Replicating disease spread in empirical cattle networks by adjusting the probability of infection in random networks.

A. J. Duncan[a,*], G. J. Gunn[b], C. Umstatter[c], R. W. Humphry[b]

[a] *Inverness College UHI, Longman Campus, 3 Longman Road, Longman South, Inverness, IV1 1SA*
[b] *Epidemiology Research Unit, SRUC (Scotland's Rural College), Drummondhill, Stratherrick Road, Inverness, IV2 4JZ*
[c] *Agroscope, Institute for Sustainability Sciences ISS, Tänikon 1, 8356 Ettenhausen, Switzerland*

[*] Corresponding author
Email address: andrew.duncan.ic@uhi.ac.uk (A. J. Duncan)

Abstract

Comparisons between mass–action or "random" network models and empirical networks have produced mixed results. Here we seek to discover whether a simulated disease spread through randomly constructed networks can be coerced to model the spread in empirical networks by altering a single disease parameter – the probability of infection. A stochastic model for disease spread through herds of cattle is utilised to model the passage of an SEIR (susceptible–latent–infected–resistant) through five networks. The first network is an empirical network of recorded contacts, from four datasets available, and the other four networks are constructed from randomly distributed contacts based on increasing amounts of information from the recorded network. A numerical study on adjusting the value of the probability of infection was conducted for the four random network models. We found that relative percentage reductions in the probability of infection, between 5.6% and 39.4% in the random network models, produced results that most closely mirrored the results from the empirical contact networks. In all cases tested, to reduce the differences between the two models, required a reduction in the probability of infection in the random network.

Keywords: Network; Mass–action; Disease; Recorded contacts; SEIR simulation

1. Introduction

The assumption of random interactions, or mass–action mixing, is a method widely used in the modelling of disease (Anderson and May, 1991; Brauer et al., 2000; De Jong et al., 1995). With cheaper and easier methods of data capture now available to record contact networks (Craft and Caillaud, 2001) homogeneously mixed networks or "random networks" have been tested against the recorded contact networks with varying results (Duncan et al., 2012; Hamede et al., 2012; Kleinlützum et al., 2013; Salathé et al., 2010). In this publication we seek to discover whether a simple model of disease spread, based on the principles of homogeneous mixing, can approximate a recorded network if the probability of infection is suitably adjusted. If this is possible, we will also investigate: whether the simplicity of the model affects the closeness of fit to the recorded network; whether there is consistency in the adjustment of the probability of infection across a variety of random network models and whether there is a relationship between the network properties, through values of network metrics, and the adjustment to the probability of infection.

Results from comparisons of simulated disease spread on random and structured network, whether recorded, empirically derived (i.e. extrapolated from empirical data) or theoretically constructed, have been mixed. Some studies have found random networks to be a suitable substitute for structured network models (Bouma et al., 1995; Dobson and Meagher, 1996; Shirley and Rushton, 2005a) whilst others have found it inadequate (Barlow, 2000; D' Amico et al., 1996; Hamede et al., 2012; Porphyre et al., 2008; Shirley and Rushton, 2005b). For

inter–herd contact networks, rather than the intra–herd networks discussed herein, it has been shown that models should be at least based on any movement data available (Vernon and Keeling, 2009). The modification of the transmission rate of disease on a random network model has been shown to provide a good representation of the results from theoretically constructed networks (Keeling, 2005). Simplified models of a complete contact network which take account of rewiring or preferential mixing show closer agreement than a mean–field model (random/mass–action mixing) when modelling Tasmanian devil facial tumour disease (Hamede et al., 2012) and it was found that the networks had highly connected animals, which would not be found in random networks. When modelling spread of influenza in high school students (Salathé et al., 2010), it was found that a small–world network (Watts and Strogatz, 1998) with a high proportion of repeated contacts fitted the recorded data best, but a homogeneous (random/mass–action) mixing model might be sufficient.

In our previous work (Duncan et al., 2012) we presented two stochastic models of the passage of an SEIR (susceptible–latent–infected–resistant) disease through herds of cattle. One model was based on a contact network constructed via continuously recorded interaction data from two herds of cattle, the other, a matching network constructed using the assumption of random mixing. Four recorded contact datasets were produced by attaching proximity data loggers (Drewe et al., 2012; Swain and Bishop-Hurley, 2007) to two separate herds of cattle during two separate recording periods. For each dataset the network constructed using the principles of random mixing had the same number of

contacts as the recorded network but these contacts were distributed randomly amongst the animals. The differences shown between the two models were that a lower proportion of simulations of the recorded network produced any disease spread when compared to those simulations of the random network and, of those that did, fewer infected animals were predicted. In this publication we seek to estimate the optimal adjustment of the probability of infection of a susceptible animal given a contact with an infectious animal so as to minimise these differences.

We constructed four types of random networks, with increasing similarities to the recorded contact network, and by adjusting the probability of infection attempted to gain the best possible approximation for the recorded network. Alongside the simulation of disease, we examined the network properties via six network metrics: assortativity, average path length, closeness, clustering, degree distribution and our own metric – the number of repeated contacts. It has been shown that assortativity can be responsible for the lowering of the epidemic threshold (Molina and Stone, 2012) and clustering to lower the reproductive number R_0 and increase the threshold of disease (Miller, 2009). We have already shown (Duncan et al., 2012) that the recorded networks had more repeated contacts, lower closeness and clustering but higher average path lengths. In this work we seek to relate any differences in these metrics to the adjustment in the probability of infection. Networks can now be constructed with algorithms, to have specific characteristics (Badham and Stocker, 2010a,b; Bansal et al., 2009; Håkansson et al., 2010). Therefore, if it were the case that a

metric value was linked to the optimal adjustment in the probability of infection, it would enable the use of specifically constructed theoretical networks in place of recorded contact networks where recording was not feasible.

