

# Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain

R. R. Kao<sup>1,\*</sup>, L. Danon<sup>2</sup>, D. M. Green<sup>1</sup> and I. Z. Kiss<sup>1</sup>

<sup>1</sup>*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*

<sup>2</sup>*Departament de Física Fonamental, Universidad de Barcelona, Marti i Franques 1, 08028 Barcelona, Spain*

Using a novel interpretation of dynamic networks, we analyse the network of livestock movements in Great Britain in order to determine the risk of a large epidemic of foot-and-mouth disease (FMD). This network is exceptionally well characterized, as there are legal requirements that the date, source, destination and number of animals be recorded and held on central databases. We identify a percolation threshold in the structure of the livestock network, indicating that, while there is little possibility of a national epidemic of FMD in winter when the catastrophic 2001 epidemic began, there remains a risk in late summer or early autumn. These predictions are corroborated by a non-parametric simulation in which the movements of livestock in 2003 and 2004 are replayed as they occurred. Despite the risk, we show that the network displays small-world properties which can be exploited to target surveillance and control and drastically reduce this risk.

**Keywords:** foot-and-mouth disease; percolation; ergodicity; disease; control

## 1. INTRODUCTION

The contact structure of a population can have consequences for disease transmission, such as when the variance in the number of potential contacts is high (Anderson & May 1991; Albert *et al.* 2000), or when transmission is localized but with occasional long-distance jumps (Watts & Strogatz 1998). While there has been significant progress in our understanding of disease transmission on networks (Watts & May 1992; Eubank *et al.* 2004; Meyers *et al.* 2005), good disease-relevant network datasets are few. Here, we analyse the network of livestock movements in Great Britain, for which the spatial, temporal and demographic characteristics are especially well described, in the context of foot-and-mouth disease (FMD). Control of the 2001 epidemic of FMD cost an estimated 8.5 million culled livestock and £4–6 billion (Anderson 2002). Initial dissemination of disease was facilitated by the rapid long-distance movement of sheep via dealers and markets (Gibbens *et al.* 2001; Kao 2002), suggestive of both small-world (Watts & Strogatz 1998) and scale-free (Albert *et al.* 2000) network properties. Mainly in recognition of the importance of these movements, legislation was introduced requiring a mandatory 20-day movement standstill (with a few clearly defined exceptions) for all large livestock from agricultural premises, following the on-movement of any other large livestock (Bourn 2003). Though the movement standstill was reduced for sheep and cattle to 13 days from 18 February 2002 in Scotland, and to 6 days from 4 March 2003 in England and Wales, it is expected that the new legislation will have

greatly reduced the risk of a large FMD epidemic, should the virus be reintroduced into Great Britain. As recorded livestock movement data includes source, destination, date and batch size (described in appendix A in the electronic supplementary material), a network analysis is ideal for testing this expectation. We predict conditions under which a large epidemic could occur and illustrate how it could be efficiently prevented, and show that the network analysis is robust when compared to an exact replay of the movement record.

## 2. MATERIAL AND METHODS

### (a) *Network interpretations*

Two issues that hinder the application of network theory to epidemiology are the typical assumptions that all potentially infectious contacts or links are the same (equally weighted), and that the network is static, i.e. that potentially infectious links are fixed over the course of an epidemic. A potentially infectious link joining two individuals (nodes)  $i$  and  $j$  will have a probability of transmission ( $p_{ij}$ ) that is more realistically time-dependent and weighted by the characteristics of  $i$  and  $j$ , the nature of the link and the direction of contact. However, analyses of weighted networks are complicated (e.g. Barrat *et al.* 2004), so here, we deconstruct the network of potentially infectious contacts by considering a fixed time-frame covering a single infectious period and discard active links with probabilities  $(1-p_{ij})$ , to create a static, directed ‘epidemiological network’ of truly infectious links. In this way, all links can be treated equally. Stochastic variability between realizations of the epidemiological network can be accounted for by repeated deconstruction and analysis.

Since all links in an epidemiological network have the same weight and the same meaning, standard network analyses become useful. Importantly, the final epidemic size for a susceptible–infectious–removed (SIR) model can be

\* Author for correspondence (rowland.kao@zoo.ox.ac.uk).

The electronic supplementary material is available at <http://dx.doi.org/doi:10.1098/rspb.2006.3505> or via <http://www.journals.royalsoc.ac.uk>.

