Title

Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park

Journal

ANNALES ZOOLOGICI FENNICI, 41(6)

ISSN

0003-455X

Authors

Cross, PC
Lloyd-Smith, JO
Bowers, JA
et al.

Publication Date

2004

Peer reviewed

Ann. Zool. Fennici 41: 879–892 ISSN 0003-455X
Helsinki 20 December 2004

Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park

Paul C. Cross[1,2], James O. Lloyd-Smith[3], Justin A. Bowers[2,4], Craig T. Hay[2,4], Markus Hofmeyr[4] & Wayne M. Getz[1,2]

[1)] *Department of Environmental Science, Policy and Management, 201 Wellman Hall #3112, University of California, Berkeley, CA 94720-3112, USA (e-mail: pcross@nature.berkeley.edu)*
[2)] *Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa*
[3)] *Biophysics Graduate Group, 299 Life Science Addition MC 3200, University of California, Berkeley, CA 94720, USA*
[4)] *Kruger National Park, P.O. Box 76, Skukuza, South Africa 1350*

Received 20 Apr. 2004, revised version received 28 Sep. 2004, accepted 23 June 2004

Cross, P. C., Lloyd-Smith, J. O., Bowers, J. A., Hay, C. T., Hofmeyr, M. & Getz, W. M. 2004: Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. — *Ann. Zool. Fennici* 41: 879–892.

Recognition is a prerequisite for non-random association amongst individuals. We explore how non-random association patterns (i.e. who spends time with whom) affect disease dynamics. We estimated the amount of time individuals spent together per month using radio-tracking data from African buffalo and incorporated these data into a dynamic social network model. The dynamic nature of the network has a strong influence on simulated disease dynamics particularly for diseases with shorter infectious periods. Cluster analyses of the association data demonstrated that buffalo herds were not as well defined as previously thought. Associations were more tightly clustered in 2002 than 2003, perhaps due to drier conditions in 2003. As a result, diseases may spread faster during drought conditions due to increased population mixing. Association data are often collected but this is the first use of empirical data in a network disease model in a wildlife population.

Introduction

Group structure is the hallmark of social species. The size and integrity of groups reflects their function, which may include vigilance against predators, sequestration and protection of resources, and alloparenting (for review *see* Dugatkin 1997). Increased susceptibility to disease is generally believed to be a cost of sociality (Alcock 1998). If the movement between groups is limited, however, group structure can act to contain disease spread. Here we investigate how the movement of individuals between groups affects disease spread using data on African buffalo (*Syncerus caffer*) and relatively simple models of disease processes.

African buffalo exist in a fission-fusion society where groups often separate and rejoin (Prins

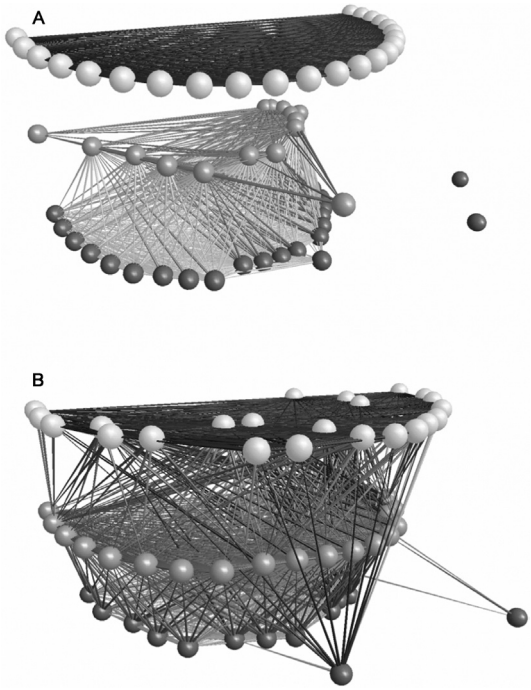

Fig. 1. Network graphs of the buffalo association data for (**A**) May 2002 and (**B**) November 2001 through October 2003. Balls represent individual buffalo and the lines represent all non-zero association values. Individuals are distributed vertically according to herd membership. Herd membership was determined by cluster analysis (e.g. the solid, dotted and dot-dashed lines in Fig. 2A refer to the black, grey and light grey, balls respectively in panel **A**).

1996, Cross *et al.* [in press]). The ability to recognize others is a prerequisite for non-random association patterns in fission-fusion societies. In this context we do not differentiate between the evaluator and cue-bearer in the recognition process, but analyze the proportion of time individuals spend together and assume that it is a function of individuals perceiving and acting upon the cues expressed by one another (Liebert & Starks 2004). We use pair-wise association data that we have collected over a two-year period in the Kruger National Park, South Africa, to characterize buffalo population structure. These data are then combined with disease models to illustrate how association patterns affect disease dynamics. Behavioral researchers often collect data on the association between individuals, but these data are seldom used in disease models. We illustrate the importance of incorporating behavioral data in models of disease dynamics, and discuss several of the unresolved questions which limit our ability to integrate data on recognition and association with disease models.

Models of disease dynamics have become increasingly important to understanding and managing disease invasion (e.g. Ferguson *et al.* 2001, Keeling *et al.* 2003, Lloyd-Smith *et al.* 2003). They reduce the complexity of a system and allow for the investigation of specific factors in ways that would not be possible using experimental methods. A tradeoff exists between realism and generality of models, however, and factors that are omitted for the sake of simplicity may play important roles in the real system. Traditionally, the network of connections between individuals has not been included in disease models (e.g. Anderson & May 1991). Modelers assumed that association was random, i.e. that every individual was equally likely to contact every other individual. More recent studies suggest, however, that the way individuals contact one another (i.e. the network structure and topology) plays an important role in determining the probability of disease invasion, the total number infected, and the speed of disease spread (Keeling 1999, Watts 1999, Newman 2002). Both traditional disease models that assume random mixing and spatial disease models that assume limited dispersal between fixed groups poorly characterize some socially structured populations, such as the African buffalo. Dynamic network models more accurately reflect connections within and between groups and the spread of disease between associating individuals. In this paper, we attempt to narrow the gap between the fields of behavior, recognition, and disease ecology by illustrating the importance of non-random association to the spread of disease using empirical data from an ongoing study of African buffalo and integrating them with simulation models of disease spread.

Social networks have been visualized in a number of different ways. Network graphs depict individuals as points and their contacts as connecting lines (Fig. 1). In the behavioral literature of animal association patterns, researchers often use cluster analyses, such as Ward's or UPGMA methods, to describe the network (e.g. White-

head 1999). Underlying these visual techniques is a contact or association matrix **A** where each matrix element a_{ij} describes the amount or type of contact between individuals *i* and *j*. Several properties of this matrix will affect the overall rate of disease spread as well as each individual's risk of infection. Most obvious are the average number of connections per individual and the strength of those connections. Less obvious, and the focus of recent work, is the way that the topology of connections (i.e. who is connected to whom) affects disease processes (e.g. Anderson *et al*. 1990, Kretzschmar & Morris 1996, Morris & Kretzschmar 1997, Boots & Sasaki 1999, Keeling 1999, Watts 1999, Kretzschmar 2000, Newman 2002, Eames & Keeling 2003, Meyers *et al*. 2003, Read & Keeling 2003).