2. Materials and Methods

2.1. Disease

The SEIR disease that is modelled through all of the network models can be described by the system of ordinary differential equations (ODEs) (Anderson and May, 1991),

$$\begin{aligned} \frac{dS}{dt} &= -\alpha\beta\frac{SI}{N}, \\ \frac{dE}{dt} &= \alpha\beta\frac{SI}{N} - \sigma E, \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \text{and} \qquad \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{1}$$

with $S + E + I + R = N$, where N is the total (constant) population size. Each susceptible animal moves from the susceptible state (S) to the latent state (E) with rate $\alpha\beta$ following a contact with an infectious animal, where α is the probability of infection from a single contact with an infectious animal and β is the average number of daily contacts per animal. The parameter σ is the rate at which those in the latent class move to the infectious class and γ the rate at which animals move from the infectious class to the resistant class.

2.2. Datasets

Four datasets were available to us. These were recorded using two herds of cattle during two recording periods. The datasets are labelled 1A, 1B, 2A and

2B with the number denoting the recording period, first or second, and the letter representing the herd. Datasets 1A and 1B were recorded during July 2009, both producing 30 complete days of usable data with both of the herds returning complete data for 29 animals. The final two datasets recorded 28 complete days of data across August and September 2009 with 2A recording data for 21 animals whilst 2B returned data for 17 animals.

2.3. Network Construction

In order to answer the question about how close the approximation to our recorded network needed to be, we constructed four types of random network. Each type of network was constructed using increasing amounts of information taken from the recorded data. Details of how all the networks were constructed follows, including details on the construction of the recorded and matched–on–day network used in our previous publication (Duncan et al., 2012). The matched–on–day network was previously referred to as a mass–action or random network but for the purposes of this paper we are using the description "matched–on–day" to demonstrate its relationship to the other types of random network we present. The information required from the recorded network and the mathematical construction for each type of random network can be seen in table 1.

2.3.1. Recorded and Matched–On–Day Networks

For each of the four datasets a contact network was established, with the nodes representing the animals, and the edges, the contacts. A contact was defined to be any recorded interaction that lasted longer than 4 minutes. Although the

term contact has been used, only close proximity of the animals can be assumed rather than actual physical contact. These networks were split into consecutive 12 hour time steps to give a manageable number of edges for each step in the later disease simulation. An identical number of random networks were constructed by taking the total number of interactions recorded in the particular 12 hour period for a particular dataset, creating the same number of random contacts and randomly allocating each of these contacts to pairs of animals in the respective herd. For each dataset and 12 hour period this gave us two networks, a recorded contact network and a random ("matched–on–day") network, with the same number of nodes and edges but with different edge distributions for each 12 hour period for each of the four datasets.

2.3.2. Additional Random Networks

For each dataset, in addition to the matched–on–day network, we constructed three other random networks: "constant–on–animal", "constant–on–day" and "matched–on–animal". For the constant–on–animal network all animals had the same number of contacts as one another for every 12 hour period. The contacts were randomly assigned amongst the animals whilst ensuring that each animal had the required number of contacts. The number of contacts per animal was calculated by averaging all the recorded contacts over the number of animals and the number of 12 hour time periods per dataset. Due to rounding, this meant that the total number of contacts for each of these networks was different from the total number of contacts in the recorded dataset they were derived from.

For the constant–on–day network, the same total number of contacts per 12 hour time period as with the constant–on–animal network was used but the contacts were allocated randomly amongst all the animals. There were no other constraints on the number of contacts an individual animal could have. The structure of this network was seen as lying between that of the constant–on–animal network and the matched–on–day network. Very little information (see table 1) from the recorded network was used in the construction of either the constant–on–animal network or the constant–on–day networks.

In the matched–on–animal network each animal had exactly the same number of contacts as in the recorded network, for each 12 hour period, but those contacts were randomly distributed amongst the other animals subject to this condition i.e. that the number of contacts each animal had was the same as the recorded network. As with the other random network, matched–on–animal networks were constructed for all four datasets.

2.4. Network Metrics

To investigate the differences between the five networks (constant–on–animal; constant–on–day; matched–on–day; matched–on–animal and recorded) six different network metrics were calculated. The first was our own metric, the number of repeated edges, chosen to quantify the observed difference in repeated contacts. The second was closeness, the inverse of the average length of the shortest paths to/from all the other vertices in the network (Csardi,

2013), and the third metric chosen was the clustering coefficient, a measure of the degree to which nodes in a network tend to cluster together (Newman, 2003). The fourth metric that we used, average path length (Strogatz, 2001), is the average number of steps along the shortest path for all possible pairs of nodes. We also calculated the average degree distribution and finally the assortativity coefficient to establish whether assortative mixing, connections between nodes that are similar, was taking place (Molina and Stone, 2012). Each of these metrics were calculated for each network and for each dataset.

2.5. Modelling Disease Spread

All the models, using recorded or any of the four random network types, were implemented as stochastic due to the small numbers of animals in each of the datasets, and hence the increased influence of individual stochastic events on the overall disease transmission process (Brauer et al., 2000). Infection was always introduced by randomly infecting a single animal at the start of each model simulation, thus this animal began the simulation in the latent state. The periods of time each animal spends in the latent and infectious states were sampled from exponential distributions with means $1/\sigma$ and $1/\gamma$. For simplicity, and because the largest dataset only contained 30 days of continuously recorded interactions, each infected animal had its length of resistance set to greater than 30 days. Both models were simulated many times and it was found that the probability densities of the number of animals in each disease state at each time point, appeared to stabilise by 5000 simulations. All results presented were produced from 5000 simulations, where each simulation was run for the number

of days contained in the respective dataset with an initially infected animal randomly chosen for each simulation.