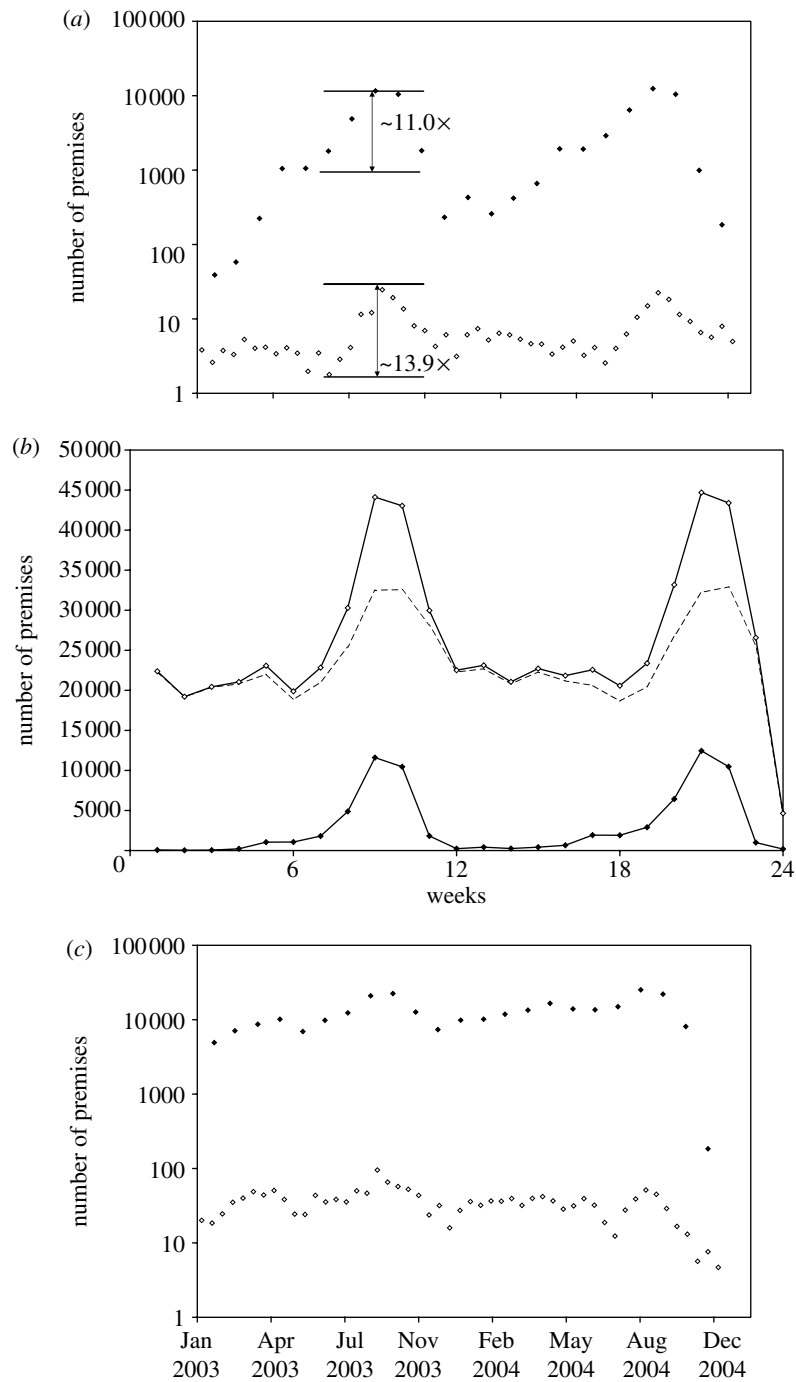


Figure 1. Comparison of giant strong component size over four-week periods from January 2003 to November 2004, with epidemic simulations. (a) GSCC size (filled diamonds) compared to FMD epidemic (open diamonds) for sheep movements only. Changes in GSCC size ( $\sim 11.0\times$ ) and epidemic size ( $\sim 13.9\times$ ) between June 2003 and September 2003 are similar (vis-a-vis change in number of movements of  $3.9\times$  (see figure A5 in the electronic supplementary material)). (b) Giant strongly connected component (filled diamonds), giant weakly connected component (open diamonds) and difference between them (dashed line) in sheep movements. (c) GSCC size (filled diamonds) compared to FMD epidemic (open diamonds) for all livestock movements (cattle, pigs and sheep).

interpreted directly in terms of the size of the epidemiological network components. In a static network, an epidemic will continue so long as an infected node can reach at least one uninfected node. A strongly connected component or strong component is a subset of a directed network in which all nodes can reach each other. A weakly connected component or weak component is a strong component plus all its sources and sinks. In a static epidemiological network, any disease starting in a strong component will infect all elements of the strong component, and will infect all sink nodes as well.

Similarly, any epidemic starting in a source node will infect the strong component plus all sinks, but not necessarily all sources. Thus, the largest or giant strongly connected component (GSCC) is an estimate of the lower bound of the maximum epidemic size, while the giant weakly connected component is an estimate of its upper bound. The distributions of strong and weak component sizes are estimates of the distribution of the final epidemic size. All components are determined using Tarjan's algorithm (Sedgewick 2002).

**(b)  $R_0$  in the epidemiological network**

The epidemiological network interpretation can be directly related to the basic reproduction ratio, or  $R_0$ . The persistence threshold associated with  $R_0$  underpins modern theoretical epidemiology (Anderson & May 1991). For  $R_0 > 1$ , a pathogen will be successful, in the sense that the introduction of a single infected individual into a wholly susceptible homogeneous population will on average infect more than one other. An equivalent interpretation can be made in terms of percolation theory for the epidemiological network. In general, a network percolation threshold (Stauffer & Aharony 1992; Moore & Newman 2000; Schwartz *et al.* 2002) can be defined by an increase in the average number of connections per node, below which there are only finite-sized strong components in an infinite network, and above which the GSCC is infinite. In a finite network, this is approximated by the 'sudden' appearance of a large GSCC as the number of connections increases, by the joining of many prior strong components. In the context of a randomly connected epidemiological network (i.e. where the probability of connection is defined by the relative node degrees), a percolation threshold is defined when each node potentially infected over the network can subsequently infect on average, at least one other node. In an undirected network, this requires two links per node, while in a directed network, it requires that there be one outward link if a node has at least one inward link; in both cases this is equivalent to  $R_0 = 1$ . In the case of proportionate mixing,

$$R_0 = \langle k_{in}k_{out} \rangle / \langle k_{in} \rangle. \quad (2.1)$$

(Schwartz *et al.* 2002 and see appendix D in the electronic supplementary material), a result consistent with standard epidemiological theory (Diekmann *et al.* 1990; Anderson & May 1991). Other factors such as negative correlation (where nodes of high out degree preferentially connect to nodes of low in degree) will of course change the expression for  $R_0$ . If connections are not random, then the additional structure in the network could lower the estimated final epidemic size. For example, a network consisting only of sets of continuous loops or necklace of nodes will have  $R_0 = 1$ , but the GSCC will be equal to the length of the longest necklace. Comparing  $R_0$  to the size of the GSCC can therefore be used to identify network structures and a percolation threshold more generally defined in terms of a value  $R_0^{crit} \geq 1$ , which defines when invasion by a novel pathogen can cause a large epidemic. A percolation value of  $R_0^{crit} > 1$  implies localizing structures, such as spatial or social clustering (i.e. higher probabilities of links between nodes for reasons that are not related to the degrees of the nodes).

