Community Structure and the Spread of Infectious Disease in Primate Social Networks

Citation

Published Version

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:8715733

Terms of Use

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Community structure and the spread of infectious disease in primate social networks

Randi H. Griffin[1] • **Charles L. Nunn[1,*]**

1. Department of Human Evolutionary Biology, Harvard University, 11 Divinity Avenue,

Cambridge, Massachusetts 02138

[*]Corresponding author, e-mail: cnunn@oeb.harvard.edu

Running title: Social Networks and Infectious Disease

Total word count: 8233

Abstract word count: 300

Introduction word count: 1307

Materials and Methods word count: 4061

Results word count: 832

Discussion word count: 1733

Number of references cited: 106

Number of figures: 3

Number of tables: 5

Electronic supplementary material: comparative data and sources, additional analyses.

Abstract Living in a large social group is thought to increase disease risk in wild animal populations, but comparative studies have provided mixed support for this prediction. Here, we take a social network perspective to investigate whether patterns of social contact within groups influence parasite risk. Specifically, increased modularity (i.e. sub-grouping) in larger groups could offset the increased disease risk associated with living in a large group. We simulated the spread of a contagious pathogen in random social networks to generate theoretically grounded predictions concerning the relationship between social network connectivity and the success of socially transmitted pathogens. Simulations yielded the prediction that community modularity (Q) negatively impacts parasite success. No clear predictions emerged for a second network metric we considered, the eigenvector centralization index (C), as the relationship between this measure and parasite success depended on the transmission probability of parasites. We then tested the prediction that Q reduces parasite success in a phylogenetic comparative analysis of social network modularity and parasite richness across 19 primate species. Using a Bayesian implementation of phylogenetic generalized least squares and controlling for sampling effort, we found that primates living in larger groups exhibited higher Q, and as predicted by our simulations, higher Q was associated with lower parasite richness for socially transmitted parasites. This suggests that increased modularity mediates the elevated risk of parasitism associated with living in larger groups, which could contribute to the inconsistent findings of empirical studies on the association between group size and parasite risk. Our results indicate that social networks may play a role in mediating pressure from socially transmitted parasites, particularly in large groups where opportunities for transmitting communicable diseases are abundant. We propose that parasite pressure in gregarious primates may have favored the evolution of behaviors that increase social network modularity, especially in large social groups.

Keywords: social networks, primates, infectious disease, parasite richness, sociality, comparative study, agent-based model.

Introduction

A fundamental goal of disease ecology is to identify host traits that influence parasitism in natural populations (Poulin 1995; Poulin and Morand 2004; Nunn and Altizer 2006). Because social interactions involving close proximity or contact provide opportunities for the transmission of many parasites, highly social hosts are expected to have higher parasite prevalence, abundance, and diversity than less social hosts (Møller et al., 1993; Altizer et al., 2003). While a number of other factors influence parasitism, including body mass, latitude and life history traits (Poulin and Morand 2004; Nunn and Altizer 2006), variation in social contact is expected to be a principle driver of variation in parasitism.

The majority of studies investigating the effects of sociality on parasitism have focused on group size as the measure of sociality, with the expectation that larger groups provide increased opportunities for parasites to spread (Møller et al., 1993). For example, Nunn et al. (2008, 2011) built a metapopulation model and found that infections spread more readily in populations composed of larger groups. Many empirical studies have found a positive association between group size and parasite risk in animals (e.g., Hoogland 1979; Wilkinson 1985; Shields and Crook 1987), although others have failed to find significant support for a positive association between these variables (e.g., Arnold and Lichenstein, 1993; Ezenwa et al. 2006; Snaith et al. 2008). In a meta-analysis of studies spanning insects, birds, and mammals, Côté and Poulin (1995) found an overall positive relationship between group size and the number of contagious parasites.

Primates are among the best-studied mammals in terms of parasites, and multiple lines of evidence suggest that infectious diseases play an influential role in primate behavior, ecology and evolution (reviewed in Nunn and Altizer 2006; Huffman and Chapman 2009). Based on comparative studies of parasite richness, prevalence, and immune system parameters, parasite risk in primates increases with increasing body mass, host population density, geographic range size, proximity to the equator, host diversification, and mating promiscuity (Nunn et al. 2000, 2003, 2004, 2005). Field studies have uncovered links between dominance rank and parasitism

within primate groups (e.g., Hausfater and Watson 1976; Hernandez et al. 2009), and habitat variation may influence parasitism among groups (e.g., Stoner 1996). Primates exhibit great variation in sociality, with group size ranging from solitary foragers, as in many nocturnal lemurs and lorises, to hundreds of individuals, as in the multi-level societies of hamadryas baboons (*Papio hamadryas*). Thus, primates present a valuable opportunity to investigate the links between sociality and parasitism (Altizer et al. 2003; Nunn and Altizer 2006).

Despite remarkable variation in sociality, studies of the relationship between group size and parasitism in primates have yielded mixed results. In the field, positive associations were found between group size and the number of intestinal protozoan species in mangabeys (*Cercocebus albigena*, Freeland 1979) and the number of nematode infections in baboons (McGrew et al. 1989). However, many comparative studies of primates have failed to detect significant associations between measures of parasitism and group size, including studies of parasite species richness (Vitone et al. 2004), white blood cell counts (Nunn et al. 2000; Nunn 2002a; Semple et al. 2002) and relative spleen size (Nunn 2002b). One comparative study found that group size was a significant predictor of parasite richness in a non-phylogenetic test, but not after controlling for phylogeny (Nunn et al. 2003). Moreover, a recent field study documented a negative association between group size and parasitism in red colobus monkeys (*Procolobus rufomitratus*, Snaith et al. 2008).

Several modeling studies have found that patterns of sociality can reduce disease risk when large groups are organized into clusters of individuals, called *modules*, in which individuals interact locally (Watve and Jog 1997; Wilson et al. 2003; Huang and Li 2007; Salathé and Jones 2010). A highly modular network can be divided into modules in which individuals interact more frequently among members of their module than with members of other modules. Mathematical models suggest that socially transmitted diseases are less likely to become established in highly modular networks because infections tend to spread quickly within modules but die out before spreading to other modules (Watve and Jog 1997; Wilson et al. 2003; Huang and Li 2007; Salathé and Jones 2010; see also Hess 1996). These findings support the

view that sociality influences the risk of parasitism in social animals, but indicate that group size is insufficient to characterize the risks associated with social living in the complex societies of many primates.

In this paper, we shift the focus away from using overall group size as a measure of sociality and disease risk, and instead investigate whether patterns of social contact within primate groups predict parasite risk across species. To achieve this goal, we integrate three methodological approaches: individual-based modeling (Kohler and Gumerman 2000; Grimm and Railsback 2005), social network analysis (Krause et al. 2007; Whitehead 2008; Wey et al. 2008), and phylogenetic comparative methods (Harvey and Pagel 1991; Nunn, 2011). We begin by presenting an agent-based susceptible-infectious-resistant (SIR) model to simulate the spread of infections across a sample of artificial networks that differ in network connectivity. In our simulations, we explore the effects of two network properties on the success of an invading parasite: network modularity (Q), or the extent to which the network is divided into subgroups, and network centralization (C), or the extent to which a small number of individuals dominate social interactions in the group.

As discussed earlier, Q is expected to limit outbreak size by containing infections within modules, and this effect has been observed in several previous modeling studies (Watve and Jog 1997; Wilson et al. 2003; Huang and Li 2007; Salathé and Jones 2010). The second property, C, is less well studied in the context of disease spread. A highly centralized network (high C) contains a small number of highly connected, or central individuals, while most individuals occupy peripheral positions in the network. By contrast, in a decentralized network (low C), individuals are more homogenous in how well connected they are in the network. In the context of epidemiology, highly central individuals are sometimes referred to as "super-spreaders," because these individuals are far more likely to receive and transmit a pathogen than peripheral individuals (Canright and Engo-Monsen 2006). However, this does not mean that centralized *networks* are necessarily more susceptible to invasion than decentralized networks. In particular, the presence of many peripheral individuals in a centralized network may increase the likelihood

that an introduced infection goes extinct before reaching a super-spreader (Lloyd-Smith et al. 2010). On the whole, we anticipated that centralization would have a negative effect on parasite success, similar to the reduced size of outbreaks on networks with greater heterogeneity in the number of contacts per individual reported by Lloyd-Smith et al. (2010)

In our simulations, average outbreak size showed a negative relationship with Q, yielding the prediction that increased modularity (high Q) limits the ability of parasites to establish in a population. Following from this, we predicted that fewer parasite species would be found in primate hosts that have more modular social networks. We investigated this by testing whether parasite richness (the number of different parasite species that are reported to infect a given host species) covaries negatively with Q across primates in the wild. We conducted the empirical analyses using comparative data from 19 anthropoid primates on social networks (comprised primarily of grooming matrices) and parasite richness counts obtained from the Global Mammal Parasite Database (Altizer and Nunn 2005).

