Linking parasitism to network centrality and the impact of sampling bias in its interpretation

Zhihong Xu[1*], Andrew J. J. MacIntosh[1], Alba Castellano-Navarro[2,3], Emilio Macanás-Martínez[2,3], Takafumi Suzumura[4], Julie Duboscq[5,6]

[1] Kyoto University Primate Research Institute, Inuyama, Japan.

[2] Ethology and Animal Welfare Section, Universidad Cardenal Herrera-CEU, CEU Universities, Valencia, Spain

[3] Leipzig University, Life Sciences, Institute of Biology, Leipzig, Germany

[4] Kyoto University Wildlife Research Center, Kyoto, Japan.

[5] CNRS-MNHN-Université de Paris, UMR7206 Eco-anthropologie, Paris, France.

[6] Dept. Behavioral Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, University of Göttingen, Göttingen, Germany

*** corresponding author: Zhihong Xu, Kyoto University Primate Research Institute, Inuyama, Japan, zhihong.xu.36r@st.kyoto-u.ac.jp**

Abstract

Group living is beneficial for individuals, but also comes with costs. One such cost is the increased possibility of pathogen transmission because increased numbers or frequencies of social contacts are often associated with increased parasite abundance or diversity. The social structure of a group or population is paramount to patterns of infection and transmission. Yet, for various reasons, studies investigating the links between sociality and parasitism in animals, especially in primates, have only accounted for parts of the group (e.g., only adults), which is likely to impact the interpretation of results. Here, we investigated the relationship between social network centrality and an estimate of gastrointestinal helminth infection intensity in a whole group of Japanese macaques (*Macaca fuscata*). We then tested the impact of omitting parts of the group on this relationship. We aimed to test: (1) whether social network centrality – in terms of the number of partners (degree), frequency of interactions (strength), and level of social integration (eigenvector) – was linked to parasite infection intensity (estimated by eggs per gram of faeces, EPG); and, (2) to what extent excluding portions of individuals within the group might influence the observed relationship. We conducted social network analysis on data collected from one group of Japanese macaques over three months on Koshima Island, Japan. We then ran a series of knock-out simulations. General linear mixed models showed that, at the whole-group level, network centrality was positively associated with geohelminth infection intensity. However, in partial networks with only adult females, only juveniles, or random subsets of the group, the strength of this relationship - albeit still generally positive - lost statistical significance. Furthermore, knock-out simulations where individuals were removed but network metrics were retained from the original whole-group network, showed that this is partly a power issue and partly a real effect of incomplete network. Our study indicates that sampling bias can thus hamper our ability to detect real network effects involving social interaction and parasitism. In addition to supporting earlier results linking geohelminth infection to Japanese macaque social networks, this work introduces important methodological considerations for research into the dynamics of social transmission, with implications for infectious disease epidemiology, population management, and health interventions.

Keywords: sociality, social network, geohelminth, knock-out simulation, parasite transmission

Introduction

Living in a group provides many benefits to individuals, but also substantial costs (Loehle, 1995). One such cost is increased exposure to parasites and a higher probability of parasite transmission from one individual to another (Freeland, 1976; Loehle, 1995; Arneberg et al., 1998; Nunn et al., 2003). This pattern may arise from higher contact rates between individuals and generally higher local densities of hosts. From the perspective of vectors of parasites, a group of animals is an aggregation of hosts, potentially more attractive than a single host (Arneberg, 2002; Nunn et al., 2003). Even for parasites that are not transmitted through direct contact between hosts, sharing of resources and space, such as food patches, sleeping sites, or home ranges, increases the likelihood that hosts encounter infectious stages of parasites in the environment, which increases the probability of transmission between hosts (Altizer et al., 2003; Nunn et al., 2015; Müller Klein, M. Heistermann, et al., 2019).

Although living in a group may generally mean greater parasite exposure, within a group, individuals exhibit different rates of interacting: social interactions are influenced by kinship ties, dominance ranks, or age, size, and sex assortativity (Kurvers et al., 2014). This creates heterogeneity in contact patterns, leading to differences between potential contacts and actual contacts, thereby influencing each individual's exposure to parasites (Altizer et al., 2003). Thus, the positive relationship that is expected to occur between increased group size and increased parasitism is not always observed (Moller et al., 1993). Accounting for social structure and the role and position of individuals within it therefore helps make sense of how parasites are transmitted within animal groups. Studying the relationship between social interactions and parasite

infection is essential for understanding how animal social behaviours might reflect adaptations to infection risk, or conversely how parasites might invade or persist in host populations.

In this endeavour, Social Network Analysis (SNA) is an efficient statistical tool that can be used to analyse groups of individuals as a whole and generate predictions about the spread of infectious diseases (Borgatti, 2005). Social networks reflect the patterns of interactions, associations, or spatial proximity between individuals. SNA can provide information about the whole structure of the network, but also about the structural importance of certain nodes (e.g. individuals) and their ties (e.g. interactions) to identify influential individuals within a network (Borgatti et al., 2009). So-called central individuals have qualitatively or quantitatively greater direct or indirect connections than less central or peripheral individuals. Central individuals are thus expected to be at once key dispersers of parasites but also at the greatest risk of acquiring parasites from others (Romano et al., 2016). For example, gidgee skinks (*Egernia stokesii*) with high refuge-sharing rates are more likely to be infected by ticks and host a more diverse blood parasite community than less-connected skinks in the same refuge-sharing network (Godfrey et al., 2009). Similarly, female Japanese macaques (*Macaca fuscata yakui*) that are better connected in their grooming networks exhibit more diverse intestinal helminth communities and greater intensity of infection with certain parasite species (MacIntosh et al., 2012). In cases such as the latter, carefully designed experiments are required to test the mechanisms linking networks to infection. But in general, such links have been demonstrated in a broad array of contexts, regardless of the type of network or parasite species under study, or even the measures of parasitism used (Altizer et al., 2003; Ezenwa et al., 2016; Rushmore et al., 2017; Briard and Ezenwa, 2021; Lucatelli et al., 2021).

82 Factors like sex, social rank, and age can also influence host exposure and susceptibility to parasites.
83 These traits might have an impact on the frequency, quality, and quantity of interactions with social partners
84 and with the environment. For instance, in species where young males disperse from their natal group,
85 males will interact with more a numerous and diverse set of individuals compared with females (Teichroeb
86 et al., 2011). In addition, these traits might be linked to variations in physiological and immunological
87 responses. As a case in point, meta-analyses in vertebrates suggest that greater parasitism is biased toward
88 high-ranking individuals (Habig et al., 2018). This may be linked to stressful social situations - such as the
89 struggle to achieve high dominance status - that trigger the production of glucocorticoids, which is essential
90 in the activation of the immune response (Cain and Cidlowski, 2017). At the same time, however,
91 dominance status can also correlate with contact networks, making it hard to disentangle the mechanisms
92 of infection using observational data (MacIntosh et al., 2012; Wooddell et al., 2020).
93 Concerning age, juveniles may have a higher chance of contracting infectious agents because of their
94 specific activity patterns, which means higher exposure to infection (e.g., greater and more diverse
95 exploration of substrates, heightened reliance on social interactions, etc.) (Sarabian, et al., 2018). Similarly,
96 juveniles are known to have a less efficient or weaker immune system than adults, which increases their
97 susceptibility to infection. (Fallon et al., 2003; East et al., 2008; Ebersole et al., 2008). Often, though, studies
98 investigating links between sociality and parasitism only include adults, perhaps because conducting
99 research on juveniles is more difficult due to more challenging identification and observation conditions.
100 Many studies, including those by some of us (MacIntosh et al., 2012; Duboscq et al., 2016; Romano et al.,
101 2016; Webber et al., 2016; Hamilton et al., 2020; Roberts and Roberts, 2020; Sandel et al., 2021), have

omitted juveniles, and this may preclude a fuller understanding of infection dynamics. As a case in point, although unrelated to infection, a previous study using knockout simulations indicated that juvenile or adolescent baboons can act as bridges between clusters of otherwise weakly connected individuals, thus changing the nature of the networks being observed (Fedurek and Lehmann, 2017).

More generally, problems of missing data in social networks are well known. Missing nodes (Kossinets, 2006; Smith and Moody, 2013; Davis et al., 2018; Gilbertson et al., 2021), especially missing central nodes (Smith et al., 2017), produce measurement bias in social network centrality because network data are relational and hence one data point (node, edge, etc) is not independent of another. Although many network measures can be estimated with incomplete information and can be robust to missing individuals (Borgatti et al., 2006), it is important to assess the impact on precision, accuracy, and bias created by subsampling networks or by carrying out analyses with partial networks (Silk et al., 2015; Silk, 2018). This impact can be accessed via comparing, correlating, or regressing original whole-group network measures with newly calculated partial network measures. At present, we are starting to better understand the effects of using partial or subsampled networks on network measures (Silk et al., 2015; Smith et al., 2017), but we have scarce knowledge about the influence that such measurement bias has on the conclusions we might draw concerning the relationship between network characteristics and other social or ecological processes (but see Silk et al., 2015 or Herrera et al., 2021). Assessing measurement bias can thus inform us about how best to carry out our observations or even how we might correct for sampling bias when data have already been collected (Silk et al., 2015; Smith et al., 2017; Hoppitt and Farine, 2018). It is thus meaningful to assess how including (or excluding) juveniles or any other often-missing subgroup(s), or even individuals at

random, might affect our interpretation of the relationship between infection and social network centrality, which is the target of this study.

