

## Epidemiological effects of group size variation in social species

Damien Caillaud, Meggan E. Craft and Lauren Ancel Meyers

*J. R. Soc. Interface* 2013 **10**, 20130206, published 10 April 2013

---

### Supplementary data

["Data Supplement"](#)

[http://rsif.royalsocietypublishing.org/content/suppl/2013/04/08/rsif.2013.0206.DC1.htm](http://rsif.royalsocietypublishing.org/content/suppl/2013/04/08/rsif.2013.0206.DC1.html)  
l

### References

[This article cites 35 articles, 10 of which can be accessed free](#)

<http://rsif.royalsocietypublishing.org/content/10/83/20130206.full.html#ref-list-1>

### Subject collections

Articles on similar topics can be found in the following collections

[bioinformatics](#) (33 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

[rsif.royalsocietypublishing.org](http://rsif.royalsocietypublishing.org)



## Research

**Cite this article:** Caillaud D, Craft ME, Meyers LA. 2013 Epidemiological effects of group size variation in social species. *J R Soc Interface* 10: 20130206.  
<http://dx.doi.org/10.1098/rsif.2013.0206>

Received: 4 March 2013  
Accepted: 19 March 2013

**Subject Areas:**  
biomathematics

**Keywords:**  
epidemic threshold, network epidemiology, susceptible–infected–removed model, social network, social structure, *Panthera leo*

**Author for correspondence:**  
Damien Caillaud  
e-mail: [dcaillaud@austin.utexas.edu](mailto:dcaillaud@austin.utexas.edu)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2013.0206> or via <http://rsif.royalsocietypublishing.org>.

# Epidemiological effects of group size variation in social species

Damien Caillaud<sup>1,5</sup>, Meggan E. Craft<sup>2</sup> and Lauren Ancel Meyers<sup>1,3,4</sup>

<sup>1</sup>Section of Integrative Biology, The University of Texas at Austin, 1 University Station C4500, Austin, TX 78712, USA

<sup>2</sup>Veterinary Population Medicine, University of Minnesota, 1988 Fitch Avenue, St Paul, MN 55108, USA

<sup>3</sup>Division of Statistics and Scientific Computation, The University of Texas at Austin, 1 University Station G2500, Austin, TX 78712, USA

<sup>4</sup>Santa Fe Institute, Hyde Park Road, Santa Fe, NM 87501, USA

<sup>5</sup>The Dian Fossey Gorilla Fund International, 800 Cherokee Avenue, Atlanta, GA 30315, USA

Contact patterns in group-structured populations determine the course of infectious disease outbreaks. Network-based models have revealed important connections between group-level contact patterns and the dynamics of epidemics, but these models typically ignore heterogeneities in within-group composition. Here, we analyse a flexible mathematical model of disease transmission in a hierarchically structured wildlife population, and find that increased variation in group size reduces the epidemic threshold, making social animal populations susceptible to a broader range of pathogens. Variation in group size also increases the likelihood of an epidemic for mildly transmissible diseases, but can reduce the likelihood and expected size of an epidemic for highly transmissible diseases. Further, we introduce the concept of *epidemiological effective group size*, which we define to be the group size of a hypothetical population containing groups of identical size that has the same epidemic threshold as an observed population. Using data from the Serengeti Lion Project, we find that pride-living Serengeti lions are epidemiologically comparable to a homogeneous population with up to 20 per cent larger prides.

## 1. Introduction

The structure of animal and human societies is often hierarchical. In their simplest form, societies have two hierarchical levels: individuals form groups, and groups form populations. In some species, such as humans, hamadryas baboons (*Papio hamadryas*), gelada (*Theropithecus gelada*) or zebras (*Equus burchelli*), populations comprise additional hierarchical levels referred to as clans, bands, herds or troops [1–4]. In such multi-level societies, individuals interact preferentially with other members of their own social clade, and may only rarely come into contact with individuals belonging to distant clades. Theoretical work has shown that variable contact rates between individuals or groups of individuals alter the spread of infectious diseases and other transmissible elements, such as information and cultural traits [5–7].

Network models have been used extensively to explore the impact of wildlife contact patterns on disease transmission [8,9]. With this approach, one can represent animal societies by networks in which individuals or groups of individuals constitute nodes (or vertices) that are connected by edges representing their social interactions. The number of connections per node is called the degree, and the distribution of degrees across all nodes in a network is called the degree distribution [10]. In network epidemiology, nodes are commonly characterized by a discrete state variable taking on the values: susceptible (S), infected (I) or removed (R) [11]. Nodes transition from one state to another depending on their own state and the states of neighbouring nodes. Network models have provided important insights into the epidemiological effects of contact heterogeneity, but are often limited to a single layer of connectivity. Hierarchically structured populations are often modelled using networks in

which nodes represent entire groups [12–16]. Such models ignore both within-group connectivity and variation in group size, both of which can influence disease transmission.

Here, we introduce a network-based model of infectious disease transmission in a two-level hierarchically structured population, and use it to investigate the impact of group size distributions on disease dynamics. We focus specifically on the mean and the variance of group sizes. If the transmission probability between two connected groups is assumed to be positively correlated with the sizes of the groups, then larger mean group sizes are expected to facilitate disease transmission. However, the variance of group size works in subtler ways. We use bond percolation methods [7,17,18] to derive expressions for (i) epidemic thresholds (above which the disease has the potential to infect a large number of individuals); (ii) expected size and variance of small outbreaks; and (iii) the probability and expected size of epidemics in two-level hierarchical networks. We show that when the mean group size is held constant, the variance of group size strongly influences transmission dynamics. Whether it facilitates or hinders disease spread depends on the transmission rate of the disease.

Using our model, we then study the impact of pride size distributions on disease transmission in a well-studied lion population (*Panthera leo*) in the Serengeti National Park, Tanzania [19]. This population has been affected by multiple infectious disease outbreaks [20,21], some of which have caused substantial mortality and could pose conservation threats, including canine distemper virus and feline immunodeficiency virus [22–25]. The mean pride size is approximately 10 individuals, whereas the variance in pride size is six times that. We show that such high pride size variance has considerable epidemiological consequences, and should be explicitly considered in epidemiological and conservation assessments of wildlife populations.

## 2. Theoretical model

Consider an infectious disease spreading through an undirected network in which nodes represent social groups of different sizes and edges represent contacts between groups (figure 1a). Disease can be transmitted between groups when infected individuals from one group come into contact with susceptible individuals from another group. In the following model, we assume that the probability of such an event depends on the sizes of the two interacting groups and that, once the disease reaches a group, all members of the group eventually become infected (the latter assumption can be relaxed, see §4).

In this section, we present and provide mathematical motivation for several key epidemiological quantities: (i) the epidemic threshold, which is a value that indicates how contagious a disease must be to cause an epidemic in a particular population, (ii) for diseases below the epidemic threshold, estimates of the mean and variance in outbreak sizes, (iii) for diseases above the threshold, the probability and expected size of an epidemic, and (iv) the epidemiological effective group size, which is a metric that indicates the epidemiological vulnerability of a population via comparison with hypothetical populations with equal-sized groups.