The value of β, the mean contact rate, used in the simulations was dependent on the dataset used, as each of the four datasets had a different average contact rate. Thus we had four values for β corresponding to our four datasets.

The disease spread through each model was a hypothetical disease with parameter values that allowed the peak of infection of an epidemic to occur within the 28 days of data available from the shortest dataset. Latent and infectious periods of six days were chosen. Using average values of $\beta = 7.987$ from our data and $R_0 = 5$ (considered reasonable), a rounded value of $\alpha = 0.1$ was calculated from

$$R_0 = \frac{\alpha\beta}{\gamma}. \tag{2}$$

As each dataset has a different value of β, the contact rate, they will also have a different value of R_0 but the characteristics specific to the disease ($\alpha = 0.1$, $1/\sigma = 6$ days and $1/\gamma = 6$ days) remain fixed across all datasets for the recorded network. For all random networks only the value of α was altered. It was assumed that when an animal became infected its behaviour did not change such that its contacts continued as normal. This is not necessarily the case (Rush et al., 2008; Wilesmith, 1998) but until there exists actual contact data for a herd with spreading disease, it is parsimonious to use the actual data that we do have.

205 *2.6. Measuring the Differences in Disease Spread*

206 The results of our previous paper (Duncan et al., 2012) were divided into two
207 parts: the proportion of 5000 simulations that produced no infection and
208 percentiles of the number of infected animals predicted by those simulations
209 that did produce infection. For all values of the disease parameters, the
210 recorded network model had a higher proportion of simulations showing no
211 infection and of those simulations that did show infection, fewer animals were
212 modelled as infected. In an attempt to minimise the differences between the
213 recorded and random network models the value of α was altered in each type of
214 random network model. The value of α was chosen because the value of β was
215 defined by the datasets and needed to be constant to maintain the continuity in
216 number of contacts between the networks and γ has a basis in other diseases
217 and was dependent on the amount of data available to us, a maximum of 30
218 days. Additionally the large uncertainty in the estimates of the probability of
219 infection for real diseases makes α an attractive candidate for adjustment in
220 random network models.

221

222 The standard value of $\alpha = 0.1$ from our previous paper (Duncan et al., 2012)
223 was used again for the recorded network model and a numerical study
224 conducted on the value of α for the various random network models. For each of
225 the 40 equally spaced values of α in the range $0.025 \leq \alpha \leq 0.4$, all random
226 network models were run with 5000 simulations. The mean absolute difference
227 in both the number of infected animals $\underset{\text{No. Inf.}}{\text{M.A.D.}}$ and in the proportion of the
228 5000 simulations showing no infection $\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}$ were calculated as shown in

equations (3) and (4). In these equations P_{rec} and P_{rand} represent the proportion of the 5000 simulations that produced no infection for the recorded and random network models respectively with $\overline{I}_{rec}$ and $\overline{I}_{rand}$ the mean number of infected animals for each model from those simulations that did produce infection. The *rand* refers to any of the four types of random network: constant–on–animal, constant–on–day, matched–on–day and matched–on–animal. Each individual time period is represented by t and T is the total number of time periods.

$$\underset{\text{No. Inf.}}{\text{M.A.D.}} = \frac{\sum_t \left|\overline{I}_{rec} - \overline{I}_{rand}\right|}{T}. \tag{3}$$

$$\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}} = \frac{\sum_t |P_{rec} - P_{rand}|}{T}, \tag{4}$$

This examination of α gave an initial estimate of where the minima occurred for each type of random network and dataset. To improve these estimates an interval of length 0.05, including this first estimate, was examined in increments of length 0.00125 for each type of network and each dataset. To get a single value for the minima, splines were fitted to these data points for the mean absolute difference in both number of infected animals and proportion of simulations showing no infection, using the smooth.spline function of CRAN R (CRAN-R, 2013) with a smoothing parameter of 0.7 which gave the closest agreement with the visual minimum of the data points. This left two values of α for each random network and dataset: one value minimising $\underset{\text{No. Inf.}}{\text{M.A.D.}}$ and a second minimising $\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}$. The arithmetic mean of these two values was calculated to leave one value α_m to minimise the differences between the recorded and random network models for each of the four random networks and

the four datasets. We conducted similar examinations to find α_m for the matched–on–day network model when we set $\alpha = 0.05$ and $\alpha = 0.2$ in the recorded network model. This sensitivity analysis was carried out to establish whether the value of α used in the recorded network model had any effect on the adjustment to find α_m.

3. Results

3.1. Network Metrics

The 5000 simulations of the random contact networks, outlined above, were stored to calculate average values for the six metrics. For each dataset the contact networks were split into 12 hour periods and the metrics calculated on each of the 5000 simulations. The results were averaged across the simulations and then over the 12 hour periods. These were then compared to the equivalent metrics calculated for the recorded network which was split into 12 hour periods after the disease simulations.

Figure 1 shows the results of the metrics in six separate plots. Each plot shows results for all networks split by the four datasets. There is no clear result from the metrics as to which of the random networks provides the closest approximation to our recorded network. The recorded network had more repeated edges and lower closeness than any of the random networks and this was consistent across all the datasets. In all but one dataset the recorded network also had higher average path length than the random networks. The more information from the recorded network used to construct the random

network – the greater the number of repeated edges in the random networks and hence closer to that of the recorded network.