**(c) The role of ergodicity**

Thus far, our description of epidemiological networks has been static; a more realistic network structure would be dynamic and unique to each epidemic, reflecting the links that exist over the repetitive infectious period of nodes taken from the respective infectious periods, from the times they became infectious. Should the pattern of behaviour underlying the network structure be unchanged over the course of an epidemic, then a network may be constructed simply by considering all possible links over all time, and choosing a subset of those links as being infectious. In this case, snapshots of the epidemiological network covering a single infectious period will give the same estimate of the GSCC and component size distribution.

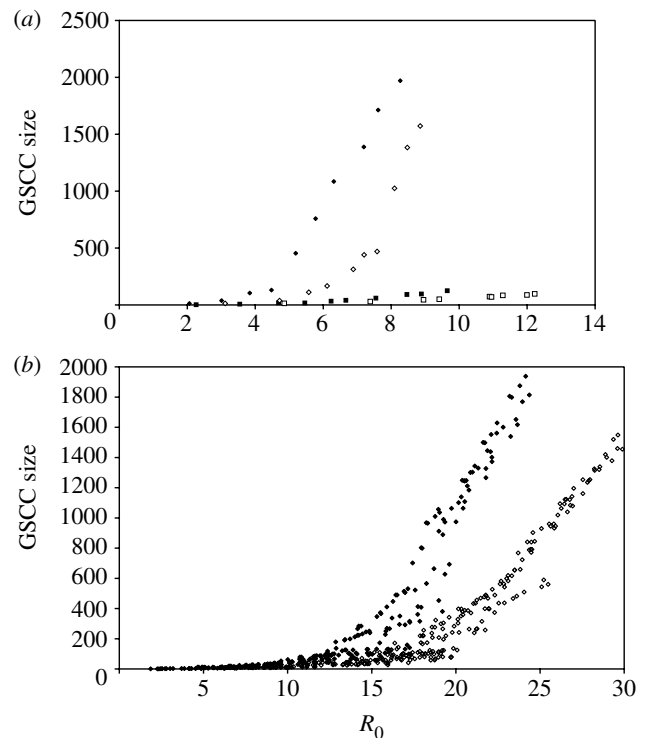


Figure 2. The growth of the giant strongly connected component (GSCC) for sheep and cattle movements. (a) Growth of the GSCC for sheep movements as a function of  $R_0$ . Shown are sizes for infectious period of 28 days starting from 19 May 2004 (filled diamonds), infectious period of 28 days starting from 5 November 2003 (open diamonds), infectious period of 7 days starting from 19 May 2004 (filled squares) and infectious period of 7 days starting from 5 November 2003 (open squares). Probabilities of transmission in all four cases range from 0.1 per batch movement to 1.0 per batch movement. (b) Establishing ergodicity: figure as in (a) but infectious periods are varied in the same fashion for farms and markets. The  $R_0$  values are generated by choosing different combinations of infectious period and probability of transmission per link. In all cases, similar  $R_0$  values produce similar GSCC sizes when comparing within four-week periods, but different sizes when comparing between them. Above  $R_0^{crit}$  the behaviour can be described by a power law,  $n_{GSCC} \propto (R_0 - R_0^{crit})^\epsilon$  where  $n_{GSCC}$  is the size of the GSCC. From 19 May 2004 (filled diamonds),  $R_0^{crit} = 4.8$ ,  $\epsilon = 4.1$ ,  $R^2 = 0.91$  on a log-log plot and from 5 November 2003 (open diamonds),  $R_0^{crit} = 12.7$ ,  $\epsilon = 5.1$ ,  $R^2 = 0.93$  on a log-log plot. The value of  $R_0^{crit}$  is chosen by identifying the local maximum in the  $R^2$  value, and power law exponent is chosen by linear regression above that value.

A consequence is that the final epidemic size of the static network is also an estimate of the final epidemic size in the dynamic network; this can be interpreted in terms of ergodicity. Ergodic theory is an extensive and complex area of interest in statistical physics (e.g. Reichl 1980). Birkhoff's 'ergodic theorem' states that if the average of thermodynamic quantities over a statistical ensemble (or set) of similar systems at equilibrium is the same as the time average for a single system, then the systems are ergodic. We can use the concept by considering an ensemble of multiple realizations of the epidemiological network. For our purposes, it is sufficient to note that a Markov chain, in this case the changing network, is considered ergodic if all possible states are accessible from any given state