### 2.1. Epidemic threshold

We will present epidemiological quantities for group-to-group contact networks that rely on two probability

distributions:  $S_0(n)$ , the probability that a randomly chosen group has size  $n$ , and  $D_0(k)$ , the probability that a randomly chosen group has degree  $k$  (where the degree of a node is the number of edges connecting it to other nodes). We assume that group size and degree are independent variables, that is, the probability that a random group has size  $n$  and degree  $k$  is simply the product  $S_0(n) \cdot D_0(k)$ .

Suppose that an entirely infected group of size  $n_1$  is connected to an entirely susceptible group of size  $n_2$ . The probability  $T$  that the disease is transmitted between an infected individual from the first group and a susceptible individual from the second group is given by

$$T = 1 - \lim_{dt \rightarrow 0} (1 - \beta dt)^{\tau/dt} = 1 - e^{-\beta\tau}, \quad (2.1)$$

where  $\beta$  is the rate of disease-causing contacts between individuals in connected groups and  $\tau$  is the duration of infectiousness. Each parameter is assumed to be identical for all individuals. The probability that at least one individual from the second group becomes infected is then

$$\theta(n_1, n_2) = 1 - e^{-\beta\tau n_1 n_2}. \quad (2.2)$$

Hereafter, we will assume that  $\beta\tau n_1 n_2$  is small compared with 1, so that  $\theta(n_1, n_2)$  can be approximated by its first-order Taylor series expansion:

$$\theta(n_1, n_2) \simeq \beta\tau n_1 n_2 \simeq T n_1 n_2. \quad (2.3)$$

As described below, this approximation allows the derivation of closed-form expressions of important epidemiological quantities, and is within 10 per cent of the exact value of  $\theta(n_1, n_2)$  as long as  $\theta(n_1, n_2) < 0.19$ , and within 20 per cent if  $\theta(n_1, n_2) < 0.37$ .

Consider an outbreak that originates with a single infected group of size  $n_1$ . This group is expected to transmit the disease along each of its outgoing edges with average probability given by

$$\theta(n_1) \simeq T n_1 \langle n \rangle, \quad (2.4)$$

where  $\langle n \rangle$  is the average group size in the population. Assuming the group has degree  $k$ , the number of outgoing edges  $m$  along which the disease is expected to be transmitted then follows a binomial distribution with parameters  $k$  and  $\theta(n_1)$ :

$$P(m, k, n_1) = \binom{k}{m} \theta(n_1)^m [1 - \theta(n_1)]^{k-m}. \quad (2.5)$$

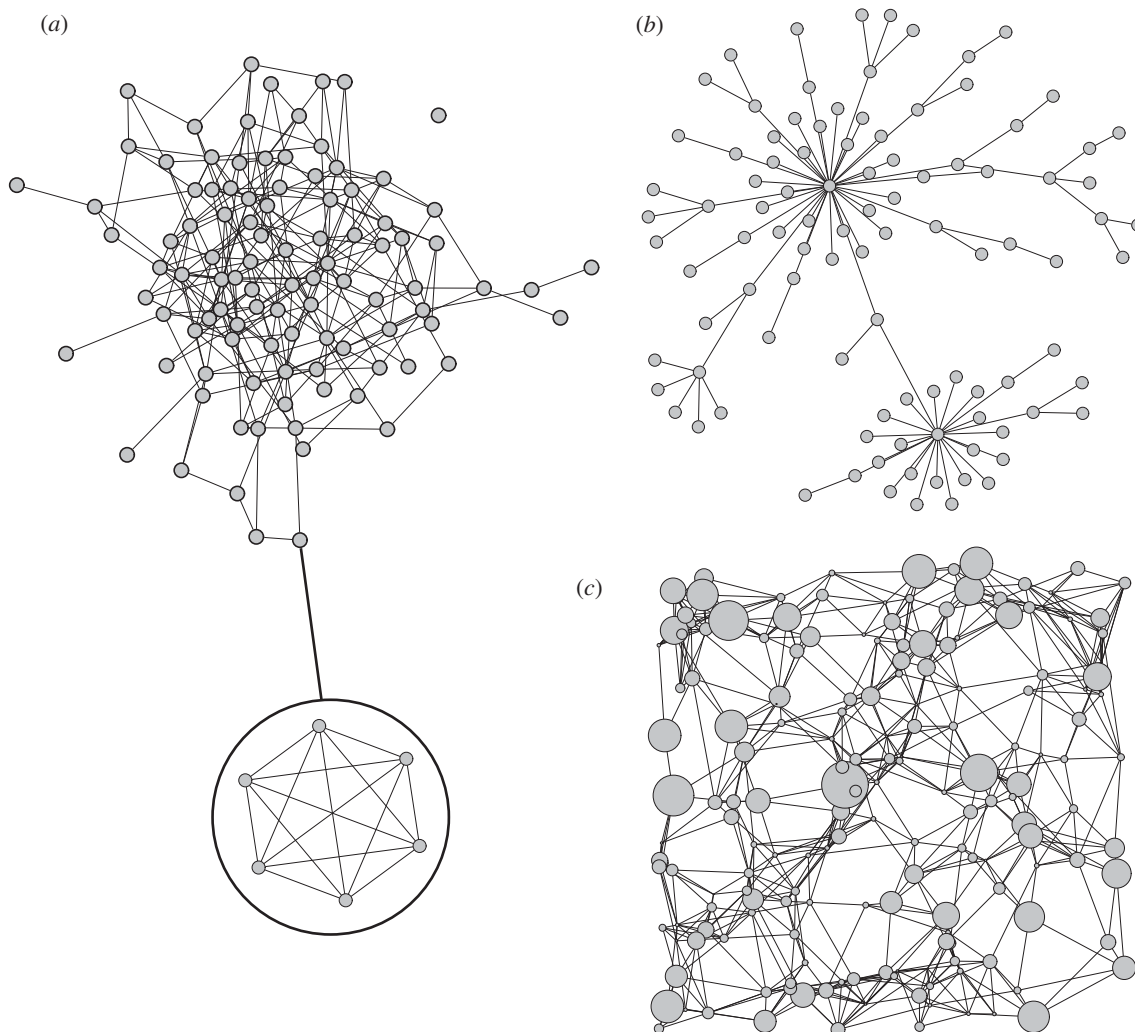
We will refer to these  $m$  edges as ‘infective edges’.

Hereafter, we will make use of probability-generating functions (PGFs), which have been used extensively to study epidemiological properties of networks [7,17,18,26]. PGFs are functions of the form

$$G(x) = \sum_{k=0}^{\infty} p_k x^k, \quad (2.6)$$

where  $p_k$  is the probability of a random variable taking discrete value  $k$ . They have interesting mathematical properties [10,17]. For example,  $G(1)$  is the sum of all  $p_k$ 's, and hence is equal to 1, and the first derivative of the PGF,

$$G'(x) = \sum_{k=0}^{\infty} k p_k x^{k-1}, \quad (2.7)$$



**Figure 1.** Hierarchical network models. (a) A 100-node group network with a Poisson degree distribution, where groups contain variable numbers of individuals and are completely intra-connected. (b) A 100-node group network with a power-law degree distribution. (c) A simulated Serengeti lion pride network with the statistical properties described in Craft *et al.* [16]. Node size indicates number of lions in the pride.

is equal to the mean of the distribution when  $x = 1$ , and can be combined with the second derivative to obtain the variance of the distribution:  $G''(1) + G'(1) - (G'(1))^2$ .