Each network shows disassortativity across all datasets. For three of the datasets the recorded network was more disassortative than all four random networks and, as with the repeated edges, the more information from the recorded network used by the random network, in general, the more disassortative they became. Generally speaking in, three metrics (average path length, average closeness and average repeated edges) increasing similarity with the recorded network was associated with the random model utilising increased information from the recorded network.

3.2. Disease Spread

A sample of the results for the mean absolute differences in both the number of infected animals and the proportion of 5000 simulations showing no infection $\left(\text{M.A.D.}_{\text{No. Inf.}} \text{ and } \text{M.A.D.}_{\text{Propn. Zero Sims.}}\right)$ can be seen in figure 2. These are the results for the matched–on–day network for all four datasets. The results for the other random networks can be seen in the supplementary information. The results for $\text{M.A.D.}_{\text{No. Inf.}}$ are shown in the solid lines using the left hand axes with the results of $\text{M.A.D.}_{\text{Propn. Zero Sims.}}$ plotted as dashed lines using the right hand axes.

For each of the datasets and across all the random networks the results were very similar with four points to note. First there is a single minimum value of α_m and the differences in $\text{M.A.D.}_{\text{No. Inf.}}$ and $\text{M.A.D.}_{\text{Propn. Zero Sims.}}$ at this value of α_m are very

small. Secondly the value of α_m is always less than the value of $\alpha = 0.1$ used in the recorded network. It is also consistent, across all networks and datasets, that the value of α that results in minimising the differences in the proportion of the 5000 simulations showing no infection is larger than the respective value of α for the difference in the number of infected animals. Finally, there are clear but not very large differences in the value of α_m for each type of network across the four datasets.

The results from the proportion of simulations with no infected animals and the values of the 25^{th}, 50^{th} and 75^{th} percentiles of the number of infected animals from those simulations showing infection are plotted for both the recorded network model ($\alpha = 0.1$; black, solid lines) and the matched–on–day network model ($\alpha_m = 0.0696$; red, dashed lines) are plotted in figure 3 for dataset 1A. Similar plots for the other random networks are shown in the supplementary information. In all cases it is clear that by adjusting α the results of simulated disease spread through the random networks are extremely close to the results from the recorded network. Using the single value of α_m provides very close agreement and it is not necessary to use both the value of α that resulted in $\underset{\text{No. Inf.}}{\text{M.A.D.}}$, and the one that gives $\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}$.

To compare the differences between the results for each of the four types of random networks the minimum values of $\underset{\text{No. Inf.}}{\text{M.A.D.}}$ and $\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}$ are shown in figure 4. These were plotted for each dataset along with the relative percentage decrease in α needed to achieve α_m. Figure 4 also shows the

differences α_m for each type of network across the four datasets. It is clear from the plot that the mean differences in number of infected animals are much less than a single animal for each of the networks. The value is dependent on the network being used in the simulation as can be seen by the consistent order of results (constant–on–day, constant–on–animal, matched–on–day and matched–on–animal). It is worth noting that the network using the least information from the recorded network, constant–on–animal, is not the poorest performing. The relative percentage decrease needed to achieve α_m is somewhere between 5.6% and 39.4% but this varies depending on the dataset and the random network used.

It is clear from the left–hand plot in figure 4 that the values of $\underset{\text{No. Inf.}}{\text{M.A.D.}}$ are dependent on the simplicity of the model. The model using the most information, the matched–on–animal network, is closest to the recorded network. However the simplest network (constant–on–animal) was numerically closer to the recorded network than the second simplest network (constant–on–day). This was also the case for $\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}$ for all but dataset 1B. The loss of representativeness that arises from choosing the simplest random network is not large.

The right–hand plot of figure 4 shows the relative percentage decrease of α needed to achieve α_m for each the random networks and for each dataset. The patterns in the adjustment are not completely consistent either with regard to the datasets or networks. There appears by eye to be a dataset effect in the

right–hand plot of figure 4. General linear regression, included in the supplementary information, suggests there is evidence of both a dataset effect and random network effect. Each factor was fairly strongly significant after the addition of the other factor, $p = 0.0007$ and $p = 0.042$ for dataset and network respectively. The mean reduction in α was 26.8% and the median reduction was 30.0%.

The exact values of α_m are shown in table 2. For three of the datasets the highest value of α_m occurred in the matched–on–animal network, the network using the most information from the recorded network. Nevertheless for dataset 2A, the matched–on–animal had the second highest value of α_m. For the first recording period (datasets 1A and 1B) the value of α_m increases as the networks use more information from the recorded network and this trend is less clear for the second recording period.

Also included in the supplementary information are plots of the differences in the proportion of 5000 simulations that produced no infection and the median number of infected animals from those simulations that did produce infection.

4. Discussion

It is clear from the simulations of disease spread that a simple homogeneous mixing model can approximate, very closely, a recorded network if the probability of infection, α, is optimally adjusted. Each of our four types of random network can approximate the recorded network and can do so for each

of the four datasets. The adjustment was consistently a reduction in α. The size of the adjustment was dependent on the dataset and random network used for the simulations. The relative percentage reduction in α ranged from 5.6% to 39.4%. The results of the sensitivity analysis shown in the supplementary information would suggest that the value of α_m as a proportion of α is negatively associated with the value of α used in the recorded network, at least for the values of α that we tested.

It has previously been shown that higher clustering tends to produce shorter path lengths within theoretical networks (Shirley and Rushton, 2005a), that clustering and assortativity can reduce epidemic size (Miller, 2009) and that increased clustering or increased assortativity can increase the likelihood of simulated disease spread occurring (Badham and Stocker, 2010a). There is however disagreement over whether clustering influences epidemics on undirected networks with regular (many repeated contacts) or random construction (Eames, 2008; Moslonka-Lefebvre et al., 2009).