Japanese macaques and their gastrointestinal parasites comprise a well-suited system for studying pathogen transmission in wildlife (MacIntosh, 2014). Japanese macaque populations are generally stable, are distributed throughout Japan, and have been subject to a long history of research, which provides a firm foundation regarding their ecology and social structure. Japanese macaques are mainly terrestrial, enabling direct observation of social behaviour and sampling for parasitological analyses. Japanese macaques are also infected by numerous gastrointestinal nematode parasites, including various geohelminths, which are amenable to study through faecal sampling (Gotoh, 2000). Helminths are commonly found in mammals (Stephens et al., 2017) and can impact host health and fitness by imposing nutritional constraints and other physiological conditions that can reduce host fitness, and in some cases may directly cause mortality (Hillegass et al., 2010).

There are three known geohelminths in Japanese macaques: *Oesophagostomum aculeatum* (Strongylida), *Strongyloides fuelleborni* (Rhabditida), and *Trichuris trichiura* (Trichocephalida) (Gotoh, 2000). These parasites are transmitted through contact with contaminated substrates, though while *O. aculeatum* and *T. trichiura* use the faecal-oral pathway, *S. fuelleborni* infects hosts percutaneously. Voided with faeces, the eggs of *O.aculeatem* and *S. Fuelleborni* hatch and develop into infectious third-stage larvae in the environment. *S. Fuelleborni* exhibits a heterogonic life cycle whereby some larvae continue to develop in the external environment and persist as free-living adults, mating and reproducing opportunistically. Others (females only) become parasitic in a host. In contrast, the eggs of *T. trichiura*

142 become embryonated but do not hatch until they have been ingested by a suitable host. Each can thus be
143 transmitted between hosts that share environmental resources, even if separated in time (Modrý et al., 2018).
144 The present study examined gastrointestinal nematode parasites infecting Japanese macaques on
145 Koshima island, Japan, to test two hypotheses: (1) to what extent social network centrality, as measured by
146 the number of partners (degree), frequency of interactions (strength), and social integration (eigenvector)
147 observed across individuals, was linked to an estimate of intestinal parasite infection intensity (eggs per
148 gram of faeces, EPG); and, (2) to what extent analyses based on partial networks produced different results
149 than analysis based on a whole-group network. First, we predicted a positive association between an
150 individual's social network centrality and its gastrointestinal nematode parasite infection intensity
151 (estimated via faecal egg counts), as was found previously in Japanese macaques (MacIntosh et al., 2012)
152 and other primates and environmentally-transmitted parasites (Grear et al., 2013; Friant et al., 2016; Wren
153 et al., 2016; Müller Klein et al., 2019). We chose multiple network metrics because, in theory, each may be
154 associated with a different aspect of transmission or may be more or less instrumental to the process
155 (MacIntosh et al., 2012; Duboscq et al., 2016; Tiddi et al., 2019). Second, we predicted that accounting for
156 only partial networks would lead to divergent results concerning the network-infection link. Partial
157 networks examined included a juvenile-only network, because of their high rates of infection (MacIntosh
158 et al., 2010), an adult female-only network, because of their core role in macaque society and their
159 prominence in previous studies of this nature, and a series of random subsets of the network. We do not
160 make more precise predictions here as we cannot *a priori* predict with confidence the effects of missing
161 individuals on network measures.

Methods

Ethical Statement

The research presented here complied with the Guidelines for the Care and Use of Nonhuman Primates established by the Primate Research Institute of Kyoto University (KUPRI), to the legal requirements of Japan, and to the American Society of Primatologists (ASP) Principles for the Ethical Treatment of Non-Human Primates. Permissions were acquired from the Field Research Committee of KUPRI, the Cooperative Research Program of the Wildlife Research Center of Kyoto University, and the City of Kushima's Agency for Cultural Affairs.

Study site and subjects

This research was conducted on Koshima island, a 0.35km^2 island located in the Sea of Hyūga in Miyazaki Prefecture, Japan (31°27′N, 131°22′E). The island is inhabited by two groups of macaques and some solitary males. The home ranges of the two groups only overlap at their edges, and macaques from different groups were never observed in proximity during the observation period. The study group, the "main" group, inhabits the western part of the island, including a sandy beach, a range of forest, and a rim of rocky beach. Provisioning and behavioural observations of Koshima macaques started in 1952, and demographic, ecological, behavioural, and life-history data are available since then (Watanabe, 2008). Previous studies on this island identified four nematode parasites infecting Koshima macaques (Horii et al., 1982; Gotoh, 2000). The study group is currently provisioned with approximately 3 kg of wheat twice weekly. At the time of the study, the main group had 47 stable group members and 5 solitary roaming males

that ranged in proximity to the group on an irregular basis. The group included 20 adult females (≥5 years old), 8 adult males (≥5 yo) (including the 5 solitary males), 13 juvenile females (between 1 and 4 yo), and 11 juvenile males (between 1 and 4 yo). Because the population has been monitored for decades, we have the exact age of each individual in the population.

Behavioural Data Collection

The observation team - including ZX, AC-N, EM-M - observed the macaques for 88 days over 3 months from March to June 2017. All individuals could be identified reliably based on a combination of facial and other physical characteristics (such as facial tattoos, scars, limps, coat colour, etc.), and allowed researchers to approach within 5m, facilitating observation and faecal sample collection. We carried out 20-minute focal animal observations, recording behavioural activities of the focal individual (including resting, feeding, locomotion, playing, giving and receiving grooming, and self-grooming) at 30-second intervals. At the same scan point, we also recorded the identity of each individual within 1m of, or in body contact with, the focal animal. Agonistic interactions of all individuals in the group were recorded *ad libitum*. To assess the dominance hierarchy, we used all agonistic interactions recorded by the observation team between November 2016 and June 2017. We observed every individual in the group, including two infants that were born in 2016, every day in a pseudorandomized order, updated daily. We avoided repeating observations of the same individual on the same day, and we balanced observations between individuals as much as possible across days and times of day (e.g., early and late morning, early and late afternoon).

Data were collected either on paper or using a tablet (Panasonic Tough-pad, with the application PRIM8 (McDonald and Johnson, 2014)). A multi-observer reliability test based on Fleiss' Kappa index (Fleiss and Cohen, 1973) indicated a good match between observers for proximity records, including the identities of all subjects involved (Kappa = 0.815). In total, we conducted 547 focal observations recording 22,066 focal scans, or 183.3 hours of focal observations, with a mean ± SD of 10.7 ± 6.6 focal follows per individual (or 11.2 ± 0.8 per individual without 4 rarely seen solitary roaming males). Table SI-1 in the Supplementary Information summarizes the attributes of and data from the study subjects.

Social data analysis

Dominance

To compute the dominance rank of all individuals, we used the Elo-rating method (Neumann and Kulik, 2014), which reflects an individual's success in agonistic interactions relative to all other members (Albers and de Vries, 2001; Neumann et al., 2011). Elo-rating is based on a sequence of agonistic interactions ordered in time. Each individual is seeded with an arbitrary rating of 1000. This rating is then updated after each agonistic interaction with a clear outcome. The rating increases/decreases based on the outcome of the interaction (won or lost), the previous ratings of both opponents, and a determined factor, k (here k = 100 (Neumann et al., 2011)). In our analyses, we aggregated all data and calculated Elo-ratings for all individuals at the end of the observation period.

220 *Social network centrality*

221 We constructed an undirected proximity network based on recorded 1m (close) proximity. The
222 parasites of interest in this study are generally thought to be transmitted primarily through contact with
223 contaminated substrates. In the context of group dynamics, resource sharing and shared space use rather
224 than direct social contact might be implicated in the transmission of such parasites. We thus assumed that
225 a close-proximity network was representative of the probability of (in)direct transmission-relevant space or
226 resource sharing among individuals in this highly social species. Our previous work with Japanese
227 macaques, as well as various other studies of other species, support this assumption (MacIntosh et al., 2012;
228 VanderWaal et al., 2013; Friant et al., 2016; Müller Klein, et al., 2019).

229 We constructed a network based on aggregate data over the whole study period (3 months). It's
230 difficult to assess how representative of transmission processes this aggregate network might be. Data from
231 livestock suggest prepatent periods – i.e., the period between acquisition of a parasite and the point at which
232 it (if female) begins shedding eggs in host faeces – of between 2 and 6 weeks for the parasite genera studied
233 here (Anderson, 2000). Thus, networks based on shorter time periods may better match the time lag between
234 infection and detection. However, such networks, given sampling limitations, would not be robust
235 representations of connections between our study subjects. We recognise the limitations of using aggregate
236 behavioural data over arbitrary time scales to infer transmission processes, and so we instead emphasize
237 that our network data should be seen as reflecting generalized patterns of interaction that can be related to
238 observed infection phenotypes of group members – themselves as an aggregate of sampling over the same
239 study period. This is an approach that appears regularly in the literature (Rimbach et al., 2015;

240 Balasubramaniam et al., 2016; Tiddi et al., 2019), as transmission events themselves are far harder and in
241 some cases currently impossible to study effectively. Moreover, our focus is on the phenomenological link
242 between networks and infection and the impacts of sampling variability on such links, rather than on
243 examining the mechanics of why the link exists in the first place, so we believe our analyses are appropriate
244 for this endeavour.