We also investigated whether Q and C covary with group size. There is reason to expect greater network structure to occur in species that live in larger social groups. One possible reason is that as group size increases, individuals inevitably face spatiotemporal constraints on opportunities for social contact. In addition, primates living in larger groups may have evolved social behaviors that lead to increased social network structure (Kudo and Dunbar 2000).

Materials and Methods

Our analyses quantify networks using community modularity (Q) and the eigenvector centralization index (C). Q measures how strongly a network is divided into subgroups, while C measures the extent to which one or a few nodes in a network have many connections (i.e., a high centrality score) relative to others. The following two sections describe these metrics.

Community modularity (Q)

Community modularity (Q) is a measure of network modularity (i.e., the extent to which the

network is divided into subgroups, or 'modules,' of locally interacting individuals) developed by

Newman and Girvan (2004). Q quantifies modularity by comparing the frequency of

interactions within modules to the frequency of interactions among modules. High Q networks

exhibit greater modularity than low Q networks. An advantage of Q is that it is straightforward

to factor weighted edges into the calculation, allowing us to preserve as much information as

possible in the empirical primate networks. Thus far, the most commonly used metric of

network modularity in primate studies is the clustering coefficient (e.g. Flack et al. 2006), but

standard implementations of this metric require binary contact networks, which erase much

useful information. Additionally, Q is independent of group size and the absolute value of edge-

weights. This latter point is particularly important for our analysis because the absolute values of

edge-weights cannot be compared among our empirical networks due to different study lengths

and behaviors recorded. A final motivation for using Q is that recent studies have set a precedent

for investigating this specific metric in the context of infectious disease dynamics (Huang & Li,

2007; Salathe & Jones, 2010).

Q does not identify the modules in a group, but rather evaluates the degree of modularity

given some prior partitioning of the network. Consider a network partitioned into k subgroups.

Define matrix $\mathbf{E}$ as a $k \times k$ symmetric matrix where the element e_{ij} is the fraction of edges in the

network with a vertex in both subgroups i and j. The column sums $a_i = \Sigma_j e_{ij}$ represent the fraction

of edges in the network with at least one vertex in subgroup i. If the edges in the network are

distributed randomly with regard to which subgroups they are assigned, then $e_{ij} = a_i a_j$. The

community modularity is defined as:

$$Q = \sum_{i=1}^{k}(e_{ii} - a_i^2) \tag{1}$$

Thus, Q represents the fraction of edges in a network that occur within modules minus the

expected value of the same quantity in a network with the same modules, but edges distributed

randomly with regard to those modules. When the distribution of edges is random with regard to

modules, the modularity approaches $Q = 0$. As modularity increases, Q approaches its limit of Q = 1. Most networks do not reach such extreme values because the maximum possible Q for a network of a given size is constrained by the number of possible modules. In practice, networks with strong modularity generally have values of Q between 0.3 and 0.7, and low modularity networks will have values of Q less than 0.3 (Newman and Girvan 2004). Note that larger networks are not *biased* toward larger values of Q, and even very small networks (e.g., 4 individuals for a network with no self-loops or isolated nodes) are able to reach values of modularity above $Q = 0.3$. Fig. 1A and Fig. 1B provide graphical representations of networks with low and high values of Q.

Before calculating Q, a network must be partitioned into modules. One way to do this is to assign vertices to modules based on characteristics that are independent of the network (e.g., age or sex). In such a scenario, Q tests whether these modules are structurally meaningful. A second approach is to identify modules empirically such that more interactions take place within modules than between them (e.g., Lusseau and Newman 2004). For this study, the latter approach was used. Modules were identified with a modularity-maximizing agglomerative algorithm that partitions the network into modules such that a higher density of edges exists within modules than between them (Clauset 2005). Modules should slow and limit the size of disease outbreaks because infections will tend to spread quickly within modules (due to many within-module interactions) but die out before spreading to other modules (due to few between-module interactions). By identifying modules empirically, we aim to identify the same modules that a socially transmitted disease would "find" as it spreads.

Eigenvector Centralization Index (C)

The eigenvector corresponding to the largest eigenvalue of an adjacency matrix provides a measure of node centrality in networks (Bonacich 1972). Eigenvector centrality (EVC) may be expressed as a sum, where γ is the largest eigenvalue of $\mathbf{A}$, a_{ij} is an edge between vertices i and j, x_i is the EVC of vertex i, and n is the number of vertices:

$$\gamma x_i = \Sigma_{j=1}^{n} a_{ij}x_j, \qquad i = 1 \dots n \qquad\qquad (2)$$

This formula shows the key difference between eigenvector centrality and other measures of centrality: the centrality of a node is proportional to the sum of the centralities of the nodes to which it connects. In contrast to metrics such as degree or strength centrality that only account for a node's direct connections, eigenvector centrality takes the entire network into account such that nodes may obtain a high centrality by being connected to many low-centrality nodes or by being connected to a smaller number of high-centrality nodes (Bonacich 2007). This property makes eigenvector centrality particularly useful for understanding how network structure affects disease spread, because an individual's risk of acquiring a socially transmitted infection depends not only on how many contacts it has, but also on how well connected its contacts are. Further, as with Q, eigenvector centrality straightfowardly incorporates information about edge-weights. Simulations have demonstrated that eigenvector centrality is a strong predictor of the power of individual nodes to spread information across a network (Canright and Engo-Monson 2006).

Global metrics have been developed to measure the centralization of entire networks. Measures of centralization reflect the tendency in a network for one or a few nodes to be more central than the others, and are generally based on the difference between the centrality of the most central node in the network and that of all the other nodes (Freeman 1979). In the present study, the centralization index described by Wasserman and Faust (1994) was modified to incorporate eigenvector centralities, as done by Kasper and Voelkl (2009). Eigenvector centrality scores were first normalized using the Euclidean norm, which ensures that the maximum centrality in the most centralized network possible, a star network, is equal to 1 (Ruhnau 2000). A star network is a network with one node having $n - 1$ connections and all other nodes having 1 connection. The eigenvector centralization index C is expressed with the following equation, where n is the number of vertices, c_n is the eigenvector centrality of vertex n, and c_{max} is the maximal eigenvector centrality occurring in the network:

$$C = \frac{\sum_{i=1}^{n}(c_{max} - c_n)}{\sum_{i=1}^{n}(1 - c_n)} \qquad\qquad (3)$$

The numerator of the equation is the sum of the differences between each of the normalized eigenvector centralities and the maximum eigenvector centrality occurring in the network. The denominator is the sum of the differences between each of the normalized eigenvector centralities and 1 (recall that when the Euclidean norm is used, 1 is the eigenvector centrality reached by the most central node in a star network). Thus, the maximum value of $C = 1$ is reached when $c_{max} = 1$, which occurs when the network is a star network, and the minimum value of $C = 0$ is reached when $c_{max} = c_n$ for all n, which occurs when the eigenvector centralities of all the nodes are the same. Fig. 1B and Fig. 1C provide graphical representations of networks with low and high values of C.

Computer Simulations

To investigate the effects of Q and C on disease spread, we generated 10,000 random networks with $n = 25$ nodes and edge probability $p = 0.20$ using the Combinatorica package in Mathematica. The value of $p = 0.20$ was selected in order to generate networks with a similar average degree (i.e., edges per node) as the empirical data set (described below). Simulated networks were unweighted, undirected and connected (i.e., no completely isolated nodes). For each random network, Q, C and average degree were calculated. Average degree was measured only to control for its effects when simulating disease spread, as it is trivial to show that infections will spread further in networks with a greater number of edges.

We implemented a simple, well-studied model of disease dynamics: an agent-based SIR model, in which individuals move between susceptible (S), infectious (I), and resistant (R) classes. The SIR model forms the basis of many epidemic models, and the certain ending point of the epidemic we simulated allowed us to focus on the average final outbreak size, R_∞, as our measure of parasite success (discussed below). Every simulation resulted in the local extinction

of the pathogen as infected individuals recovered and demographic processes such as births and deaths did not change the pool of susceptibles.

