easter Bush Campus,
Edinburgh, Scotland, UK
Full list of author information
is available at the end of the
article

Abstract

Introduction: The objective of this study is to show the importance of interspecies links and temporal network dynamics of a multi-species livestock movement network. Although both cattle and sheep networks have been previously studied, cattle-sheep multi-species networks have not generally been studied in-depth. The central question of this study is how the combination of cattle and sheep movements affects the potential for disease spread on the combined network.

Materials and methods: Our analysis considers static and temporal representations of networks based on recorded animal movements. We computed network-based node importance measures of two single-species networks, and compared the top-ranked premises with the ones in the multi-species network. We propose the use of a measure based on contact chains calculated in a network weighted with transmission probabilities to assess the importance of premises in an outbreak. To ground our investigation in infectious disease epidemiology, we compared this suggested measure with the results of disease simulation models with asymmetric probabilities of transmission between species.

Results: Our analysis of the temporal networks shows that the premises which are likely to drive the epidemic in this multi-species network differ from the ones in both the cattle and the sheep networks. Although sheep movements are highly seasonal, the estimated size of an epidemic is significantly larger in the multi-species network than in the cattle network, independently of the period of the year. Finally, we demonstrate that a measure based on contact chains allow us to identify around 30% of the key farms in a simulated epidemic, ignoring markets, whilst static network measures identify less than 10% of these farms.

Conclusion: Our results ascertain the importance of combining species networks, as well as considering layers of temporal livestock movements in detail for the study of disease spread.

Keywords: Livestock movements, Multi-species, Network analysis, Temporal network, Stochastic simulations

Introduction

Infectious diseases in livestock are of great concern as they pose an economic burden, compromise animal health and welfare, and threaten human health by contributing to the emergence of new zoonotic diseases. Mathematical models of infectious disease spread are useful tools to help us understand the drivers of an outbreak, and inform policy decisions. In a world where pandemics are becoming more likely (Morse 2001; Jones et al. 2008; Madhav et al. 2017), the usefulness of modelling techniques is well recognised (Colizza et al. 2007; Dye and Gay 2003; Lessler et al. 2014). Presently, models of infectious disease spread at a population scale are typically based on two phenomena: (i) the infection dynamic, which depends on the characteristics of the disease itself (transmission rate, infectious period, etc.), and (ii) the contact patterns allowing for disease transmission, depending on the transmission routes of the disease.

Here our interest is in the transmission of infectious livestock diseases, where the movements of live animals between farms are known to be one of the main transmission routes (Fèvre et al. 2006). Better knowledge and understanding of contact patterns is a key element for building realistic and useful models. However, detailed models can be computationally costly, and require a substantial amount of data in order to be fitted properly.

When outbreaks occur, policy makers need rapid and robust information to define their strategy and support decisions at the early stages of the epidemic, when data are still limited. A better understanding of the structure of the livestock movement network and its characteristics is therefore useful, both to understand their role in the spread of endemic diseases such as bovine Tuberculosis (Boehm et al. 2009; Palisson et al. 2016; Brooks-Pollock et al. 2014; Green et al. 2008), or BVD (Tinsley et al. 2012)), and to inform policies to control a newly introduced disease in an early stage. The 2001 FMD epidemic provided considerable incentive to study and use livestock movements for network analysis (Ortiz-Pelaez et al. 2006; Christley et al. 2005; Kao et al. 2006; Robinson and Christley 2007; Robinson et al. 2007; Kiss et al. 2006; Vernon and Keeling 2009; Volkova et al. 2010). Analysis of contact networks has proven useful to help identify key actors in terms of disease spread.

Although most infectious diseases can affect several host species (Taylor et al. 2001), network analysis studies have generally focused on single species contact networks. Notable exceptions include Boehm et al. (2009), Nöremark et al. (2011), Kao et al. (2006), and Mohr et al. (2018). Practically, aggregation of movement data from different species is often difficult, because (i) data are recorded separately, often stored in different databases, and possibly managed by different administrative authorities; and (ii) the databases might have different formats or contain different levels of information, and therefore need to be homogenised before use.

The cattle and sheep farming systems are strongly linked in Scotland, because on approximately half of the cattle farms, sheep are also raised. This allows ample opportunity for transmission of diseases between the two species, with FMD and bluetongue virus (BTV) being notable examples of diseases affecting both species. As a consequence, mixed-species farms can link groups of farms that would not be in contact in the network if the species were considered separately. From the network point of view, this might also have consequences for metrics describing the general structure, as well

as the ranking of importance of nodes. It is therefore crucial to explore the multi-species network in order to highlight and quantify potential consequences for disease risk.

Livestock movement network analyses have been performed mostly on static networks, where most analytic results are available (Newman 2018). A static network assumes that the change in the set of contacts are negligible over the course of the epidemic (Enright and Kao 2018). In reality, livestock movements have an inherent temporal component, that are highly relevant to transmission, as they occur on a daily basis and constitute discrete events. As well as being intermittent, movements of livestock are not necessarily consistent over time (Bajardi et al. 2011). A number of studies have shown that dynamic network analyses of livestock movements outperform those from static network analyses, when the aim is an in-depth understanding of disease spreading processes (Lentz et al. 2016; Vidondo and Voelkl 2018; Rossi et al. 2017), or predictions of epidemic risks (Valdano et al. 2015). The study of the dynamics of the cattle-sheep network in Scotland is of interest, because as well as the general dynamics of livestock networks we consider two farming systems which have distinct seasonalities and varying trading behaviours, and the interaction between these systems.

The aim of this work is to understand how the sheep and cattle movement networks interact, and the implication for understanding disease spread. We first analyse the static movement networks, and compare the results of the single- and multi-species networks. Secondly we analyse and compare the cattle, sheep, and multi-species dynamic networks. Finally, we compared results of the static and dynamic network analyses with a disease simulation model explicitly incorporating the temporal dynamics of the network. The dynamic network analysis exhibited important differences between the single-species and the multi-species networks, providing evidence that the premises driving epidemics would not be the same in the single-species and the multi-species networks. These results would have important consequences for disease control. In addition, we showed that dynamic network measures outperform static network measures to identify the most important farms in the network.

Materials and methods

Cattle movement data were obtained from the Cattle Tracing System (CTS), Animal Plant and Health Agency (APHA), and sheep movement data were retrieved from ScotEID, the livestock traceability system for Scotland managed by the Scottish Agricultural Organization Society (SOAS) on behalf of the Scottish Government. We considered movements within Scotland only: between premises, which can be farms, markets or shows. Our interest is in the control of an outbreak after introduction, and therefore movements to or from outside of Scotland were ignored. Births, deaths, and movements to slaughterhouses were also ignored, because the length of the period considered in the study (i.e. four weeks) is short compared to the turnover in the population. Some characteristics of the data are summarised in Table 1.

Overall the sheep population is larger accounting for 6.83 million heads, while the cattle were 1.76 million. There were slightly more sheep farms than cattle farms; of these, 6,039 farms raised cattle and sheep on the same premises (i.e. 50% of the sheep farms, and 56% of the cattle farms). In addition to the farms, the data include 26 auction markets.