More traditional disease models assume that an individual's risk of infection depends upon the global state of the population. From a network perspective, however, an individual's risk of infection depends on the number of connections they have and which of those connected individuals are infected. Network models of disease have largely focused on sexually transmitted diseases (STDs), often simulated on static networks with uniform connection strengths (e.g. Kretzschmar & Morris 1996, Keeling 1999, Watts 1999), though recent work has included analytic approximations (Ferguson & Garnett 2000, Bauch & Rand 2001, Eames & Keeling 2002), exact solutions (Newman 2002), non-sexually transmitted diseases (Meyers *et al*. 2003), an STD network with changing connections (Eames & Keeling 2004), and an empirically-derived urban social network (Eubank *et al*. 2004). However, our extension of this network approach to cover airborne diseases in animal systems, with a network based on empirical data, raises two novel issues. Is variance in the connection strengths an important factor? How does the dynamic nature of the network affect disease spread?

Early attempts to model stochastic disease dynamics in networks were probably limited by computational power. Current efforts are impeded by scarcity of empirical data on the structure of human and animal networks and the limited communication between behavioral researchers and epidemiologists. In practice, several critical issues make it difficult to determine the network structure of a population: (1) contacts may be difficult to define and differ between diseases, (2) people are usually bad at estimating their contacts while animals are often difficult to observe, (3) large populations require researchers to choose a sample of individuals or a small portion of the network and then extrapolate, (4) connections between individuals change over time, and (5) on longer time scales individuals enter and leave the network due to birth and death processes or migration. All of these issues affect our ability to produce an unbiased estimate of the network structure and none of them are easily solved. For this reason, relatively few empirically-based social networks exist in the disease ecology literature (but *see*, Woodhouse *et al*. 1994, Edmunds *et al*. 1997, Wallinga *et al*. 1999, Liljeros *et al*. 2001, Jolly & Wylie 2002, Eubank *et al*. 2004). In this study, we address issues two and four above, in the context of data that we have collected on associations among individuals in the African buffalo (*Syncerus caffer*) population in the Kruger National Park (KNP) that is in the midst of a bovine tuberculosis (BTB) epidemic (Rodwell *et al*. 2000). More importantly, we demonstrate that association data collected routinely by behavioral researchers are applicable and valuable to studies of disease dynamics.

African buffalo typically occur in breeding herds of approximately 30 to 1000 individuals, and adult males move between breeding herds in bachelor groups of 2–30 individuals. Previous researchers concluded that buffalo herds were relatively stable units, and although herds sometimes separated they did not associate with neighboring groups (Sinclair 1977, Mloszewski 1983, Prins 1996). Furthermore, females, subadults, and juveniles were assumed not to move between herds (Sinclair 1977, Mloszewski 1983, Prins 1996). Recent studies based upon larger sample sizes of radio-collared individuals, however, suggest that females and subadults do move between groups and the structure of herds may not be stable (Halley 2002, Cross *et al*. [in press]). Radio-tracking data from our study in the KNP of over 123 radio-collared individuals since November 2000 indicate that herds frequently separate and reunite, and females move

to different areas via splinter groups. Further, because herd membership changes, the definition of a herd becomes more nebulous over time. The framework developed in this study presents a rigorous definition of a herd based upon association data.

In this study, we begin by describing the association pattern and network structure of 64 radio-tracked buffalo from November 2001 to October 2003. We then use the association data in a stochastic disease model to investigate three questions: (1) How does the incorporation of non-random association data affect our predictions about the speed and intensity of a disease outbreak in the buffalo population in the Kruger National Park? (2) Does the variance in the frequency of contact between individuals affect disease dynamics? (3) How does duration of infectiousness affect the degree of population structure experienced by the disease process? At the moment, we do not have empirical disease data that we could compare with the model predictions, but work on this aspect is ongoing. To the authors' knowledge, this study is the first application of a network disease model to a wildlife population using empirical data to create the social network.

Methods

Association data

Field data were collected during an ongoing study in the Satara Region of the Kruger National Park. The study area contained 4–12 buffalo herds, depending upon the amount of herd fragmentation, and roughly 3000 buffalo. Buffalo were darted from helicopters and fitted with radio-collars in four sessions: November 2000 ($N = 6$), April 2001 ($N = 27$), August ($N = 51$) and November 2001 ($N = 12$). To simplify the analysis, we restricted the data to sightings of 64 radio-collared buffalo that survived from November 2001 to October 2003. This restriction allowed for a 'complete' dataset where individuals were present for the duration of the study period.

We monitored buffalo herds, on foot and from vehicles, approximately 2–3 times per week throughout the year from distances ranging from 50–1000 m. If an individual was missing for over one month, we located it from an aircraft. If a herd split during the day, only the first sighting was used for that day. Since all marked individuals had radio-collars and herds were usually separated by several kilometers, we could determine which individuals were in a herd without visually sighting all individuals.

Following Whitehead and DuFault (1999), we considered two individuals to be associating if they were located in the same herd. This one/zero metric of association was used to calculate the proportion of samples in which two individuals were seen together (i.e. the simple ratio index). Ideally, when using association data in disease models the distance cutoff used to determine whether two individuals are in the same group should depend on how infectivity decreases with increasing distance. We assumed that buffalo were associating when they were in the same herd and the probability of infection between a particular susceptible-infected dyad was proportional to the time they spent in the same herd. This definition implicitly assumes that herds are sufficiently well mixed that within-herd transmission is equal between all dyads and sufficiently separated that between-herd transmission is nonexistent. Non-random association, however, may also play a role within herds when herds are large and diseases are transmitted only over very short distances. Unpublished data suggest that buffalo frequently move between the front, middle, and back portions of the herd, but it remains to be determined whether this is sufficient for our assumption of a well-mixed herd to be valid. Recognition and its role in determining association patterns may be important both within and between herds depending upon the distance over which individuals are infectious. In this study, however, we investigate the role of different association patterns at the herd-level only.

We constructed association indices between all possible dyads using the simple-ratio association index calculated from varying windows of time. Except where noted otherwise, association indices were calculated using monthly data or all the data from November 2001 to October 2003. Over the 24-month study period, there were 16

occasions when a pair of individuals was not seen during a month. In these cases, we assumed that the individuals were not associating during this period. Since this is less than 0.04% of the association indices used, this assumption probably had a negligible effect on our numerical simulations. When yearly estimates for 2002 and 2003 were calculated they were based upon data from November 2001 to October 2002 and November 2002 to October 2003, respectively, to coincide with the onset of the wet season.