The SIR model was built in MATLAB (version 7.8). At the start of a simulation, a random individual in the network was selected for the first infection and all other individuals were susceptible. In each subsequent time step, the probability that a susceptible individual became infected was equal to the per-contact infection probability, β, multiplied by the probability that the individual interacted with an infectious individual. The underlying social network determined the probability of social interactions occurring between infectious and susceptible individuals. If there was no edge between a pair of individuals, then the probability that they interacted was zero. If there was an edge, then those two individuals shared an interaction. If one individual was infectious and the other was susceptible, then the probability of a transmission event was equal to the per-contact transmission probability, β. Infected individuals remained infectious for 10 days before becoming resistant and non-infectious. Each simulation ran until one of two criteria was reached: either the infection spread throughout the group, or the infection went extinct (i.e., no infectious individuals remained in the population). Once the simulation ended, we recorded the final outbreak size. For each network, we estimated R_∞ by running 1000 independent simulations from random starting points and recording the average outbreak size as R_∞. The measure R_∞ is akin to R_0, the basic reproductive number that determines parasite success. Like R_0, R_∞ is influenced by network structure and measures a group's vulnerability to an outbreak following the introduction of an infection (Diekmann et al. 1998; Keeling 1999). Because greater susceptibility to disease outbreaks allows increased opportunities for parasites to become established in a group in the long-term, social network characteristics that lead to higher R_∞ in our simulations are also expected to lead to greater parasite richness in hosts. Thus, we use the relationship between R_∞ and network metrics in our simulations to predict patterns of parasite richness across primate species with different social networks.

An initial set of simulations was carried out in which the per-contact transmission probability β varied randomly between 0 and 1 across simulations. Because we observed a non-linear relationship between β and R_∞ (Fig. 2), we ran our final analysis four times with β held constant at each of four intermediate values of β: 0.1, 0.2, 0.3 and 0.4. At each value of β, we estimated R_∞ for each of the 10,000 random networks (recall that R_∞ is estimated by recording the average of 1000 independent simulation runs on the network). Thus, each of our 10,000 random networks is associated with its calculated network metrics (Q, C and average degree) along with four estimates of R_∞ (i.e., one mean R_∞ from 1000 simulations for each of the values of β).

All simulated data, including network metrics and R_∞, was standardized to have a mean of zero and standard deviation of 1. The effect of network metrics on R_∞ was assessed using a multiple regression with network metrics (Q, C and average degree) as predictor variables and R_∞ as the dependent variable. Since there are four estimates of R_∞ for each network, four separate regression analyses were conducted, as this proved more effective for assessing the impact of network structure on R_∞ at different values of β. We focused on assessing the relative effects of predictor variables based on standardized regression coefficients, and thus do not provide p-values, as these depend on the sample size of our simulations. However, we provide t-statistics for those wishing to convert the results to more standard statistical significance tests.

We used the Akaike Information Criterion (AIC, Akaike 1974) to assess whether to include interactions in the statistical model. Models with interaction terms gave a substantially lower AIC than those without interactions (e.g., for β =0.2, AIC with interactions = -8902, AIC without interactions = -5071). We therefore present results from analyses that included interaction terms. We used the "car" package in R to calculate variance inflation factors for all variables in the model. Although there were correlations among network metrics ($Q \sim C = 0.31$; $Q \sim$ Ave. Degree $= -0.72$; $C \sim$ Ave. Degree $= -0.41$), variance inflation factors were always less than 10, indicating that collinearity is not an issue (Marquardt 1970; Petraitis et al. 1996).

Primate Social Networks

We compiled interaction matrices for haplorrhine primate species from published studies and personal communications with experts on different primate species (Table 1). Because we focused on the relationship between patterns of social contact and exposure to parasites, only interactions involving close proximity were considered, including grooming, play and contact sitting. We focused on data from free-ranging groups of primates in the wild. We included some captive groups that that were taken as a unit from a wild or free-ranging group (although subsequent births may have occurred) and represented a relatively typical composition of age and sex classes and group size for the species in the wild (based on Rowe 1996). We focused on collecting data from long-term studies over periods of social stability; thus, our networks capture stable relationships among group members and will not reflect fission-fusion sociality or other forms of temporal variation in social contact across the study period. By recording weighted measures of interaction, we reduce the sensitivity of our networks to rare events during the study period.

We compiled a total of 19 primate social interaction matrices and calculated Q and C for each network (Table 1; electronic supplements A and B). Sixteen matrices were found in the literature and three were obtained through personal communications. Twelve groups were wild, three were free ranging, and four were captive. Most matrices contained grooming data ($n = 16$), with the remaining matrices based on close proximity ($n = 1$), contact sitting ($n = 1$), and socio-positive interactions involving contact ($n = 1$). Network analysis was performed with the Combinatorica package in Mathematica 7.0. Interaction matrices were depicted as weighted networks: each individual in the matrix was a node in the network and the weight of an edge between any two nodes was equal to the value of the cell in the matrix corresponding to the interactions between those two individuals. Values within matrices were normalized so that edges within networks are comparable, but the absolute values of edges cannot be compared between networks. However, our metrics Q and C are independent of the absolute values of edges (i.e. every edge in the network can be multiplied by a constant without changing Q or C);

thus, Q and C can be compared among networks. Undirected networks were used under the assumption that the direction of social interactions was not relevant to the spread of parasites. Whenever the original interaction matrix for a social group provided directional interactions (e.g. individual A grooms individual B), interactions were summed to create a symmetric matrix and thus an undirected network. Infants were removed from the networks because infants interact almost exclusively with their mothers and therefore do not add to the assessment of community network structure. The matrices used for analyses are available from the authors upon request.

Parasite Richness

Parasite richness counts, defined as the number of parasite species documented for a given host, were obtained from the *Global Mammal Parasite Database*, an online database of infectious diseases reported in wild primate populations (Nunn and Altizer, 2005). Richness is a general measure of parasitic pressure that has been examined in a wide range of host-parasite systems (Poulin and Morand 2004), including primate and other mammalian hosts (Poulin 1995; Morand and Harvey 2000; Nunn we al. 2003; Lindenfors et al. 2007; Bordes and Morand 2009). Alternative measures of parasitism, such as prevalence, focus on a single parasite, and are difficult to compare across species due to heterogeneity in environmental conditions, measurement methods, and stochastic extinction of parasites from some populations but not others (resulting in many estimates of zero prevalence). By contrast, parasite richness accounts for the impact of infection with multiple parasite species on host fitness, which is ubiquitous in natural host populations (Poulin & Morand, 2004; Bordes & Morand, 2009).

We recorded two different measures of parasite richness. For our main measure, we limited the count to highly contagious pathogens that are transmitted through close proximity or direct physical contact (Pederson et al. 2005). This measure, close-transmission parasite richness (CLOSE) explicitly reflects our prediction that network structure influences the spread of contagious parasites. Additionally, we measured total parasite richness (TOTAL) because it was used previously in a comparative study relating several variables, including measures of host

sociality (group size and host density), to parasite diversity (Nunn et al. 2003, see electronic supplement C). By comparing the relative effects of modularity and group size on the two measures of parasite richness, we can test whether network modularity influences CLOSE to a greater extent than it influences TOTAL, which is expected if the effect of social contact patterns on parasite transmission is driving the relationship. Conversely, group size may show a relatively stronger association with TOTAL, which include parasites that may experience a greater degree of density-dependent transmission (e.g. vector borne, fecally-transmitted, and environmentally transmitted).

When a parasite is not found in a particular host this may reflect one of two possibilities: either the parasite does not occur in the host, or the host has not been sampled sufficiently (Walther et al. 1995). For this reason, we estimate the total research effort that has been directed at a given primate host species by counting the number of publications involving that species. We include this variable in our general linear models to control for sampling effort in our parasite counts under the assumption that more research effort on a host species will lead to more of its parasites being discovered. Citation counts were obtained from PrimateLit (http://primatelit.library.wisc.edu), which provides a comprehensive compilation of bibliographic information for scientific publications on primates from 1940 to the present. Following Nunn et al. (2003), the citation count was not restricted to studies of parasitism, but included all published studies for each primate host species. Citation counts were log-transformed to meet the assumption of linearity for the regression analysis.

Phylogenetic comparative methods

We conducted phylogenetic comparative analyses to assess phylogenetic signal and to investigate relationships among network structure, group size and parasite richness. When data points are non-independent due to shared phylogenetic history, Type I error rates (i.e., false positives) are elevated in analyses that fail to control for phylogeny (Martins and Garland 1991; Purvis et al. 1994; Harvey and Rambaut, 1998). Various aspects of primate social systems are

thought to correlate with primate phylogeny (Di Fiore and Rendall 1994). We therefore investigated whether the individual variables (and residuals from the regression models, Revell 2010) show evidence for phylogenetic non-independence.