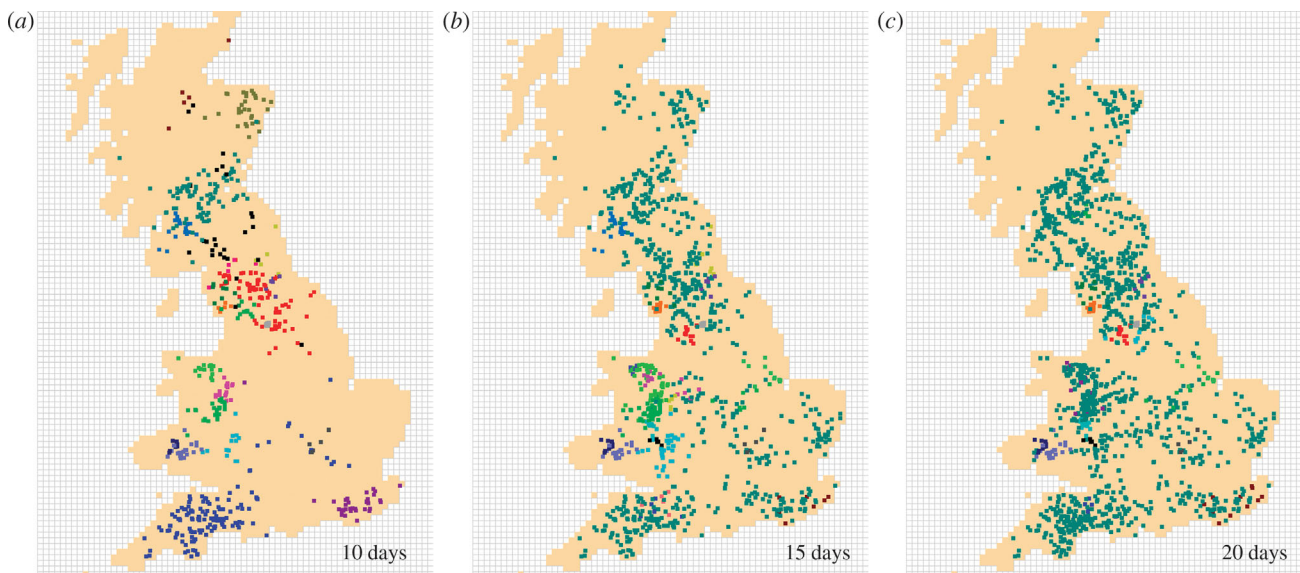


Figure 3. Geographical distribution of strongly connected components (by colour) in sheep movements from 19 May 2004, assuming all movements are infectious, showing all components of size greater than four, as the length of the infectious period increases from 10 to 20 days. Component colours are only conserved across the three periods if the component is conserved; e.g. if the ‘blue’ component is absorbed by the ‘red’ component, then blue may be re-used.

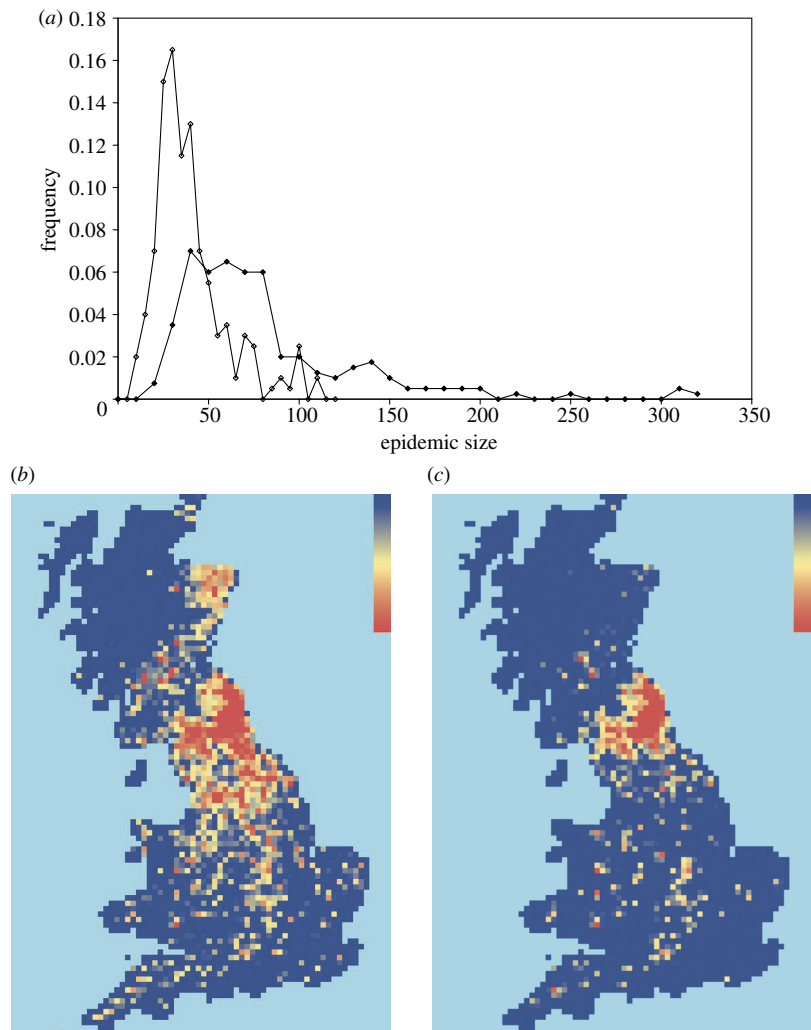


Figure 4. (a) Distribution of number of IPs for epidemics seeded in Northumberland with 10 sheep farms, simulating early conditions during the 2001 epidemic, for movement of sheep only (open diamonds) and allowing for movements of all livestock (filled diamonds). In both cases, random local transmission up to 10 km is allowed. Two hundred simulations of epidemics are run for 21 days, assuming transmission parameters for the 2001 epidemic, but starting in September 2004, when sheep movements are at their highest. (b) Risk map for all of GB allowing for all movements of livestock, showing average number of IPs per simulated epidemic in 100 km<sup>2</sup> grid squares. Maximum density of IPs is 0.25 per 100 km<sup>2</sup> (red). (c) As in (b), but only allowing for movement of sheep.