We analyzed the association data in three ways, independent of the disease model. First we constructed dendrograms of the association matrices using the unpaired group averaging method (UPGMA), which had the highest cophenetic correlation coefficient as compared with dendrograms that we constructed using Ward's weighted, complete linkage, and single linkage clustering methods (Romesburg 1984). Second, we calculated a measure of the network clustering using the following metric, proposed by Keeling (1999): From the association matrix the total number of connected triples (three individuals sharing at least two connections) is calculated. Some proportion of these triples will be closed loops, or triangles, where all three individuals are connected.

Keeling's clustering coefficient, ϕ, is the ratio of triangles to triples in the network. When ϕ equals one then all triples form closed loops and one's neighbor's neighbor is always one's own neighbor as well. When ϕ is small, individuals have few associates in common. We calculated ϕ by first converting the association matrix to a matrix of zeros and ones where all non-zero associations were changed to one. Finally, we calculated the percentage of connections between individuals that changed from one month to the next, where a connection was defined as any association index greater than zero. All cluster analyses were coded in MATLAB 6.1 (MathWorks, Inc.).

Disease modelling

The model presented here is a stochastic individually-based elaboration of a discrete time Susceptible-Infected-Recovered (SIR) epidemic model (Anderson & May 1991, Kermack & McKendrick 1991), with time-steps of one month. A variable $x_i(t)$ is used to represent the state of an individual i at time t and equals zero when the individual is susceptible and one when the individual is infected. We assume that the probability for the disease to be transmitted between an infected individual i and a susceptible individual j is a function of the transmission coefficient β and the association coefficient $a_{ij}(t)$, where $a_{ij}(t)$ equals the proportion of time individuals i and j spent together over the period $[t-1,t]$. Additionally, we assume that infected individuals recover with constant probability γ per time-step and do not become susceptible again. Under these assumptions, our model takes the form:

$$\begin{aligned}&\mathrm{Prob}\left\{x_i(t+1)=1 \middle| x_i(t)=0\right\}\\&=\exp\left(-\beta\sum_{j=1}^{n}a_{ij}(t)x_j(t)\right)\quad t=0,1,2,...,T-1\end{aligned}, \tag{1}$$

$$\mathrm{Prob}\left\{x_i(t+1)=0 \middle| x_i(t)=1\right\}=\gamma, \tag{2}$$

where T is the length of the simulation in months and n is the total number of individuals in the population. In all simulations, we calculated $a_{ij}(t)$ values using the radio-tracking association data from the 64 buffalo for which we had a complete dataset during the study period. We started each simulation with one infected individual. Qualitative results were insensitive to the choice of this individual, except for a few individuals who were isolated from the others during the initial phase of the study period. In most cases, we simulated the model for 24 months to match the amount of radio-tracking data that was available. For those cases where we simulated the model for longer periods of time, we used the same association data repeatedly. We compared the results of this model with a mean-field equivalent to highlight the importance of incorporating the association data. The mean-field model assumed that all individuals were associated according to the grand mean of the association matrix for that month, so the overall force of infection for the mean-field model equaled that of the models using association data.

Using the buffalo association data and SIR model described above we investigated the

impact on disease dynamics of two aspects of association patterns: topology (i.e. who is connected to whom) and variation in the frequency of connections. Further, we investigated how the effects of association patterns depended on disease characteristics by varying the transmission coefficient β and probability of recovery γ to simulate "fast" versus "slow" diseases, while holding the basic reproductive number (R_0) constant. The basic reproductive number is the expected number of infections caused by the first case in a completely susceptible population, and is a measure of the growth potential of an epidemic (Anderson & May 1991).

We investigated the importance of topology by randomly rewiring connections in both static and dynamic network simulations, using the following algorithm. Let δ represent the proportion of network connections that were randomly re-assigned, such that $\delta = 0$ corresponds to the original network structure and $\delta = 1$ corresponds to a network with all connections randomly reassigned. Both the static and dynamic simulations started with the association matrix from November 2001. In the static network simulations, a proportion δ of the connections were rewired at the beginning of the simulation, then associations between individuals were assumed to be constant over time. In the dynamic network simulations, a proportion δ of the connections were rewired before every time-step.

Next, we investigated the importance of variation in connection frequency or strength. Previous research on network disease models typically assumed that the strengths of connections between individuals are equal (i.e. an 'unweighted' network) and thus the variance of connection strengths is zero. The variance in the amount of time spent together, however, may be biologically meaningful and have important consequences for disease dynamics. In dynamic networks, connection strength may vary within a dyad over time or among dyads. Assuming that individuals contact one another with a constant probability that is related to their time-averaged association index $\overline{a}_{ij}$, then the variance in contact frequency within the dyad is related to a binomial probability density function and decreases as $\overline{a}_{ij}$ approaches zero or one. We investigate the effects of increasing the variance of the time-averaged association indices $\overline{a}_{ij}$ among dyads, which results in lower variance within dyads. High variance among dyads would be analogous to some individuals having two sets of associates, those that they spend time with often and those that they do not. One might hypothesize that weak connections are less significant for disease transmission, and hence that systems with a high variance in connection strength may be less permeable to disease spread. On the other hand, the disease may spread more rapidly amongst those individuals that are tightly associated.

Let $\mathbf{A}_t$ represent the matrix of association coefficients a_{ij} that is based on data from the period $[t-1,t]$ and $\mathbf{A}_{\text{avg}}$ equal the time-average of $\mathbf{A}_t$ over the study period. Elements of $\mathbf{A}_{\text{avg}}$ between (but not including) zero and one indicate that a pair of individuals spent only a portion of the study period together. We introduced a parameter α to represent the relative amount to increase or decrease the time-averaged connection strength for each pair of individuals. For each matrix element $\overline{a}_{ij}$ in $\mathbf{A}_{\text{avg}}$, we drew a random variable z from a uniform distribution between zero and one, and calculated a new element $\overline{a}'_{ij}$ as:

$$\overline{a}'_{ij} = \begin{cases} \overline{a}_{ij} + \alpha\left(1 - \overline{a}_{ij}\right) & \text{IF} \quad \overline{a}_{ij} > z \\ \overline{a}_{ij} - \alpha\left(\overline{a}_{ij}\right) & \text{OTHERWISE} \end{cases} \tag{3}$$

Thus, we randomly increased or decreased each element of $\mathbf{A}_{\text{avg}}$ some proportion α of the distance between $\overline{a}_{ij}$ and zero or one. Specifically, when $\alpha = 0$ the time-averaged connection strength has its original value, and when $\alpha = 1$ all connections in the time-averaged association matrix are either zero or one. Using this algorithm we could increase the variance in time-averaged connection strengths while maintaining the expected mean connection strength. We also preserved the topology of connections from the original matrix $\mathbf{A}_{\text{avg}}$, except when α equals one and some connections were removed entirely. We applied this algorithm once to $\mathbf{A}_{\text{avg}}$ at the beginning of each simulation to create $\mathbf{A}'_{\text{avg}}$. Then for each timestep of the simulation we created an association matrix $\mathbf{A}'_t$ of zeros or ones, using $\overline{a}'_{ij}$ in the altered association matrix $\mathbf{A}'_{\text{avg}}$ as the probability that each connection exists.