Table 1 General characteristics of the setting in figures

Network	Nodes	Movements	Animals moved	Distinct movements
Cattle [mixed]	10,731 [6,039]	89,963	591,933	33,271
Sheep [mixed]	12,078 [6,039]	67,453	2,406,062	26,514
Multi-species	16,758	–	–	53,051

The number of nodes describe the total number of farms raising cattle or sheep reporting at least one animal movement during 2016. The number of movements (batches), the total number of animal moved per species, as well as the headcount of cattle and sheep in Scotland are shown. The number of distinct movements correspond to the number of unique pair of origin and destination in each network

We constructed networks by considering each premises as a node, and animal movements between two premises as a directed link. If one movement of an animal between two premises occurred during the period considered, we assigned a permanent link between these two premises in the static network. The links were weighted depending on the number of animals moved and the probability of an animal being infected:

$$1 - (1 - \mu)^n \tag{1}$$

where μ is the probability of an animal being infected, and n the number of animals moved. The probabilities depend on the type of movement and the species. We used the parameter values estimated by Kao et al. (2006) in the 2001 FMD epidemic in GB:

$\mu_1 = 0.02$ for a sheep movement between two farms;
$\mu_2 = 1$ for a cattle movement between two farms;
$\mu_3 = 0.004$ for a sheep movement from a market;
$\mu_4 = 0.02$ for a cattle movement from a market.

These weights are relevant for an infectious disease similar to FMD, where the infectiousness of sheep is lower than that of cattle (Geering 1967; Gibson and Donaldson 1986; Sørensen et al. 2000; Ferguson et al. 2001).

In the dynamic network each link was annotated with a time variable equal to the date of the animal movement (i.e. we assume these movements occur on a single day).

Static network analysis

We considered the static networks in successive 4-week periods. This allows us to highlight (i) short-term changes in the network structure, which would be relevant for the control of a fast-spreading disease, and (ii) temporal variation according to the season. Livestock movements are generally seasonal, depending on the species and type of production. In Scotland the cattle network typically shows two peaks; the largest is observed in spring, and the second largest in autumn (Robinson and Christley 2006), whereas the sheep network has one main period of high trading activity around September (Kiss et al. 2006). These peaks can be seen in Fig. 1, which shows the number of cattle and sheep moved in each 4-week period of the year. The combined network is represented in Fig. 2, during the Spring and the Autumn peak, 5th adn 10th 4-week periods of the year respectively.
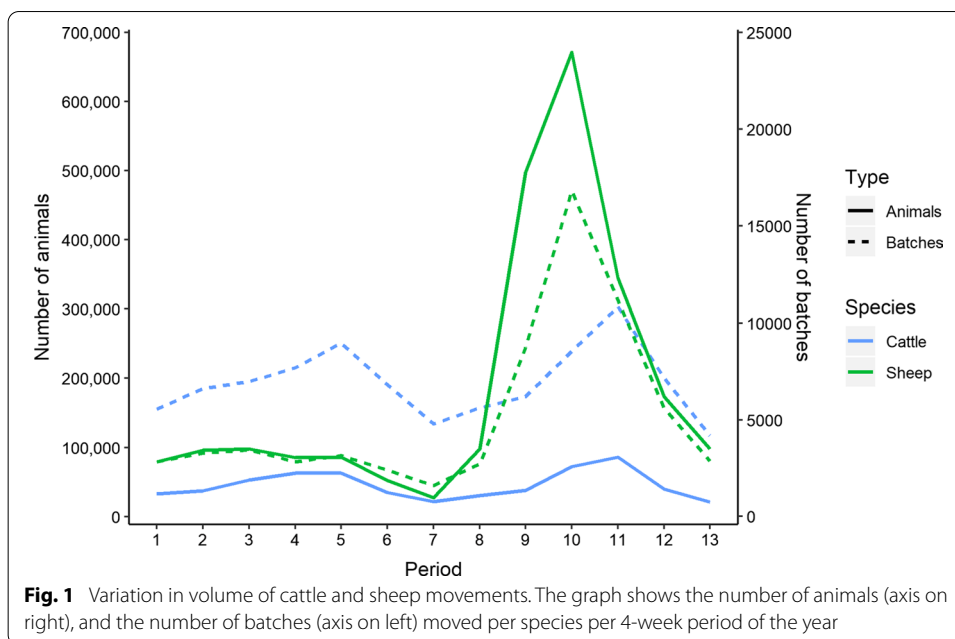
Fig. 1 Variation in volume of cattle and sheep movements. The graph shows the number of animals (axis on right), and the number of batches (axis on left) moved per species per 4-week period of the year

Table 2 Network analysis terminology

Measure	Definition
Average Path Length	Average length of the shortest path between all pairs of nodes of the network (Watts and Strogatz 1998)
Betweenness	Frequency with which a node is in the shortest path between pairs of nodes (Freeman 1978)
Clustering Coefficient	Number of triplets of nodes all connected to each other (closed triplets) over the total number of triplets in the network (Watts and Strogatz 1998)
Component	Subset of nodes of the network for which a path exists between any pair of nodes (Newman 2010)
Giant Weakly Connected Component	Largest component of a directed network, when the directionality of edges is ignored (Newman 2010)
Giant Strongly Connected Component	Largest subset of nodes for which a directed path exists between all pairs of them
Degree	Number of links a node has (Freeman 1978)
Diameter	Length of the shortest path between the two most distant nodes of the network (Wasserman and Faust 1994)
Edge density	Proportion of links between nodes that actually exists in the network, calculated as the number of links, divided by the possible number of links (Wasserman and Faust 1994)
PageRank	A variant of Eigenvector Centrality, primarily used for directed networks: measure of a node's importance while giving consideration to the importance of its neighbors in a directed network (Newman 2010)

We examined the overall characteristics of each network by calculating the average path length, clustering coefficient, edge density, component structure (number of components and sizes of the giant strongly and weakly connected components (GSCC and GWCC respectively) and diameter (definitions in Table 2). These measures were calculated for the single-species networks and the multi-species network, for each 4-week period of the year 2016.

