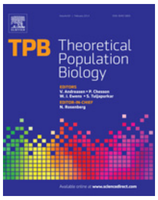
Theoretical Population Biology 98 (2014) 11–18

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb



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, United Kingdom
[b] Epidemiology Research Unit, SRUC (Scotland's Rural College), Drummondhill, Stratherrick Road, Inverness, IV2 4JZ, United Kingdom
[c] Agroscope, Institute for Sustainability Sciences (ISS), Tänikon 1, CH-8356, Ettenhausen, Thurgau, Switzerland

ARTICLE INFO

Article history:
Received 22 April 2014
Available online 16 September 2014

Keywords:
Network
Mass-action
Disease
Recorded contacts
SEIR simulation

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.

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

---

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

http://dx.doi.org/10.1016/j.tpb.2014.08.004
0040-5809/© 2014 Elsevier Inc. All rights reserved.

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 network, 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),

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

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 useable 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 four 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 h 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 h 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 h 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 h 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 h period. The contacts were randomly assigned amongst the animals whilst ensuring that each