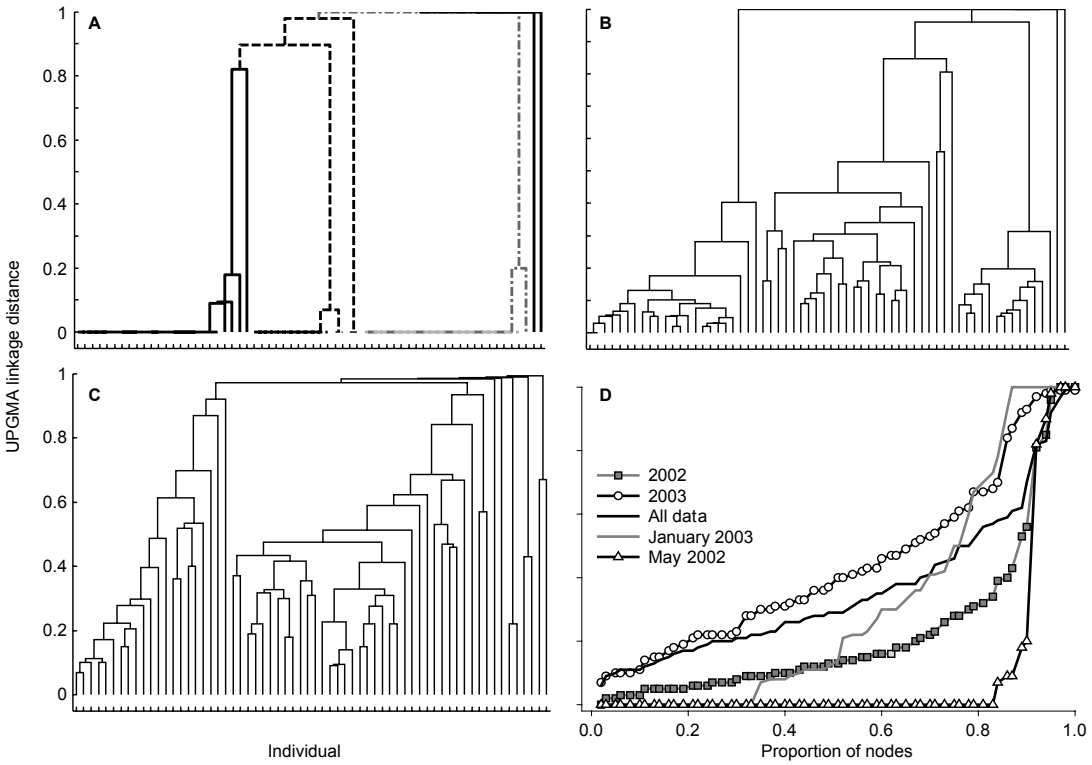

Fig. 2. UPGMA cluster analyses of 64 radio-collared buffalo in the Kruger National Park for different periods of data collection: (**A**) May 2002, (**B**) November 2001 through October 2002, (**C**) and November 2002 through October 2003. Buffalo with higher association indices are linked at lower linkage distances. In panel **A** the solid, dashed, and dot-dashed lines show three distinct herds present in May 2002 (Fig. 1). Panel **D** compares the overall structure of the dendrograms by showing the linkage distance required to include a given proportion of nodes (*see* text).

Thus we reconstructed an association history for the population in which pairs of individuals were assumed to be completely associated or isolated within time-steps, but the time-averaged association strength for each pair was determined by $\mathbf{A}'_{avg}$. High values of α corresponded to increased variance among pairs in time-averaged connection strength, but decreased variability in the existence of particular connections through time. Low values of α, which yielded more intermediate values of $\bar{a}_{ij}$ in the altered association matrix $\mathbf{A}'_{avg}$, corresponded to lower variation between pairs but higher temporal variation for each pair. Note that this approach is distinct from altering the variation in association indices within each time-step. To explore the role of variance in connection strength, we experimented with different $\mathbf{A}_{avg}$ matrices (e.g. using all data or just the data from particular months). All simulations were conducted in MATLAB 6.1 (MathWorks, Inc.).

Results

Association data

The time frame used to calculate association indices has a large effect on the apparent structure of the system and thus conclusions about the ability of a disease to spread through that system. We visualize the network structure using 3-D network graphs and dendrograms (Figs. 1 and 2). The network graph of May 2002 (Fig. 1A) indicates two distinct groups and two outlying individuals all of whom become well connected when considering contacts over the whole period from November 2001 to October 2003 (Fig. 1B). In general, association networks become more connected over longer time frames due to the movement of individuals between groups, thereby allowing diseases to spread among groups. Dendrograms illustrate the hierarchy of

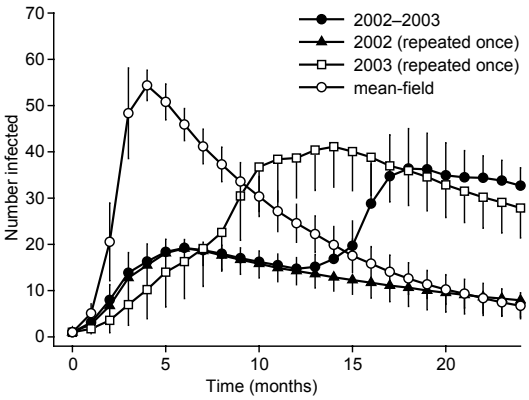

Fig. 3. Mean and standard deviations of the number of infected individuals for 50 runs of the disease model using monthly association data from the entire study period (closed circles), 2002 (closed triangles), 2003 (open squares), or a mean-field model (open circles) where all individuals were connected but the force of infection per month was the same. All simulations used a transmission coefficient β of 0.3 and recovery probability γ of 0.1. For simulations using one year of data, the same association matrices were used again the second year.

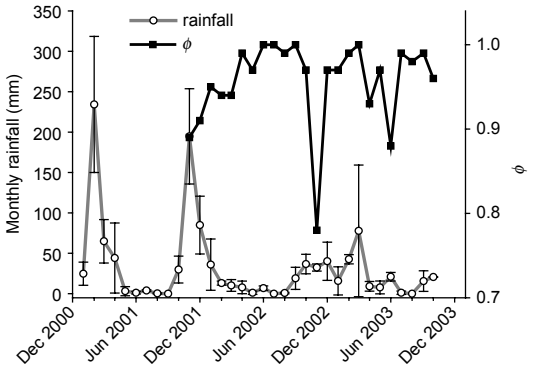

Fig. 4. Monthly rainfall (mm) and Keeling's clustering coefficient ϕ of the association data during the course of the study. Mean and standard deviations of the monthly rainfall were calculated using four rainfall stations in the Satara region of the Kruger National Park. Rainfall data from 2000 to 2001 are shown to contrast the below-average wet season in 2002–2003.

associations in the buffalo network (Fig. 2). Individuals that are more often together, similar to highly related species on a phylogeny, are joined at lower linkage distances on the dendrogram. As in the network graphs, dendrograms based upon a month of data were generally more tightly clustered than those based upon a year of data (Fig. 2).