We then calculated node centrality measures for all premises of the network, using the geometric mean degree, betweenness and PageRank (definitions in Table 2). In our case, *degree centrality* corresponds to the number of trading partners a farmer has. Because our network is directed, we differentiate in-degree (denoted $degree_{in}$), i.e. number of premises a farmer buys animals from, and out-degree (denoted $degree_{out}$), i.e. number of premises a farmer sells animals to. The geometric mean of the degree $\sqrt{degree_{in} \times degree_{out}}$ (denoted $GM - Deg$), accounts for the risk of introducing the disease as well as spreading it further. *Betweenness centrality* is the frequency with which a premises is in the shortest path between pairs of premises in the network. Identifying high-betweenness premises is useful from a disease control point of view because these premises represent bridges, which can accelerate the epidemic by spreading diseases to previously unexposed communities of farms. *PageRank centrality* is based on an algorithm used by Google to rank web pages in their search engine (Page et al. 1999). PageRank centrality can capture useful information relevant to diffusion processes, such as epidemics, in networks (Bucur and Holme 2019; Kandhway and Kuri 2017). Data manipulation and analysis have been conducted in R (R Core Team 2019); the 'igraph' package (Csardi and Nepusz 2006) was used for the network analysis.
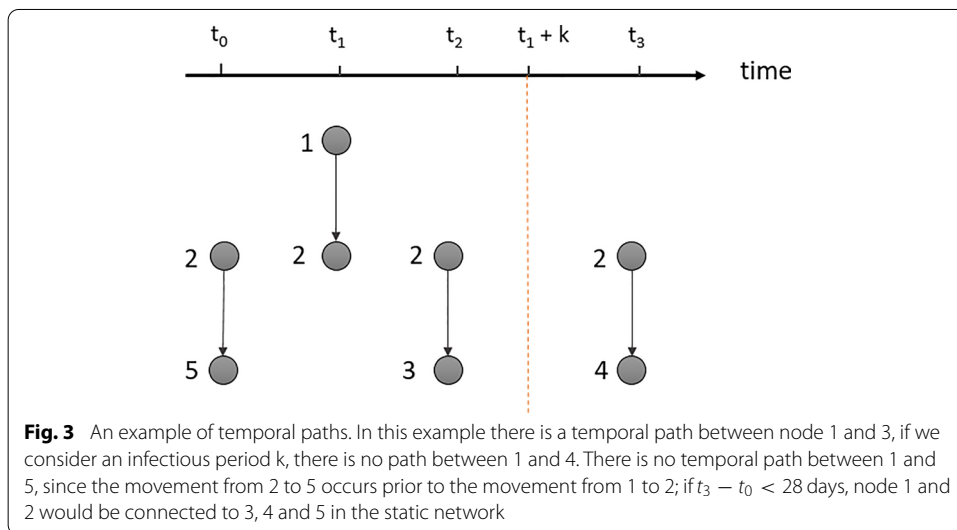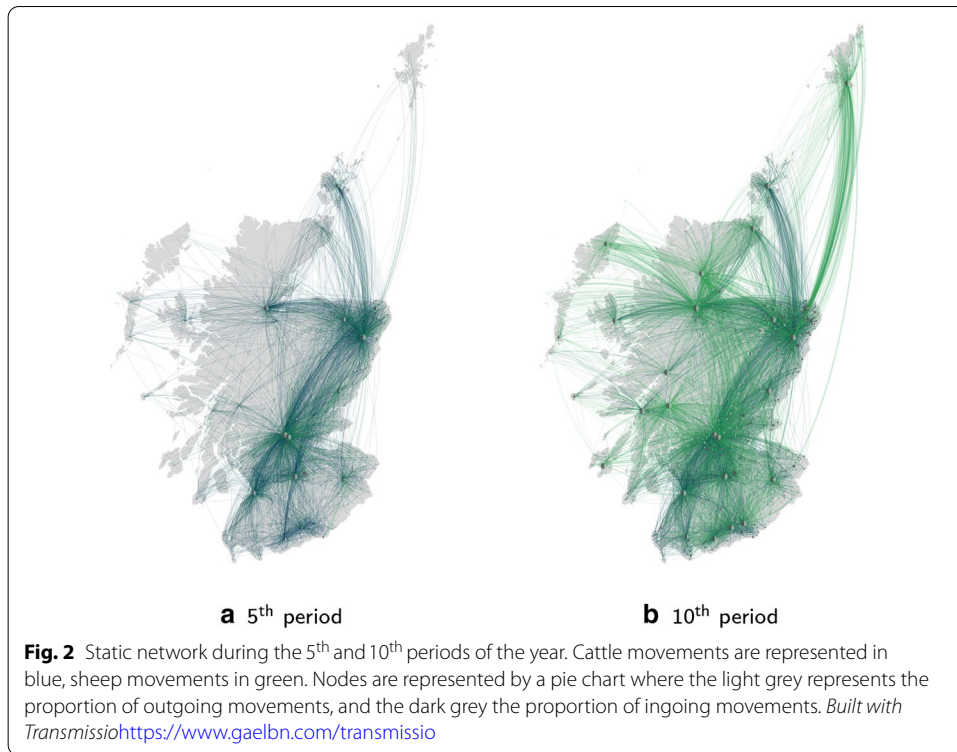
We used these measures to rank the premises in each 4-week period for the single-species and multi-species networks respectively. The premises which showed the highest value (i.e. ranked first) was removed, and the measure was computed again. We focused on the top 100 premises in each network, and refer to these as the *risky premises*. These premises could be targeted for control strategies in the first stages of an epidemic.

We compared the set of risky premises from the multi-species network, with the set of risky premises in the cattle or sheep network by looking at the intersection. The size of the intersection in the set of risky premises between single-species and multi-species networks serves as a measure of how wrong one would be if considering only one species or the other, instead of the combination of both in the context of an outbreak where both species would be involved in the epidemiology.
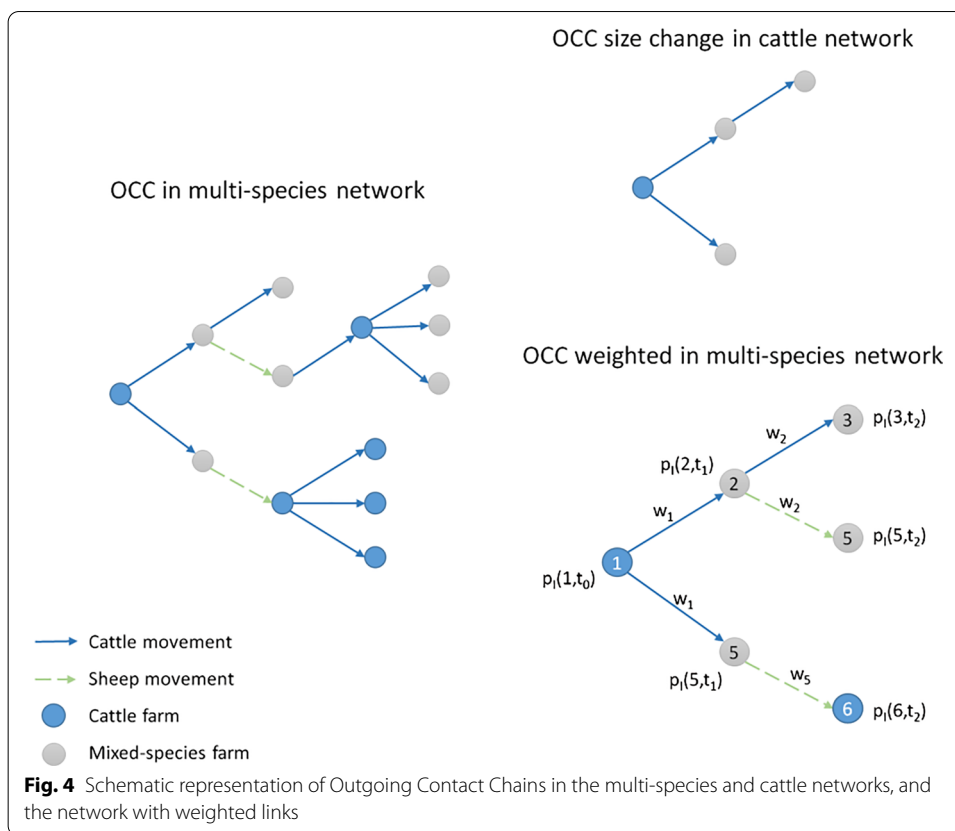
Dynamic network analysis

Livestock movements for trade are occasional and not necessarily recurrent over time. Animal movements occur and are recorded on a daily basis, giving the network a temporal dimension. Thus, it is a system where network dynamics are both likely to be important and are well recorded. In the dynamic network, links are considered as an origin, a destination, and a date of occurrence. Two nodes are in contact if there exists a temporally logical path between them (see Fig. 3).