Suppose the first infected group in an outbreak is chosen randomly from an entirely susceptible population. The PGF of the number of other groups infected by that initial group is given by

$$G_0(x) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} S_0(n) D_0(k) \sum_{m=0}^k P(m, k, n) x^m. \quad (2.8)$$

Making use of the relation

$$\sum_{m=0}^k P(m, k, n) x^m = \sum_{m=0}^k \binom{k}{m} [x\theta(n)]^m [1 - \theta(n)]^{k-m} = [1 + (x-1)\theta(n)]^k, \quad (2.9)$$

we obtain

$$G_0(x) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} S_0(n) D_0(k) [1 + (x-1)\theta(n)]^k. \quad (2.10)$$

Let us now determine the PGF of the distribution of the number of infective edges leaving a group reached by following a random infective edge, denoted  $G_1(x)$ . Intuitively, groups that are highly connected and/or large are more

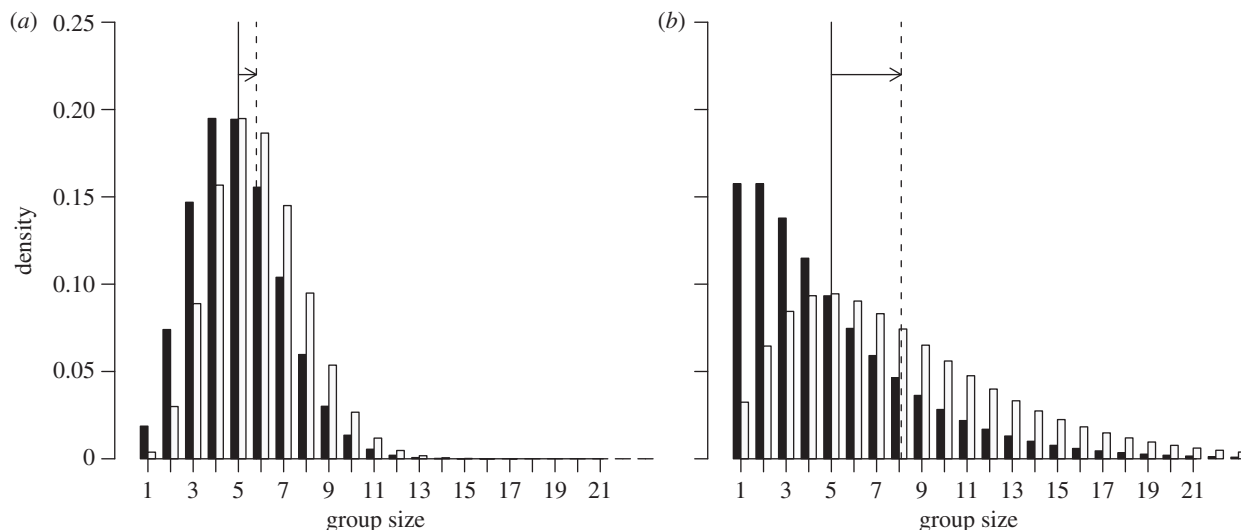
likely to become infected than their less connected and/or smaller counterparts. The probability of reaching a group of degree  $k$  by following a random infective edge is proportional to  $k \cdot D_0(k)$ . Similarly, by approximation (2.3), the probability of reaching a group of size  $n$  by following a random infective edge is proportional to  $n \cdot S_0(n)$ .  $G_1(x)$  is therefore proportional to

$$g_1(x) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} n S_0(n) k D_0(k) \sum_{m=0}^{k-1} P(m, k-1, n) x^m = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} n S_0(n) k D_0(k) [1 + (x-1)\theta(n)]^{k-1}. \quad (2.11)$$

$G_1(x)$  is obtained by normalizing  $g_1(x)$ , which makes the probabilities sum to one:

$$G_1(x) = \frac{g_1(x)}{g_1(1)}. \quad (2.12)$$

$G_1(x)$  generates the distribution of the number of infective edges leaving a group reached by following a random infective edge. During the course of an outbreak, an infected group will infect on average  $G'_1(1)$  susceptible groups. If  $G'_1(1) > 1$ , then an outbreak is expected to continue increasing in size, and may grow into an epidemic that is expected to scale with the size of the network (described in detail



**Figure 2.** Size distribution of infected groups at the beginning of an outbreak. Black bars: group size distribution for the whole population, here assumed to follow a negative binomial (as in equation (3.1)), with mean  $\langle n \rangle = 5$  and variance (a)  $\sigma_n^2 = 4$  or (b)  $\sigma_n^2 = 16$ . White bars: distribution of infected group sizes after a few transmission events (equation (2.15)). The arrows represent the differences between the means of the two distributions. The higher the variance  $\sigma_n^2$ , the higher the mean size of infected groups, and, consequently, the more easily disease spreads in the population.

below). The epidemic threshold  $T_c$ , beyond which epidemics are possible, can be determined by solving

$$\begin{aligned}
 G_1'(1) &= 1 \\
 \Leftrightarrow \frac{\sum_{k=0}^{\infty} \sum_{n=1}^{\infty} k(k-1) D_0(k) n S_0(n) \theta(n)}{\sum_{k=0}^{\infty} \sum_{n=1}^{\infty} k D_0(k) n S_0(n)} &= 1 \\
 \Leftrightarrow \frac{T_c \langle n \rangle \sum_{k=0}^{\infty} k(k-1) D_0(k) \sum_{n=1}^{\infty} n^2 S_0(n)}{\sum_{k=0}^{\infty} k D_0(k) \sum_{n=1}^{\infty} n S_0(n)} &= 1 \\
 \Leftrightarrow T_c \langle n^2 \rangle \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} &= 1 \\
 \Leftrightarrow T_c &= \left( \frac{1}{\langle n^2 \rangle} \right) \left( \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right). \quad (2.13)
 \end{aligned}$$

The second moment of the group size distribution,  $\langle n^2 \rangle$ , is a function of the mean and variance of the group size distribution:  $\langle n^2 \rangle = \langle n \rangle^2 + \sigma_n^2$ . In agreement with equation (2.13), numerous studies have already shown that larger mean group sizes ( $\langle n \rangle$ ) typically promote the spread of infectious disease outbreaks [27–32]. Here, we focus instead on the role of the variance of the group size distribution ( $\sigma_n^2$ ).

As the mean and variance of the group size distribution increase, the epidemic threshold decreases. In other words, the epidemiological vulnerability of the population increases. This stems from the assumption that between-group transmission is proportional to the sizes of interacting groups, making larger groups more vulnerable to infection. Figure 2 depicts the approximate probability distribution of the size of newly infected groups after a few transmission events, as given by

$$S_1(n) = \frac{n S_0(n)}{\sum_{u=1}^{\infty} u S_0(u)} = \frac{n S_0(n)}{\langle n \rangle}, \quad (2.14)$$

with mean

$$\langle n_{\text{inf}} \rangle = \frac{\langle n^2 \rangle}{\langle n \rangle} = \langle n \rangle + \frac{\sigma_n^2}{\langle n \rangle}. \quad (2.15)$$

This equation tells us that the expected size of infected groups early in an outbreak is higher than the average group size,

and that it increases with the variance in group size. Because the between-group transmission rate is proportional to the size of the infected group, then the early epidemiological vulnerability of a population also increases with variance in group size. Thus, variance in group size enhances transmission via the presence of large groups (albeit relatively rare) that have the unfortunate combination of high probability of catching disease from other groups and high probability of spreading disease to other groups. Note that  $S_1(n)$  does not give the probability distribution for groups infected throughout the entire epidemic, but applies to the early stages of transmission. The sizes of groups infected very early and later in the epidemic tend to be lower than those given by  $S_1(n)$ .