245 We built a symmetrical matrix of undirected social proximities, in which dyadic proximity data were
246 scaled by the number of scans collected from each individual of the corresponding dyad. The proximity
247 matrix was then analysed with the package "igraph" (Csardi and Nepusz, 2006) in R Statistical Software
248 v.4.0.5 (R. Development Core Team, 2021). For each individual, we computed: (1) degree (how many
249 partners an individual had); (2) strength (how many times an individual was in proximity with others); and,
250 (3) eigenvector centrality (the sum of the centralities of an individual's neighbours; high centrality can be
251 achieved by having either a large degree or being connected to associates with a high degree (or both))
252 (Farine and Whitehead, 2015).

253

254 Parasite Collection and Analysis

255 We opportunistically collected between 1 and 4 fresh faecal samples from each individual (3.5±1.6)
256 during observations whenever a clearly identified individual was seen defecating. We put the entire faeces
257 into a sealable plastic bag and then into a cooler bag with ice until we brought all samples to the lab for
258 processing. Within 12h of collection, we weighed and homogenized about 1-2 grams of faeces into 3.5ml
259 of a sodium acetate-acetic acid-formaldehyde (SAF) solution for the preservation of parasite eggs and

larvae. After the fieldwork period, we sent all samples to Kyoto University's Primate Research Institute, where we conducted further analysis.

We processed all parasitological samples following previous work with Japanese macaque parasites using a modified formalin-ethyl acetate sedimentation protocol to concentrate helminth eggs from faeces (MacIntosh et al., 2010). We added 5 drops of Triton X-100 to the faecal mixture before filtration to better isolate eggs from other components of faeces. To filter samples, we used a 330μm Saran™ mesh (Asahi Kasei, Japan). To reduce the amount of formaldehyde consumed in the process, we used saline instead of formaldehyde for all intermediate washing steps during concentration. For all centrifugation steps, we spun our samples at 3000 rpm for 5 minutes.

After processing, we again suspended the retained faecal pellet in SAF and stored it at room temperature until analysis. For parasitological examination, we used a volumetric method based on McMaster microscopy to estimate the intensity of parasitic infection in each individual through counting the number of eggs observed per gram of faecal sediment (EPG) (Modrý et al., 2018). We drew aliquots from the faecal suspension, kept homogeneous during reading using a magnetic stirrer, and viewed them in a McMaster chamber with a 10x objective lens. Parasite eggs were identified by their specific morphology and size (Modrý et al., 2018). We repeated this procedure five times per sample and used the mean count to calculate EPG from each sample based on the amount of sediment viewed per volume of suspension, with the following formula: $EPG = \frac{\frac{total\ number\ of\ eggs}{number\ of\ chambers}}{\frac{W_{sediment}}{V_{solution}} \times V_{chamber}}$

In the formula, *total number of eggs* reflects the number of eggs the examiner counted in all chambers. The *number of chambers* refers to the number of chambers examined in the sample (here, 5). $W_{sediment}$ refers

to the weight of the faecal pellet after filtration, and $V_{solution}$ refers to the volume of solution that was used to dilute the sample (in this case, between 10ml and 50ml depending on the thickness of the faecal suspension, which significantly affects viewing). $V_{chamber}$ refers to the volume of suspension contained in each counting chamber (0.15ml for a single McMaster chamber). EPG was then rounded to the nearest integer.

EPG is a standard index of parasite infection intensity. Although it is not always indicative of true worm infection intensity (Gillespie, 2006; Modrý et al., 2018), numerous studies have found significant correlations between EPG and the number of reproductive female worms infecting a given host (De Bont et al., 2002; Mohammed et al., 2016). The method is commonly used to measure domestic animal health and in wildlife parasitology, particularly in cases where parasite population estimation is needed but it is neither advisable nor possible to sample individuals via other - more invasive or destructive - means.

Statistical Analyses

All statistical analyses were conducted using R v4.0.5 (R. Development Core Team, 2021). We built a dataset that included one line for each parasite species in each faecal sample of each individual as the unit of analysis, i.e., three times (for the three parasite species) the number of faecal samples analysed. In addition, each sample was associated with an individual and all of its attributes, including its centrality measures (degree, strength, and eigenvector), Elo-rating, sex, and age, along with the sample's collection date. EPG (rounded to the nearest integer) was used as the response variable in the statistical models described below. We also scaled and centred to 0 the centrality measures, age and Elo-rating.

300 *Testing for associations between network centrality and parasite infection intensity*

301 Considering that all parasites of interest are transmitted via contact with contaminated substrates, we
302 decided to include EPG from different parasite species in one model as the response variable, accounting
303 for variation among parasites with a control factor "parasite species". In doing so, we accept the limitation
304 that we are not testing for specific associations related to each parasite type, assuming instead that similar
305 mechanisms may drive links between social behaviour and infection, should such links exist. We set
306 parasite species identity as a fixed effect in the model to at least account for variation in EPG by species.
307 We modelled rounded EPG as overdispersed count data. Parasites typically exhibit an aggregated
308 distribution across their hosts that is best approximated by a negative binomial distribution (Crofton, 1971;
309 Shaw et al., 1998; Poulin, 2007). We confirmed that the EPG distribution indeed followed a negative
310 binomial distribution using the "*fitdistr"* function from the R package "MASS" (Venables and Ripley,
311 2002). We then built generalized linear mixed models with a negative binomial error structure using the
312 "glmmTMB" package (Brooks et al., 2017).

313 The three centrality measures we investigated were related to each other (Spearman's Correlation of
314 degree and strength: $r = 0.42$, $p = 0.002$; degree and eigenvector: $r = 0.47$, $p < 0.001$; strength and
315 eigenvector: $r = 0.72$, $p < 0.001$). Moreover, eigenvector centrality itself is derivative of degree and strength.
316 Thus, including these metrics in the same model would violate the assumption of data independence, and
317 the covariance between them could make the interpretability of coefficients challenging. Therefore, we
318 constructed one model for each centrality measure. Age, sex, and Elo-ratings were set as control fixed
319 effects as they are known to affect EPG but are not factors of interest here. EPG can also vary considerably

320 over time and sample to sample within an individual (Wood et al., 2013). To account for such effects, we
321 collected multiple faecal samples per individual and set the individual ID and date of sample collection as
322 random effects in the models.

323 The model residuals showed substantial heteroscedasticity and overdispersion, so we added a zero-
324 inflation term to account for the substantial number of zeros in the EPG data (124 instances or 24% of the
325 517 total data points). The zero-inflated models showed better residual distributions and significantly lower
326 AIC compared with standard negative binomial models (degree: 7268.8<7727.0; strength: 7254.5<7718.0;
327 eigenvector centrality: 7260.7<7722.1). We also assessed whether other assumptions of the models were
328 met, such as lack of multicollinearity (variance inflation factor < 3) and overly influential cases (Cook's
329 distance < 4/N). Lastly, we conducted likelihood ratio tests (LRT) to evaluate the statistical strength of our
330 models against an informed null model containing only the control factors age, sex, Elo-rating, parasite
331 species, and random effects.

332 Note that we did not carry out any permutation or randomisation procedure typically associated with
333 non-independent data such as social network centrality measures to account for autocorrelation induced by
334 the nature of the data (Hart et al., 2021; Weiss et al., 2021; Farine and Carter, 2022). In this study, centrality
335 metrics were treated as predictor variables, not as response variables. The statistical issue of non-
336 independence arises when network measures are the response variable because it is important that residuals
337 are independent, which relates to links between the response and the predictor variable(s). In this study, we
338 include network centrality metrics as predictors like other individual traits such as rank or age (which can

also be considered inter-dependent; yet, we do not habitually control for it). Here, mixed models should be able to account for dependencies in the data if they are well specified, as we believe they are.

Testing the effect of knockout simulations on the relationship between centrality and parasite infection intensity

To have a point of comparison with previous studies and to gain insight into the influence of only examining certain age-sex classes on the relationship between centrality and infection intensity, we ran two more analyses. We first constructed partial social networks including only (1) adult females (20, or 38.5% of the group), which mimics what was done in previous studies, or (2) juveniles (28, or 53.8% of the group). We calculated Spearman's correlation coefficients to compare between the network metrics from original and partial networks. Then, we built models with the same structure as those described above but with network metrics calculated from partial networks. These models went through all the checks as previously described.

Removing individuals from the network itself might already affect social network structure and thus change the observed result. In a second step, therefore, we conducted randomized knock-out simulations: we randomly removed 5%, 10%, 25%, and 50% of individuals from the network, re-calculated the network metrics, and re-integrated them into the models for calculation over 1000 iterations. After each simulation, the model parameters of interest (here estimates and confidence intervals) were saved and the results were compared with the results from the original models. We chose those percentages because they represent biologically relevant cases of missing individuals during data collection or within a study design, like

359 solitary roaming males or peripheral females (e.g., <10% of the group), or when taking into account only
360 certain age-sex classes, here adult females or juveniles, effectively removing everyone else (e.g., >25% of
361 the group, depending on group composition and demography).