In order to assess the importance of premises in the dynamic network, we calculated temporal Outgoing Contact Chains (OCC) and Ingoing Contact Chains (ICC), which are derived from the reachability, as described by Holme (2005). Contact chains (CC) were used in the context of diseases in livestock systems by Dubé et al. (2008) under the name of infection chain. Here we used the method previously described by Konschake et al. (2013), where the OCC is defined as the number of premises that can be temporally reached from a primary infected node, considering an infectious period of k days. The ICC is the number of nodes from which a particular node can be temporally reached, accounting for the considered infectious period. We considered

Ruget *et al. Appl Netw Sci*   (2021) 6:15

Page 7 of 20

Fig. 2 Static network during the 5th and 10th periods of the year. Cattle movements are represented in blue, sheep movements in green. Nodes are represented by a pie chart where the light grey represents the proportion of outgoing movements, and the dark grey the proportion of ingoing movements. *Built with Transmissio*https://www.gaelbn.com/transmissio

Fig. 3 An example of temporal paths. In this example there is a temporal path between node 1 and 3, if we consider an infectious period k, there is no path between 1 and 4. There is no temporal path between 1 and 5, since the movement from 2 to 5 occurs prior to the movement from 1 to 2; if $t_3 - t_0 < 28$ days, node 1 and 2 would be connected to 3, 4 and 5 in the static network

an infectious period of seven days, consistent with a fast-spreading FMD-like disease. In other words, the OCC of a premises corresponds to the largest possible epidemic size if the outbreak started in this premises; and the ICC of a premises is proportional to its probability of being infected if an epidemic starts somewhere in the network. We used a method based on a Breadth-First-Search algorithm to calculate the contact chains for limited periods of four weeks. Starting from a designated node, we traverse the network by exploring all the neighbor nodes at the present depth prior to moving

Fig. 4 Schematic representation of Outgoing Contact Chains in the multi-species and cattle networks, and the network with weighted links

on to the nodes at the next depth level. We chose to compute the measure for a period of four weeks because: (i) we are interested in the early stage of the epidemic before the outbreak is detected and a movement ban applied; (ii) this makes our results comparable with the results of the static network analysis which had been performed for the same periods.

Unweighted in- and out-going contact chains

We first considered unweighted links to avoid making assumptions about the characteristics of the disease. This corresponds to the worst case scenario where the probability of transmission is certain given a link between premises. We compared the sets of risky premises according to the geometric mean of their contact chains ($GM - CC$), defined as $\sqrt{ICC \times OCC}$, in the different networks, i.e. comparing the top hundred risky premises in the single-species network and the multi-species network. This measure has been proven useful to assess the infection potential for fast spreading disease (Rossi et al. 2017). We also looked at the changes in the set of risky premises according to geometric mean degree and geometric mean contact chain sizes for the same network, to understand the difference between considering a static or dynamic network.

In order to highlight potential shifts in estimated risk between the multi-species and the cattle systems, we looked at the difference in maximal epidemic size between these two systems, by quantifying the change in the OCC of cattle premises taking into account the movements of both species or cattle movements only (see schematic

Ruget *et al. Appl Netw Sci*   (2021) 6:15

Page 9 of 20

representation in Fig. 4). The maximal size of an epidemic is a critical parameter, often used in epidemiological studies to quantify the potential impact of an outbreak. Because we computed the OCC for a limited period of 28 days, the OCC is the potential size of the epidemic after 28 days of uncontrolled spread. We calculated for all cattle premises the factor by which their OCC was multiplied in the multi-species network; we called this factor the *multiplication factor*, defined as:

$$\frac{OCC_M}{OCC_C} \tag{2}$$

where OCC_M and OCC_C are the OCC in the multi-species and cattle networks respectively.

Weighted in- and out-going contact chain in a multi-species network

Assuming that all movements are equally important—regardless of the species type, the number of animals, or the characteristics of the premises—neglects important and potentially useful information affecting the spread of a disease. We therefore also calculated weighted Outgoing Contact Chains (OCC_w) where the weights are equal to Formula 1 and correspond to the probability of transmission given that the node is infected. We consider a network defined as a set of nodes V, and the set of edges E $j \xrightarrow{t,w_j} i$ where $i, j \in V$, t is a time, w_j is a weight. We denote the probability of being infected for a node i at time t, $p_I(i, t)$, the complementary probability of not being infected $p_{NI}(i, t)$. The probability of disease transmission for a movement from j to i at time t is consequently equal to $p_I(j, t-1) \times w_j$.

We adapted the algorithm used in the previous section, using a similar method to the one proposed by Enright and Kao (2016). In the initial conditions, all nodes are susceptible, except one root node u. At each discrete time step, we identify all edges E $j \xrightarrow{t,w_j} i$, where j has a non null probability of being infected. The probability of not being infected for the nodes i is updated by multiplying it by the probability for the edge $j \xrightarrow{t,w_j} i$ to not transmit infection, which is $1 - p_I(j, t-1) \times w_j$. We keep track of the probability of not having been infected so far, to consider cases of multiple potential infections. We present this algorithm as pseudo code in Algorithm 1.

Algorithm 1: Algorithm to calculate OCC_w

Result: Average size of the epidemic after 28 days: OCC_w

For all $i \in V$, $p_I(i, t_0) = 0$, except a root node u for which $p_I(u, t_0) = 1$;

for $t \in [t_0 + 1, t_0 + 27]$ **do**

  for $i \in V$ **do**

    for $j \in V$ *such that* $(j \xrightarrow{t,w} i) \in E$ **do**

      $p_{NI}(i,t) = p_{NI}(i, t-1) \times [\prod_j (1 - p_I(j, t-1) \times w_j)];$

    end

  end

end

$OCC_w = \sum_i (1 - p_{NI}(i, t_0 + 27))$

return OCC_w

Likewise, we calculated the weighted ICC (ICC_w) for all premises in the multi-species network and the different periods of the year. We ranked premises according to the geometric mean of weighted contact chains ($GM - CC_w$ defined as $\sqrt{ICC_w \times OCC_w}$), expecting this ranking to be relevant to the prioritisation of control strategies.

Disease simulations

To investigate agreement between the network analysis results and a more realistic situation, we stochastically simulated transmission of a fast-spreading disease in both cattle and sheep. The simulation is based on a Susceptible-Infected-Recovered (SIR) metapopulation model, compatible with an immunising infection. The time step is one day, to take into account the daily recorded animal movements, and disease transmission is frequency-dependent. We considered an infection with asymmetric transmission risk, where the rate of effective contacts β has the highest value between cattle, and the lowest from sheep to cattle (Table 3). The contact rate between sheep, and from cattle to sheep have intermediate values. The parameter values were chosen arbitrarily within the range of plausibility for a fast-spreading disease like FMD (Keeling 2005). Parameter values are given in Table 3.

For simplicity, we simulated epidemics starting only in premises having an OCC size greater than 100 premises, that is, premises that can potentially lead to an epidemic of 100 premises or more. The simulations were run for a limited period of four weeks, starting at the first day of each 4-week period of the year. We used the SimInf package (Widgren et al. 2019) in R to perform 100 simulations per seed, and recorded the size of the epidemic after four weeks, as well as the number of times a premises was involved in the outbreak over all simulations for each period.

We defined an indicator of the *epidemic risk* for each premises and each period as, $ER = N_E \times N_I$, where N_E is the average size of the epidemic at four weeks, and N_I is the number of times the premises is infected during the epidemic and is proportional to the probability of getting infected.

To evaluate the performance of network measures in identifying the most important farms, we compared the 100 premises with the highest *ER* according to the simulations with the 100 most risky premises according to the different measures ($GM - Deg$,

Table 3 Daily rates for the parameters in the simulation model

Parameter	Value	Definition
β_{CC}	0.2	Effective contact rate between cattle
β_{SC}	0.19	Effective contact rate between susceptible sheep and infectious cattle
β_{SS}	0.17	Effective contact rate between sheep
β_{CS}	0.15	Effective contact rate between susceptible cattle and infectious sheep
γ	0.14	Recovery rate

betweenness, PageRank, $GM - CC$, and $GM - CC_w$). For this comparison we considered only farms in the ranking, because markets and shows are already known to be high risk and would be targeted first for control measures.