## 2.2. Mean and variance of outbreak sizes below the epidemic threshold

Functions  $G_0(x)$  and  $G_1(x)$  can be used to determine the distribution of the total number of nodes that become infected in an outbreak starting from a single infected group [33]. Newman *et al.* [34] and Newman [17] demonstrated that this distribution is generated by the PGF  $H_0(x)$  defined by the following system of equations:

$$\begin{aligned}
 H_0(x) &= x G_0(H_1(x)) \\
 \text{and} \quad H_1(x) &= x G_1(H_1(x)). \quad (2.16)
 \end{aligned}$$

The self-consistent form of the second equation makes closed-form expressions of  $H_0(x)$  and  $H_1(x)$  difficult to obtain. However, the mean outbreak size  $\langle s \rangle = H_0'(1)$  can be written as a simple function of the mean and mean-squared degree ( $\langle k \rangle$  and  $\langle k^2 \rangle$ ), mean and mean-squared group size ( $\langle n \rangle$  and  $\langle n^2 \rangle$ ), and transmission rate  $T$  (see the electronic supplementary material, S1.1):

$$\begin{aligned}
 \langle s \rangle &= 1 + \frac{G_0'(1)}{1 - G_1'(1)} \\
 &= 1 + \frac{T \langle n \rangle^2 \langle k \rangle}{1 - (T(\sigma_n^2 + \langle n \rangle^2)(\langle k^2 \rangle - \langle k \rangle)/\langle k \rangle)}. \quad (2.17)
 \end{aligned}$$



The second moment of the group size distribution,  $\langle n^2 \rangle$ , is equal to  $\sigma_n^2 + \langle n \rangle^2$ , where  $\sigma_n^2$  is the variance of the group size distribution. The derivatives of  $\langle s \rangle$  with respect to  $\sigma_n^2$  and  $\langle n \rangle$  are both positive real numbers. The mean outbreak size is hence an increasing function of both the mean and variance of group sizes. The variance of outbreak sizes,  $\sigma_s^2 = H_0''(1) + H_0'(1) - (H_0'(1))^2$ , can also be derived from equation (2.16):

$$\sigma_s^2 = \frac{1}{(G_1'(1) - 1)^3} [(G_1'(1) - 1)(G_0''(1) + G_0'(1)) \times (1 + G_1'(1) - G_0'(1)) - G_0'(1)G_1''(1)] \quad (2.18)$$

with

$$\begin{cases} G_0'(1) = T\langle n \rangle^2 \langle k \rangle \\ G_1'(1) = T\langle n^2 \rangle \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \\ G_0''(1) = (T\langle n \rangle)^2 \langle n^2 \rangle (\langle k^2 \rangle - \langle k \rangle) \\ G_1''(1) = T^2 \langle n \rangle \langle n^3 \rangle \frac{\langle k^3 \rangle - 3\langle k^2 \rangle + 2\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \end{cases}$$

From this expression, we find that the variance in outbreak size, which indicates the unpredictability of outbreaks, is also an increasing function of both the mean and variance of group sizes.

### 2.3. Epidemic size and probability of an epidemic

The expression for  $\langle s \rangle$  in equation (2.17) diverges when  $1 - G_1'(1) = 0$  and thus is valid only when  $G_1'(1) < 1$ , i.e. when  $T > T_c$  (equation (2.13)). Above this epidemic threshold, outbreaks can either have a small, finite size or become epidemic, that is, have a large expected size that scales with the size of the network. By definition,  $H_0(x)$  is the PGF of the size of finite outbreaks. Thus,  $H_0(1)$  is the expected fraction of the network that, if infected, would yield small, finite-sized outbreaks. Following Newman [17], the expected size of an epidemic is thus given by

$$S = 1 - G_0(H_1(1)), \quad (2.19)$$

where  $H_1(1)$ , the probability that the node at the end of a random edge does not get infected during an epidemic, can be obtained by solving

$$H_1(1) = G_1(H_1(1)). \quad (2.20)$$

Note that  $S$  is both the proportion of the network affected by the epidemic and the probability that an epidemic occurs (as opposed to a small outbreak), under the assumption that the first infected group is randomly determined [33].

In the epidemic regime, the mean size of small outbreaks that do not reach epidemic proportions is given by (see the electronic supplementary material, S1.2)

$$\langle s \rangle = \frac{H_0'(1)}{H_0(1)} = \frac{1}{1 - S} \left( 1 - S + \frac{G_0'(H_1(1))H_1(1)}{1 - G_1'(H_1(1))} \right). \quad (2.21)$$

Note that in the non-epidemic regime,  $S = 0$  and  $H_1(1) = 1$ , which makes equations (2.21) and (2.17) equivalent.

### 2.4. Epidemiological effective group size

To further quantify the epidemiological significance of group size variance, we compare populations with variable group sizes to hypothetical populations with constant group sizes. The equations above tell us that all else being equal (same transmission rate, same degree distribution, same mean group size), a population with heterogeneous group sizes will be more vulnerable to epidemics (i.e. have a lower epidemic threshold) than a population with homogeneous group sizes.

As the concept of epidemic threshold is specific to the field of epidemiology, the comparison of two epidemic thresholds may only be meaningful for a specialized audience. Instead, let us introduce a more meaningful quantity. For a given network with variable-sized groups, we ask which homogeneous network (with identical-sized groups) has the exact same epidemic threshold, when the mean group size is allowed to change, but not the degree distribution. In other words, if we reassigned all individuals in the original heterogeneous network to equal-sized groups (leaving the number of groups and degree distribution unchanged), then the population would become less vulnerable to epidemics. By how much would we then have to increase the size of those groups to reduce the epidemic threshold back to its original value? This unit for this quantity is a 'number of individuals per group', which makes it easily understandable by non-epidemiologists.

Because all groups in the hypothetical equivalent population have the same size,  $n_e$ , then the second moment of the group size distribution is simply  $\langle n_e^2 \rangle = n_e^2$ . Using equation (2.13), we see that the group size in this hypothetical equivalent population,  $n_e$ , must satisfy

$$\left( \frac{1}{n_e^2} \right) \left( \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right) = \left( \frac{1}{\langle n^2 \rangle} \right) \left( \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right). \quad (2.22)$$

Because the degree distributions are identical, the mean size of the groups in the homogeneous network is given by

$$n_e = \sqrt{\langle n \rangle^2 + \sigma_n^2}. \quad (2.23)$$

This quantity is conceptually analogous to the 'effective population size' used in population genetics [35] and defined as the size of the ideal population that would undergo the same amount of random genetic drift as the actual population [36]. Hence, we call this quantity the epidemiological effective group size (EEGS). Notably, it depends only on the original group size distribution, and not on either the degree distribution or the transmission rate of the disease.