362 Knock-out simulations not only disrupt the network structure and change the centrality metrics
363 calculated, but they also reduce the number of individuals included in the model, reducing sample size and
364 possibly statistical power. We therefore further analysed the effect of reducing the number of individuals
365 in the networks using the following approach. We first sub-sampled the complete network as before (i.e.
366 adult female-only data or juvenile-only data), but instead of re-constructing the network and re-calculating
367 metrics based on this reduced set of individuals, we ran the statistical models using the centrality metrics
368 for those individuals from the original whole-group network. We then re-ran the same randomized knock-
369 out procedure, but this time retaining the centrality metrics from the original whole-group network. In this
370 way, we are able to determine to what extent any differences between the subsampled networks and the
371 original network resulted from real changes in network structure versus issues caused by reduced statistical
372 power (i.e. smaller sample size).

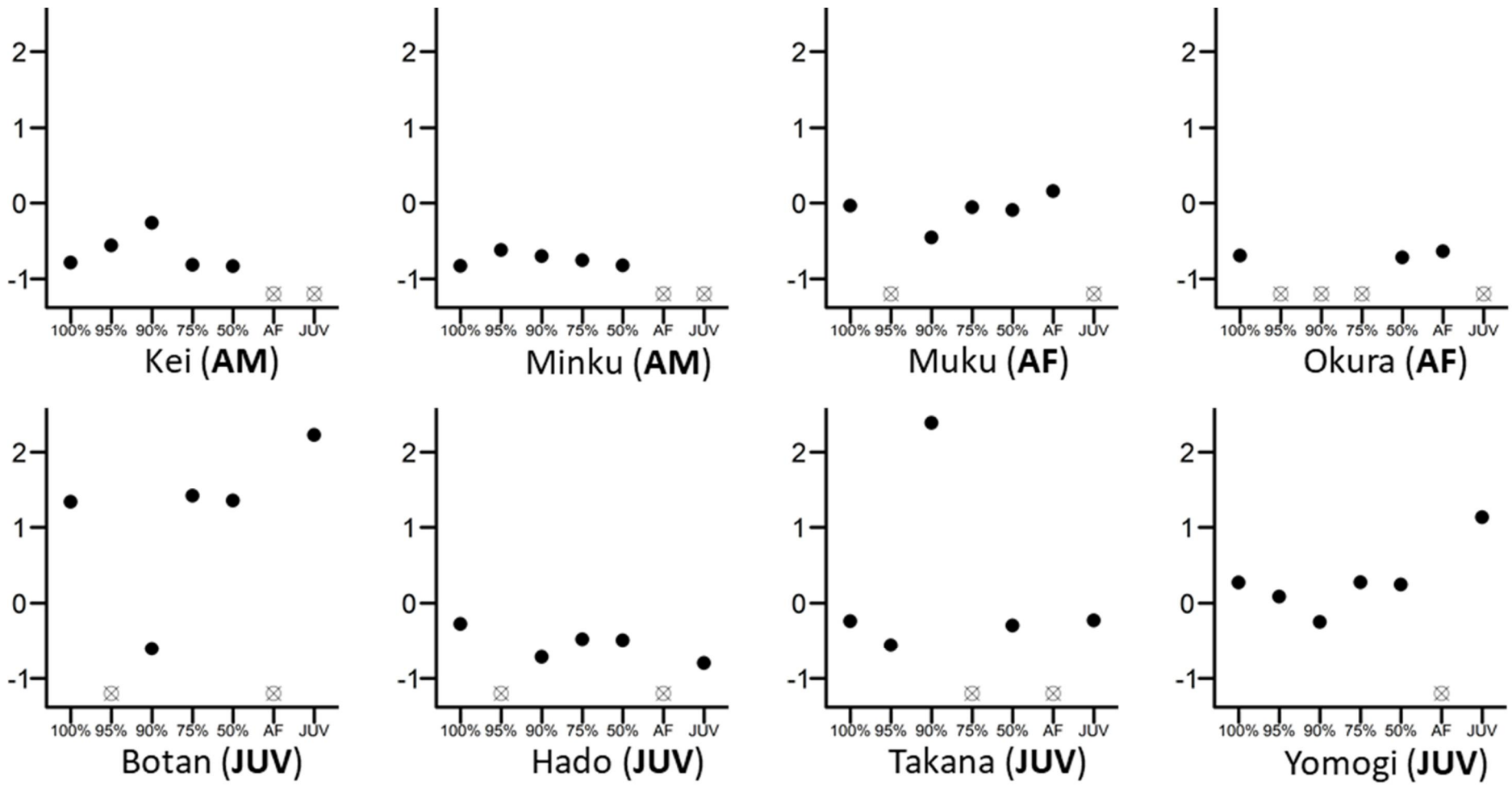

Figure 1: Differences in the eigenvector centralities of eight randomly chosen individuals from the whole group. The 100% mark on the x-axis reflects the whole-group network, while 95%, 90%, 75%, 50% are values retained from networks where 5%, 10%, 25% or 50% of individuals were randomly removed. Individual figures are marked as AM (adult male), AF (adult female) or JUV (juvenile). All eigenvector centrality values were scaled and centred to 0 within their own network for direct comparison. Empty crossed dots mean that the individual was either removed randomly from that network or that the individual does not belong in this particular network; for example, the individual is an adult female and cannot figure into the juvenile-only network (e.g. Okura (AF)).

Results

Social network metrics

The whole-group network includes 52 nodes with 480 connections between nodes, and is based on 14,420 proximity interactions. In the whole-group network, individuals on average had 14 (± 5) social partners (degree centrality), an average proximity strength centrality of 0.44 ± 0.21 and an average eigenvector centrality of 0.32 ± 0.29 (Table 2). The adult female-only network included 20 nodes with 87 connections between nodes, and was based on 2,403 proximity interactions. The juvenile-only network included 24 nodes with 139 connections, and was based on 2,316 proximity interactions (see Table 1 for details).

Table 1. Summary of network metrics from whole-group or partial networks in this study.

	Whole-group (mean±SD)	Adult females (whole-group network) (mean±SD)	Juveniles (whole-group network) (mean±SD)	Adult females (adult female-only network) (mean±SD)	Juveniles (juvenile-only network) (mean±SD)
Degree	13.80±5.48	13.75±4.22	15.33±4.93	6.20±2.76	8.83±2.79
Strength	0.44±0.21	0.55±0.20	0.43±0.16	0.21±0.11	0.15±0.09
Eigenvector	0.32±0.29	0.39±0.32	0.33±0.29	0.33±0.32	0.25±0.29

Parasitological results

We observed all three of the expected geohelminths infecting our subjects during the study period: *Oesophagostomum aculeatum*, *Strongyloides fuelleborni*, *Trichuris trichiura*. Most individuals were infected by all parasites (see Table 2 for details).

Table 2. Summary of prevalence and infection intensity (EPG) of three geohelminths examined in this study.

		O. aculeatum	*T.trichiura*	*S.fuelleborni.*
Whole-group data	Prevalence	100%	90%	90%
	Average EPG	1380.72±863.28	775.09±989.16	3080.10±3776.99
Adult females-only	Prevalence	100%	90%	85%
	Average EPG	1172.38±578.25	345.21±383.47	682.33±918.725
Juveniles-only	Prevalence	100%	96%	100%
	Average EPG	1736.90±975.41	1282.26±1195.90	5827.03±3811.14

Network-infection relationship

Whole-group networks

Likelihood ratio tests demonstrated that the models with strength and eigenvector centrality as fixed effects, but not with degree centrality, had significantly higher explanatory value than their respective informed null models (Table 3). Strength and eigenvector centralities were positively related to infection intensity, such that more central individuals had higher geohelminth EPG (Table 3). In addition, age was negatively related to infection, with younger individuals showing higher EPG. Sex and dominance rank were not related to infection intensity (Figure 2).