Results

The number of sheep movements was consistently higher than the number of cattle movements (Fig. 1). The highest volume of trading activity in the Scottish network occurs in late summer to early autumn, when the sheep movement volumes peak. Overall, most of the recorded movements went through markets, accounting for 75% of the trading operations for cattle, and 93% for sheep.

Static network analysis

The sheep network is more dense than the cattle and the multi-species networks, whereas the cattle network is generally more clustered (Table 4). As expected, average path length is longest in the cattle network, and shortest in the multi-species network. These shorter paths between pairs of premises enable faster spread of diseases. The sheep network is generally more connected, with typically only a single weakly connected component, while the cattle network is sparser, counting on average 187.1 components during any 4-week snapshot. The multi-species network is less fragmented than the cattle-movement network; this indicates that sheep movements connect different components of mixed or cattle premises in the multi-species network, which were disconnected in the cattle network.

The variation in sizes of the components follows the same pattern as the seasonality of the movements (Fig. 5), i.e. the sizes of the strongly and weakly connected components show two peaks in the cattle network during the 5th and 11th 4-week periods of the year, whereas the sizes of both components sharply increase around the 10th period of the year in the sheep network. Moreover, the membership to components across periods is mostly consistent, with over 60% of the premises constituting the GWCC of the cattle networks being the same in periods 5 and 11, when cattle movements peak. In the multi-species network, over 70% of the premises in the GWCC are remaining the same between these periods. In addition, the size of the GSCC in the multi-species network is always larger than the sum of the GSCC in the cattle and the sheep networks. This highlights how interconnected the two farming systems are. This is important, because the

Ruget *et al. Appl Netw Sci*   (2021) 6:15

Page 12 of 20

Table 4 Static network measures for the 4-week animal movement networks in the year 2016 (mean [min, max])

	Edge Density ($\times 10^{-3}$)	Clustering ($\times 10^{-3}$)	Mean Path Length	Diameter	No. of Components
Multi-Species	2.3 [1.7, 3.1]	0.88 [0.6, 1.7]	4.2 [4.1, 4.7]	12 [10, 14]	138.6 [58, 218]
Cattle	3.2 [2.5, 3.7]	1.5 [0.9, 2.6]	4.5 [4.2, 5.4]	13 [11, 16]	187.1 [133, 275]
Sheep	4.9 [2.2, 9.0]	0.13 [0.01, 0.55]	4.32 [2.8, 4.9	9.8 [6, 13]	1.8 [1, 4]

Average values are calculated over the thirteen 4-week periods of the year 2016
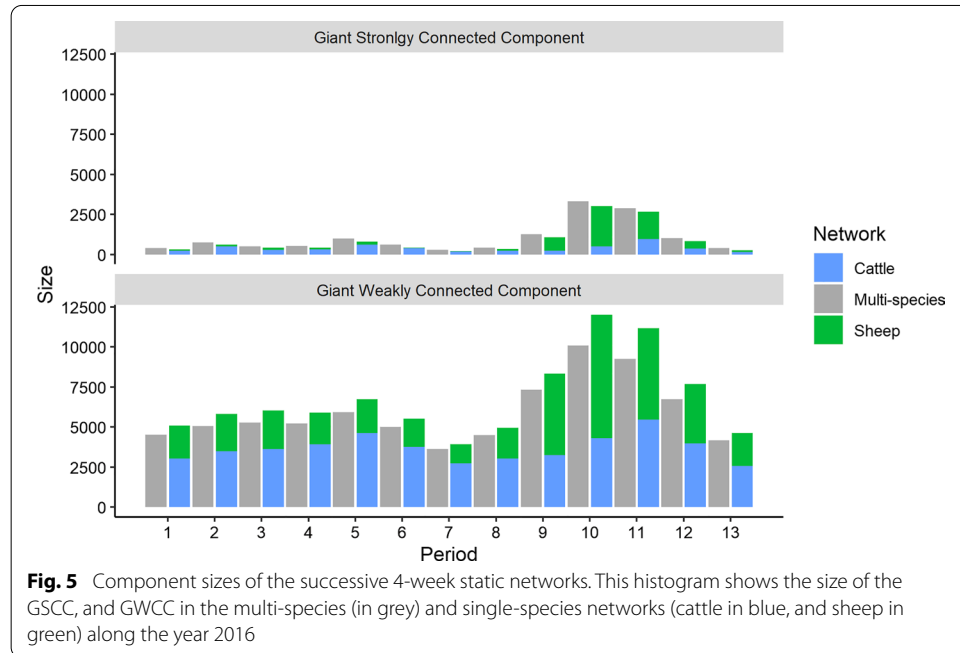
Fig. 5 Component sizes of the successive 4-week static networks. This histogram shows the size of the GSCC, and GWCC in the multi-species (in grey) and single-species networks (cattle in blue, and sheep in green) along the year 2016

size of the GSCC corresponds to a lower bound on the maximum number of nodes that a newly introduced infectious agent might reach.

Figure 6 shows on the one hand that the three measures used are correlated, consistently identifying mostly the same premises as risky; but on the other hand that, for all periods of the year, the most risky premises in the multi-species network are more similar to the ones in the cattle network. The graph also shows, during the 9th and 10th periods, an increase in the number of identical risky premises between the sheep and the multi-species systems, whilst a decrease in this number is observed between the multi-species and cattle network. The majority (more than 96%, for all 4-week periods) of risky premises in the multi-species network were also considered most risky in the cattle network, for all three network measures considered. However, only around 20% of the risky premises in the multi-species network were also identified as most risky in the sheep network, in any 4-week period.

Dynamic network analysis

In the dynamic network analysis, the risky premises are the top 100 premises with the largest $GM - CC$. The set of identified risky premises in the multi-species network

Fig. 6 Comparison of the single-species and multi-species static network analysis results. This graph details the variation in the number of identical risky premises between the multi-species and each of the single-species networks (cattle in blue, and sheep in green) along the year for the different static network measures
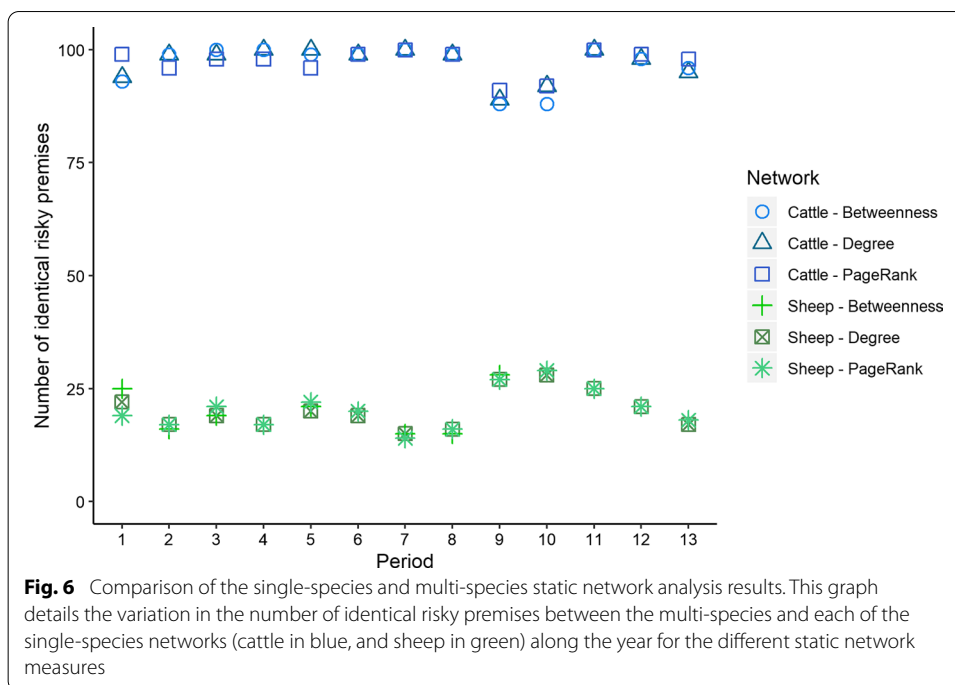
are substantially different from the sets in the cattle or the sheep movement networks (Table 5). On average only 47.2% of the most risky premises in the multi-species network are most risky in the cattle networks as well, and 32.4% in the sheep networks.