To investigate phylogeny, we used methods that incorporate evolutionary history by representing the error term of the statistical model as a variance-covariance matrix that reflects the phylogenetic relationships among the species (Freckleton et al. 2002). For each of our models, we estimated the parameter λ, which scales the internal branches of the phylogeny and serves as a measure of phylogenetic signal (Freckleton et al. 2002). The parameter λ generally lies between 0 and 1. When $\lambda = 0$, this corresponds to a non-phylogenetic test because all internal branches are set to 0, resulting in a tree with no phylogenetic structure (i.e., a "star phylogeny"; Felsenstein 1985; Pagel and Lutzoni 2002). Values of λ greater than 0 represent increasing phylogenetic signal, with $\lambda = 1$ indicating that the given branch lengths adequately account for variation in the trait under a Brownian motion model of evolution. Phylogenetic relationships and branch lengths are never known with certainty; therefore results should not be conditioned on a single phylogenetic hypothesis (Huelsenbeck et al. 2000; Pagel and Lutzoni 2002). Instead of relying on a single tree for the analyses, we used a sample of 100 dated phylogenies from a recent Bayesian inference of primate phylogeny (*10kTrees*, Version 2; Arnold et al. 2010; http://10ktrees.fas.harvard.edu/).

We used similar phylogenetic methods to incorporate phylogeny when testing the predictions. Using the same sample of 100 dated phylogenies, statistical parameters were sampled from a Bayesian posterior probability distribution using the program BayesTraits (Pagel and Meade, 2007). BayesTraits uses Markov Chain Monte Carlo (MCMC) to sample regression coefficients and λ, with one of the trees randomly selected in each iteration of the chain. We ran each MCMC chain for 3,050,000 iterations, sampling parameter values every 100 iterations and discarding the first 50,000 iterations as burnin. We used uniform priors for regression coefficients ranging from -100 to 100 and adjusted the "ratedev" parameter to obtain average acceptance rates between approximately 20% and 30% (Pagel and Meade 2007).

We ran all analyses three times to ensure convergence to the same distribution of parameter estimates. Although it is not possible to adjust the prior in the regression model for continuously varying data in BayesTraits, we also ran analyses using a standard phylogenetic generalized least squares approach based on maximum likelihood and a consensus tree, which produced congruent results.

In testing predictions, we first investigated whether group size was related to network modularity and centralization, as measured by Q and C, in our empirical primate networks. We then investigated the effects of modularity, group size and sampling effort on our measures of parasite richness. The analysis was conducted once with CLOSE richness as the dependent variable, and again with TOTAL richness as the dependent variable. Due to the relatively small number of data points (n = 19), we limited this model to those predictor variables for which we had the strongest predictions based on the simulation results and theoretical considerations. Thus, centralization was not included in our main results because our simulations yielded ambiguous predictions for the relationship between centralization and parasite richness (see Results and electronic supplement D). Average degree was not included because the metric is not comparable across the empirical primate networks; although weighted edges can be added together to create a measure of "weighted average degree," absolute edge weights are not comparable between networks in our study because, in contrast to Q and C, weighted degree depends on the *absolute* value of edges rather than the *relative* values of edges within the network. However, one might predict a positive relationship between average degree and parasite richness if species that tend to have more social partners per individual are also generally more social. Thus, for completeness, we ran an additional analysis including centralization, average degree and body mass as predictor variables. Our results were unchanged by the inclusion of these additional variables (Electronic supplement E).

To assess support for our predictions, we calculated the percentage of samples from the MCMC analysis in which a regression coefficient was in the predicted direction (positive or negative) and report those percentages, along with the mean estimates for λ and the regression

coefficients. If an independent variable has no effect on the dependent variable, we expect its coefficient will be equally represented as positive or negative (i.e., 50% of samples will support the prediction). Percentages closer to 100% reflect greater support for a prediction. We again used the "car" package in R to calculate variance inflation factors for all variables, which is justifiable because maximum likelihood estimates of λ were close to zero. Variance inflation factors were less than 2 for all predictor variables, suggesting that collinearity is unlikely to impact the output from our multiple regression models (Marquardt 1970; Petraitis et al. 1996).

Results

Computer Simulations

Table 2 provides the means and ranges of network metrics for the randomly generated networks and for estimates of R_∞ for different values of β. With R_∞ as our response variable, we investigated the effects of Q and C, while controlling for average degree. Our model was a strong predictor of R_∞ in all four sets of simulations, explaining between 94-98% of variation in R_∞ depending on the value of β (Table 3). Controlling for average degree, we found that increases in Q resulted in lower R_∞ in all of our models.

The effects of C on R_∞ were less straightforward, with the effect being positive when $\beta = 0.1$ and $\beta = 0.2$, and negative when $\beta = 0.3$ and $\beta = 0.4$ (Table 3). To explain the dependency of C on the value of β, we ran simulations on a few networks at different values of β and tracked the infection status of individual nodes. By doing this, we were able to determine the contribution of individual nodes to the average size of outbreaks at different values of β. At the level of individual nodes, we found that eigenvector centrality was a very strong predictor of the probability that an individual became infected and transmitted the infection to many individuals in a given simulation (see electronic supplement D). Importantly, we found that networks with high C exhibited a greater proportion of highly *isolated* nodes as well as highly *central* nodes. This helps to explain why the effects of C on R_∞ are dependent on β. At lower values of β, rapid extinction was extremely common in all networks, but the presence of central individuals in

highly centralized networks drove the positive effect of C on R_∞ by contributing to rare outbreaks. Conversely, at higher values of β, R_∞ tended to be large in all networks, but isolated individuals in highly centralized networks drove the negative relationship between C and R_∞ by frequently evading infection and resulting in rapid extinction (electronic supplement D).

Because C is dependent on β, we lack a clear prediction for the influence of C on parasite richness, as parasite species counts include parasites with a variety of unknown per-contact transmission probabilities. Thus, we focused our empirical tests on Q, with the expectation that higher values of Q reduce parasite success, and thus Q should covary negatively with parasite richness.

Community Modularity (Q) and Parasitism in Primates

The 19 primate social networks displayed wide variation in Q. A group of five brown howler monkeys (*Alouatta guariba*) had the lowest modularity ($Q = 0$) and 25 Guinea baboons (*Papio papio*) had the highest modularity ($Q = 0.57$). The mean modularity was $Q = 0.27$ with most networks falling between $Q = 0.2$ and $Q = 0.4$, indicating a moderate amount of community structure (Newman and Girvan 2004). We also discovered wide variation in C across primate networks. The group of five brown howler monkeys (*Alouatta guariba*) had the lowest score ($C = 0.10$), and a group of seven saddleback tamarins (*Saguinus fuscicollis*) had the highest score ($C = 0.87$). Mean C was 0.58, with most networks falling between $C = 0.4$ and $C = 0.8$, suggesting that primate social networks tend to be moderately centralized on a few individuals.

Controlling for phylogeny, regressions of Q and C against group size revealed that group size is positively related to Q (posterior support of 99.1%), while C showed no clear relationship with group size (posterior support of 53.9%; Table 4). Phylogenetic signal (measured as the mean λ in our regression analyses) was low to intermediate for both Q (mean $\lambda = 0.33$) and C (mean $\lambda = 0.37$; Table 4).

Our comparative test of social network structure and parasitism examined close-transmission parasite richness (CLOSE), with Q, group size and sampling effort as predictor

variables (Table 5; Fig. 3). Overall, this regression model explained substantial variation in CLOSE (mean $R^2 = 0.68$) and a low to intermediate level of phylogenetic signal (mean $\lambda = 0.35$). As expected, strong support was found for a positive relationship between sampling effort and CLOSE (posterior support of ~100%). As predicted by our simulations, we found strong evidence for a negative effect of Q on CLOSE (posterior support of 96.4%). We found only a very weak indication of a positive relationship between group size and CLOSE (posterior support of 72.6%).

We also investigated the effects of the same predictor variables on total parasite richness (TOTAL, Table 5). As compared to the analysis of CLOSE, this regression model explained less of the variation in TOTAL (mean $R^2 = 0.45$, mean $\lambda = 0.33$). Again, sampling effort was strongly supported as a predictor of parasite richness (posterior support of 99.9%). However, support for a negative effect of network modularity on TOTAL was markedly weaker (posterior support of 80.5%) compared to its effect on CLOSE, while support for a positive influence of group size on TOTAL increased, although it was still relatively weak (posterior support of 86.4%).

Discussion

We combined theoretical and empirical approaches to investigate the links among group size, social network structure and parasitism in nonhuman primates. Most social network analysis studies have investigated networks with hundreds or thousands of individuals and many thousands of interactions (e.g. 16,881 e-mail addresses and 57,029 e-mails in Newman et al. 2002), while fewer studies have focused on smaller networks with tens or hundreds of individuals that are typical of many social animals (Vital and Martins 2009). In contrast, our simulations focused on disease dynamics in relatively small networks, such as those that characterize most wild primate groups. In addition, our study is the first to incorporate empirical data on primate social networks into a comparative analysis of parasite diversity across species,

thereby providing the first empirical test of hypotheses involving the relationship between parasitic pressure and social network structure.