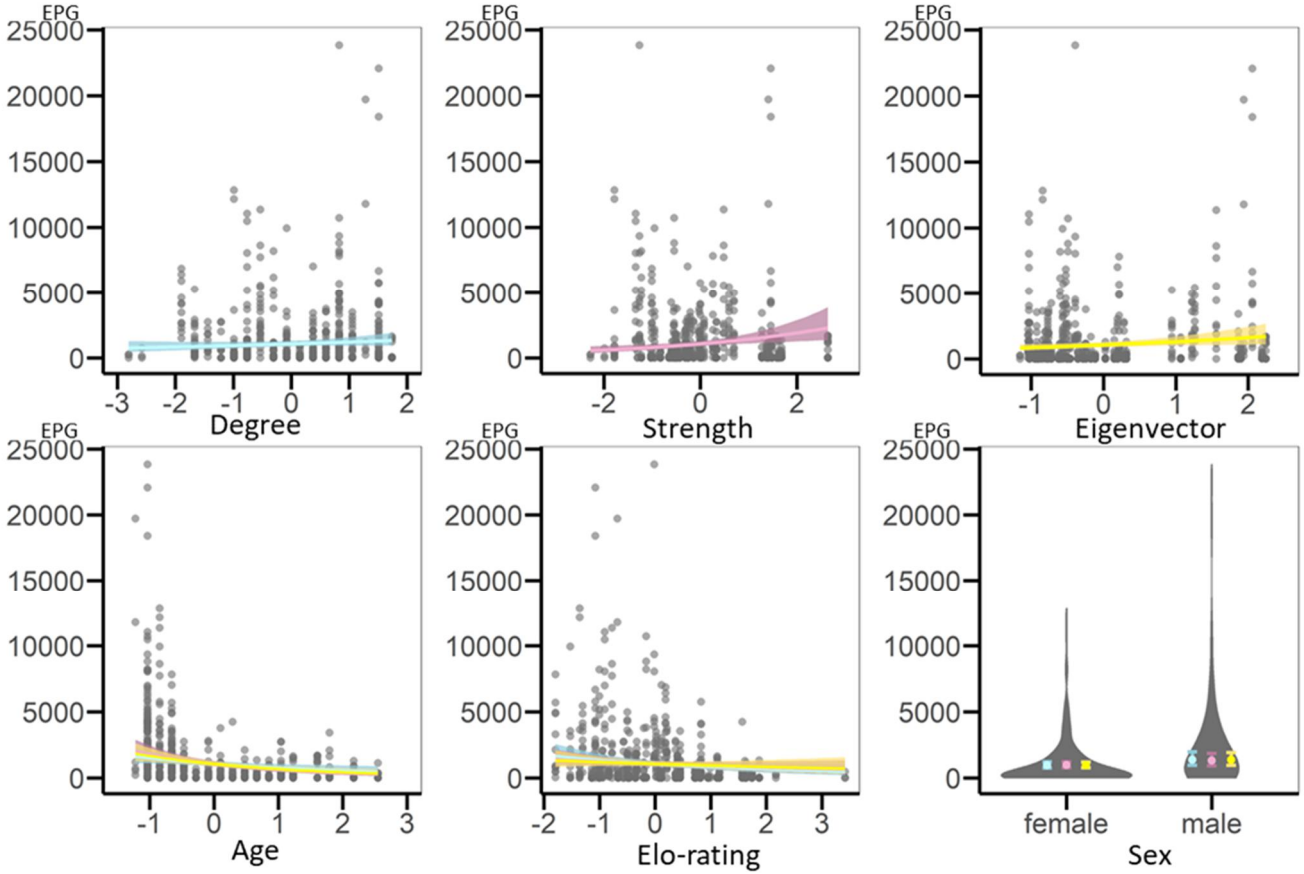

Figure 2: Regression plots representing results of GLMMs testing for variation in geohelminth infection intensity (EPG) based on the whole-group network. On the x-axis are the scaled variables, on the y-axis is infection intensity (EPG). Each circle is a data point. Each network metric is included in its own model and is represented by its own colour, blue for the model with degree, pink for the model with strength and yellow for the model with eigenvector centrality. The regression plots representing shared fixed effects (age and Elo-rating) show lines from each of the models in the respective different colours. Note that results are very similar and thus lines overlap. The plot for sex is a violin plot representing the spread of the data (violins) and the estimates with standard errors (filled circles and error bars) shown in each model's respective colour.

Table 3. Results of GLMMs testing for variation in geohelminth infection intensity (EPG) among a whole group of Japanese macaques on Koshima Island. LRT refers to likelihood ratio test, here shown with Chi-square χ^2 tests, degrees of freedom (d.f.) and p-values (p). Estimates (β) are given with standard errors (SE) and p-values (p). Variance is given with standard deviation (SD). Results in bold are statistically significant ($p < 0.05$). For clarity, results from models that failed to outperform their null counterparts are presented in full in the table SI-4.

Centrality metric of focus	*Degree*	*Strength*	*Eigenvector*
LRT against null (χ^2; d.f.; p)	1.940; 2; 0.379	16.491; 2; <0.001	10.238; 2; 0.006
Fixed effects		*Estimate β* ± SE [p]	
Intercept	/	6.88 ± 0.12 [<0.001]	6.89 ± 0.13 [<0.001]
Centrality metric	/	**0.29 ± 0.09 [0.003]**	**0.21 ± 0.09 [0.025]**
Age	/	**-0.53 ± 0.13 [<0.001]**	**-0.45 ± 0.13 [<0.001]**
Sex (male)*	/	0.34 ± 0.21 [0.110]	0.29 ± 0.22 [0.199]
Elo-rating	/	-0.16 ± 0.11 [0.152]	-0.13 ± 0.12 [0.286]
Species (Stro)**	/	**0.65 ± 0.13 [<0.001]**	**0.65 ± 0.13 [<0.001]**
Species (Tri)***	/	**-0.67 ± 0.12 [<0.001]**	**-0.67 ± 0.12 [<0.001]**
Random effects		*Variance*	
Collection Date	/	0.02	0.02
Individual ID	/	0.24	0.26

** Sex (male) shows model coefficients for males compared with the reference value, females*
*** Species (Stro) shows model coefficients for the parasite species S. fuelleborni compared with the reference value, O. aculeatum*
**** Species (Tri) shows model coefficients for the parasite species T. trichiura compared with the reference value, O. aculeatum*

Female- and juvenile-only networks

When comparing measures calculated from partial versus whole-group networks, in both adult female-only and juvenile-only networks, most measures derived from partial-networks were strongly to moderately correlated with the original whole-group network measures. Only strength derived from the adult female-only network was not correlated with strength derived from the original whole-group network (Spearman's rank correlation; adult females: $r_{degree} = 0.71$, $p < 0.001$, $r_{strength} = 0.07$, $p = 0.772$,

435 $r_{eigenvector} = 0.76$, $p < 0.001$; juveniles: $r_{degree} = 0.87$, $p < 0.001$, $r_{strength} = 0.80$, $p < 0.001$, $r_{eigenvector} = 0.69$,
436 $p < 0.001$).

437 Using data from adult females only, likelihood ratio tests demonstrated that the models with
438 strength, but not those with eigenvector centrality or degree centrality as main predictors, had higher
439 explanatory value than their respective informed null models (Table 4). Neither age nor rank was related
440 to nematode infection intensity (Table 4, Figure SI-2). Using data from juveniles only, likelihood ratio
441 tests demonstrated that none of the models had a higher explanatory value than their respective informed
442 null models (Table 5, Figure SI-3). The same models after data reduction (i.e. with centrality measures
443 derived from the whole-group network) showed similar results (in terms of estimate values in the model)
444 than those after network reduction (i.e. based on recalculated measures from sub-sampled networks),
445 except for juvenile-only networks, which gave much lower model estimates overall (see Table SI-7,
446 Table SI-8). These results suggest that the lack of association between centrality and infection in network
447 subsets may arise from a combination of a change in the network itself and a loss of statistical power
448 with decreasing sample size. In the case of adult females, it seemed that regardless of how the network
449 was built (i.e. with more or fewer interactions/connections), results stayed more or less the same, but in
450 the case of juveniles, model estimates were much lower with more interactions/connections included.
451

Table 4. Results of GLMMs testing for variation in geohelminth infection intensity (EPG) among adult female Japanese macaques on Koshima Island. LRT refers to the likelihood ratio test, here shown with Chi-square χ^2 tests, degrees of freedom (d.f.) and p-values (p). Estimates (β) are given with standard errors (SE) and p-values (p). Results in bold are statistically significant ($p < 0.05$). For clarity, results from models that failed to outperform their null counterparts are presented in full in table SI-5).

Centrality metric of focus	*Degree*	*Strength*	*Eigenvector*
LRT against null (χ^2; d.f.; p)	3.616; 2; 0.164	**7.424; 2; 0.024**	4.214; 2; 0.122
Fixed effects		*Estimate β ± SE [p]*	
Intercept	/	6.76 ± 0.13 [<0.001]	/
Centrality metric	/	**0.32 ± 0.12 [0.006]**	/
Age	/	-0.09 ± 0.13 [0.477]	/
Elo-rating	/	-0.06 ± 0.13 [0.622]	/
Species (Stro)*	/	**-1.00 ± 0.29 [<0.001]**	/
Species (Tri)**	/	**-1.16 ± 0.18 [<0.001]**	/
Random effects		*Variance*	
Collection Date	/	0.00	/
Individual ID	/	0.15	/

** Species (Stro) shows model coefficients for the parasite species S. fuelleborni compared with the reference value, O. aculeatum*
*** Species (Tri) shows model coefficients for the parasite species T. trichiura compared with the reference value, O. aculeatum*

Table 5. Results of GLMMs testing for variation in geohelminth infection intensity (EPG) among juvenile Japanese macaques on Koshima Island. LRT refers to the likelihood ratio test, here shown with Chi-square χ^2 tests, degrees of freedom (d.f.) and p-values (p). For clarity, results from models that failed to outperform their null counterparts are presented in full in table SI-6).