Theoretical networks constructed with many repeated contacts show slower disease spread than random networks (Eames, 2008). This is also shown by both our earlier work (Duncan et al., 2012) and further demonstrated by random networks constructed here. In general, our random networks with lower repeated contacts, i.e. the simpler networks (contact–on–animal and contact–on–day) required smaller values of α_m suggesting that disease spreads quicker through them.

As all the random networks are derived from the recorded network and the average degree distributions are either extremely close to one another or identical, we can gain little insight from degree distribution. However, degree distribution alone has been shown to not provide enough information for prediction of disease spread (Ames et al., 2011; Boily et al., 2007).

We found no clear relationship between the values of the metrics and the values of α_m and formal inferential statistics are not possible given the sample size. Any inferential statistical relationship will, however, depend on a large number of herds being assessed in the same manner.

One of the largest differences between the recorded network and the random networks is the number of repeated edges. One possible reason for the high number of repeated edges in the recorded network was that the herds were constructed of cows with calves at foot. Of the repeated edges recorded, 15% to 30%, depending on dataset, were between a cow and her calf. These repeated edges could also be a reason for the increased disassortativity found in the recorded network. Assortative mixing would normally entail cows contacting cows and calves contacting calves. With young calves present in the herd, the disassortative mixing, resulting from cow contacting calf, would seem probable. Assortativity has been shown to decrease epidemic size (Miller, 2009) and we have found that $\alpha_m < 0.1$ for all networks and datasets, showing that the recorded network produces slower disease spread than the random networks.

The age of the calves may also explain why in the first recording period (datasets 1A and 1B) the value of α_m increases as the random networks approach the recorded network. In the second recording period, where the calves were a little older, there is not such a clear pattern.

It has recently been shown that indirect, environmental or faecal, contact may aid the spread of disease in herds of cattle (Kleinlützum et al., 2013). These factors cannot be taken into account with the data available to us. Likewise we only have proximity data with which to construct our contact networks. We do not know the extent of the contacts and how likely each one is to spread disease. However, the only way to gather such data would be to film the animals at all times and to monitor real life spread of infection. Even those studies which attempt to take such things into account by observing animals and categorising the contacts by strength (Norton et al., 2012) are still summarising the contact networks as they extrapolate their networks from the observed data.

5. Conclusion

We have shown that it is possible to closely model disease spread through a network of recorded contacts with a network of randomly allocated contacts by adjusting the probability of infection. The adjustment in probability of infection is consistently a reduction and there appears to be a dataset effect in the value of the reduction. The exact values in adjustment varies between 5.6% and 39.4% and as yet, with only four datasets, we have no clear relationship between the network properties and the adjustment in the probability of infection.

Recommended reductions in α should not be made until further intra–herd contact data becomes available. Importantly, the simplest network, requiring least information to construct, performed reasonably well by giving a close match to disease spread in the recorded network. This is important because it suggests that in the absence of real contact data a good approximation to disease spread could be made if the correct adjustment in the probability were known.

6. Acknowledgements

The Scotland's Rural College (SRUC) receives financial support from the Scottish Government within WP6.1 of the RESAS 2011–2016 research programme. The authors would like to acknowledge the link between Inverness College UHI and the SRUC thus enabling this work to be carried out.

References

Ames, G. M., George, D. B., Hampson, C. P., Kanarek, A. R., McBee, C. D., Lockwood, D. R., Achter, J. D., Webb, C. T., 2011. Using network properties to predict disease dynamics on human contact networks. Proceedings of the Royal Society B: Biological Sciences 278, 3544–3550.

Anderson, R., May, R., 1991. Infectious Diseases of Humans. Oxford University Press, Oxford, UK.

Badham, J., Stocker, R., 2010a. The impact of network clustering and assortativity on epidemic behaviour. Theoretical Population Biology 77, 71–75.

Badham, J., Stocker, R., 2010b. A spatial approach to network generation for three properties: degree distribution, clustering coefficient and degree assortativity. Journal of Artificial Societies and Social Simulation 13 (1), 1–13.

Bansal, S., Khandelwal, S., Meyers, L. A., 2009. Exploring biological network structure with clustered random networks. BMC Bioinformatics 10, 405.

Barlow, N. D., 2000. Non-linear transmission and simple models for bovine tuberculosis. J. Anim. Ecol. 69 (4), 703–713.

Boily, M., Asghar, Z., Garske, T., Ghani, A., Poulin, R., 2007. Influence of selected formation rules for finite population networks with fixed macrostructures: implications for individual-based model of infectious diseases. Mathematical Population Studies 14 (4), 237–267.

Bouma, A., Dejong, M. C. M., Kimman, T. G., 1995. Transmission of pseudorabies virus within pig-populations is independent of the size of the population. Prev. Vet. Med. 23 (3-4), 163–172, doi:10.1016/0167-5877(94)00442-L.

Brauer, F., van den Driessche, P., Wu, J. (Eds.), 2000. Mathematical Epidemiology. Lecture Notes in Mathematics. Spinger.

Craft, M., Caillaud, D., 2001. Network models: An underutilized tool in wildlife epidemiology? Interdis. Perspec. Inf. Dis.ID 676949, doi: 10.1155/2011/676949.

CRAN-R, 2013. Smooth.spline r documentation. Last accessed 30/09/2013.

URL

http://stat.ethz.ch/R-manual/R-patched/library/stats/html/smooth.spline.html

Csardi, G., 2013. Package 'igraph'. Last accessed 03/09/2013.

URL http://cran.r-project.org/web/packages/igraph/igraph.pdf

D' Amico, V., Elkinton, J. S., Dwyer, G., Burand, J. P., Buonaccorsi, J. P., 1996. Virus transmission in gypsy moths is not a simple mass action process. Ecol. 77 (1), 201–206.