Although for most of the year, the set of risky premises in the multi-species network is more similar to the one in the cattle network (Fig. 7), during the 9th and 10th periods of the year, the trend reverses with the risky premises in the multi-species network becoming more similar to the ones in the sheep network. In addition, over all periods on average 29% of the risky premises in the multi-species network are not identified as risky in any of the single-species networks. This suggests that some premises with the largest ICC and OCC in the multi-species network exhibit large contact chains only through combination of cattle and sheep movements. The proportion of risky premises of this kind can be as high as 72% during the 11th period.

These results differ considerably from those of the static analysis. We compared the sets of risky premises according to $GM - Deg$ in the static networks, to the risky premises according to $GM - CC$ in the dynamic networks (grey cells in Table 5). The percentage similarity between the risky premises according to these two measures is low, with on average only around 17% of the premises being the same.

Overall, 99.5% of the cattle premises considered have a larger OCC when including sheep movements, and therefore a multiplication factor greater than 1. Very few cattle premises see their OCC unchanged when sheep movements are considered, even during the period of low activity in the sheep network (Fig. 8). Sheep movements contribute to the construction of significantly larger OCC in most cases: over all periods of the year, half of the premises (54%) see their OCC multiplied by at least 2. As expected, large increases in OCC is more consistent during the period of high activity

Ruget *et al. Appl Netw Sci* (2021) 6:15

Page 14 of 20

Table 5 Table showing average results in dynamic network analysis and comparing with static network analysis results

	GM-CC	GM-CC/GM-Deg
Multi-species	100	16.1 (± 2.4)
Cattle	47.2 (± 16.1)	16.8 (± 3.3)
Sheep	32.4 (± 12.3)	17.9 (± 9.2)

Number of identical *risky premises* identified in both single- and multi-species networks , and by dynamic and static measures. Values are averages over periods of the year, with standard deviations in brackets. GM-CC: geometric mean Contact Chain; GM-Deg: geometric mean degree
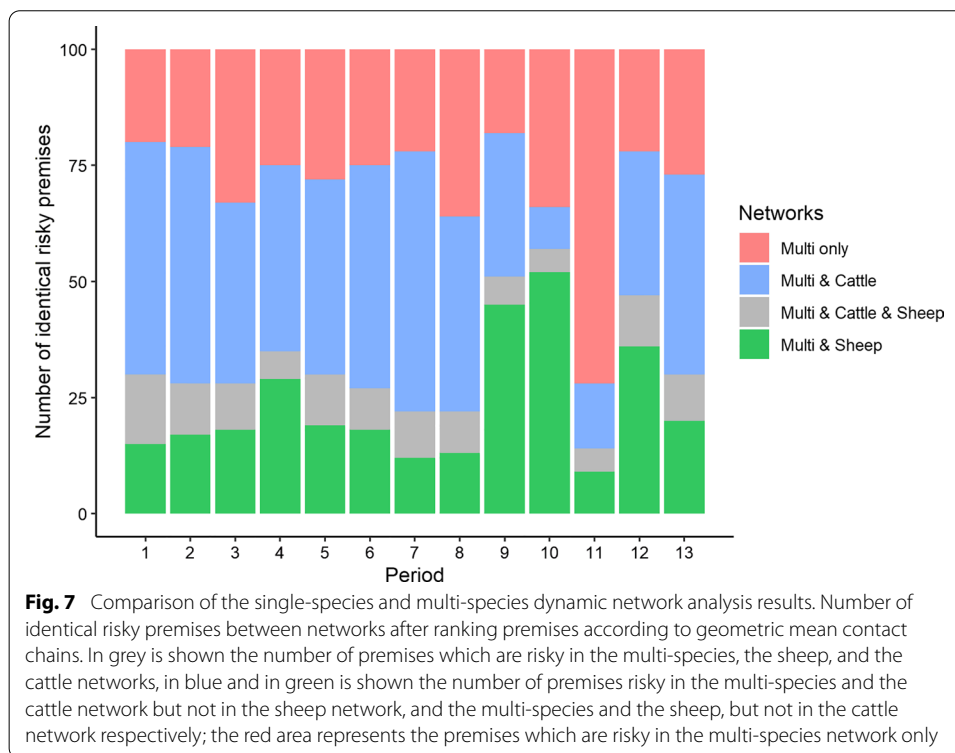
Fig. 7 Comparison of the single-species and multi-species dynamic network analysis results. Number of identical risky premises between networks after ranking premises according to geometric mean contact chains. In grey is shown the number of premises which are risky in the multi-species, the sheep, and the cattle networks, in blue and in green is shown the number of premises risky in the multi-species and the cattle network but not in the sheep network, and the multi-species and the sheep, but not in the cattle network respectively; the red area represents the premises which are risky in the multi-species network only

in the sheep network, i.e. 10th period of the year. During this period, OCC of cattle premises are multiplied by an average of 8.9.
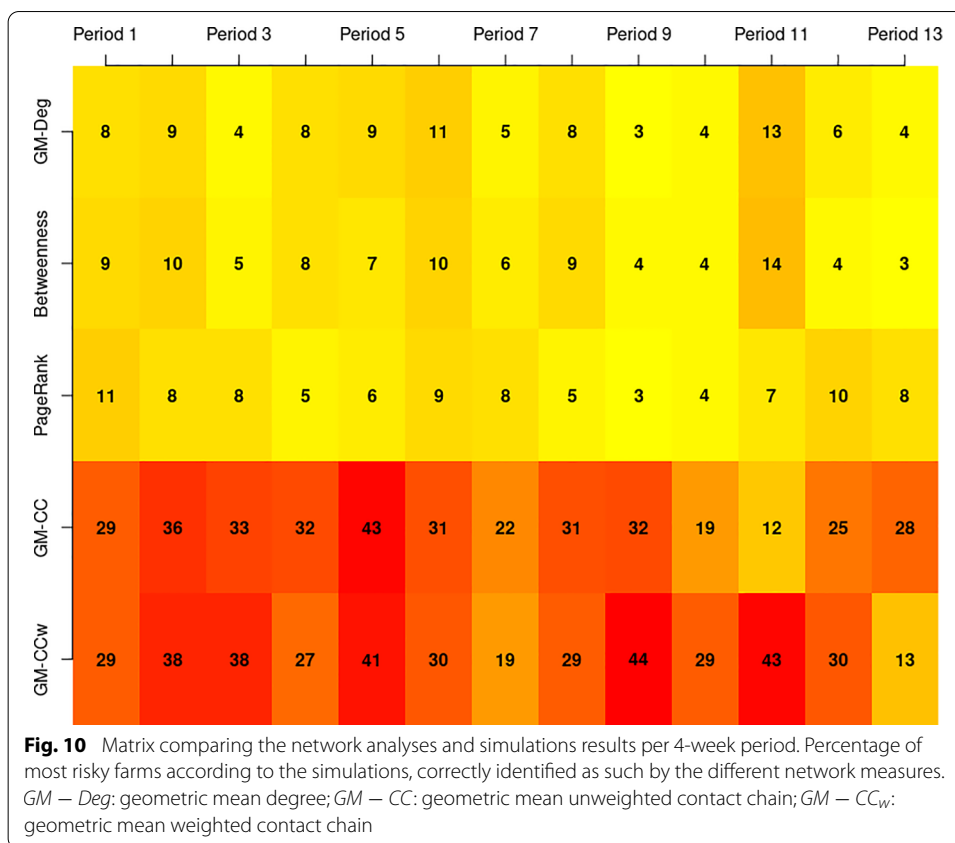
Almost all cattle premises see their OCC increased in the multi-species network (Fig. 9), but the range of OCC values as well as the range of increase are variable. The OCC values are similar in the 4th and 10th periods are similar for the cattle network, but differ considerably for the multi-species network, exhibiting sharper increases of the OCC during the 10th period. During the 10th period, on average 2513 additional premises can be reached in the multi-species network, which would not be reached through cattle movements only.