The goal of our simulations was to generate theoretically grounded predictions for the relationship between network structure and disease risk. Our agent-based SIR model confirmed that greater modularity reduces parasite success (Watve and Jog 1997; Wilson et al. 2003; Huang and Li 2007; Salathé and Jones 2010; see Table 3). We found that the effect of network centralization (i.e., C, the tendency for a few individuals to dominate social interactions) was dependent on the per-contact transmission probability (β), such that the relationship was positive at lower values of β and negative at higher values of β (see electronic supplement D and Lloyd-Smith et al. 2010). As parasite richness includes parasites spanning a range of transmission probabilities, most of which are unknown, we lack a clear prediction for the effect of network centralization on parasite richness. However, the prediction for network modularity was clear: greater modularity should reduce the ability of socially transmitted pathogens to spread through groups, resulting in reduced prevalence, abundance and diversity of socially transmitted parasites.

The empirical component of our study aimed at testing the prediction that social network modularity reduces parasitic pressure on hosts. Using phylogenetic Bayesian regression models, we found strong evidence that modularity results in lower parasite richness for CLOSE parasite richness (Table 5). Importantly, the link between network modularity and parasite richness is much stronger for CLOSE richness than for TOTAL richness, as expected if the former result is driven by the influence of social contact on infectious disease dynamics (Table 4). Group size showed a somewhat different relationship to parasite richness than modularity did, with suggestive evidence for a positive effect of group size on TOTAL parasite richness, but a greatly reduced effect when only CLOSE parasites were counted (Table 5). This indicates that social network modularity is a stronger predictor of pressure from socially transmitted parasites, while group size may be more important for non-socially transmitted parasites, such as those that spread through fecal-oral routes. One consequence of this result is that social networks should

be considered in studies investigating links between sociality and socially transmitted parasites. Additionally, the transmission mode of parasites must be considered when making predictions linking host sociality to patterns of parasitism. In light of this, TOTAL parasite richness is probably of limited value to studies investigating links between parasitism and host sociality.

We also investigated whether group size is related to measures of network structure in nonhuman primates. Although network centralization was unrelated to group size, we observed strong evidence for a positive relationship between group size and network modularity (Table 4). This raises the possibility that increased social network modularity in larger groups could contribute to the failure of some comparative studies to find the expected relationship between group size and parasite richness in social hosts. Three factors could lead to increased modularity in larger groups. First, spatiotemporal constraints could contribute to this relationship because in larger groups, individuals are farther apart from one another and face time constraints that limit their ability to interact with all individuals in the group (Dunbar and Dunbar 1988; Dunbar 1992). Second, sub-grouping in primate groups probably occurs as a consequence of non-random social interactions in larger groups through alliance formation, which is driven by competition within groups over food, mates and other resources (van Schaik 1989; van Hoof and van Schaik 1992). Finally, we propose that parasitism could represent a selection pressure favoring behaviors that increase social network modularity in large groups. For instance, selection could favor individuals who interact with only a few other individuals because this reduces exposure to new parasites, and these non-random interactions could translate to more modular social networks. At the group level, even if modular social networks arise for other reasons, parasitic pressure could select for more modular groups because they are less vulnerable to invading pathogens than groups with homogenous patterns of social contact (see also Loehle 1995).

Knowledge of host and parasite characteristics that influence disease risk and transmission can inform conservation efforts to prevent or slow infectious disease outbreaks or cross-species transmission of generalist parasites to endangered species or humans. Our study

shows that a global feature of social networks – community modularity – is expected to impact the likelihood of establishment of a newly introduced parasite, as might occur following a cross-species transmission event. Although reports of direct parasite-induced extinction are rare (Smith et al. 2006), a plethora of anecdotal evidence suggests a high potential for parasites to cause population declines in wild animals, including primates (Nunn and Altizer 2006). One comparative study of mammals found that parasites transmitted by close contact are more likely to cause extinction risk than parasites with other transmission modes (Pederson et al. 2007).

Two assumptions of our simulation model and empirical test are worth discussing. First, our simulations assumed that the social behavior of individuals is independent of infection status. However, primates could avoid others when they exhibit signs of infection (Freeland 1976; Loehle 1995; Nunn and Altizer 2006), or sickness behaviors involving increased sleep and rest may lead infected individuals to be less social (Hart 1990). Thus, the presence of certain parasites could alter host social interactions, and thus network properties. These effects could be modeled in extensions of our study. Second, we assumed for our empirical test that grooming interactions provide opportunities for parasite transmission, which is paradoxical considering that grooming also removes parasites. However, only 3.6% of the CLOSE parasites in our database are arthropods that might be removed during grooming (Electronic supplement C). Thus, while grooming may result in the removal of some parasites, it may provide a greater opportunity for sharing other socially transmitted pathogens among hosts, thereby increasing the diversity of parasites found on each host (Nunn and Altizer 2006). Our study suggests that grooming networks do impact patterns of parasitism. Future studies could investigate this hypothesis more directly, for instance by combining genetic data on parasites with social network analysis of their hosts (e.g., Keele et al. 2009). If parasites tend to spread within social modules before spreading to other modules, then parasites found within modules should be more closely related to one another than parasites in different modules.

Several limitations of our comparative analyses are worth noting. First, the sample size was small both in overall magnitude ($n = 19$) and in relation to the number of predictor variables

(~1:6.3 ratio of predictors to observations). When too many predictor variables are included in a regression model, the Type I error rate is elevated (Tabachnick and Fidell 1989). A second limitation is the potential for error in our estimates of parasite richness and social network connectivity. Parasite richness is largely a function of sampling effort, and our method of controlling for sampling effort was based on the number of citations rather than actual sampling effort devoted to sampling host species for parasites (which is unknown). Similarly, error in the measures of network structure may have been present because the metrics failed to capture the temporal features of real-world social networks and did not account for dispersal or fission-fusion social dynamics, which are prominent aspects of social organization in many primate species. Lastly, the measures of network characteristics and parasitism came from different social groups, raising the possibility that intraspecific variation in these characteristics may reduce the linkage among them in a broad comparative test. The extent of intraspecific variation in social network structure is difficult to assess given the scarcity of data, and is an area in need of further research. Future studies could also investigate the relationship between parasite risk and social network connectivity *within* species to determine whether effects similar to those presented in this study also exist intraspecifically.

We could discern phylogenetic signal in our models, yet the signal was weak (mean λ ranged from 0.24 to 0.37, compared to an expected value of $\lambda=1$ under Brownian motion evolution). Previously, parasite diversity in primates has been linked to primate phylogeny (Nunn et al. 2003), and phylogeny has been associated with qualitative aspects of primate social organization including group size and community substructure (Di Fiore and Rendall 1994). If we assume that the host traits investigated in this study (i.e., parasite diversity and social network structure) are actually related to primate phylogeny, the low levels of phylogenetic signal can be explained by two factors. First, the sample size in the present study was relatively small, including only 19 species, while Nunn et al. (2003) included 101 species and Di Fiore and Rendall (1994) included 37. Perhaps phylogenetic signal would become more detectable with a larger sample size (see Freckleton et al. 2002). Second, error in estimating parasite richness and

network connectivity may have weakened the association between these host traits and primate phylogeny (Ives et al. 2007).

In summary, improved knowledge of the distribution of primate parasites in relation to social networks should assist ecologists and evolutionary biologists in developing more comprehensive models of socioecology and evolution, and also informs conservation biology. In this study, we combined empirical and theoretical approaches to study the relationships among social network modularity, group size and parasitic pressure on primate hosts. We found that social network modularity increases with group size in primates, and has a negative impact on parasite risk. Our results suggest that future work on the links between primate sociality and disease ecology would benefit from incorporating data on social networks. Conversely, the transmission mode of parasites is expected to determine whether group size or social network structure plays a greater role in mediating parasitic pressure, and "total parasite richness" may therefore be of limited value to studies of host sociality and parasitism.

Acknowledgments

Literature Cited

Ahumada JA. 1992. Grooming behavior of spider monkeys (*Ateles geoffroyi*) on Barro Colorado Island, Panama. Int J Primatol 13 (1): 33-49.

Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19(6): 716-723.

Altizer S, Nunn CL, Thrall PH et al. (2003) Social organization and parasite risk in mammals: integrating theory and empirical studies. Ann Rev of Ecol Evol Syst 34: 517-547.