De Jong, M. C. M., Diekmann, O., Heesterbeek, H., 1995. How does transmission of infection depend on population size? Publications of the Newton Institute; Epidemic models: Their structure and relation to data, 84–94.

Dobson, A., Meagher, M., 1996. The population dynamics of brucellosis in the yellowstone national park. Ecol. 77 (4), 1026–1036.

Drewe, J., Weber, N., Carter, S., Bearhop, S., Harrison, X., Dall, S., McDonald, R. a., 2012. Performance of proximity loggers in recording intra and inter–species interactions: A labratory and field based validation study. PLoS One 7, 1–9'.

Duncan, A., Gunn, G., Lewis, F., Umstatter, C., Humphry, R., 2012. The influence of empirical contact networks on modelling diseases in cattle. Epidemics 4, 117–123.

Eames, K., 2008. Modelling disease spread through random and regular contacts in clustered populations. The Journal of Theoretical Biology 73 (1), 104–111.

Håkansson, N., Jonsson, A., Lennartsson, J., Lindström, T., Wennergren, U., 2010. Generating structure specific networks. Advances in Computer Systems 13 (2), 239–250.

Hamede, R. K., Bashford, J., Jones, M., McCallum, H., 2012. Simulating devil facial tumour disease outbreaks across empirically derived contact networks. Journal of Applied Ecology 49, 447–456.

Keeling, M., 2005. The implications of network structure for epidemic dynamics. Theo. Pop. Biol. 67 (1), 1–8.

Kleinlützum, D., Weaver, G., Schley, D., 2013. Within–group contact of cattle in dairy barns and the implications for disease transmission. Research in Veterninary Science 95, 425–429.

Miller, J. C., 2009. Percolation and epidemics in random clustered networks. Physical Review E 80, 020901(R).

Molina, C., Stone, L., 2012. Modelling the spread of disease in clustered networks. Journal of Theoretical Biology 315, 110–118.

Moslonka-Lefebvre, M., Pautasso, M., Jeger, M., 2009. Disease spread in small–size directed networks: epidemic threshold, correlation between links to and from nodes, and clustering. Journal of Theoretical Biology 260 (3), 402–411.

Newman, M., 2003. The structure and function of complex networks. SIAM Review 45, 167–256.

Norton, E., Benaben, S., Mbotha, D., Schley, D., 2012. Seasonal variations in physical contact amongst domestic sheep and the implications for disease transmission. Livestock Science 145 (1), 34–43.

Porphyre, T., Stevenson, M., Jackson, R., McKenzie, J., 2008. Influence of contact heterogeneity on tb reproduction ratio r_0 in a free–living brushtail possum *Trichosurus vulpecula* population. Veterinary Research 39, 31–43.

Rush, J., de Passille, A., von Keyserlingk, M., Weary, D., 2008. The Welfare of Cattle Volume 5. Springer, Dordrecht, The Netherlands.

Salathé, M., Kazandjieva, M., Lee, J., Levis, P., Feldman, M., Jones, J., 2010. A high-resoultion human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences 107 (51), 22020–22025.

Shirley, M., Rushton, S., 2005a. The impacts of network topology on disease spread. Ecological Complexity 2, 287–299.

Shirley, M., Rushton, S., 2005b. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot–and–mouth disease epidemic. Epidemiology and Infection 133, 1023–1032.

Strogatz, S. H., 2001. Exploring complex networks. Nature 410 (6825), 268–276.

Swain, D. L., Bishop-Hurley, G. J., 2007. Using contact logging devices to explore animal affiliations: Quantifying cow-calf interactions. Applied Animal Behaviour Science 102 (1-2), 1–11, doi:10.1016/j.applanim.2006.03.008.

Vernon, M. C., Keeling, M. J., 2009. Representing the uk's cattle herd as static

and dynamic networks. Proceedings of the Royal Society B: Biological Sciences 276, 469–476.

Watts, D., Strogatz, S., 1998. Collective dynamics of 'small–world' networks. Nature 393, 440–442.

Wilesmith, J., 1998. Manual on bovine spongiform encephalopathy, 34.

Table 1: Descriptions of how the four random networks relate to the recorded network and how much information from the recorded network was necessary to create them.

Information needed to construct random network	Random Network	Mathematical Comparison
Total number of contacts, number of animals, total number of time periods	constant–on–animal	$\sum_{j} x_{i,j,t} = k \quad \forall i,\ t$
Total number of contacts, number of animals, total number of time periods	constant–on–day	$\sum_{i,j;i>j} x_{i,j,t} = kN \quad \forall t$
Total number of contacts per time period, number of animals	matched–on–day	$\sum_{i,j;i>j} x_{i,j,t} = \sum_{i,j;i>j} r_{i,j,t} \forall t$
Total number of contacts per animal per time period, number of animals	matched–on–animal	$\sum_{j} x_{i,j,t} = \sum_{j} r_{i,j,t} \forall i,\ t$

where:

$x_{i,j,t}$ = a simulated contact between animals i and j during time period t with $i \neq j$

$r_{i,j,t}$ = a recorded contact between animals i and j during time period t with $i \neq j$

$$k = \text{round}\left(\frac{\sum_{i,j,t;i>j} r_{i,j,t}}{NT}\right)$$

N = Total population size (Number of animals)

T = Total number of timeperiods

Table 2: Values of α_m, the value of the probability of infection α, used to minimise the differences between the recorded and random network models for each of the four types of random networks - for each of the four datasets. A value of α = 0.1 was used for the recorded model across all simulations.