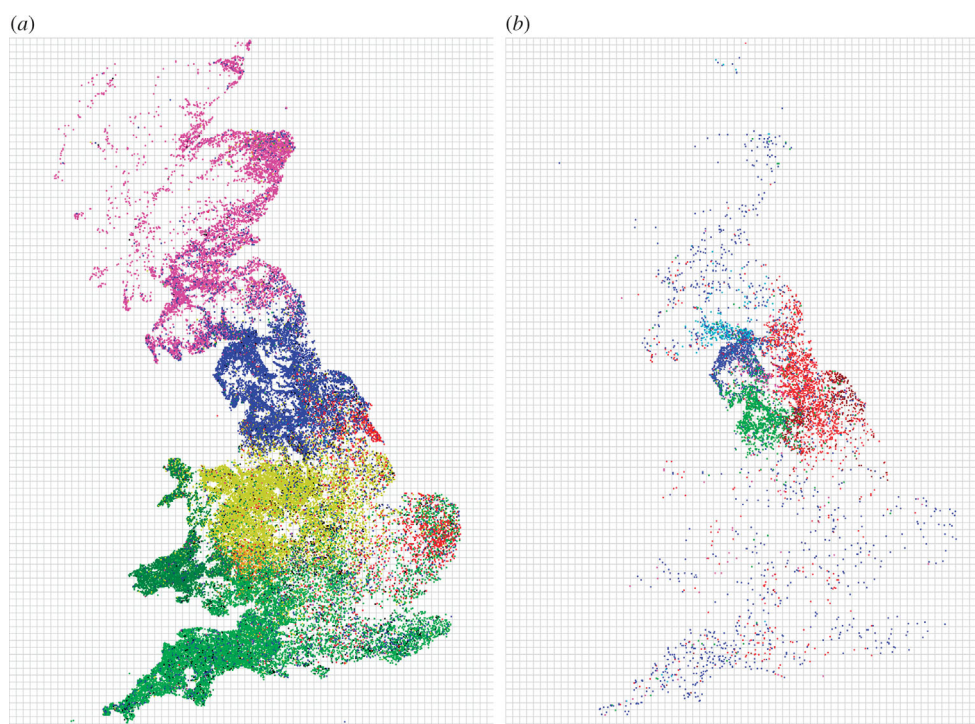


Figure 5. Community structures (by colour). Shown are (a) the six largest communities using all sheep movements in the four-week period starting from 19 May 2004, and (b) the result of applying the algorithm to the largest of the communities, to identify subcommunities. In both cases, community structure is highly regional though with some community members widely distributed.

(e.g. Reichl 1980), clearly true for static networks. For a dynamic network, the connection is less straightforward, but can be visualized by considering an underlying demographic network that is constructed by a set of fixed rules that do not depend on the prior state of the network. Since disease transmission depends only on local information (i.e. what a single infectious node can infect over a single infectious period), the unique dynamic epidemiological network developed by tracking an evolving epidemic will also have the same epidemiological properties as the statistical ensemble, provided the system is ergodic. This can be illustrated by considering the result of an epidemic simulation via two routes—first, constructing a static network and then running an epidemic over the already constructed network, or second constructing the network dynamically as the epidemic progresses by infecting a single node, identifying its links and determining which nodes it infects, then for each of those second generation infected nodes, identifying their links and determining which nodes they infect, and continuing this process until the epidemic ends. If the two processes give the same estimates for epidemiological quantities, e.g. the GSCC-based estimate of the final epidemic size, in terms of  $R_0$ , then the system is ergodic. Conversely, if the system is not ergodic, then there will be structural changes in different realizations of the epidemiological network, and so these would have to be considered when estimating epidemiological quantities such as the final epidemic size.

#### (d) Simulations

We corroborate the predictions of the network analysis by a ‘non-parametric’ simulation of the movements, i.e. a direct replay of the movements as they are recorded. A detailed description of the simulations is given in appendix B in the

electronic supplementary material; in brief, simulations are seeded with a single infected farm, active on a given day. The movement record is then replayed day by day, assuming that on farms all off-movements occur before all on-movements. A farm is assumed to become infected when at least one infected animal is moved onto the farm. The probability of a farm becoming infected is weighted by the number of animals and the species of animals moved. Weightings were chosen so that the mean probability that a movement from an infected premise (IP) contained at least one infected animal was consistent with epidemiological estimates from the 2001 epidemic. Markets ceased to be infectious the day after the first off-movement following infection, and re-enter the susceptible state. While FMD virus transmission characteristics are highly variable (Haydon *et al.* 2004), all epidemiological parameters were chosen to be consistent with experimental transmission data and movement-based transmission during the 2001 epidemic, prior to the imposition of a national movement ban on 23 February (Gibbens *et al.* 2001; Gibbens & Wilesmith 2002; Hughes *et al.* 2002; Alexandersen *et al.* 2003a,b; Gloster *et al.* 2003; Mansley *et al.* 2003; Wilesmith *et al.* 2003). Further details are given in appendix C in the electronic supplementary material. In order to consider the impact of ‘breaking’ the network structure, additional, random transmission is implemented, creating new IPs at a rate and over a distance consistent with the local transmission that occurred during the 2001 epidemic (Kao 2003).

### 3. RESULTS

The distribution of movements per premises and animals per movement are shown in figure A1 in the electronic supplementary material. Both show evidence of scale-free properties (i.e. probability distribution  $\propto k^{-\lambda}$ , where  $k$  is the number of on or off movements per premises, and  $\lambda$  is

a constant). Because of the high variance of these distributions (infinite in the limit of very large populations), high values of  $R_0$  are possible even when the transmission rate is low (Albert *et al.* 2000). As expected, markets are very active, with large numbers of animals passing through them. While pig and cattle movements may result in FMD transmission, the 6- and 13-day movement standstills would allow for clinically infected animals to appear in large numbers in any infected pig or cattle herd, thus shortening the time to detection; in contrast, because the disease is difficult to detect in sheep, movements become much more important, as was the case in 2001 (Kao 2002). Thus, we consider first the network of sheep movements. The GSCC size for the network of sheep movements varies considerably over the course of a year. The pattern of GSCC variation shows evidence of percolation behaviour, with changes in the size of the GSCC (figure 1) varying by a factor of 11, qualitatively different from changes in the number of movements, which only differs by a factor of less than 4 (see figure A2 in the electronic supplementary material). Based on GSCC sizes, under the current movement regime, the network is relatively safe from disease propagated in the sheep population in the winter, the principle species and the timeframe for the 2001 epidemic of FMD in the UK (Kao 2002). However, there is a substantially greater risk from late spring to early autumn. We corroborate our predictions using a non-parametric simulation of disease spread, where infection is seeded on random farms, the record of all livestock movements is replayed chronologically, and movements from infectious premises potentially result in transmission (full description of the non-parametric simulations are given in appendix B in the electronic supplementary material). These simulations showed threshold behaviour that is similar to the behaviour revealed by the GSCC analysis (figure 1a), and they corroborate the risk of a large epidemic in late summer and early autumn. While transmission from or to the giant weak component could potentially change this scenario, in this case the giant weak component retains the threshold behaviour (figure 1b). In contrast, consideration of transmission across all movements (pigs, sheep and cattle) show less seasonal variation, and no evidence of percolation behaviour (figure 1c), though of course the epidemics are also larger.