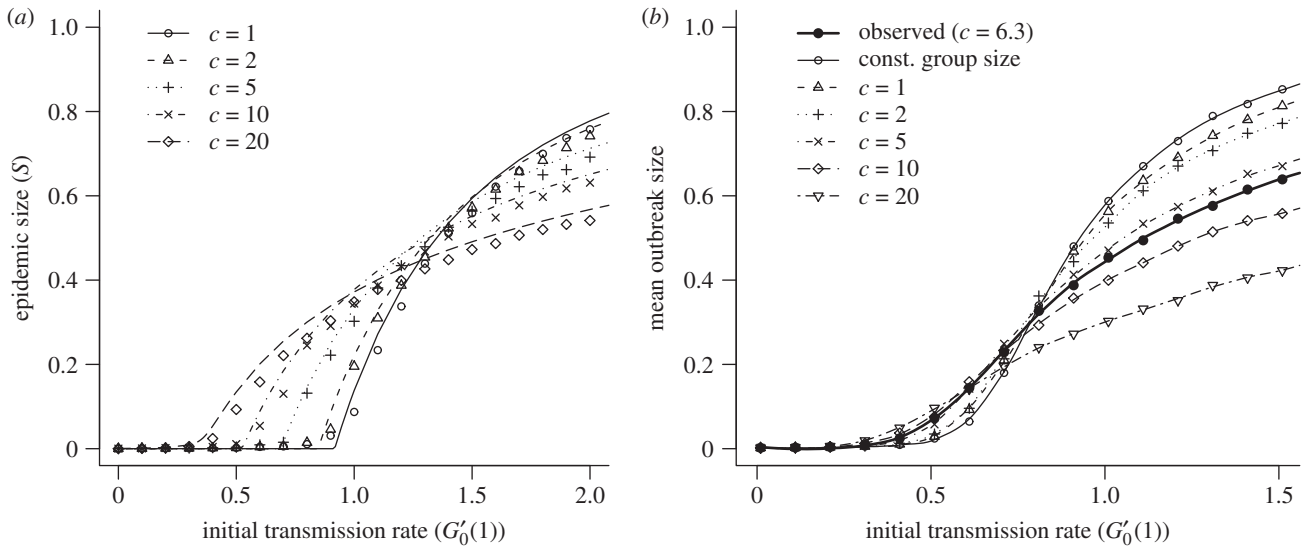
## 3. Examples

We provide two examples of application of this model. The first example is theoretical. We present exact solutions for infinite networks where group sizes follow a negative-binomial distribution and edges are distributed randomly between nodes. Erdős & Rényi [37] have shown that such networks have Poisson degree distributions (figure 1a). We then extend these results to networks with power-law (figure 1b) and exponential degree distributions. As a second example, we use simulations to investigate the impacts of group size distribution for an actual biological system, pride-living Serengeti lions [14,16].

### 3.1. Negative binomial group size distribution

Consider a network composed of an infinite number of social groups, where group sizes follow a discrete distribution  $S_0(n)$  in which mean and variance can be tuned independently. As an example, we will investigate the negative binomial distribution shifted one unit to the right to avoid groups of size zero:

$$\begin{cases} S_0(n) = 0, & \text{if } n \leq 0 \\ S_0(n) = \binom{n+x-2}{x-1} (1-p)^{n-1} p^x, & \text{if } n > 0. \end{cases} \quad (3.1)$$



**Figure 3.** (a) Epidemic size (or probability of an epidemic,  $S$ ) in relation to the disease contagiousness and the group size coefficient of dispersion  $c$  (assuming negative binomial group size distribution with  $\langle n \rangle = 10$ ), as predicted by the theoretical model. The contagiousness is expressed as the initial transmission rate, i.e. the expected number of secondary cases resulting from the primary infection of a randomly chosen group:  $G'_0(1) = \langle n \rangle^2 \langle k \rangle T$ . The lines display values calculated numerically using equations (2.19) and (2.20) for networks with Poisson degree distributions ( $\langle k \rangle = 30$ ). Each dot corresponds to the average of 100 stochastic simulations on networks including 10 000 vertices. The slight discrepancy between theoretical and simulated values is due to approximation (1.3). (b) Mean outbreak size in relation to the disease contagiousness, simulated in Serengeti lion networks with  $\langle n \rangle = 10.3$ . Each dot corresponds to the average of 100 simulations. Symbols correspond to different values of the group size coefficient of dispersion  $c$ , and also to the case of a constant group size.

This probability distribution has mean given by

$$\langle n \rangle = \sum_n n S_0(n) = 1 + x \left( \frac{1}{p} - 1 \right). \quad (3.2)$$

To help with interpretation, we will henceforth replace parameter  $x$  with  $p(\langle n \rangle - 1)/(1 - p)$  and parameter  $p$  with  $1/c$ . This gives the variance of the distribution a simpler expression:

$$\sigma_n^2 = \sum_n S_0(n) (n - \langle n \rangle)^2 = c(\langle n \rangle - 1). \quad (3.3)$$

Parameter  $c \in (1, +\infty)$  is thus equal to  $\sigma_n^2/(\langle n \rangle - 1)$  and measures the overdispersion of the distribution. When  $c$  approaches 1,  $S_0(n)$  tends towards

$$\lim_{c \rightarrow 1} S_0(n) = (\langle n \rangle - 1)^{n-1} \frac{e^{-(\langle n \rangle - 1)}}{(n - 1)!}, \quad (3.4)$$

which is a Poisson distribution shifted one unit to the right. Assuming the network as a Poisson degree distribution implies that  $\langle k^2 \rangle = 2\langle k \rangle^2$ , which gives

$$\frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} = \langle k \rangle^{-1}. \quad (3.5)$$

The epidemic threshold can be derived by solving equation (2.13):

$$\begin{aligned} T_c &= \left( \frac{1}{\langle n^2 \rangle} \right) \left( \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right) \\ &= [\langle k \rangle (\langle n^2 \rangle + \sigma_n^2)]^{-1} \\ &= [\langle k \rangle (\langle n^2 \rangle + c(\langle n \rangle - 1))]^{-1}. \end{aligned} \quad (3.6)$$

$T_c$  is a decreasing function of  $\langle k \rangle$ ,  $\langle n \rangle$  and  $c$ . The epidemic threshold is more easily reached when the mean degree, the mean group size and the coefficient of dispersion of group sizes are high (see figure 3a and electronic supplementary material, figure S2.1a).

The EEGS (equation (2.23)) for such degree distribution is given by

$$n_e = \sqrt{\langle n \rangle^2 + c(\langle n \rangle - 1)}, \quad (3.7)$$

and the difference between  $n_e$  and the actual mean group size  $\langle n \rangle$ ,  $\delta = n_e - \langle n \rangle$ , is an increasing function of  $\langle n \rangle$ , which asymptotes to the constant value  $c/2$ . Therefore, when the mean group size is large,  $n_e$  varies between  $n + 1/2$  (in the case of a Poisson group size distribution) and  $+\infty$  (in the case of highly overdispersed group size distribution (i.e.  $c \rightarrow +\infty$ )).