	α_m per dataset			
Network	1A	1B	2A	2B
constant–on–animal	0.0645	0.0684	0.0705	0.0944
constant–on–day	0.0649	0.0770	0.0606	0.0757
matched–on–day	0.0695	0.0830	0.0664	0.0844
matched–on–animal	0.0799	0.0915	0.0765	0.0886

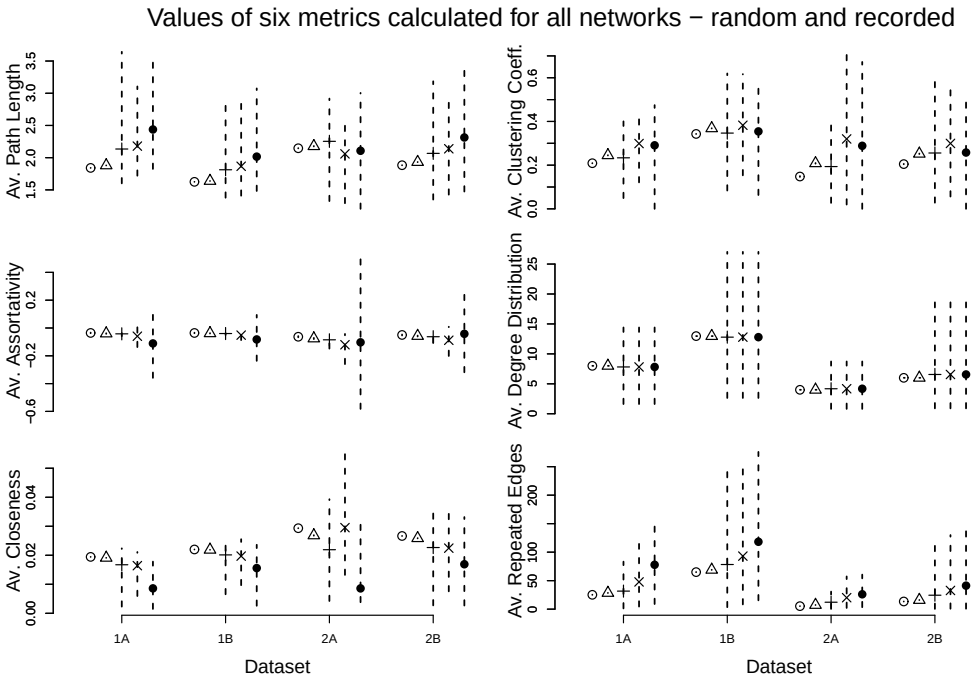

Figure 1: The average values of all six metrics calculated for each of the five networks. The symbols ∘, △, +, × and • denoting results from the constant–on–animal, constant–on–day, matched–on–day, matched–on–animal and recorded networks respectively. The vertical dashed lines represent the 95% percentiles for each metric.

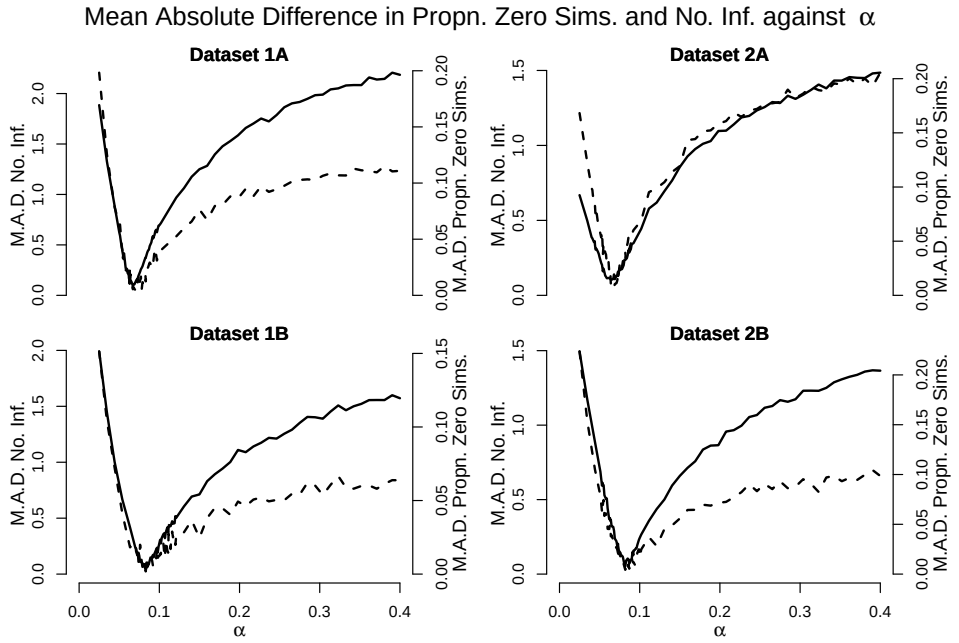

Figure 2: Plots of the mean absolute difference in the number of infected animals $\left(\underset{\text{No. Inf.}}{\text{M.A.D.}}\right)$ (left–hand axis, solid line) and mean absolute difference in the proportion of the 5000 simulations showing no infection $\left(\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}\right)$ (right–hand axis, dashed line) against α for all four datasets. $\alpha = 0.1$ was used in the recorded network model.