Arnold C, Matthews LJ, Nunn CL (2010) The 10kTrees Website: A New Online Resource for Primate Phylogeny. Evol Anthro 19: 114-118.

Arnold W, Lichtenstein AV (1993) Ectoparasite loads decrease the fitness of alpine marmots (Marmotamarmota) but are not a cost of sociality. Behav Ecol 4: 36-39.

Boese GK. 1975. Social behavior and ecological considerations of West African baboons (*Papio papio*). In: Tuttle RH, editor, Socioecology and Psychology of Primates. The Hague: Mouton Publishers, p. 208.

Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. J of Math Sociol 2: 113-120.

Bonacich P (2007) Some unique properties of eigenvector centrality. Social Networks 29(4): 555-564.

Bordes F, Morand S (2009) Parasite diversity: an overlooked metric of parasite pressure? Oikos 118*:* 801-806.

Butovskaya ML, Kozintsev AG, Kozintsev BA. 1994. The structure of affiliative relations in a primate community: allogrooming in stumptailed macaques (*Macaca arctoides*). Hum Evol 9(1): 11-23.

Canright GS, Engo-Monson K (2006) Spreading on Networks: A Topographic View. Complexus 3: 131-146.

Chiarello AG. 1995. Grooming in brown howler monkeys, *Alouatta fusca*. Am J Primatol 35: 73-81.

Clauset A (2005) Finding local community structure in networks. Phys Rev E 72: 026132.

Cooper MA, Bernstein IS, Hemelrijk CK. 2005. Reconciliation and relationship quality in Assamese macaques (*Macaca assamensis*). Am J Primatol 65: 269-282.

Côté IM, Poulin R (1995) Parasitism and group size in social animals: a meta-analysis. Behav Ecol 6(2): 159-165.

Di Fiore A, Rendall D (1994) Evolution of social organization: A reappraisal for primates by using phylogenetic methods. Proc Nat Acad Sci USA 91: 9941-9945.

Diekmann O, De Jong MCM, Metz JJ (1998) A Deterministic Epidemic Model Taking Account of Repeated Contacts between the Same Individuals. J of Applied Prob 35(2): 448-462.

Dunbar RI, Dunbar EP. 1976. Contrasts in Social Structure among Black-and-White Colobus Groups. Anim Behav 24: 84-92.

Dunbar RIM, Dunbar P (1988) Maternal time budgets of gelada baboons. Animal Behaviour 36:970-980.

Dunbar RIM (1992) Time: a hidden constraint on the behavioural ecology of baboons. Behavioral Ecology and Sociobiology 31(1):35-49.

Ezenwa VO, Price SA, Altizer S et al. (2006) Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. OIKOS 115: 526-536.

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125: 1-15.

Flack JC, Girvan M, de Waal FBM, Krakauer DC. 2006. Policing stabilizes construction of social niches in primates. Nature 439: 426-429.

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: A test and review of evidence. Am Nat160: 712-726.

Freeland WJ. 1976. Pathogens and the evolution of primate sociality. Biotropica 8:12-24.

Freeland WJ (1979) Primate social groups as biological islands. Ecol 60: 719-728.

Freeman LC (1979) Centrality in Social Networks: Conceptual Clarification. Social Networks 1: 215-239.

Grimm V, Railsback SF (2005) Individual Based Modeling in Ecology. Princeton University Press, Princeton, NJ.

Hart BL (1990) Behavioral adaptations to pathogens and parasites: five strategies. Neuroscience and Biobehavioral Reviews 14:273-294.

Harvey PH, Pagel MS (1991) The Comparative Method in Evolutionary Biology. Oxford University Press, Oxford.

Harvey PH, Rambaut A (1998) Phylogenetic extinction rates and comparative methodology. Proc Royal Soc London B: Bio Sci 265: 1691-1696.

Hausfater G, Watson DF (1976) Social and Reproductive Correlates of Parasite Ova Emissions by Baboons. Nature 262: 688-689.

Hernandez AD, Macintosh AJ, Huffman MA (2009) Primate parasite ecology: patterns and predictions from an ongoing study of Japanese macaques. Huffman MA, Chapman CA (eds) Primate Parasite Ecology: the Dynamics and Study of Host-Parasite Relationships. Cambridge University Press, Cambridge.

Hess G (1996) Disease in metapopulation models: Implications for conservation. Ecol 77: 1617-1632.

Hoogland L (1979) Aggression, ectoparasitism, and other possible costs of prairie dog (Sciuridae, Cynomys spp.) coloniality. Behav 69: 1-35.

Huang W, Li C (2007) Epidemic spreading in scale-free networks with community structure.

J of Statistical Mechanics, P01014.

Huelsenbeck J P, Rannala B, Masly JP (2000) Accommodating phylogenetic uncertainty in evolutionary studies. Science 288: 2349-2350.

Huffman MA, Chapman C (2009) Primate Parasite Ecology: the Dynamics and Study of Host-Parasite Relationships (Studies in biological and evolutionary anthropology series). Cambridge University Press.

Hunkeler C, Bourliere F, Bertrand M. 1972. Le comportement social de la Mone de Lowe (*Cercopothecus campbelli lowei*). Folia Primatol 17: 218-236.

Ives AR, Midford PE, Garland T (2007) Within-species variation and measurement error in phylogenetic comparative methods. Systematic Biology 56(2):252-270.

Izawa K. 1980. Social behavior of the wild black-capped capuchin (*Cebus apella*). Primates 21 (4): 443-467.

Kaplan JR, Zucker E. 1980. Social organization in a group of free-ranging patas monkeys. Folia Primatol 34: 196-213.

Kasper C, Voelkl B (2009) A social network analysis of primate groups. Primates 50(4): 343-356.

Keele B, Jones J, Terio K, Estes J, Rudicell R, Wilson M, Li Y, Learn G, Beasley T, Schumacher-Stankey J (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460:515-519.

Keeling MJ (1999) The effects of local spatial structure on epidemiological invasions. Proc Royal Soc London B 266: 859-867.

Kohler TA, Gumerman GJ (2000) Dynamics of Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes. Oxford University Press, Oxford.

Krause J, Croft DP, James R (2007) Social network theory in the behavioural sciences: potential applications. Behav Ecol Socioecol 62(1): 15-27.

Kudo H, Dunbar RIM (2000) Neocortex size and social network size in primates. Anim Behav 62: 711-722.

Lindenfors P, Nunn CL, Jones KE et al. (2007) Parasite species richness in carnivores: effects of host body mass, latitude, geographic range and population density. Global Biol Biogeo 16: 496-509.

Lloyd-Smith JO, Schreiber SJ, Kopp PE et al. (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438(17): 355-359.

Loehle C (1995) Social barriers to pathogen transmission in wild animal populations. Ecology 76:326-335.

Lottker P, Huck M, Zinner DP, Heymann EW. 2007. Grooming relationships between breeding females and adult group members in cooperatively breeding moustached tamarins (*Saguinus mystax*). Am J Primatol 69: 1159-1172.

Lusseau D, Newman MEJ (2004) Identifying the role that individual animals play in their social network. Proc Royal Soc London B: Biol Sci 271: S477-S481.

Marquardt DW (1970) Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics 12(3): 591-612.

Martins EP, Garland T (1991) Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. Evol 45: 534-557.

McGrew WC, Tutin CEG, Collins DA et al. (1989) Intestinal parasites of sympatric Pan troglodytes and Papio spp. at two sites: Gombe (Tanzania) and Mt. Assirik (Senegal). Am J Primatol 17: 147-155.

Møller AP, Dufva R, Allander K (1993) Parasites and the evolution of host social behavior. Advan Study Behav 22: 65-102.

Morand S, Harvey PH (2000) Mammalian metabolism, longevity and parasite species richness. Proc R Soc Lond B Biol Sci 267: 1999-2003.

Newman MEJ, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. Phys Rev E 66:1-4.

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(026113): 1-15.

Nunn C L (2011) The Comparative Approach in Evolutionary Anthropology and Biology. University of Chicago Press, Chicago.

Nunn CL (2002a) A comparative study of leukocyte counts and disease risk in primates. Evol 56: 177-190.

Nunn CL (2002b) Spleen size, disease risk and sexual selection: a comparative study in primates. Evol Ecol Res 4: 91-107.

Nunn CL, Altizer S (2005) The global mammal parasite database: an online resource for infectious disease records in wild primates. Evol Anthro 14: 1-2.

Nunn CL, Altizer S (2006) Infectious Diseases in Primates: Behavior, Ecology and Evolution. Oxford University Press, New York.

Nunn CL, Altizer S, Jones KE et al. (2003) Comparative Tests of Parasite Species Richness in Primates. Am Nat 162(5): 597-614.