Buffalo appeared more highly clustered in 2002 than 2003 (cf. Fig. 2B and C). In 2002, three groups are apparent, but two of these appeared to have merged in 2003. Further, as compared with 2002, in 2003 a lower proportion of nodes (where two individuals/groups are joined in the dendrogram) were included for a given linkage distance, indicating looser associations between individuals and groups (Fig. 2D). Also, considerable variation exists between months, with May 2002 and January 2003 respectively representing the low and high connectivity extremes of monthly association data (Fig. 2A and D). Although the network graph for May 2002 (Fig. 1A) shows only two groups, the dendrogram, which accounts for the relative strength of associations, appears to show three groups (Fig. 1A). Over the entire study period the frequency of mixing events between herds resulted in a well-connected network (Fig. 1B).

Simulation results

To show the effects of the clustering patterns in Fig. 2, we simulated SIR disease dynamics (β = 0.3, γ = 0.1) using monthly association matrices from either the entire dataset, one year of data, or a mean-field model (Fig. 3). The mean-field model where all individuals were connected predicts much faster spread of disease even though the force of infection was the same as the model using all the association data (cf. closed circles to open circles; Fig. 3). When we used only November 2001 through October 2002 (i.e. year 2002), then repeated the same values to simulate months 13 through 24, the disease was limited to only one herd and did not infect as many individuals as compared with our using the data from 2003 (Fig. 3). Thus, the tighter clustering of buffalo in 2002 as compared with that in 2003 (Fig. 2) translated into a population that is less permeable to disease invasion. When we used the monthly data for the whole period, there was a second pulse of infections starting around month 14, presumably due to the disease moving into a new herd (Fig. 3). Coincident with the second pulse of infections was a wet season with

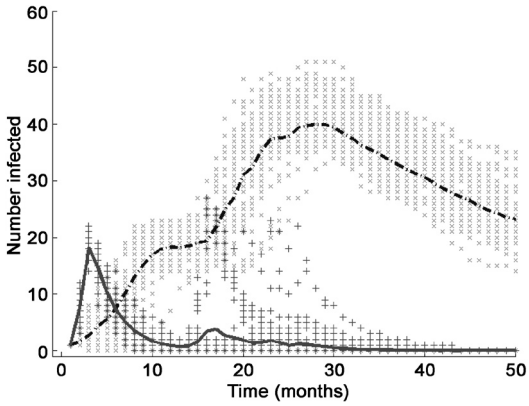

Fig. 5. The number of individuals infected as a function of the speed of the disease. Slow diseases (dotted line and x's; $\beta = 0.04$, $\gamma = 0.03$) allow for more switching of infectious individuals between groups than faster diseases (solid line and +'s; $\beta = 0.4$, $\gamma = 0.3$), and hence for greater overall disease spread. Simulations used monthly association values from November 2001 to October 2003. Symbols show particular model runs, and lines represent the mean of 50 runs.

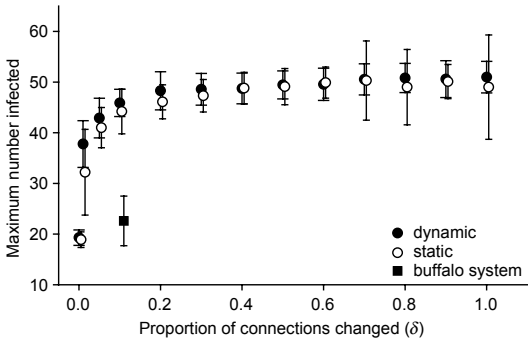

Fig. 6. The maximum number of individuals infected at any point in time after 50 time steps depends upon the amount δ of random rewiring of the association network at the beginning of each simulation (static) or cumulatively every time-step (dynamic). Dynamic and static simulations started with association data from November 2001; the buffalo system data point was based on all of the association data (i.e. unmanipulated). Disease parameters were $\beta = 0.3$, $\gamma = 0.2$. Error bars represent the standard deviations from 50 stochastic simulations. (For clarity, δ values of the static simulations were increased slightly before plotting.)

below-average rainfall (November 2002 to February 2003). This also coincided with a marked decrease in the clustering of the association data in November 2002 as indicated by Keeling's clustering coefficient ϕ (Fig. 4).

Using all the monthly association data we simulated two SIR-type diseases with the same basic reproductive number ($R_0 = \beta/\gamma$) but different infectious periods. A faster-moving disease, with high infectiousness and rapid recovery ($\beta = 0.4$, $\gamma = 0.3$), is more likely to fade out in populations where the frequency of movement between groups is low, because it may burn out in the local population before sufficient connections to other groups are made (Fig. 5). A slower-moving disease with the same R_0 ($\beta = 0.04$, $\gamma = 0.03$) is less infectious but persists longer, increasing the probability of transmission to other groups of individuals (Fig. 5). Thus, slower diseases effectively integrate over a longer period and the network becomes more fully connected (Fig. 1).

Next, to assess the impact of the network structure on the spread of disease, we analyzed the system using only particular months of the association data and either randomly rearranging network connections or increasing the variance in connection strengths. Random rearrangement (or "rewiring") of the network connections involves establishing a new contact between two randomly chosen individuals and removing a contact between two other individuals. Random rewiring had a non-linear effect upon disease dynamics (Fig. 6). Only a small amount of rewiring ($\delta < 0.2$) creates a network that behaves like a randomly wired network ($\delta = 1.0$). Furthermore, because only a small amount of rewiring had a large impact, the static simulations (i.e. rewiring the matrix once at the beginning of the simulation) yielded similar results to the dynamic simulations (i.e. cumulative random rewiring of the matrix over time), though for very small δ the dynamic rewiring showed greater effects as expected (Fig. 6). Other output variables, such as the number of susceptible individuals remaining at the end of the simulation, showed results similar to those in Fig. 6.

To place these random rewiring simulations into context, we plot the percentage of the topology that remains the same over increasing time lags for the monthly empirical data and dynamic rewiring simulations (Fig. 7). The percentage of similar topology of the empirical data appears to be similar to randomly rewiring 10% of the

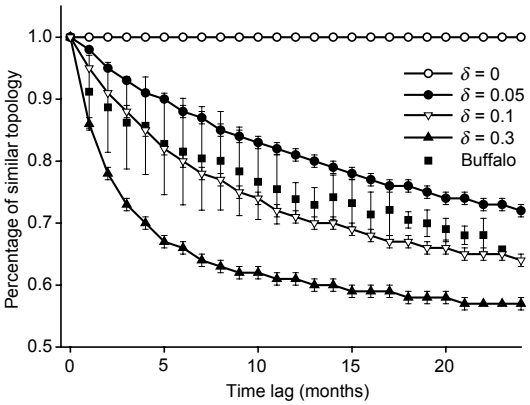

Fig. 7. Random rewiring of a proportion of connections δ in the association matrix each time-step decreases the similarity between association matrices over time. Buffalo data (squares) represent the mean and SD of the similarity between all available association matrices that are separated by a time lag of one to 23 months, whereas other lines represent the mean and SD of 50 simulations of the model starting with November 2001 association data using different δ values.