In the case of more dispersed degree distributions, the results are slightly different. Following several authors [17,33, 34,38,39], we investigated the following degree distribution:

$$D_0(k) = \begin{cases} 0, & \text{for } k \leq 0 \\ Ck^{-\alpha}e^{-k/\kappa}, & \text{for } k > 0 \end{cases} \quad (3.8)$$

which is a power-law distribution with an exponential cut-off. Constant  $C$ , ensuring  $\sum_k D_0(k) = 1$ , is equal to  $\text{Li}_\alpha(e^{-1/\kappa})$ , where  $\text{Li}_\alpha$  is the polylogarithm function, defined as

$$\text{Li}_\alpha(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^\alpha}. \quad (3.9)$$

When  $\alpha = 0$ ,  $D_0(k)$  becomes an exponential with scale parameter  $\kappa$ ; and when  $\kappa \rightarrow 0$ , it approaches a pure power-law distribution of exponent  $\alpha$ . According to equation (2.13), the epidemic threshold is proportional to

$$\frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} = \frac{\text{Li}_{\alpha-1}(e^{-1/\kappa})}{\text{Li}_{\alpha-2}(e^{-1/\kappa}) - \text{Li}_{\alpha-1}(e^{-1/\kappa})}, \quad (3.10)$$

which is an increasing function of  $\alpha$  and a decreasing function of  $\kappa$  (see the electronic supplementary material, figure S2.1b). This result is identical to that obtained with networks composed of simple nodes [17].

Quantity  $S$  (the epidemic size or, equivalently, the probability of an epidemic) can be computed numerically for different values of the coefficient of dispersion of the group size distribution using equations (2.19) and (2.20). The effect of  $c$  depends on the contagiousness of the disease, as shown for a set of networks with Poisson degree distributions (figure 3a). When the contagiousness is low, variation in group size increases the mean epidemic size. Just as group size variability increases epidemiological vulnerability early in an outbreak via enhanced infection to and transmission from large groups, it also leads to increases in sizes of infected groups and resulting probabilities of transmission throughout epidemics of mildly transmissible diseases. However, when the disease is highly contagious, increasing the variance of the group size distribution leads to smaller epidemics. The crossing of the lines in figure 3 may seem surprising, as it shows that networks with the lowest epidemic thresholds (highest vulnerability) and large expected epidemic sizes for mildly contagious diseases are expected to experience the smallest epidemics for highly contagious diseases, and vice versa. For networks with heterogeneous group sizes, the flip side of having some large groups that increase the epidemiological vulnerability of the population as a whole is also having small groups that are unlikely to become infected. Recall that the transmission probability,  $\theta(n_1, n_2)$ , is proportional to the product of the sizes of the two connected groups. Small groups thus are relatively protected from infection, and may result in stochastic termination of transmission chains.

Interestingly, the quantity  $H_1(1)$  computed using equation (2.20) can also be used to derive the probability of a group of size  $n$  and degree  $k$  remaining uninfected during an epidemic:

$$P(\overline{\text{inf}}|n, k) = [1 - \theta(n) + \theta(n) \cdot H_1(1)]^k, \quad (3.11)$$

where  $1 - \theta(n)$  is the probability of an edge remaining uninfected, and  $\theta(n) \cdot H_1(1)$  is the probability of an infective edge that is not connected to the part of the network affected by the epidemic. Using Bayes theorem, we can then calculate the group size and degree distributions of uninfected and infected groups during an epidemic:

$$\left. \begin{aligned} P(n, k|\overline{\text{inf}}) &= \frac{P(\overline{\text{inf}}|n, k)S_0(n)D_0(k)}{(1 - S)} \\ \text{and } P(n, k|\text{inf}) &= \frac{[1 - P(\overline{\text{inf}}|n, k)]S_0(n)D_0(k)}{S} \end{aligned} \right\} \quad (3.12)$$

As expected, groups that become infected during an epidemic have on average a larger size and a larger degree than groups that remain uninfected (see the electronic supplementary material, figure S2.2).

### 3.2. Disease spread in Serengeti lions

Craft *et al.* [14,16] have described the structure and properties of the contact patterns among approximately 180 lion prides in Serengeti National Park, Tanzania, using detailed demographic and behavioural data collected since 1965 on a subset of the population [19]. They developed an algorithm to simulate stochastic networks with statistical properties similar to the real pride network [16]. They also showed that pride size and degree did not correlate, as assumed by our model [16]. We used this algorithm to generate hypothetical pride networks (figure 1c).

The distribution of pride sizes in these networks was assumed to be the shifted negative binomial distribution defined in equation (3.1). Maximum-likelihood estimates of parameters  $\langle n \rangle$  and  $c$  were calculated from composition data on 21 prides recorded in 1992. We estimated a mean pride size of 10.3 individuals (females and cubs older than three months), and a coefficient of dispersion of 6.3. A likelihood-ratio test indicated that the negative binomial fitted the data significantly better than a Poisson distribution (see the electronic supplementary material, figure S2.3). This test was highly significant ( $\chi^2 = 26.2$ , d.f. = 1,  $p = 3 \times 10^{-7}$ ).

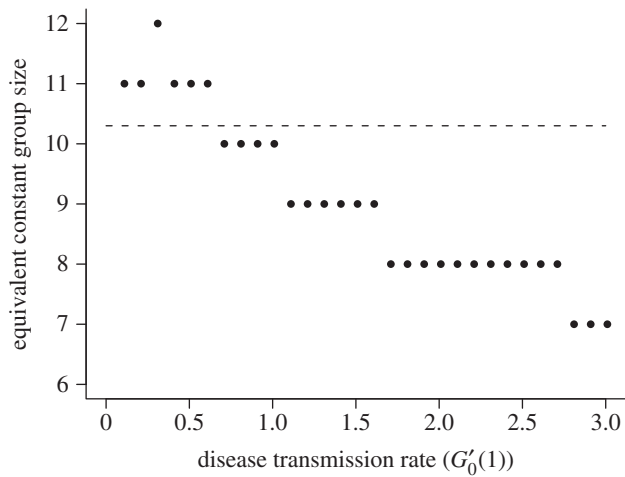
We performed chain-binomial simulations of disease outbreaks through the hypothetical Serengeti lion populations, keeping mean pride size  $\langle n \rangle$  constant and varying the value of  $c$  between 1 and 20 [40]. The lion pride network is relatively small, making it difficult to distinguish between small, non-epidemic outbreaks and full epidemics. Thus, rather than estimating epidemic thresholds and epidemic sizes from the simulations, we instead computed final outbreak sizes, defined as the proportion of prides that became infected. The results, displayed in figure 3b, are qualitatively comparable to the analytical results obtained with a Poisson network of infinite size (figure 3a). When the contagiousness is low (i.e. when outbreaks affect on average less than approx. 20–30% of the prides), group size variability promotes larger outbreaks. When the contagiousness increases, however, the trend reverses and increased group size variability leads to smaller outbreaks. Thus, the epidemiological impact of group size variability can be positive or negative, depending on the infectious properties of the spreading disease.

Because the Serengeti lion population forms a small, finite-sized network, we cannot estimate its epidemic threshold and the EEGS, as defined above. However, an analogous quantity can be calculated using simulations. For 30 different transmissibility values  $T$  (corresponding to 30 different initial transmission rates  $G'_0(1)$  ranging from zero to three), we simulated disease outbreaks in hypothetical networks with homogeneous pride sizes,  $n$ , ranging from five to 20 individuals per pride. For each of the 480 combinations of  $T$  and  $n$ , we simulated 100 outbreaks and calculated the mean outbreak size. For each  $G'_0(1)$ , we then determined the homogeneous pride size that yielded the closest mean outbreak size to the empirical population (figure 4). For diseases expected to have total attack rates of less than 23 per cent of all prides, the observed lion population was found epidemiologically equivalent to populations in which all prides include 11 or 12 individuals (figure 4), that is, populations with 10–20% more individuals than the empirical population.