Centrality metric of focus	*Degree*	*Strength*	*Eigenvector*
LRT against null (χ^2; d.f.; p)	0.120; 2; 0.942	2.491; 2; 0.288	1.643; 2; 0.440

Random partial networks

Random removals of individuals comprising 5, 10, 25, and 50% of the group, following which networks were reconstructed with remaining individuals, all led rapidly to a decrease in the strength of

472 the observed relationship between infection and centrality (except for degree, in which case no
473 relationships were observed in both original network and partial networks) (Table 6). In other words,
474 estimates describing that observed relationship approached zero in these models and did not reach
475 statistical significance (Figure SI-9). Moreover, these results do not appear to be caused by a loss of
476 statistical power, because the majority of results from models in which knockouts were conducted but
477 centrality measures were retained from the whole-group network were much closer to the original results
478 obtained with all individuals present (Table SI-10, Figure SI-11).

479

480 Table 6: Results of 1000 random removals of individuals (% given in the first column) run with observed
481 network data and used to (re)model the relationship between network centrality and infection intensity.
482 The heading "Number of model results equivalent to observed" indicates the number of simulation
483 model results out of 1000 simulations that gave an equivalent result to models based on the whole-group
484 network, i.e. for strength and eigenvecetor, positive estimates and 95% confidence intervals not
485 overlapping zero, or for degree, a near-zero estimate with 95% confidence intervals overlapping zero.
486

Centrality measure	Percentage of individuals removed	Number of model results equivalent to observed
Degree	5%	98.0%
	10%	96.6%
	25%	95.4%
	50%	95.9%
Strength	5%	3.2%
	10%	4.2%
	25%	3.9%
	50%	4.7%
Eigenvector	5%	12.6%
	10%	7.1%
	25%	3.8%
	50%	5.3%

487

488 **Discussion**

489 Empirically and theoretically, increased sociality is oftentimes linked to increased parasitism,
490 although this relationship is not always linear (Freeland, 1976; Altizer et al., 2003; Griffin and Nunn,
491 2012; Godfrey, 2013; Nunn et al., 2015, 2015; Romano et al., 2016; Rushmore et al., 2017; White et al.,
492 2017; Briard and Ezenwa, 2021; Lucatelli et al., 2021). Our study found positive correlations between
493 geohelminth infection intensity and two out of three social network centrality measures examined in a

494 Japanese macaque group. Specifically, the frequency with which individuals were in close proximity to
495 one another (network strength), as well as their further social integration (eigenvector centrality), but
496 not the number of individuals with which they shared proximity (degree centrality), were positively
497 associated with infection intensity of geohelminths. These results partially align with the results of
498 previous work in a different population of Japanese macaques (MacIntosh et al., 2012), with both
499 investigations suggesting that eigenvector centrality, whether in grooming or proximity networks, is
500 associated with nematode parasitism in Japanese macaques.

501 Our results also align with a recent meta-analysis of studies investigating parasites in host social
502 networks, wherein strength and eigenvector centrality showed stronger associations with parasitism than
503 did any of betweenness, closeness, or degree centrality (Briard and Ezenwa, 2021). It therefore seems
504 important to consider what these various centrality measures represent in terms of host socio-biology
505 and in terms of social or ecological processes. Exposure and transmission risk of fast and directly spread
506 parasites such as lice or certain viruses and bacteria might depend simply on the number of associations
507 individuals have (degree), whereas slower-spreading/developing parasites or those requiring additional
508 developmental stages in the environment may be more dependent on the actual time hosts spend in
509 contact or in spatial proximity with one another (strength), or the overall degree of social integration
510 they have (eigenvector centrality) (Silk et al., 2017; Briard and Ezenwa, 2021; Lucatelli et al., 2021).
511 By whatever mechanism, our results support those of review studies and meta-analyses showing that,
512 across many different sociality measures and parasites with different life histories and modes of
513 transmission, the relationship between sociality and parasitism is generally positive (Altizer et al., 2003;
514 Rushmore et al., 2017; Briard and Ezenwa, 2021; Lucatelli et al., 2021).

515 However, we also demonstrated that the choice of animals within a study group (i.e., sampling bias)
516 can impact the relationship that is observed between network centrality and parasitism, a fact that is
517 often neglected or ignored in wildlife studies. Indeed, numerous studies have focused on specific age-
518 sex classes rather than the whole social group (MacIntosh et al., 2012; Duboscq et al., 2016; Romano et
519 al., 2016; Webber et al., 2016; Hamilton et al., 2020; Roberts and Roberts, 2020; Sandel et al., 2021).
520 In our study, considering partial networks – e.g., those including only a subset of the group like adult
521 females or juveniles – yielded divergent results compared to those analyses that included the whole-

522 group network. Statistical relationships originally observed between network centrality and infection
523 intensity weakened considerably, except for network strength in the adult female model, although the
524 direction of the effect remained positive in all cases for both strength and eigenvector centrality. Thus,
525 sampling bias can lead to misdirected conclusions about the relationships between sociality and
526 parasitism, and potentially any other network process for that matter.

527 From an epidemiological perspective, every individual can be responsible for the transmission of a
528 pathogen, so neglecting them in the analysis may reduce the reliability of the data. In our study, as in
529 previous work with Japanese macaques (MacIntosh et al., 2010), juveniles exhibited higher geohelminth
530 infection intensities than adults. Furthermore, within the juvenile class, younger individuals and those
531 that were male shed more parasite eggs than did older individuals or those that were female (Table SI-
532 5). These results concur with the general findings that young animals and males are disproportionately
533 infected by parasites (Khan et al., 2010; Hinney et al., 2011; Habig and Archie, 2015). Hypotheses
534 explaining this difference include age/sex-associated hormonal profiles (Roberts et al., 2001), variable
535 levels of immunity (Roberts et al., 2001; Fallon et al., 2003; East et al., 2008; Ebersole et al., 2008),
536 different probabilities of exposure to contaminated items (Sarabian, Belais, et al., 2018; Sarabian, Curtis,
537 et al., 2018), and differing social positions within the group (Grassi, 2002; Cords et al., 2010; Liao et
538 al., 2018). Considering such age and sex biases in infection, along with the variable activity patterns (in
539 terms of social interactions and ranging behaviour) performed by juveniles and males, neglecting them
540 as study subjects may have important consequences for the conclusions reached about links between
541 sociality and infection. Moreover, because individuals form connected nodes in networks and are thus
542 not independent of one another, excluding them from the network fundamentally changes its overall
543 structure (Fedurek and Lehmann, 2017). Indeed, missing nodes in general dramatically change network
544 structure and affect the metrics calculated (Kossinets, 2006; Smith and Moody, 2013; Silk et al., 2015;
545 Smith et al., 2017; Silk, 2018). Such effects may percolate through the processes linked to network
546 structure, such as the transmission of information or pathogens (Godfrey et al., 2010).

547 That said, in our study at least, most of the network metrics calculated from both the adult female-
548 only network and the juvenile-only network were rather well correlated with those calculated from the
549 whole-group network. Among these correlations, degree and strength from the juvenile-only network

were closer to those from the whole-group network than were the same measures from the adult female-only network. And, interestingly, strength from the adult female network was not correlated with the whole-group metric. In contrast, eigenvector centralities from the adult female-only network were closer to those in the whole-group network than were eigenvector centralities from the juvenile-only network. A possibility would be that each juvenile is likely to interact with a small number of adults, i.e., adult female kin, and most of their contacts are between juveniles. Therefore, the inclusion or exclusion of adult individuals in a juvenile network may have comparatively less influence on their degree or strength centrality than the reverse. In contrast, each adult female may have multiple offspring, and this number can vary across individuals, with the potential to significantly change the number of contacts a female has, and the frequency with which they interact. With regards to eigenvector centrality, this measure represents the integration of an individual within its whole network of connections. Since females are known to form the core of Japanese macaque groups (Yamagiwa and Hill, 1998), it may be that their patterns of interaction dominate the emergent social network structure, such that properties of female networks linked to social integration (as in eigenvector centrality) do not change as much with the inclusion/exclusion of other subgroups.

In stark contrast to these targeted knockouts, our random knockouts quickly erased any relationship found in the original network between infection and centrality, for all metrics observed. This was evidenced both by the complete absence of statistical significance in the knockout models, which was rapidly lost even at just 5% random reduction in individuals in the network, and by the magnitude of the model estimates, which quickly approached zero at just 5% random network reduction. Random removal of individuals in a network then, much more so than targeted removal of subsets, might cause greater disruption in network structure, at least as far as it concerns processes occurring on networks, like infection. It has been shown that targeted removal of central individuals from a network has greater effects on resultant network measures than random removal (Smith and Moody, 2013; Smith et al., 2017). We did not test the effect of removing specific individuals in our study, so cannot draw comparison here, but that is a topic worthy of future consideration. Furthermore, it has been shown that, as the proportion of individuals excluded from a network increases, the probability that one or more central nodes are excluded also increases, leading to stronger effects on the resultant metrics (Silk et al.,

578 2015; Smith et al., 2017; Silk, 2018). However, since our random knockouts resulted in a complete loss
579 of the infection-centrality link at just 5% removed, this is unlikely to be a key factor explaining our
580 results.