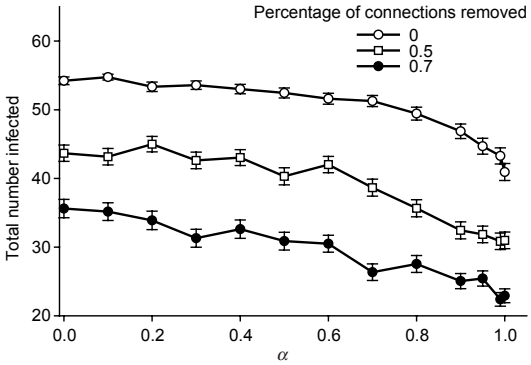

Fig. 8. The total number of individuals infected after 50 time-steps decreases with increasing variability in the time-averaged connection strength between pairs and decreasing temporal variability of connections within pairs (α) and the proportion of connections that are removed from the network. *See* Methods for a description of how α increases the variance in association indices. Error bars indicate the standard errors of 200 simulations using September 2003 as the association matrix, $\beta = 0.3$, and $\gamma = 0.1$.

connections (i.e. $\delta \approx 0.1$) in the November 2001 data (i.e. about 10% of the buffalo network changed per month; Fig. 7). However, for reasons we discuss later, the disease dynamics of simulations using $\delta = 0.1$ differ greatly from those based on the actual association data, which yield results more similar to runs with $\delta \leq 0.01$ runs (Fig. 6).

The variance in the connection strength among dyads was only important under certain circumstances. Using data from September 2003 and a fast SIR disease ($\beta = 0.3$, $\gamma = 0.1$) as an example, Fig. 8 illustrates how the total number of individuals infected over the entire epidemic was a function of α (a proxy for the variance in connection strength; *see* Methods) and the connectivity of the network. We conducted the same analysis as in Fig. 8 for other disease characteristics and found that α had less effect upon disease dynamics for slower diseases (e.g. $\beta = 0.03$, $\gamma = 0.01$). Further, α had little effect when we used data from months where buffalo herds were either very well connected or very weakly connected (data not shown). Finally, dropping a certain proportion of network connections had a much larger impact than increasing α (Fig. 8).

Discussion

Early work on disease modeling assumed instantaneous random mixing between all individuals. More recently researchers have begun to account for non-random host mixing patterns, often using a static network (e.g. Watts 1999, Newman 2003) although one recent study incorporated changing connections within a background network that is static (Eames & Keeling 2004). Pair-formation models of sexually transmitted diseases include dynamic contacts but usually not the fully non-random mixing of network structure (Dietz & Hadeler 1988, Kretzschmar & Morris 1996, Morris & Kretzschmar 1997, Lloyd-Smith *et al.* 2004). To the authors' knowledge, empirically-derived, fully dynamic contact networks have not been used as a substrate for investigating disease dynamics, but *see* Corner *et al.* (2003) for an example of a more experimental approach to social networks in a wildlife disease system. Our analysis shows how disease dynamics depend on the topology of connections between individuals, the dynamic nature of these connections, and the variance in the frequency of contacts between individuals. Association data similar to those presented here are often collected by behavioral researchers (e.g. Myers 1983, Smolker *et*

al. 1992, Brager *et al*. 1994, Whitehead 1999, Szykman *et al*. 2001), but have not been combined with models of disease dynamics even though they may have important impacts upon the spread of disease (Fig. 3).

Analysis of the association data suggests that the buffalo population was more tightly clustered in 2002, but with more connections between groups in 2003 (Fig. 2). From the perspective of a disease, this translates into a more permeable population in 2003 (Fig. 3). More specifically, many buffalo were moving to new areas and groups in November 2002, which resulted in a less tightly clustered population at the same time when rainfall was well below average (Fig. 4). Rainfall totals for November 2002 through January 2003 were only 34% of the long-term average for those months. Results from the model then present a testable prediction that dry conditions facilitate more rapid spatial spread of disease in the buffalo population due to increased herd-switching.

This presents a worrying scenario for southern Africa, where precipitation is predicted to remain stable or decline while temperatures are likely to increase due to global climate change (Hulme *et al*. 2001). As a result, vegetation conditions are likely to decline, which may result in an increase in the amount of population mixing as wildlife is forced out of previously habitable areas. Many studies have linked climate change with altered disease distributions or dynamics (e.g. Epstein 2002, Patz & Khaliq 2002). However, evidence for the impact of climate change on animal behavior, which in turn affects disease dynamics, is rare.

Previous studies of African buffalo have suggested that herds are relatively static (Sinclair 1977, Prins 1996). We show, however, that buffalo herds are very dynamic in the KNP and that over time the population becomes well-connected (Fig. 1). As a result, over time it becomes difficult to define what is a herd. As suggested by a reviewer, we propose a definition of a herd based upon the association data, such that a herd is defined as the set of maximal complete subgraphs of the network where all individuals, or vertices, are associated with a_{ij} greater than some threshold (for a particular time period). Here a complete subgraph is defined as a subset of the vertices of the network such that for every pair of vertices in the subset, there is an edge connecting them and the set of those edges and vertices are a complete network and a subset of the original network. We will explore the implications of this definition for our buffalo data in a future publication.

In considering effects of association patterns on disease, one issue of critical importance is accounting for the relative time scales of the disease and host mixing patterns. This interplay has important implications for the spread of disease, the way association data should be collected, and the evolution of both host and parasite. Fast diseases, such as rabies, measles, and Ebola hemorrhagic fever, with a short duration of infectivity encounter a more structured population because the amount of mixing between groups decreases as the time frame decreases (Figs. 2D and 5). Thus, association data should have higher time resolution if they are intended to give insight into the spread of diseases with fast dynamics. More specifically, association data should be collected at least as frequently as the duration of infection for the disease of interest, to capture population mixing on timescales relevant to the disease. Furthermore, association data should not be averaged over too long a period. If association matrices are constructed using data from a time frame longer than the duration of infection within an individual, then the network is biased in favor of too many connections between individuals. As an extreme example compare the connectivity of the networks based upon one or 24 months of data (Fig. 1).