## 4. Discussion

Our analyses demonstrate the epidemiological importance of variation in group size in social species. As the variance of the group size distribution increases, the model demonstrates that the epidemic threshold decreases and that both the mean and variance of small outbreak sizes should increase. Above the epidemic threshold, that is, for diseases capable of causing large epidemics, the effect of group size variability depends on the transmissibility of the disease. For mildly contagious diseases, more variable group sizes promote larger epidemics, whereas for more highly contagious diseases, the effect reverses and group size variability inhibits epidemics.





**Figure 4.** Effect of the variance of Serengeti lion pride sizes on epidemic sizes. The y-axis shows the group size of a hypothetical homogeneous population with constant group sizes in which epidemics have the same average size as the observed lion population, in relation to disease transmission rate in the observed population.  $G'_0(1) = \langle n \rangle^2 \langle k \rangle T$  is the expected number of secondary cases resulting from the primary infection of randomly chosen group. The dotted line marks the observed mean pride size ( $\langle n \rangle = 10.3$ ). Variability in group sizes increases the mean outbreak size only when  $G'_0 < 0.7$ , which corresponds to an expected outbreak size less than 23% of the population. For diseases above this threshold, variability in group sizes decreases the expected outbreak size.

These findings have important implications for disease-control strategies, including vaccination, quarantine and culling, which often target social groups that are likely to decrease the epidemic threshold. Prior studies have highlighted the importance of targeting groups with high numbers of interacting neighbours or who occupy a central location in the network [41]. Our study suggests that factoring the group size distribution into epidemiological assessments can improve both understanding of disease dynamics and efforts to prevent and mitigate outbreaks. For example, vaccination efforts should perhaps be targeted to reduce the variation in the number of susceptible individuals per group (perhaps by vaccinating a fraction of individuals in large groups) rather than simply reducing the total number of susceptible individuals population-wide or immunizing all individuals in a few groups. For conservation biologists, this work also suggests that assessments of disease risk in endangered animal populations should carefully consider variation in both contact patterns and group sizes.

The effect of the hierarchical structure of a population on the transmission of infectious diseases can similarly apply to the transmission of information. In human populations, the transmission of rumours and cultural traits between families or communities is thus probably influenced by the variance in social group size. One can also predict that the transmission of computer viruses between small networks of interconnected computers (e.g. intranets) is more likely to become epidemic when the size of subnetworks is more variable.

We have introduced EEGS as a simple, intuitive metric for quantifying the epidemiological impact of group size variation and the epidemiological vulnerability of a population. As long as the network structure of the population is sufficiently random (lacks clustering or other local structure),

EEGS can be estimated using only the mean and variance of the group size distribution. EEGS is independent of degree distribution and indicates which homogeneous network with same-sized groups has comparable epidemiological properties. The original network and the EEGS homogeneous network share the same epidemic threshold and thus are vulnerable to the same suite of pathogens. However, the two networks will not necessarily experience outbreaks of similar magnitude. Analogous metrics can also be computed by equating the mean outbreak size (as illustrated here with the Serengeti lion network) or the probability of an epidemic instead of the epidemic threshold. However, such metrics can generally only be computed numerically or using Monte Carlo simulations.

The model makes several simplifying assumptions. First, it assumes that between-group transmission is an increasing function of the size of the groups. This happens, for example, when all the individuals of two interacting groups come into contact, or when movements of potentially infected individuals between connected social groups are proportional to the size of the groups (as in the gravity model [42,43]). This may also be the case when disease vectors such as mosquitoes are more likely to detect large social groups than small social groups (e.g. malaria in primates [44,45]). It also assumes that when disease reaches a group, all individuals in the group become infected. This may not hold when the within-group disease transmission rate or the within-group connectivity is low. The model can be modified to handle such scenarios by replacing the group size distribution with the distribution of within-group outbreak sizes. This distribution can, for example, be inferred using stochastic SIR compartmental models for highly intra-connected groups [46] or network-based models for sparser groups. Within-group network structures may be sensitive to stochastic effects and yield highly variable within-group outbreak sizes, which will, in turn, affect the epidemiological vulnerability of the larger population. In these cases, the dynamics of infectious diseases will depend on the two-level within- and between-group contact network.

Second, the model assumes that group sizes are randomly distributed across the landscape. In reality, group sizes may depend on the quality of locally available resources, with large groups occurring around resource-rich areas and small groups found in more marginal areas [47]. If group sizes are spatially clustered, then the epidemiological impact of the group size variability may be different, depending on the size and spatial distribution of these clusters.

Finally, we assume that group size and degree are not correlated. If there is a positive correlation between the two, that is, larger groups have more inter-group contacts, then we expect the effects of group size variability to be amplified. The epidemic threshold would then be lower and epidemic size for high transmission rates would be smaller.

In conclusion, group size variability strongly impacts disease transmission in hierarchical populations. This pertains not only to group-living wild animals such as the Serengeti lions, but more generally to structured populations, including patch-structured wild plants, cultivated crops, herd-structured domestic livestock and community-structured humans.

We thank Rosalind Eggo and Alexandre Courtiol for useful discussions and suggestions. We also thank three anonymous reviewers for their useful comments. This work was supported by NSF grant (no. OISE-0804186) to M.E.C. and NSF grant (no. DEB-0749097) to L.A.M.