Disease simulations

The static network measures identified at most 11% of the top 100 farms involved in the simulation, and their performance is very poor in most time periods (Fig. 10).

Fig. 8 Multiplication factor distribution considering only cattle premises having an OCC of more than 10 premises

Fig. 9 Multiplication factor of cattle premises in the multi-species network for three periods of the year. The graph shows the multiplication factor in log scale according to the OCC in the cattle network. Only premises with an OCC of at least ten premises are shown

Dynamic network measures offer a clear improvement. The static measures identified on average only 7% of the 100 farms potentially most important in the epidemic, whereas the dynamic measures identified on average over 30% of the main actors of an outbreak (after markets and shows).

Fig. 10 Matrix comparing the network analyses and simulations results per 4-week period. Percentage of most risky farms according to the simulations, correctly identified as such by the different network measures. *GM − Deg*: geometric mean degree; *GM − CC*: geometric mean unweighted contact chain; *GM − CC_w*: geometric mean weighted contact chain

Discussion

The first objective of this study was to explore the Scottish cattle-sheep multi-species network characteristics to determine if this network was substantially different than the single-species ones. The temporal multi-species network exhibited significant differences in its structure, compared to the temporal cattle network. We found that more than a half of the risky premises in the multi-species network, were not identified as risky in the cattle network. If the cattle network was used to identify risky premises in a context of a disease involving sheep and cattle, these premises would be missed. More importantly, a number of risky premises are identified as such in the multi-species network only: 72% in the 11th period of the year, indicating that their risk is derived from interaction between the two farming systems. These differences indicate that, not only are the risks associated with multi-species epidemics higher, the premises likely to be driving those risks are also different. These differences are not captured by a static network representation of the system, and underlines the importance of temporality in livestock movement networks.

In the temporal network, most OCC of cattle premises were significantly larger in the multi-species network than in the cattle network, for all periods of the year. By constructing longer contact chains, interaction between the farming systems increases risk of larger epidemics throughout the year, and not only during the period of intensive trading in the sheep farming system, as one may expect.

Ruget *et al. Appl Netw Sci*   (2021) 6:15

Page 17 of 20

Finally, when disease spread was simulated in the multi-species system, the temporal measures performed better at identifying the most important farms than the static network measures. The inability of static descriptors to reliably predict outbreak risk is expected (Vidondo and Voelkl 2018), when the pattern of contacts changes over timescales that are short compared to disease generation times. The measures based on contact chains take into account the more important aspects of the movement networks, such as temporal paths, which are relevant in the occurrence of an epidemic (Holme 2005; Lentz et al. 2016). Usefulness of a measure based on ingoing and outgoing contact chains for assessing disease risk was already confirmed by other studies (Frössling et al. 2012; Vidondo and Voelkl 2018). We used similar metrics, but took into account transmission probabilities. Our weighted contact chains performed better than the simpler contact chains. Here we used weightings in line with the 2001 FMD epidemic in the UK, which substantially improved the predictive power of the metric. This shows that even a mildly informed choice of transmission probability per link, without prior knowledge on the disease parameters, gives a better prediction of risk than an unweighted network.

Our work reinforces the importance of incorporating differences in transmission probability where they exist within a system (e.g. differences between host-species in a multi-species system (Dobson 2004)). Significant variation can lead to critically different disease dynamics (Lloyd-Smith et al. 2005) which must be captured for predictive modelling. In our weighted network, the weights incorporate variation due to both the volume and species traded, which allows our network metrics to explicitly include this information. Similar methods based on contact chains have included more explicit simulation rather than metric calculation: e.g. Knific et al. (2020) reports work which filters a weighted temporal network and then simulates scenarios with differing transmission probabilities, thus somewhat decoupling the disease model from the underlying network. Our metric differs in that it incorporates transmission probabilities directly into network and metric, and is thus most useful when consider a particular pathogen with known transmission probabilities that vary by category of edge.

Future work is possible to improve the applicability of this approach to FMD, to broaden its use to other diseases and systems, and to improve the computational performance of the method. While we demonstrated the importance of using a multi-species network to understand transmission of an FMD-like disease, additional work remains to be done. Cattle and sheep are not the only species vulnerable to FMD: in a future FMD epidemic, network layers including other species (e.g. pigs) could be included, as well as non-trade layers that could incorporate transmission risk due to shared equipment or human movements.

Because characteristics of the disease are included in our weighted temporal network construction, a number of adaptations would be needed to apply our work to other diseases. In particular, the time of aggregation has important consequences for the network's interaction with the pathogen (Bajardi et al. 2011). The time scale in our approach should therefore be adapted to correspond to both the infectious period of a disease and the time scale of network dynamics (Kao et al. 2007).

While we have implemented our methods with reasonably efficient code, our focus has been on assessing the usefulness of the approaches as opposed to producing code optimised for speed and memory requirement. Computational performance could likely be

improved, and further work may be required to deploy these approaches on very large or very dense networks.

Weighted contact chains can be a powerful tool to inform decisions in the early stages of an epidemic because it only relies on animal movement data that are immediately available. As well as being easy and fast to compute, it is deterministic, which means the metric can be calculated in a single computational run. In addition, the method proposed showed that weighting the network with reasonable transmission probabilities helps to improve the prediction of risk, which could aid decision making in the early stages of an epidemic when disease parameters are still unknown.

Author details
[1] The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, Scotland, UK. [2] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland G12 8QQ, UK. [3] INRAE, Oniris, BIOEPAR, 44300 Nantes, France. [4] School of Computing Science, University of Glasgow, Glasgow, Scotland G12 8QQ, UK. [5] Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Scotland, UK.