Using equation (2.1), we now compare  $R_0$  to the GSCC size in the sheep movement data in more detail for the four-week period from 19 May 2004, the critical period when the sheep network first appears to exceed the percolation threshold (figure 1). Though the formula of equation (2.1) assumes proportionate mixing, higher order correlations appear to have only a small effect on the value of  $R_0$ ; the best estimate of  $R_0$  is found by directly calculating the dominant eigenvalue of the full contact matrix  $M$  (Diekmann *et al.* 1990), where the elements of the matrix  $m_{ij}$  are 0 if there is no contact from node  $i$  to  $j$ , and 1 if there is contact. While it is not feasible to calculate this quantity in all cases (i.e. when the GSCC is very large), during the periods of fewer movements and therefore smaller GSCCs, the difference between the two values is small (e.g. for the four-week period from 19 May 2004, the true value of  $R_0$  is 14.1 compared to  $R_0 = 13.3$  computed with equation (2.1)). This is generally consistent for all the smaller GSCCs (i.e. less than 2000

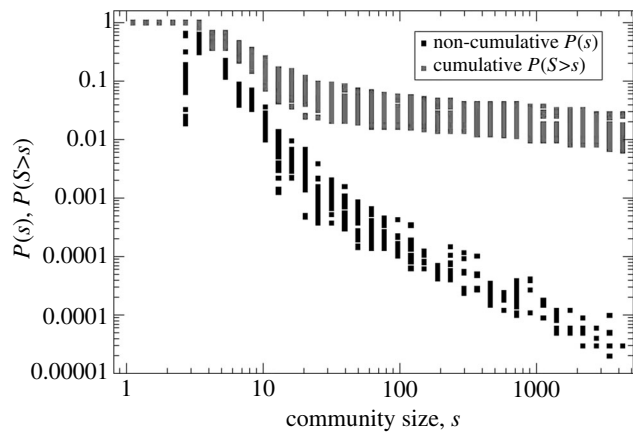


Figure 6. The distributions (cumulative and non-cumulative) of community sizes for all 24 four-week periods starting from 1 January 2003. The distribution of sizes is consistent across all periods, suggesting that local patterns of movement remain similar, even though the size of the GSCC changes dramatically over the course of the year (figure 1).

nodes) throughout the timeframe of the data though there are larger discrepancies when  $R_0$  is large.

For the sheep network, particularly important in the 2001 epidemic, there is no risk of a large epidemic until  $R_0$  is well above one (in this case,  $R_0 \approx 4$ ); the threshold location is also dependent on the underlying disease parameters (figure 2a). Markets are the most highly connected nodes and therefore the greatest contributors to  $R_0$ , but have a fixed infectious period of 1 day, as they are assumed to hold no resident livestock and therefore rely on transient contact for transmission. As there is the possibility of environmental contamination, this is a lower estimate for the infectious period. However, direct movement of sheep from single infected herds dispersed at markets was the major source of long-distance spread (Mansley *et al.* 2003). If the market infectious period is scaled with the farm infectious period (figure 2b), the same  $R_0$  gives a similar GSCC size over the four-week time-scale (an ergodic scale, allowing comparison of the static network properties with transmission dynamics), but not when comparing between seasons (a non-ergodic scale, where such a comparison is not valid). The components are spatially clustered; the growth of the GSCC as  $R_0$  increases appears to follow a scale-free distribution, and represents both accumulation of newly active premises, and the absorption of smaller components (figure 3), each typically centred on one or more markets. The absorption of smaller components and scale-free growth of the GSCC are consistent with a percolation threshold (Stauffer & Aharony 1992).

The correspondence between the GSCC and epidemic size is robust when all movements and local transmission are added at levels appropriate for FMD in 2001; of epidemiological concern is that FMD virus introduced in sheep during the peak risk time (September 2004) have similar size and geographical extent to the 2001 epidemic prior to the imposition of the nationwide livestock movement ban, when seeded in a fashion similar to 2001, and allowed to spread for a similar length of time (figure 4). The pattern of spread is typified by many short-range transmission events, with occasional long-distance jumps resulting in many transmission events from a single node, usually a market.

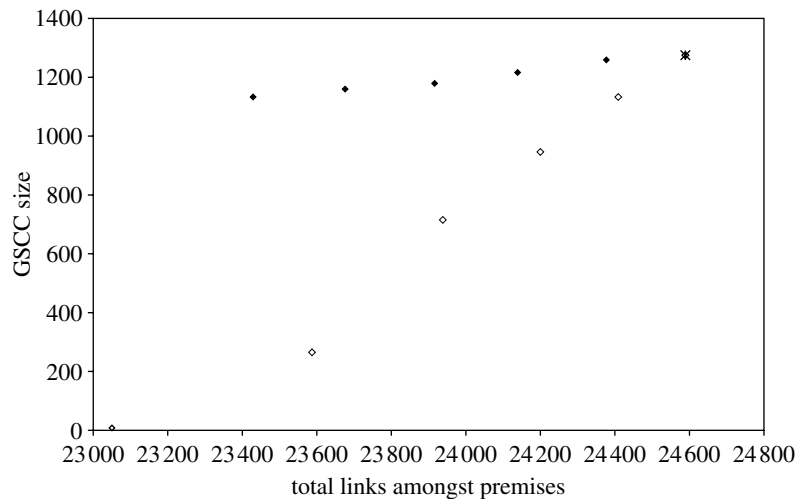


Figure 7. Targeted control of disease transmission on the sheep movement network. Comparison of GSCC size with no removal of links (\*) to GSCC size under the effect of random removal of links (filled diamonds) and targeted removal of farm-to-market links (open diamonds), where the farm has previously bought from a market. Data from four-week period starting from 19 May 2004, total links = 24 589, of which 1539 are the targeted farm-to-market following market–farm links.