## References

- Kummer H. 1968 *Social organization of Hamadryas baboons*. Chicago, IL: University of Chicago Press.
- Dunbar R. 1986 The social ecology of gelada baboons. In *Ecological aspects of social evolution: birds and mammals* (eds RW Wrangham, DI Rubenstein), pp. 332–351. Princeton, NJ: Princeton University Press.
- Stammach E. 1987 Desert, forest and montane baboons: multilevel-societies. In *Primate societies* (eds B Smuts, D Cheney, R Seyfarth, R. Wrangham, T Struhsaker), pp. 112–120. Chicago, IL: University of Chicago Press.
- Rubenstein D, Hack M. 2004 Natural and sexual selection and the evolution of multi-level societies: insights from zebras with comparisons to primates. In *Sexual selection in primates: new and comparative perspectives* (eds P Kappeler, C van Schaik), pp. 266–279. New York, NY: Cambridge University Press.
- Krützen M, Mann J, Heithaus M, Connor R, Bejder L, Sherwin W. 2005 Cultural transmission of tool use in bottlenose dolphins. *Proc. Natl Acad. Sci. USA* **102**, 8939–8943. (doi:10.1073/pnas.0500232102)
- Lloyd-Smith J, Schreiber S, Kopp P, Getz W. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
- Meyers L, Newman M, Pourbohloul B. 2006 Predicting epidemics on directed contact networks. *J. Theor. Biol.* **240**, 400–418. (doi:10.1016/j.jtbi.2005.10.004)
- Craft M, Caillaud D. 2011 Network models: an underutilized tool in wildlife epidemiology? *Interdiscip. Perspect. Infect. Dis.* **2011**, 676949. (doi:10.1155/2011/676949)
- Tompkins D, Dunn A, Smith M, Telfer S. 2011 Wildlife diseases: from individuals to ecosystems. *J. Anim. Ecol.* **80**, 19–38. (doi:10.1111/j.1365-2656.2010.01742.x)
- Newman M. 2010 *Networks: an introduction*. Oxford, UK: Oxford University Press.
- Keeling M, Rohani P. 2008 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press.
- Woolhouse M, Shaw D, Matthews L, Liu W, Mellor D, Thomas M. 2005 Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* **1**, 350–352. (doi:10.1098/rsbl.2005.0331)
- Davis S, Trapman P, Leirs H, Begon M, Heesterbeek J. 2008 The abundance threshold for plague as a critical percolation phenomenon. *Nature* **454**, 634–637. (doi:10.1038/nature07053)
- Craft M, Volz E, Packer C, Meyers L. 2009 Distinguishing epidemic waves from disease spillover in a wildlife population. *Proc. R. Soc. B* **276**, 1777–1785. (doi:10.1098/rspb.2008.1636)
- Salkeld D, Salathé M, Stapp P, Jones J. 2010 Plague outbreaks in prairie dog populations explained by percolation thresholds of alternate host abundance. *Proc. Natl Acad. Sci. USA* **107**, 14 247–14 250. (doi:10.1073/pnas.1002826107)
- Craft M, Volz E, Packer C, Meyers L. 2011 Disease transmission in territorial populations: the small-world network of Serengeti lions. *J. R. Soc. Interface* **8**, 776–786. (doi:10.1098/rsif.2010.0511)
- Newman M. 2002 Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128. (doi:10.1103/PhysRevE.66.016128)
- Meyers L, Pourbohloul B, Newman M, Skowronski D, Brunham R. 2005 Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* **232**, 71–81. (doi:10.1016/j.jtbi.2004.07.026)
- Packer C, Hilborn R, Mosser A, Kissui B, Borner M, Hopcraft G, Wilmshurst J, Mduma S, Sinclair A. 2005 Ecological change, group territoriality, and population dynamics in Serengeti lions. *Science* **307**, 390–393. (doi:10.1126/science.1105122)
- Craft M. 2010 Ecology of infectious diseases in Serengeti lions. In *Biology and conservation of wild felids* (eds D Macdonald, A Loveridge), pp. 263–281. Oxford, UK: Oxford University Press.
- Packer C, Altizer S, Appel M, Brown E, Martenson J, O'Brien S, Roelke-Parker M, Hofmann-Lehmann R, Lutz H. 1999 Viruses of the Serengeti: patterns of infection and mortality in African lions. *J. Anim. Ecol.* **68**, 1161–1178. (doi:10.1046/j.1365-2656.1999.00360.x)
- Roelke-Parker M *et al.* 1996 A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* **379**, 441–445. (doi:10.1038/379441a0)
- Cleaveland S, Mlengeya T, Kaare M, Haydon D, Lembo T, Laurenson M, Packer C. 2007 The conservation relevance of epidemiological research into carnivore viral diseases in the Serengeti. *Conserv. Biol.* **21**, 612–622. (doi:10.1111/j.1523-1739.2007.00701.x)
- Munson L, Terio K, Kock R, Mlengeya T, Roelke M, Dubovi E, Summers B, Sinclair A, Packer C. 2008 Climate extremes promote fatal co-infections during canine distemper epidemics in African lions. *PLoS ONE* **3**, e2545. (doi:10.1371/journal.pone.0002545)
- Troyer J *et al.* 2011 Fiv diversity: Fivple subtype composition may influence disease outcome in African lions. *Vet. Immunol. Immunopathol.* **143**, 338–346. (doi:10.1016/j.vetimm.2011.06.013)
- Volz E, Meyers L. 2007 Susceptible–infected–recovered epidemics in dynamic contact networks. *Proc. R. Soc. B* **274**, 2925–2934. (doi:10.1098/rspb.2007.1159)
- Altizer S *et al.* 2003 Social organization and parasite risk in mammals: integrating theory and empirical studies. *Annu. Rev. Ecol. Syst.* **34**, 517–547. (doi:10.1146/annurev.ecolsys.34.030102.151725)
- Caillaud D, Levvero F, Cristescu R, Gatti S, Dewas M, Douadi M, Gautier-Hion A, Raymond M, Menard N. 2006 Gorilla susceptibility to ebola virus: the cost of sociality. *Curr. Biol.* **16**, R489–R491. (doi:10.1016/j.cub.2006.06.017)
- Côté I, Poulin R. 1995 Parasitism and group size in social animals: a meta-analysis. *Behav. Ecol.* **6**, 159–165. (doi:10.1093/beheco/6.2.159)
- Møller A, Merino S, Brown C, Robertson R. 2001 Immune defense and host sociality: a comparative study of swallows and martins. *Am. Nat.* **158**, 136–145. (doi:10.1086/321308)
- Nunn C, Altizer S. 2006 *Infectious diseases in primates: behavior, ecology and evolution*. New York, NY: Oxford University Press.
- Woodroffe R *et al.* 2009 Social group size affects *Mycobacterium bovis* infection in European badgers (*Meles meles*). *J. Anim. Ecol.* **78**, 818–827. (doi:10.1111/j.1365-2656.2009.01545.x)
- Newman M, Watts D, Strogatz S. 2002 Random graph models of social networks. *Proc. Natl Acad. Sci. USA* **99**, 2566–2572. (doi:10.1073/pnas.012582999)
- Newman M, Strogatz S, Watts D. 2001 Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118. (doi:10.1103/PhysRevE.64.026118)
- Wright S. 1938 Size of population and breeding structure in relation to evolution. *Science* **87**, 430–431.
- Lande R, Barrowclough GF. 1987 Effective population size, genetic variation, and their use in population management. In *Viable populations for conservation* (ed. ME Soulé), pp. 87–124. Cambridge, UK: Cambridge University Press.
- Erdős P, Rényi A. 1959 On random graphs, I. *Publicationes Mathematicae (Debrecen)* **6**, 290–297.
- Amaral L, Scala A, Barthélémy M, Stanley H. 2000 Classes of small-world networks. *Proc. Natl Acad. Sci. USA* **97**, 11 149–11 152. (doi:10.1073/pnas.200327197)
- Callaway D, Newman M, Strogatz S, Watts D. 2000 Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471. (doi:10.1103/PhysRevLett.85.5468)
- Bailey N. 1957 *The mathematical theory of epidemics*. London, UK: Griffin.
- Bansal S, Pourbohloul B, Meyers L. 2006 A comparative analysis of influenza vaccination programs. *PLoS Med.* **3**, e387. (doi:10.1371/journal.pmed.0030387)
- Erlander S, Stewart N. 1990 *The gravity model in transportation analysis: theory and extensions*. Utrecht, The Netherlands: VSP.
- Xia Y, Bjørnstad O, Grenfell B. 1991 Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164**, 267–281. (doi:10.1086/422341)
- Davies C, Ayres J, Dye C, Deane L. 1991 Malaria infection rate of Amazonian primates increases with body weight and group size. *Funct. Ecol.* **5**, 655–662. (doi:10.2307/2389485)
- Nunn C, Heymann E. 2005 Malaria infection and host behavior: a comparative study of neotropical primates. *Behav. Ecol. Sociobiol.* **59**, 30–37. (doi:10.1007/s00265-005-0005-z)
- Daley D, Gani J. 2001 *Epidemic modelling: an introduction*. Cambridge, UK: Cambridge University Press.
- Mosser A, Fryxell J, Eberly L, Packer C. 2009 Serengeti real estate: density versus fitness-based indicators of lion habitat quality. *Ecol. Lett.* **12**, 1050–1060. (doi:10.1111/j.1461-0248.2009.01359.x)