References
Bajardi P, Barrat A, Natale F, Savini L, Colizza V (2011) Dynamical patterns of cattle trade movements. PloS one 6(5):e19869
Boehm M, Hutchings MR, White PC (2009) Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine tb among badgers and cattle. PLoS One 4:4
Brooks-Pollock E, Roberts GO, Keeling MJ (2014) A dynamic model of bovine tuberculosis spread and control in great britain. Nature 511(7508):228–231
Bucur D, Holme P (2019) Beyond ranking nodes: Predicting epidemic outbreak sizes by network centralities. arXiv preprint arXiv:1909.10021
Christley RM, Robinson SE, Lysons R, French NP (2005) Network analysis of cattle movement in Great Britain. Proc Soc Vet Epidemiol Prev Med 234–243
Colizza V, Barrat A, Barthelemy M, Valleron A-J, Vespignani A (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med 4:1
Csardi G, Nepusz T (2006) The igraph software package for complex network research. Int J Complex Syst :1695. http://igraph.sf.net
Dobson A (2004) Population dynamics of pathogens with multiple host species. Am Nat 164(S5):S64–S78
Dubé C, Ribble C, Kelton D, McNab B (2008) Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in ontario, canada. Transboundary Emerg Diseases 55(9–10):382–392

Dye C, Gay N (2003) Modeling the sars epidemic. Science 300(5627):1884–1885

Enright J, Kao RR (2016) A fast algorithm for calculating an expected outbreak size on dynamic contagion networks. Epidemics 16:56–62

Enright J, Kao RR (2018) Epidemics on dynamic networks. Epidemics 24:88–97

Ferguson NM, Donnelly CA, Anderson RM (2001) The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions. Science 292(5519):1155–1160

Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1(3):215–239

Frössling J, Ohlson A, Björkman C, Håkansson N, Nöremark M (2012) Application of network analysis parameters in risk-based surveillance-examples based on cattle trade data and bovine infections in sweden. Prev Vet Med 105(3):202–208

Fèvre EM, Bronsvoort BMdC, Hamilton KA, Cleaveland S (2006) Animal movements and the spread of infectious diseases. Trends in Microbiology, 14(3):125–131. https://doi.org/10.1016/j.tim.2006.01.004.http://www.sciencedirect.com/science/article/pii/S0966842X06000175

Geering W (1967) Foot-and-mouth disease in sheep. Aust Vet J 43(11):485–489

Gibson C, Donaldson A (1986) Exposure of sheep to natural aerosols of foot-and-mouth disease virus. Res Vet Sci 41(1):45–49

Green DM, Kiss IZ, Mitchell AP, Kao RR (2008) Estimates for local and movement-based transmission of bovine tuberculosis in british cattle. Proc R Soc B Biol Sci 275(1638):1001–1005

Holme P (2005) Network reachability of real-world contact sequences. Phys Re E 71(4):046119

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P (2008) Global trends in emerging infectious diseases. Nature 451(7181):990–993

Kandhway K, Kuri J (2017) Using node centrality and optimal control to maximize information diffusion in social networks. IEEE Trans Syst Man Cybernet Syst 47(7):1099–1110

Kao RR, Danon L, Green DM, Kiss IZ (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

Kao RR, Green DM, Johnson J, Kiss IZ (2007) Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the uk. J R Soc Interface 4(16):907–916

Keeling MJ (2005) Models of foot-and-mouth disease. Proc R Soc B Biol Sci 272(1569):1195–1202

Kiss IZ, Green DM, Kao RR (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

Knific T, Ocepek M, Kirbiš A, Lentz HHK (2020) Implications of cattle trade for the spread and control of infectious diseases in slovenia. Front Vet Sci 6:454. https://doi.org/10.3389/fvets.2019.00454 (**ISSN 2297-1769**)

Konschake M, Lentz HHK, Conraths FJ, Hövel P, Selhorst T (2013) On the robustness of in- and out-components in a temporal network. PLOS ONE 8(2):e55223. https://doi.org/10.1371/journal.pone.0055223

Lentz HH, Koher A, Hövel P, Gethmann J, Sauter-Louis C, Selhorst T, Conraths FJ (2016) Disease spread through animal movements: a static and temporal network analysis of pig trade in germany. PloS ONE 11:5

J. Lessler, I. Rodriguez-Barraquer, D. A. Cummings, T. Garske, M. Van Kerkhove, H. Mills, S. Truelove, R. Hakeem, A. Albarrak, N. M. Ferguson, et al. Estimating potential incidence of mers-cov associated with hajj pilgrims to saudi arabia, 2014. *PLoS currents*, 6, 2014

Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438(7066):355–359

Madhav N, Oppenheim B, Gallivan M, Mulembakani P, Rubin E, Wolfe N (2017) Pandemics: risks, impacts, and mitigation. In: Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition. The International Bank for Reconstruction and Development/The World Bank,

Mohr S, Deason M, Churakov M, Doherty T, Kao RR (2018) Manipulation of contact network structure and the impact on foot-and-mouth disease transmission. Prev Vet Med 157:8–18

Morse SS (2001) Factors in the emergence of infectious diseases. In: Plagues and politics, pp 8–26. Springer

Newman M (2010) Networks. Mar 2010. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Newman M (2018) Networks. Oxford University Press, Oxford

Nöremark M, Håkansson N, Lewerin SS, Lindberg A, Jonsson A (2011) 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

Ortiz-Pelaez A, Pfeiffer D, Soares-Magalhaes R, Guitian F (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. Prev Vet Med 76(1–2):40–55

Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab

Palisson A, Courcoul A, Durand B (2016) Role of cattle movements in bovine tuberculosis spread in france between 2005 and 2014. PLoS One 11:3

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Robinson S, Christley R (2007) Exploring the role of auction markets in cattle movements within great britain. Prev Vet Med 81(1–3):21–37

Robinson S, Everett M, Christley R (2007) Recent network evolution increases the potential for large epidemics in the british cattle population. J R Soc Interface 4(15):669–674

Robinson SE, Christley RM (2006) Identifying temporal variation in reported births, deaths and movements of cattle in britain. BMC Vet Res 2(1):11

Rossi G, De Leo GA, Pongolini S, Natalini S, Zarenghi L, Ricchi M, Bolzoni L (2017) The potential role of direct and indirect contacts on infection spread in dairy farm networks. PLoS Comput Biol 13:1

Sørensen JH, Mackay D, Jensen C, Donaldson AI (2000) An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. Epidemiol Infection 124(3):577–590

Ruget *et al. Appl Netw Sci*   (2021) 6:15

Page 20 of 20

Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. Philos Trans R Soc Lond Ser B Biol Sci 356(1411):983–989

Tinsley M, Lewis FI, Brülisauer F (2012) Network modeling of bvd transmission. Vet Res 43(1):11

Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V (2015) Predicting epidemic risk from past temporal contact data. PLoS Comput Biol 11(3):e1004152

Vernon MC, Keeling MJ (2009) Representing the uk's cattle herd as static and dynamic networks. Proc R Soc B Biol Sci 276(1656):469–476

Vidondo B, Voelkl B (2018) Dynamic network measures reveal the impact of cattle markets and alpine summering on the risk of epidemic outbreaks in the swiss cattle population. BMC Vet Res 14(1):88

Volkova VV, Howey R, Savill NJ, Woolhouse ME (2010) Sheep movement networks and the transmission of infectious diseases. PloS one 5:6

Wasserman S, Faust K et al (1994) Social network analysis: methods and applications, vol 8. Cambridge University Press, Cambridge

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440

Widgren S, Bauer P, Eriksson R, Engblom S (2019) SimInf: An R package for data-driven stochastic disease spread simulations. J Stat Softw 91(12):1–42. https://doi.org/10.18637/jss.v091.i12

Publisher's Note