While a serious national emergency is possible, network properties could be exploited for efficient surveillance or disease control. Community structures, i.e. groups of premises more closely connected to each other in the network, are identified using an algorithm that maximizes the difference between the true network, and an equivalent randomly connected network (Newman 2004), with the algorithm modified for directed networks (see appendix D in the electronic supplementary material). The identified communities are highly regionalized, corroborating the way that individual components of the epidemiological network grow as  $R_0$  increases (figure 5). Community size distributions taken in four-week snap shots over the whole timeframe (figure 6) are consistent for all periods, suggesting that localized movements retain the same pattern, but greater number of bridging contacts between communities exist to create a larger GSCC during the late summer to mid-autumn period. Considering only the subsystem of farm-to-market, market-to-farm and farm-to-farm movements which are most likely to be responsible for transmission, direct farm-to-farm movements contribute only 6123 of 52 637 movements in the four-week period from 19 May 2004, suggesting a largely bipartite structure to the network, with the majority of other moves either farm-to-market (35 115 moves), or market-to-farm (11 129 moves). We hypothesize that farms linking together different markets are largely responsible for percolation. To test this, we selectively delete farm-to-market moves where the immediately prior movement onto that farm was from a different market. These movements are few, for example representing only 3825 of the 52 637 sheep movements (1539 of 24 589 in the demographic network GSCC). However, preventing transmission via these movements collapses the GSCC. The dramatic effect of removing ‘bridging’ links compared to the removal of random links (figure 7) is reminiscent of small-world effects (Watts & Strogatz 1998), and suggests that increased surveillance, awareness and biosecurity enforcement targeted at community-bridging movements would be valuable, particularly during the critical months.

#### 4. DISCUSSION

The cattle data have been extensively audited (Bourn 2003), and similar analyses of the sheep and pig movements are ongoing. However, some inaccuracies are inevitable. Nevertheless, unless there are underlying biases in which movements are recorded, the observed patterns of behaviour are unlikely to change. As the majority of discrepancies are probably the result of under-reporting or mis-reporting rather than over-reporting, our analysis would probably underestimate the timeframe of risk for FMD since additional movements can only increase the risk of passing over the percolation threshold.

We have considered the stochastically generated epidemiological network of contacts, and related this network to  $R_0$ , showing via the ergodic hypothesis when structural considerations are epidemiologically important. While there have been other studies on the effect of spatial heterogeneity on epidemic success (e.g. Kao 2003; Boots *et al.* 2004), to our knowledge, this is the first detailed analysis of the relationship between disease epidemiology and contact structure on such a well-characterized network. While prior extensive analyses have often provided detailed results, they have used largely inferred datasets for actual contact, since it is notoriously difficult to find more detailed data for most populations. Here, with precisely defined contact, we can compare directly the replayed series of potentially infectious movements as they occurred, and show that the simpler static network analyses still provide meaningful predictions. Both the network analysis and simulations show overall patterns of spread that are consistent with the 2001 epidemic, with widespread dissemination driven by markets and a few individuals trading between them (i.e. dealers). That the changes in legislation have decreased the risk of a large epidemic is encouraging, as is the prediction that good targeted surveillance could reduce the size of any possible epidemic.

Prior analyses of networks in epidemiology have analysed the characteristics of the social or demographic network (e.g. distribution of contacts per node and clustering of contacts) separately from epidemiological

factors such as the transmission rate and incubation period (e.g. Meyers *et al.* 2005). However, such an approach leads to conceptual and technical problems. Disease and network characteristics cannot always be easily separated—for example, social behaviour of individuals may be related to their health status (e.g. the relationships amongst drug use, needle sharing and an individual's health), or in the case of FMD, probability of transmission depends on both the nature of the connected nodes (commercial versus purebred sheep flocks, or farms versus markets) and the links (different species and number of livestock moved) and thus their susceptibility and transmissibility. While the demographic contact structure is important when ascertaining the viability or efficiency of intervention strategies (e.g. Kiss *et al.* 2005, 2006), what matters for evaluating the efficacy of control is the number of new infections caused. Reducing the social or demographic network to an epidemiological network means that all nodes can be treated in the same fashion, and then a statistical analysis of underlying demographic and physiological risk factors used to determine the relative importance of all node types.

While we have analysed the network of livestock movements, the use of the epidemiological network is more generally relevant. Network models are more common for diseases with well-defined contact characteristics such as sexually transmitted diseases, but less so where the number of potentially infectious contacts is large. However, for all transmissible diseases, epidemiological contacts are sparse and therefore epidemiological network analyses are applicable. For a contact structure with random mixing,  $R_0 = 1$  is the percolation threshold. Otherwise, heterogeneities in network structure may result in another threshold for  $R_0 \gg 1$ . Estimates of  $R_0$  for various infectious diseases from measles to poliomyelitis have an average value of order 10, and a minimum of 4 (Anderson & May 1991). While  $R_0$  estimates are notoriously model dependent, one would expect that there would be more estimates with low  $R_0$ . One possible explanation is the effect of community structure; while  $R_0 < 1$  is a robust condition for the failure of an invading pathogen, community structure complicates the relationship between  $R_0$  and pathogen success.