581 Now, in real world cases, missing individuals from a network is rarely a random process, especially
582 if significant study or observation effort is made over extended time periods. Instead, there are usually
583 systematic reasons for missing certain individuals, for example those that are difficult to observe, such
584 as roaming adult males or young, unrecognisable juveniles. This is one reason why we focused on
585 removing specific sex/age classes from our data set. It is notable that, in our study, complete age/sex
586 subsets preserved the infection-centrality link far better than the random knockouts did. This might
587 suggest that, while the full network is always preferred, including complete subsets might retain much
588 more of the information present in the original network than would taking a random sample of
589 individuals. It will be interesting to examine these effects in additional species of hosts and parasites,
590 especially those with different network structures or modes of transmission, to further disentangle the
591 impact of social dynamics and sampling on network processes like parasite transmission. Ultimately,
592 whether individuals are omitted intentionally or accidentally, our results highlight the potential impacts
593 of partial networks on the conclusions drawn.

594 One final point to consider is the potential reasons behind the different results obtained in the
595 knock-out simulations: the act of removing nodes itself, i.e., a data reduction issue leading to statistical
596 power issues, or the actual disruption in the network structure caused by removing nodes and their
597 perceived relationships. We posited that if the model centrality estimates with data reduction (without
598 network metric recalculation) were of the same magnitude as the model centrality estimates with
599 network restructuring (and recalculation), then results can be seen as independent of the way the
600 measures are calculated. If so then, results have little to do with missing interactions which might be
601 rather indicative of a power issue (less data points). If we look at the change in centrality estimates from
602 all models regardless of whether they outperformed their null counterparts (Table SI-5 & Table SI-7;
603 Table SI-6 & Table SI-8), we obtained contrasting outcomes: in adult females, results are unchanged
604 for degree and strength but for eigenvector centrality, model estimates with data reduction are lower
605 than estimates with network reduction. In juveniles, model estimates with data reduction are several-

606 fold lower than estimates with network reduction. Therefore, we cannot distinguish whether the
607 divergent results were caused by disruption of the network or because removal of specific set of data
608 points reduced statistical power of the model when including only specific age-sex classes. If we look
609 at the change in centrality estimates in the random knockouts, however, the results show clear
610 differences: estimates from simulations with network restructuring centered on zero, while estimates
611 from simulations with data reduction centered on original (expected) results (Figure SI-9, Figure SI-11).
612 Thus, disruption of the network caused by missing nodes is the main reason why divergent results were
613 obtained after random removal of individuals.

614 Despite our results and the conclusions drawn from them, we do not intend to devalue the results
615 of previous studies solely on these grounds. Partial networks can still provide robust network metrics,
616 providing that there is little imbalance as to who gets excluded (Borgatti et al., 2006; Silk et al., 2015).
617 In our study, we focused only on excluding broad categories of individuals or used pseudo-random
618 knockouts, and we did not control for the specific social characteristics of the removed individuals. It is
619 further possible that excluding particular individuals – super-spreaders, keystone individuals, policing
620 individuals, alpha individuals – would have a much larger effect on the processes that occur on social
621 networks than would excluding individuals with a less influential role. For instance, knocking out
622 policers in a social group of pigtailed macaques greatly destabilized the structure of several social
623 networks, setting the stage for reduced group cohesion (Flack et al., 2006). This is also reminiscent of
624 work investigating the effect of targeting specific individuals for health interventions or information
625 transmission: depending on which individual is targeted, information has a greater or lesser likelihood
626 of spreading in the community (e.g. Kim et al., 2015). Similar work has looked at the spread of an
627 infectious disease or a parasite depending on which individual gets it first, and what position they occupy
628 in the social network (e.g. Romano et al., 2016). So, knowing the characteristics of individuals that play
629 important roles in the sociality-parasitism relationship can help practitioners develop strategies, for
630 example, for disease control (Rushmore et al., 2014; Romano et al., 2016) and captive population
631 management (McCowan et al., 2008). The key, however, is to ensure that the assumptions made are the
632 result of careful consideration of the completeness and thus representativeness of the data at hand.
633

Acknowledgements

We are grateful to the Cooperative Research Program of Kyoto University's Wildlife Research Center and the city of Kushima's Agency for Cultural Affairs for permitting this research. We thank Christof Neumann for statistical and programming support, accepting that all analytical decisions – and outstanding errors – are ours alone. We thank Dr. Zhang Peng for supporting this research and collaboration through Sun Yat-Sen University.

Funding

At the time of the study, ZX was supported by a scholarship from Sumitomo Corporation. AJJM was supported by a Grant-in-Aid from the Japan Society for the Promotion of Science (JSPS: 16H06181). AC-N and EM-M were additionally supported by the German Research Foundation (AM 409/4-1 to Federica Amici, who we also thank) and the Universidad Cardenal Herrera-CEU (INDI 15/12, CEINDO 16/17). JD was supported by a postdoctoral fellowship from the JSPS.

Conflict of interest disclosure

The authors declare they have no conflict of interest relating to the content of this article.

Author contributions

ZX: Conceptualization; Data curation; Formal analysis; Writing – original draft.

AJJM: Conceptualization; Methodology; Project administration; Supervision; Writing – review and editing.

AC-N: Methodology; Investigation.

EM-M: Methodology; Investigation.

TS: Resources.

JD: Methodology; Supervision; Formal analysis; Writing – review and editing.

Data, script and code availability

The dataset (SP 1) and scripts are available on Git repository:

662 “https://github.com/XuZhihong-cn/Network_Centrality_analysis_and_KO-simulation”

663 The Supplementary information is available on Biorxiv:

664 “https://www.biorxiv.org/content/10.1101/2021.06.07.447302v1.supplementary-material”

665

666

667 **References**

668 Albers, P. C. H., and H. de Vries, 2001, Elo-rating as a tool in the sequential estimation of dominance

669 strengths: 489–495 p., doi:10.1006/anbe.2000.1571.

670 Altizer, S. et al., 2003, Social organization and parasite risk in mammals: Integrating theory and empirical

671 studies: 517–547 p., doi:10.1146/annurev.ecolsys.34.030102.151725.

672 Anderson, R. C., 2000, Nematode Parasites of Vertebrates: Their Development and Transmission: CABI Pub.

673 Arneberg, P., 2002, Host population density and body mass as determinants of species richness in parasite

674 communities: comparative analyses of directly transmitted nematodes of mammals: 88–94 p.

675 Arneberg, P., A. Skorping, B. Grenfell, and A. F. Read, 1998, Host densities as determinants of abundance in

676 parasite communities: 1283–1289 p., doi:10.1098/rspb.1998.0431 %/ The Royal Society.

677 Balasubramaniam, K., B. Beisner, J. Vandeleest, E. Atwill, and B. McCowan, 2016, Social buffering and

678 contact transmission: Network connections have beneficial and detrimental effects on Shigella infection

679 risk among captive rhesus macaques: doi:10.7717/peerj.2630.

680 Borgatti, S. P., 2005, Centrality and network flow: 55–71 p., doi:10.1016/j.socnet.2004.11.008.

681 Borgatti, S. P., K. M. Carley, and D. Krackhardt, 2006, On the robustness of centrality measures under

682 conditions of imperfect data: 124–136 p., doi:10.1016/j.socnet.2005.05.001.

683 Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca, 2009, Network Analysis in the Social Sciences: 892–

684 895 p., doi:10.1126/science.1165821.

685 Briard, L., and V. O. Ezenwa, 2021, Parasitism and host social behaviour: a meta-analysis of insights derived

686 from social network analysis: 171–182 p., doi:https://doi.org/10.1016/j.anbehav.2020.11.010.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Machler, and B. M. Bolker, 2017, glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling: 378–400 p.

Cain, D. W., and J. A. Cidlowski, 2017, Immune regulation by glucocorticoids: 233–247 p., doi:10.1038/nri.2017.1.

Cords, M., M. J. Sheehan, and L. S. Ekernas, 2010, Sex and Age Differences in Juvenile Social Priorities in Female Philopatric, Nondespotic Blue Monkeys: 193–205 p., doi:10.1002/ajp.20768.

Crofton, H. D., 1971, QUANTITATIVE APPROACH TO PARASITISM: 179- p., doi:10.1017/s0031182000071420.

Csardi, G., and T. Nepusz, 2006, The igraph software package for complex network research: 1695 %\ 2018-03–01 16:58:00 p.

Davis, G. H., M. C. Crofoot, and D. R. Farine, 2018, Estimating the robustness and uncertainty of animal social networks using different observational methods: 29–44 p., doi:10.1016/j.anbehav.2018.04.012.

De Bont, J., D. J. Shaw, and J. Vercruysse, 2002, The relationship between faecal egg counts, worm burden and tissue egg counts in early Schistosoma mattheei infections in cattle: Acta Tropica, v. 81, no. 1, p. 63–76, doi:10.1016/s0001-706x(01)00198-x.

Duboscq, J., V. Romano, C. Sueur, and A. J. J. MacIntosh, 2016, Network centrality and seasonality interact to predict lice load in a social primate: 13 p., doi:10.1038/srep22095.

East, M. L., G. Wibbelt, D. Lieckfeldt, A. Ludwig, K. Goller, K. Wilhelm, G. Schares, D. Thierer, and H. Hofer, 2008, A HEPATOZOON SPECIES GENETICALLY DISTINCT FROM H. CANIS INFECTING SPOTTED HYENAS IN THE SERENGETI ECOSYSTEM, TANZANIA: 45–52 p., doi:10.7589/0090-3558-44.1.45 %/ Wildlife Dis Assoc.