Similar to Watts's (1999) work on static networks we found that only a small amount of random rewiring is necessary to make the empirical buffalo network behave as a randomly wired network. Further, there was little difference between rewiring the network once at the beginning of the simulation or once each timestep (Fig. 6). This is probably due to limited number of groups in the network. Only a small number of connections are necessary to make it a 'small-world' graph where there are very few degrees of separation between any two individuals and additional changes to the network have little additional impact (Fig. 6). Interestingly, the model predicts that fewer individuals would be

infected when using empirical association data as compared with the simulated data that approximates the amount of change in the association matrix ($\delta \approx 0.1$) seen in the empirical dataset (Fig. 6). This suggests that although the empirical network is changing every time step, these changes are not random and a proportion of the population remains inaccessible to the disease (Fig. 6). Thus the population is less permeable to disease invasion than one would expect if movement of individuals between groups was random. This amount of movement is higher than what might have been expected from previous studies of African buffalo (Sinclair 1977, Prins 1996), however, given the intensive radio-tracking conducted in this study probably more accurately describes the fluid structure of the population in the KNP (Cross *et al*. [in press]).

Somewhat surprisingly, we found that increasing the variance in connection strength among dyads had only minor effects upon disease dynamics in this system (Fig. 8). This suggests that for the range of networks we investigated, disease dynamics are very similar in a system where individuals spend their time equally with all associates or spend a large portion of their time with only a few individuals and a little time with many others. This conclusion, however, may be limited to well-connected networks with a limited number of groups. Further, we did not investigate the effects of increasing the variation in contact frequency within a pair of individuals. More work is necessary to determine the generality of these results and in what contexts networks of weighted connections behave differently than unweighted networks.

Future directions

Despite the known importance of association patterns to disease dynamics, empirical data are lacking, especially for airborne diseases. First of all, disease ecology is a relatively new field, and although a number of studies have investigated how association patterns may affect disease dynamics (e.g. Keeling 1999, Eames & Keeling 2003, Boots *et al*. 2004), few studies have attempted to integrate empirical data and disease models. Secondly, an accurate depiction of a social network is very difficult to generate. We believe that future research should focus on the following questions, which currently limit our ability to apply empirical data to disease models: (1) What defines a contact for airborne diseases? (2) What are the appropriate time and spatial scales to sample a network of animals? (3) How does one scale-up a sample of a network to represent an entire population? (4) Given that population dynamics are an important factor in disease dynamics, how does one allow for births, deaths and changes of association patterns while maintaining the overall properties of a network? Studies on the tradeoffs faced by individuals in social systems with regard to disease and behavior could also be enlightening (e.g. Adamo *et al*. 2001, Boots & Knell 2002). In particular, individuals may modify their behavior to decrease their contact with others when they are at higher disease risk. If empirical data were available, they could be incorporated into the network model framework by adjusting the association indices between infectious and susceptible individuals.

In conclusion, the way that individuals associate with one another has a large effect upon the spread and dynamics of a disease. Although there are a few questions that need to be answered before empirically-derived social networks can be widely applied, the methods presented here provide a flexible framework for combining behavioral data with models of disease dynamics. Finally, our results suggest that a critical interplay exists between the time scales over which social interactions take place and those associated with a particular disease. An assessment of how likely it is that a given disease become an epidemic requires that we pay attention to social interaction and disease time scales, which then determines the appropriate temporal resolution for the collection of data determining the structure of social networks.

Acknowledgements

This research was funded by the United States National Science Foundation Ecology of Infectious Disease Grant DEB-0090323. Many thanks to Stephan Ferreira, Julie Wolhuter, Kutani Bulunga, Augusta Mabunda, and Fernando Muhlovo for their help collecting field data. Drs. Douw Grobler, Lin-

Mari de Klerk, Peter Buss, KNP Game Capture, and State Veterinary technicians assisted in buffalo capture operations. We are grateful to Molly Smith for useful comments on previous versions of this paper, Neo Martinez for assisting with the network graphs, and two anonymous reviewers, one of which suggested using the association data as a more rigorous definition of a herd.

References

Adamo, S. A., Jensen, M. & Younger, M. 2001: Changes in lifetime immunocompetence in male and female *Gryllus texensis* (formerly G-integer): trade-offs between immunity and reproduction. — *Anim. Behav.* 62: 417–425.

Alcock, J. 1998: *Animal behavior: an evolutionary approach*. — Sinauer Associates, Inc. Massachusetts.

Anderson, R. M. & May, R. M. 1991: *Infectious diseases of humans: dynamics and control*. — Oxford University Press, Oxford.

Anderson, R. M., Gupta, S. & Ng, W. 1990: The significance of sexual partner contact networks for the transmission dynamics of HIV. — *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 3: 417–429.

Bauch, C. & Rand, D. A. 2001: A moment closure model for sexually transmitted disease transmission through a concurrent partnership network. — *Proc. R. Soc. Lond. B* 267: 2019–2027.

Bengis, R. G. 1999: Tuberculosis in free-ranging mammals. — In: Fowler, M. E. & Miller, R. E. (eds.), *Zoo and wild animal medicine*: 101–114. W. B. Sauders Company, Philadelphia.

Bengis, R. G., Kriek, N. P., Keet, D. F., Raath, J. P., De Vos, V. & Huchzermeyer, H. 1996: An outbreak of bovine tuberculosis in a free-living African buffalo (*Syncerus caffer* Sparrman) population in the Kruger National Park: A preliminary report. — *Onderstepoort J. Vet. Res.* 63: 15–18.

Boots, M. & Knell, R. J. 2002: The evolution of risky behaviour in the presence of a sexually transmitted disease. — *Proc. R. Soc. Lond. B* 269: 585–589.

Boots, M. & Sasaki, A. 1999: 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. — *Proc. R. Soc. Lond. B* 266: 1933–1938.

Boots, M., Hudson, P. J. & Sasaki, A. 2004: Large shifts in pathogen virulence relate to host population structure. — *Science* 303: 842–844.

Brager, S., Wursig, B., Acevedo, A. & Henningsen, T. 1994: Association patterns of bottlenose dolphins (*Tursiops truncatus*) in Galvestion Bay, Texas. — *J. Mammal.* 75: 431–437.

Corner, L. A. L., Pfeiffer, D. U. & Morris, R. S. 2003: Social network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). — *Prev. Vet. Med.* 59: 147–167.

Cross, P. C., Lloyd-Smith, J. O., & Getz, W. M. [in press]: Disentangling association patterns in fission-fusion societies using African buffalo as an example. — *Animal Behaviour.*

Dietz, K. & Hadeler, K. P. 1988: Epidemiological models for sexually-transmitted diseases. — *J. Math. Biol.* 26: 1–25.

Dugatkin, L. A. 1997: *Cooperation among animals*. — Oxford University Press, Oxford.

Eames, K. T. D. & Keeling, M. J. 2002: Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. — *Proc. Natl. Acad. Sci. U.S.A.* 99: 13330–13335.

Eames, K. T. D. & Keeling, M. J. 2003: Contact tracing and disease control. — *Proc. R. Soc. Lond. B* 270: 2565–2571.

Eames, K. T. D. & Keeling, M. J. 2004: Monogamous networks and the spread of sexually transmitted diseases. — *Math. Biosci.* 189: 115–130.