We thank Prof. J. Wilesmith, Dr M. B. Gravenor and two anonymous referees for useful comments. DEFRA and the Scottish Executive, with special thanks to Mr J. T. Bell, Miss C. Autran, Mrs B. Dyer and Mr K. Duffy, have provided the data. R.R.K. and I.Z.K. are funded by the Wellcome Trust, D.M.G. by DEFRA, and L.D. by the Generalitat de Catalunya.

## REFERENCES

- Albert, R., Jeong, H. & Barabasi, A. L. 2000 Error and attack tolerance of complex networks. *Nature* **406**, 378–382. (doi:10.1038/35019019)
- Alexandersen, S., Kitching, R. P., Mansley, L. M. & Donaldson, A. I. 2003a Clinical and laboratory investigations of five outbreaks of foot-and-mouth disease during the 2001 epidemic in the United Kingdom. *Vet. Rec.* **152**, 489–496.
- Alexandersen, S., Quan, M., Murphy, C. & Zhang, Z. 2003b Studies of quantitative parameters of virus excretion and transmission in pigs and cattle experimentally infected with foot-and-mouth disease virus. *J. Comp. Pathol.* **129**, 268–282. (doi:10.1016/S0021-9975(03)00045-8)
- Anderson, I. 2002 *Foot & Mouth Disease 2001: Lessons to be Learned Inquiry Report*. London, UK: The Stationary Office.
- Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. 2004 The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA* **101**, 3747–3752. (doi:10.1073/pnas.0400087101)
- Boots, M., Hudson, P. J. & Sasaki, A. 2004 Large shifts in pathogen virulence relate to host population structure. *Science* **303**, 842–844. (doi:10.1126/science.1088542)
- Bourn, J. 2003 In *Identifying and tracking livestock in England* vol. HC 1144 (ed. NAO). London, UK: The Stationary Office.
- Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.
- Eubank, S., Guclu, H., Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z. & Wang, N. 2004 Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184. (doi:10.1038/nature02541)
- Gibbens, J. C. & Wilesmith, J. W. 2002 Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain. *Vet. Rec.* **151**, 407–412.
- Gibbens, J. C., Sharpe, C. E., Wilesmith, J. W., Mansley, L. M., Michalopoulou, E., Ryan, J. B. & Hudson, M. 2001 Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet. Rec.* **149**, 729–743.
- Gloster, J., Champion, H. J., Sorensen, J. H., Mikkelsen, T., Ryall, D. B., Astrup, P., Alexandersen, S. & Donaldson, A. I. 2003 Airborne transmission of foot-and-mouth disease virus from Burnside Farm, Heddon-on-the-Wall, Northumberland, during the 2001 epidemic in the United Kingdom. *Vet. Rec.* **152**, 525–533.
- Haydon, D. T., Kao, R. R. & Kitching, R. P. 2004 The UK foot-and-mouth disease outbreak—the aftermath. *Nat. Rev. Microbiol.* **2**, 675–681. (doi:10.1038/nrmicro960)
- Hughes, G. J., Mioulet, V., Haydon, D. T., Kitching, R. P., Donaldson, A. I. & Woolhouse, M. E. 2002 Serial passage of foot-and-mouth disease virus in sheep reveals declining levels of viraemia over time. *J. Gen. Virol.* **83**, 1907–1914.
- Kao, R. R. 2002 The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends Microbiol.* **10**, 279–286. (doi:10.1016/S0966-842X(02)02371-5)
- Kao, R. R. 2003 The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease. *Proc. R. Soc. B* **270**, 2557–2564. (doi:10.1098/rspb.2003.2546)
- Kiss, I. Z., Green, D. M. & Kao, R. R. 2005 Disease contact tracing in random and clustered networks. *Proc. R. Soc. B* **272**, 1407–1414. (doi:10.1098/rspb.2005.3092)
- Kiss, I. Z., Green, D. M. & Kao, R. R. 2006 Infectious disease control using contact tracing in random and scale-free networks. *Proc. R. Soc. Interface* **3**, 55–62. (doi:10.1098/rsif.2005.0079)
- Mansley, L. M., Dunlop, P. J., Whiteside, S. M. & Smith, R. G. 2003 Early dissemination of foot-and-mouth disease virus through sheep markets in February 2001. *Vet. Rec.* **153**, 43–50.
- Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M. & Brunham, R. C. 2005 Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* **232**, 71–81. (doi:10.1016/j.jtbi.2004.07.026)



- Moore, C. & Newman, M. E. 2000 Epidemics and percolation in small-world networks. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **61**, 5678–5682.
- Newman, M. E. 2004 Fast algorithm for detecting community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **69**, 066133.
- Reichl, L. E. 1980 *A modern course in statistical physics*. Austin, TX: University of Texas Press.
- Schwartz, N., Cohen, R., Ben-Avraham, D., Barabasi, A. L. & Havlin, S. 2002 Percolation in directed scale-free networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 015104.
- Sedgewick, R. 2002 Algorithms in C. In *Graph algorithms*, 3rd edn, pp. 197–208. Boston, MA: Addison-Wesley.
- Stauffer, D. & Aharony, A. 1992 *Introduction to percolation theory*. Boca Raton, FL: CRC Press.
- Watts, C. H. & May, R. M. 1992 The influence of concurrent partnerships on the dynamics of HIV/AIDS. *Math. Biosci.* **108**, 89–104. (doi:10.1016/0025-5564(92)90006-I)
- Watts, D. & Strogatz, S. 1998 Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
- Wilesmith, J. W., Stevenson, M. A., King, C. B. & Morris, R. S. 2003 Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. *Prev. Vet. Med.* **61**, 157–170. (doi:10.1016/j.prevetmed.2003.08.002)