Nunn CL, Altizer SM, Sechrest W, Cunningham A. 2005. Latitudinal Gradients of Disease Risk in Primates. Diversity and Distributions 11:249-256.

Nunn CL, Altizer S, Sechrest W, Jones KE, Barton RA, Gittleman JL (2004) Parasites and the Evolutionary Diversification of Primate Clades. Am Nat 164: S90-S103.

Nunn CL, Gittleman JL, Antonovics J (2000) Promiscuity and the primate immune system. Science 290: 1168-1170.

Nunn CL, Thrall PH, Stewart K (2008) Emerging infectious diseases and animal social systems. Evol Ecol 22: 519-543.

Nunn CL, Thrall PH, Leendertz FH, Boesch C (2011) The spread of fecally transmitted parasites
in socially-structured populations. PLoS One in press.

Pagel M, Lutzoni F (2002) Accounting for phylogenetic uncertainty in comparative studies of
evolution and adaptation. In: Lässig M, Valleriani A (eds) Biological Evolution and
Statistical Physics. Springer-Verlag, Berlin, pp 148-161.

Pagel M, Meade A (2007) BayesTraits (www.evolution.rdg.ac.uk). Reading, UK.

Pederson A, Altizer S, Poss M, Cunningham AA, Nunn CL (2005). Patterns of host specificity
and transmission among parasites in wild primates. Int J of Parasitol 35: 547-657.

Pederson A, Jones K, Nunn CL et al. (2007) Infectious diseases and extinction risk in wild
mammals. Conserv Biol 21: 1269-1279.

Petraitis PS, Dunham AE, Niewlarowski PH (1996) Inferring multiple causality: the limitations
of path analysis. Funct Ecol 10: 421- 431.

Poirier FE. 1969. The Nilgiri langur (*Presbytis johnii*) troop: its composition, structure, function
and change. Folia Primatol 10: 20-47

Poulin R (1995) Phylogeny, ecology, and the richness of parasite communities in vertebrates.
Ecol Monographs 65: 283-302.

Poulin R, Morand S (2004) Parasite Biodiversity. Smithsonian Institution Press, Washington,
D.C.

Purvis A, Gittleman JL, Luh H (1994) Truth or consequences: effects of phylogenetic accuracy on two comparative methods. J Theor Biol 167: 293-300.

Revell L. 2010. Phylogenetic signal and linear regression on species data. Methods in Ecology and Evolution 1: 319–329.

Rowe N (1996) The Pictorial Guide to the Living Primates. Pogonias Press, East Hampton, New York.

Ruhnau B (2000) Eigenvector centrality- a node centrality? Social Networks 22: 357-365.

Sade SD. 1971. Sociometrics of Macaca mulatta I. Linkages and cliques in grooming matrices. Folia Primatol 18: 196-223.

Salathé M, Jones JH (2010) Dynamics and Control of Diseases in Networks with Community Structure. PLoS Comp Biol 6(4): e1000736.

Semple S, Cowlishaw G, Bennett PM (2002) Immune system evolution among anthropoid primates: parasites, injuries and predators. Proc Royal Soc B: Biol Sci 269: 1031-1037.

Shields WM, Crook JR (1987) Barn swallow coloniality- a net cost for group breeding in the Adirondacks. Ecol 68: 1373-1386.

Smith KF, Sax DF, Lafferty, KD (2006) Evidence for the role of infectious disease in species extinction and endangerment. Conservation Biology 20:1249-1357.

Snaith T, Chapman C, Rothman J et al. (2008) Bigger groups have fewer parasites and similar cortisol levels: a multi-group analysis in red colobus monkeys. Am J Primatol 70: 1072-1080.

Stoner KE (1996) Prevalence and intensity of intestinal parasites in mantled howling monkeys (Alouatta palliata) in northeastern Costa Rica: Implications for conservation biology. Conserv Biol 10: 539-546.

Strier KB. 1992. Causes and Consequences of Non-Aggression in the Woolly Spider Monkey, or Miriqui. In Silverberg J, Gray JP, editors, Aggression and Peacefulness in Humans and Other Primates. Oxford University Press, USA, p. 109.

Sugiyama Y. 1971. Characteristics of the social life of bonnet macaques (*Macaca radiata*). Primates 12: 247-266

Tabachnick BG, Fidell LS (1989) Using Multivariate Statistics (2nd. Ed.). Harper and Row, Cambridge, MA.

Takahashi H, Furuichi T. 1998. Comparative study of grooming relationships among wild Japanese macaques in Kinkazan A troop and Yakushima M troop. Primates 39(3): 365-374

van Hooff JARAM, van Schaik CP (1992) Cooperation in competition: the ecology of primate bonds. In: Waal AHHaFBMd, editor. Coalitions and Alliances in Humans and Other Animals. Oxford: Oxford University Press. p 357-389.

van Schaik CP (1989) The ecology of social relationships amongst female primates. In: Standen V, Foley RA, editors. Comparative Socioecology. Oxford: Blackwell. p 195-218.

Vital C, Martins EP (2009) Using Graph Theory Metrics to Infer Information Flow Through Animal Social Groups: A Computer Simulation Analysis. Ethology 115: 347-355.

Vitone ND, Altizer SM, Nunn CL (2004) Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. Evol Ecol Res 6: 1-17.

Vogt JL. 1978. The social behavior of a marmoset (*Saguinus fuscicollis*) Group II: Behavior Patterns and Social Interaction. Primates 19 (2): 287-300

Walther BA, Cotgreave P, Gregory RD et al. (1995) Sampling effort and parasite species richness. Parasitol Today 11: 306-310.

Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.

Watve MG, Jog MM (1997) Epidemic Diseases and Host Clustering: An Optimum Cluster Size Ensures Maximum Survival. J Theor Biol 184(2): 165-169.

Wey T, Blumstein DT, Shen W et al. (2008) Social network analysis of animal behaviour: a promising tool for the study of sociality. Anim Behav 75: 33-344.

Whitehead H (2008) Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis. University of Chicago Press, Chicago, IL.

Wilkinson GS (1985) The social-organization of the common vampire bat. 1. Pattern and cause of association. Behav Ecol Sociobiol 17: 111-121.

Wilson K, Knell R, Boots M et al. (2003) Group living and investment in immune defense: an interspecific analysis. J Anim Ecol 72: 133-143.

Table 1: Summary of empirical network data.

Species	Group Size	Study Type	Data	Source
Alouatta guariba	5	Wild	Grooming	Chiarello (1995)
Ateles geoffroyi	15	Free Ranging	Grooming	Ahumada (1992)
Brachyteles arachnoides	22	Wild	Close Proximity	Strier (1992)
Cebus apaella	12	Wild	Grooming	Izawa (1980)
Cercopithecus campbelli	15	Wild	Grooming	Hunkeler et al. (1972)
Colobus guereza	8	Wild	Grooming	Dunbar & Dunbar (1976)
Erythrocebus patas	19	Free Ranging	Grooming	Kaplan & Zucker (1980)
Macaca arctoides	19	Captive	Grooming	Butovskaya et al. (1994)
Macaca assamensis	19	Wild	Grooming	Cooper et al. (2005)
Macaca fuscata	21	Wild	Grooming	Takahashi & Furuichi (1998)
Macaca mulatta	16	Wild	Grooming	Sade (1971)
Maraca radiata	16	Wild	Grooming	Sugiyama (1971)
Macaca tonkeana	25	Free Ranging	Contact Sitting	Thierry (unpublished)
Pan paniscus	19	Wild	Grooming	Furuichi (unpublished)
Pan troglodytes	24	Wild	Grooming	Arnold & Whiten (unpublished)
Papio papio	25	Captive	Grooming	Boese (1975)
Saguinus fuscicollis	7	Captive	Grooming	Vogt (1978)
Saguinus mystax	6	Wild	Grooming	Lottker et al. (2007)
Trachypithecus johnii	10	Captive	Socio-positive	Poirier (1969)

Table 2: Means and ranges of network metrics from artificial (simulated) networks.

	Community Modularity (Q)	Eigenvector Centralization Index ($\underline{C}$)	Average Degree	R_∞ when $\beta = 0.1$	R_∞ when $\beta = 0.2$	R_∞ when $\beta = 0.3$	R_∞ when $\beta = 0.4$
Minimum	0.19	0.16	2.16	0.45	1.1	2.1	3.625
Mean	0.35	0.38	3.75	1.325	5.1	10.775	15.575
Maximum	0.59	0.73	5.68	4.325	12.8	19.375	22.175

Note: $N = 10,000$ random networks, β = per-contact transmission probability, R_∞ = average outbreak size (estimated at four different values of β) across 1000 simulations on each network.