Edmunds, W. J., O'Callaghan, C. J. & Nokes, D. J. 1997: Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. — *Proc. R. Soc. Lond. B* 264: 949–957.

Epstein, P. R. 2002: Biodiversity, climate change, and emerging infectious diseases — In: Aguirre, A. A., Ostfeld, R. S., Tabor, G. M., House, C. & Pearl, M. C. (eds.), *Conservation medicine: ecological health in practice*: 27–39. Oxford University Press, Oxford.

Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczal, Z. & Wang, N. 2004: Modelling disease outbreaks in realistic urban social networks. — *Nature* 429: 180–184.

Ferguson, N. M. & Garnett, G. P. 2000: More realistic models of sexually transmitted disease transmission dynamics — Sexual partnership networks, pair models, and moment closure. — *Sex. Transm. Dis.* 27: 600–609.

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001: Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. — *Nature* 413: 542–548.

Halley, D. J., Vanderwalle, M. E., Mari, M. & Taolo, C. 2002: Herd-switching and long distance dispersal in female African buffalo *Syncerus caffer*. — *East African Wildlife Society* 40: 97–99.

Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. 2001: African climate change: 1900–2100. — *Climate Research* 17: 145–168.

Jolly, A. M. & Wylie, J. L. 2002: Gonorrhoea and chlamydia core groups and sexual networks in Manitoba. — *Sex. Transm. Infect.* 78: I145–I151.

Keeling, M. J. 1999: The effects of local spatial structure on epidemiological invasions. — *Proc. R. Soc. Lond. B* 266: 859–867.

Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G. & Grenfell, B. T. 2003: Modelling vaccination strategies against foot-and-mouth disease. — *Nature* 421: 136–142.

Keet, D. F., Kriek, N. P., Penrith, M. L., Michel, A. & Huchzermeyer, H. 1996: Tuberculosis in buffaloes (*Syncerus caffer*) in the Kruger National Park: spread of the disease to other species. — *Onderstepoort J. Vet. Res.* 69: 239–244.

Kermack, W. O. & McKendrick, A. G. 1991: Contributions

to the mathematical theory of epidemics. 1. [Reprinted from *Proceedings of the Royal Society* 115a(1927): 700–721). — *Bull. Math. Biol.* 53: 33–55.

Kretzschmar, M. 2000: Sexual network structure and sexually transmitted disease prevention — A modeling perspective. — *Sex. Transm. Dis.* 27: 627–635.

Kretzschmar, M. & Morris, M. 1996: Measures of concurrency in networks and the spread of infectious disease. — *Math. Biosci.* 133: 165–195.

Liebert, A. E. & Starks, P. T. 2004: The action component of recognition systems: a focus on the response. — *Ann. Zool. Fennici* 41: 747–764.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Aberg, Y. 2001: The web of human sexual contacts. — *Nature* 411: 907–908.

Lloyd-Smith, J. O., Galvani, A. P. & Getz, W. M. 2003: Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. — *Proc. R. Soc. Lond. B* 270: 1979–1989.

Lloyd-Smith, J. O., Getz, W. M. & Westerhoff, H. V. 2004: Frequency-dependent incidence in models of sexually transmitted diseases: portrayal of pair-based transmission and effects of illness on contact behaviour. — *Proc. R. Soc. Lond. B* 271: 625–634.

Meyers, L. A., Newman, M. E. J., Martin, M. & Schrag, S. 2003: Applying network theory to epidemics: Control measures for Mycoplasma pneumoniae outbreaks. — *Emerg. Infect. Dis.* 9: 204–210.

Mloszewski, M. J. 1983: *The behavior and ecology of the African buffalo*. — Cambridge University Press, Cambridge.

Morris, M. & Kretzschmar, M. 1997: Concurrent partnerships and the spread of HIV. — *AIDS* 11: 641–648.

Myers, J. P. 1983: Space, time and pattern of individual associations in a group-living species: sanderlings have no friends. — *Behav. Ecol. Sociobiol.* 12: 129–134.

Newman, M. E. J. 2002: Spread of epidemic disease on networks. — *Physical Review E* 66. DOI:10.1103/PhysRevE.66.016128.

Newman, M. E. J. 2003: The structure and function of complex networks. — *Siam Review* 45: 167–256.

Patz, J. A. & Khaliq, M. 2002: Global climate change and health: challenges for future practitioners. — *JAMA* 287: 2283–2284.

Prins, H. H. 1996: *Ecology and behaviour of the African buffalo*. — Chapman and Hall, London.

Read, J. M. & Keeling, M. J. 2003: Disease evolution on networks: the role of contact structure. — *Proc. R. Soc. Lond. B* 270: 699–708.

Rodwell, T. C., Whyte, I. J. & Boyce, W. M. 2001: Evaluation of population effects of bovine tuberculosis in free-ranging African buffalo (*Syncerus caffer*). — *J. Mammal.* 82: 231–238.

Rodwell, T. C., Kriek, N. P., Bengis, R. G., Whyte, I. J., Viljoen, P. C., De Vos, V. & Boyce, W. M. 2000: Prevalence of bovine tuberculosis in African buffalo at Kruger National Park. — *J. Wildl. Dis.* 37: 258–264.

Romesburg, H. C. 1984: *Cluster analysis for researchers*. — Lifetime Learning Publications, Belmont, CA.

Sinclair, A. R. E. 1977: *The African buffalo*. — University of Chicago Press, Chicago.

Smolker, R. A., Richards, A. F., Connor, R. C. & Pepper, J. W. 1992: Sex differences in patterns of association among Indian Ocean bottlenose dolphins. — *Behaviour* 123: 38–69.

Szykman, M., Engh, A. L., Van Horn, R. C., Funk, S. M., Scribner, K. T. & Holekamp, K. E. 2001: Association patterns among male and female spotted hyenas (*Crocuta crocuta*) reflect male mate choice. — *Behav. Ecol. Sociobiol.* 50: 231–238.

Wallinga, J., Edmunds, W. J. & Kretzschmar, M. 1999: Perspective: human contact patterns and the spread of airborne infectious diseases. — *Trends Microbiol.* 7: 372–377.

Watts, D. J. 1999: *Small worlds: the dynamics of networks between order and randomness*. — Princeton University Press, Princeton.

Whitehead, H. 1999: Testing association patterns of social animals. — *Anim. Behav.* 57: F26–F29.

Whitehead, H. & DuFault, S. 1999: Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. — *Advances in the Study of Behavior* 28: 33–74.

Woodhouse, D. E., Rothenberg, R. B., Potterat, J. J., Darrow, W. W., Muth, S. Q., Klovdahl, A. S., Zimmerman, H. P., Rogers, H. L., Maldonado, T. S., Muth, J. B. & Reynolds, J. U. 1994: Mapping a social network of heterosexuals at high risk for HIV infection. — *AIDS* 8: 1331–1336.