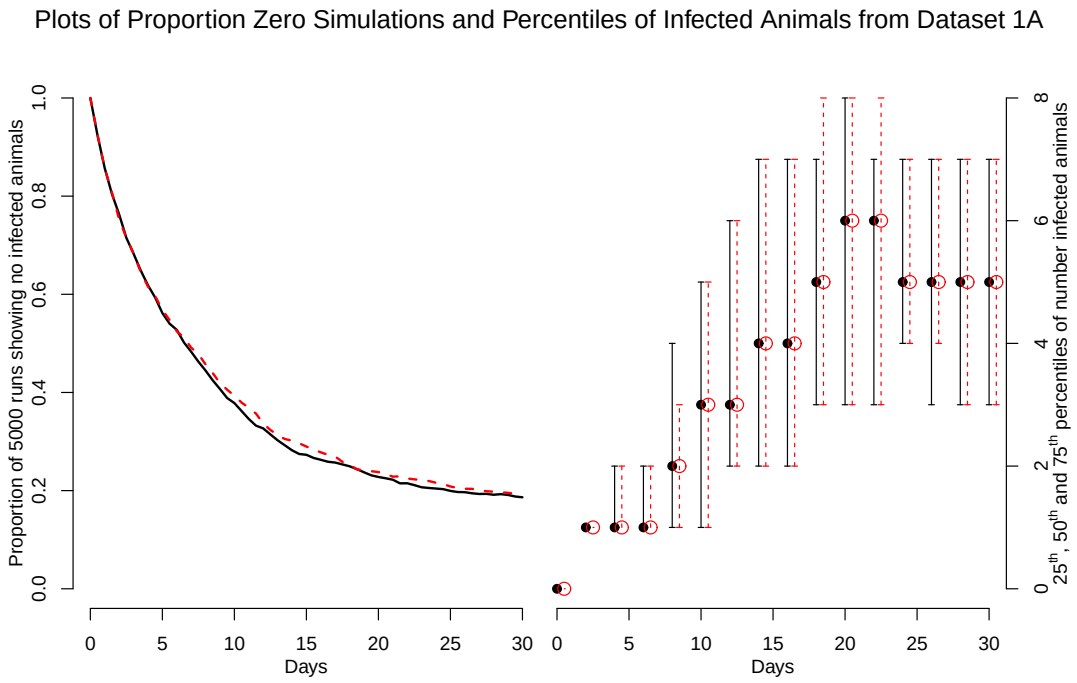

Figure 3: Left–hand plot: Proportion of 5000 simulations that produced no infection for the recorded network model with $\alpha = 0.1$ (black, solid line) and the adjusted random network model with $\alpha = \alpha_m$ (red, dashed line). Right–hand plot: The 25^{th}, 50^{th} and 75^{th} percentiles of the number of infected animals from those simulations that did produce infection for the recorded network model with $\alpha = 0.1$ (black, solid line) and the adjusted random network model with $\alpha = \alpha_m$ (red, dashed line). Dataset 1A was used for both models.

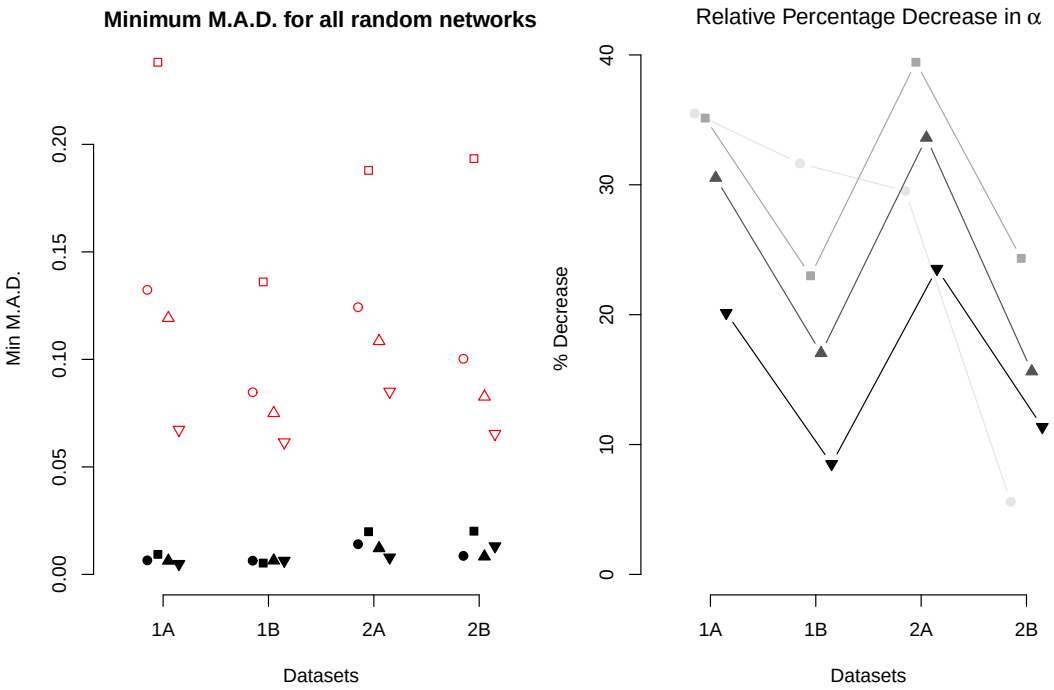

Figure 4: Left–hand plot: The values of the mean absolute difference in the number of infection animals $\left(\underset{\text{No. Inf.}}{\text{M.A.D.}}\right)$ (unfilled, red symbols) and the mean absolute difference in the proportion of 5000 simulations showing no infection $\left(\underset{\text{Propn. Zero Sims.}}{\text{M.A.D.}}\right)$ (filled, black symbols) for α_m plotted for each of the four random networks. Right–hand plot: The relative percentage decrease in α to achieve α_m from the value of $\alpha = 0.1$ used in the recorded network. The shading denotes the amount of information from the recorded needed to construct the random network, lightest representing the least information and the darkest representing the most information. In both plots the symbols ○, □, △ and ▽ represent the constant–on–animal, constant–on–day, matched–on–day and matched–on–animal networks.