Table 3: Results of multiple regression of simulation output.

Independent Variables	$\beta = 0.1$ (R^2 = 0.94)		$\beta = 0.2$ (R^2 = 0.98)		$\beta = 0.3$ (R^2 = 0.98)		$\beta = 0.4$ (R^2 = 0.96)	
	Estimate	t-Statistic	Estimate	t-Statistic	Estimate	t-Statistic	Estimate	t-Statistic
Community modularity (Q)	-0.08	-20.0	-0.10	-39.42	-0.081	-35.7	-0.05	-13.74
Eigenvector centralization index (C)	0.06	22.86	0.02	8.51	-0.03	-19.46	-0.06	-27.08
Average Degree (D)	0.92	215.87	0.91	344.05	0.91	386.85	0.90	262.54
$Q*C$	-0.04	-10.6	-0.03	-11.99	0.01	5.78	0.04	11.77
$Q*D$	-0.1	-41.22	-0.08	-53.93	0.03	22.59	0.13	69.36
$C*D$	-0.03	-6.507	-0.04	-17.4	-0.01	-3.75	0.03	9.67

Note: The dependent variable, R$_\infty$, was estimated at four different values of β. Columns for each β represent separate multiple regressions for each estimate of R$_\infty$. $N = 10{,}000$ data points for each regression (i.e., averages from 1000 simulations on each of 10000 networks). The sign of the estimate and t-statistic indicates whether the relationship is positive or negative. Estimates are standardized, such that larger values (in absolute magnitude) indicate larger effects of the variable on outbreak size.

Table 4. Relationships between group size and network metrics in primates (empirical results).

	% Positive Estimates	Mean Estimate	Mean λ	95% Credible Interval for λ	R^2
Community Modularity (Q)	99.1	0.012	0.33	0.02 to 0.8	0.3
Eigenvector Centralization Index (C)	53.9	0.0012	0.37	0.02 to 0.92	< 0.01

Model: network property (Q or C) ~ group size. The parameter λ is a measure of phylogenetic signal and is taken into account when obtaining other parameter estimates.

Table 5. Predictors of parasite richness in primates (empirical results).

	Close-Transmission Parasite Richness		Total Parasite Richness	
	% In Predicted Direction	**Mean Estimate**	**% In Predicted Direction**	**Mean Estimate**
Community Modularity (Q)	96.44 (-)	-20.41	80.46 (-)	-31.16
Group Size	72.56 (+)	0.14	86.41 (+)	0.96
Sampling Effort	100 (+)	5.02	99.89 (+)	9.99
Phylogenetic Signal (λ)		0.35		0.33
R^2		0.68		0.45

Model: parasite richness ~ community modularity (Q) + group size + sampling effort
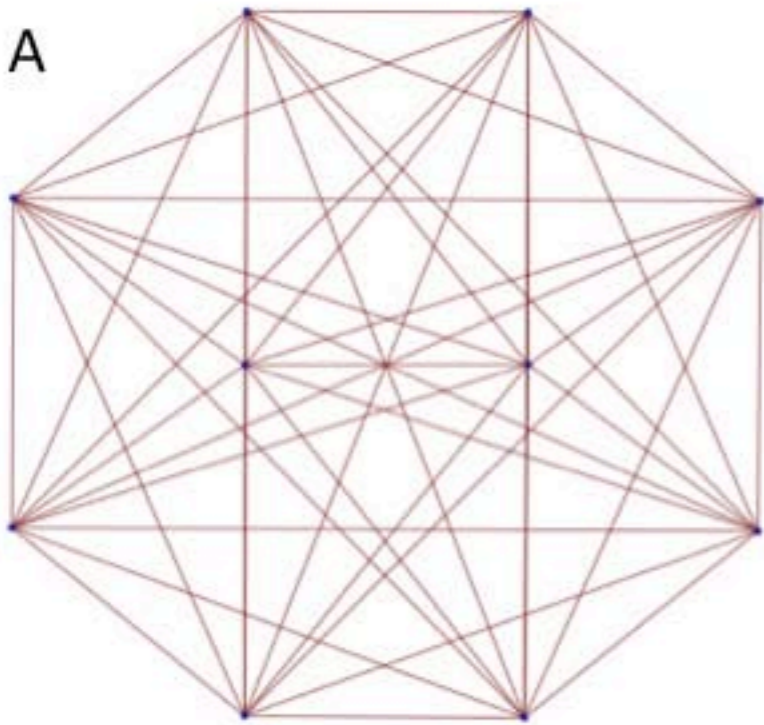
Figure Legends

Fig. 1 Graphical representations of networks with extreme values of community modularity (Q) and the eigenvector centralization index (C). Network A is a complete network (i.e. an edge between every pair of nodes) with 10 nodes. Since there are no subgroups in Network A, $Q = 0$, and since every node has the same number of edges, $C = 0$. Network B also has 10 nodes, but is divided into two clear subgroups and thus has a high modularity, $Q = 0.409$. Network C again has 10 nodes, but is a "star network," which defines the maximum value of $C = 1$; one node, the most central node, has $n - 1$ connections, while all other nodes have 1 connection.
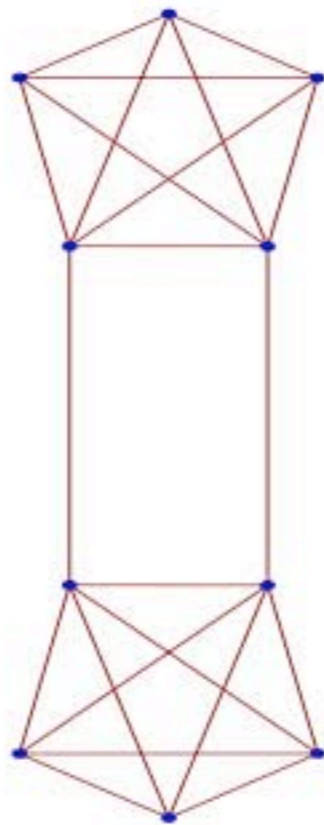
Fig. 2 Non-linear relationship between average outbreak size (R_∞) in simulations and the per-contact transmission probability (β) across the full range of β.

Fig. 3 Distributions of regression coefficients predicting the richness of closely-transmitted parasites of primates (panels A to C) and λ (panel D) (empirical results). Histograms represent the distribution of values sampled from the Bayesian (MCMC) analysis, controlling for phylogenetic uncertainty. For regression coefficients, an absence of an effect could be indicated by a histogram centered on zero, while increasingly strong effects are indicated by a departure from zero (e.g. for sampling effort).
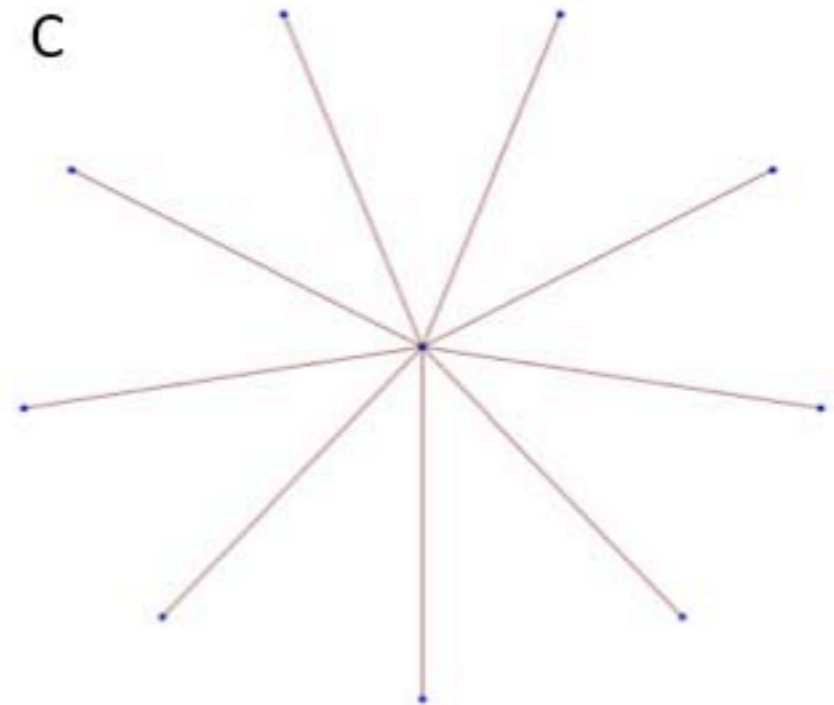
R_∞ as a Function of β

A — Distribution of Estimates for Q

B — Distribution of Estimates for Group Size

C — Distribution of Estimates for Sampling Effort

D — Distribution of Estimates for λ