Ebersole, J. L., M. J. Steffen, J. Gonzalez-Martinez, and M. J. Novak, 2008, Effects of Age and Oral Disease on Systemic Inflammatory and Immune Parameters in Nonhuman Primates: 1067–1075 p., doi:10.1128/cvi.00258-07.

Ezenwa, V. O., E. A. Archie, M. E. Craft, D. M. Hawley, L. B. Martin, J. Moore, and L. White, 2016, Host behaviour–parasite feedback: an essential link between animal behaviour and disease ecology: Proceedings of the Royal Society B: Biological Sciences, v. 283, no. 1828, p. 20153078, doi:10.1098/rspb.2015.3078.

713 Fallon, P. G. et al., 2003, Juvenile Rhesus Monkeys Have Lower Type 2 Cytokine Responses than Adults after
714 Primary Infection with Schistosoma mansoni: 939–945 p., doi:10.1086/368130 %/ University of Chicago
715 Press.
716 Farine, D. R., and G. G. Carter, 2022, Permutation tests for hypothesis testing with animal social network
717 data: Problems and potential solutions: Methods in Ecology and Evolution, v. 13, no. 1, p. 144–156,
718 doi:10.1111/2041-210X.13741.
719 Farine, D. R., and H. Whitehead, 2015, Constructing, conducting and interpreting animal social network
720 analysis: 1144–1163 p., doi:10.1111/1365-2656.12418.
721 Fedurek, P., and J. Lehmann, 2017, The effect of excluding juveniles on apparent adult olive baboons (Papio
722 anubis) social networks: e0173146 p., doi:10.1371/journal.pone.0173146.
723 Flack, J. C., M. Girvan, F. B. M. de Waal, and D. C. Krakauer, 2006, Policing stabilizes construction of social
724 niches in primates: 426–429 p., doi:10.1038/nature04326.
725 Fleiss, J. L., and J. Cohen, 1973, The Equivalence of Weighted Kappa and the Intraclass Correlation
726 Coefficient as Measures of Reliability: 613–619 p., doi:10.1177/001316447303300309.
727 Freeland, W. J., 1976, Pathogens and the evolution of primate sociality: 12–24 p.
728 Friant, S., T. E. Ziegler, and T. L. Goldberg, 2016, Primate reinfection with gastrointestinal parasites:
729 Behavioural and physiological predictors of parasite acquisition: 105–113 p.,
730 doi:10.1016/j.anbehav.2016.04.006.
731 Gilbertson, M. L. J., L. A. White, and M. E. Craft, 2021, Trade-offs with telemetry-derived contact networks
732 for infectious disease studies in wildlife: 76–87 p., doi:10.1111/2041-210x.13355.
733 Gillespie, T. R., 2006, Noninvasive assessment of gastrointestinal parasite infections in free-ranging
734 primates: 1129–1143 p., doi:10.1007/s10764-006-9064-x.
735 Godfrey, S. S., 2013, Networks and the ecology of parasite transmission: A framework for wildlife
736 parasitology: 235–245 p., doi:https://doi.org/10.1016/j.ijppaw.2013.09.001.
737 Godfrey, S. S., C. M. Bull, R. James, and K. Murray, 2009, Network structure and parasite transmission in a
738 group living lizard, the gidgee skink, Egernia stokesii: 1045–1056 p., doi:10.1007/s00265-009-0730-9.

739 Godfrey, S. S., J. A. Moore, N. J. Nelson, and C. M. Bull, 2010, Social network structure and parasite infection
740 patterns in a territorial reptile, the tuatara (Sphenodon punctatus): International Journal for Parasitology,
741 v. 40, no. 13, p. 1575–1585, doi:10.1016/j.ijpara.2010.06.002.
742 Gotoh, S., 2000, Regional differences in the infection of wild Japanese macaques by gastrointestinal
743 helminth parasites: 291–298 p., doi:10.1007/bf02557598.
744 Grassi, C., 2002, Sex differences in feeding, height, and space use in Hapalemur griseus: 677–693 p.,
745 doi:10.1023/a:1014934103832.
746 Grear, D. A., L. T. Luong, and P. J. Hudson, 2013, Network transmission inference: host behavior and parasite
747 life cycle make social networks meaningful in disease ecology: Ecological Applications: A Publication of the
748 Ecological Society of America, v. 23, no. 8, p. 1906–1914, doi:10.1890/13-0907.1.
749 Griffin, R. H., and C. L. Nunn, 2012, Community structure and the spread of infectious disease in primate
750 social networks: 779–800 p., doi:10.1007/s10682-011-9526-2.
751 Habig, B., and E. A. Archie, 2015, Social status, immune response and parasitism in males: a meta-analysis:
752 17 p., doi:10.1098/rstb.2014.0109.
753 Habig, B., M. M. Doellman, K. Woods, J. Olansen, and E. A. Archie, 2018, Social status and parasitism in male
754 and female vertebrates: a meta-analysis: 13 p., doi:10.1038/s41598-018-21994-7.
755 Hamilton, D. G., M. E. Jones, E. Z. Cameron, D. H. Kerlin, H. McCallum, A. Storfer, P. A. Hohenlohe, and R. K.
756 Hamede, 2020, Infectious disease and sickness behaviour: tumour progression affects interaction patterns
757 and social network structure in wild Tasmanian devils: 7 p., doi:10.1098/rspb.2020.2454.
758 Hart, J. D. A., M. N. Weiss, L. J. N. Brent, and D. W. Franks, 2021, Common Permutation Methods in Animal
759 Social Network Analysis Do Not Control for Non-independence: bioRxiv, p. 2021.06.04.447124,
760 doi:10.1101/2021.06.04.447124.
761 Herrera, J. P., J. Moody, and C. L. Nunn, 2021, Predictions of primate–parasite coextinction: Philosophical
762 Transactions of the Royal Society B: Biological Sciences, v. 376, no. 1837, p. 20200355,
763 doi:10.1098/rstb.2020.0355.
764 Hillegass, M. A., J. M. Waterman, and J. D. Roth, 2010, Parasite removal increases reproductive success in a
765 social African ground squirrel: 696–700 p., doi:10.1093/beheco/arq041.

Hinney, B., N. C. Wirtherle, M. Kyule, N. Miethe, K. H. Zessin, and P. H. Clausen, 2011, Prevalence of helminths in horses in the state of Brandenburg, Germany: 1083–1091 p., doi:10.1007/s00436-011-2362-z.

Hoppitt, W. J. E., and D. R. Farine, 2018, Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups: 227–238 p., doi:https://doi.org/10.1016/j.anbehav.2017.08.029.

Horii, Y., I. Imada, T. Yanagida, M. Usui, and A. Mori, 1982, Parasite changes and their influence on the body weight of Japanese monkeys (Macaca fuscata fuscata) of the Koshima troop: 416–431 p., doi:10.1007/BF02381324.

Khan, M. N., M. S. Sajid, M. K. Khan, Z. Iqbal, and A. Hussain, 2010, Gastrointestinal helminthiasis: prevalence and associated determinants in domestic ruminants of district Toba Tek Singh, Punjab, Pakistan: 787–794 p., doi:10.1007/s00436-010-1931-x.

Kim, D. A., A. R. Hwong, D. Stafford, D. A. Hughes, A. J. O'Malley, J. H. Fowler, and N. A. Christakis, 2015, Social network targeting to maximise population behaviour change: a cluster randomised controlled trial: 145–153 p., doi:10.1016/s0140-6736(15)60095-2.

Kossinets, G., 2006, Effects of missing data in social networks: 247–268 p., doi:10.1016/j.socnet.2005.07.002.

Kurvers, R. H. J. M., J. Krause, D. P. Croft, A. D. M. Wilson, and M. Wolf, 2014, The evolutionary and ecological consequences of animal social networks: emerging issues: Trends in Ecology & Evolution, v. 29, no. 6, p. 326–335, doi:10.1016/j.tree.2014.04.002.

Liao, Z. J., S. Sosa, C. F. Wu, and P. Zhang, 2018, The influence of age on wild rhesus macaques' affiliative social interactions: 10 p., doi:10.1002/ajp.22733.

Loehle, C., 1995, SOCIAL BARRIERS TO PATHOGEN TRANSMISSION IN WILD ANIMAL POPULATIONS: 326–335 p., doi:10.2307/1941192.

Lucatelli, J., E. Mariano-Neto, and H. F. Japyassú, 2021, Social interaction, and not group size, predicts parasite burden in mammals: Evolutionary Ecology, v. 35, no. 1, p. 115–130, doi:10.1007/s10682-020-10086-6.

MacIntosh, A. J. J., 2014, Ecology and Epidemiology of Nematode Infection in Japanese Macaques Building an Empirical Model: 23–51 p., doi:10.2354/psj.30.003.

MacIntosh, A. J., A. D. Hernandez, and M. A. Huffman, 2010, Host age, sex, and reproductive seasonality affect nematode parasitism in wild Japanese macaques: 353–64 p., doi:10.1007/s10329-010-0211-9.

MacIntosh, A. J. J., A. Jacobs, C. Garcia, K. Shimizu, K. Mouri, M. A. Huffman, and A. D. Hernandez, 2012, Monkeys in the Middle: Parasite Transmission through the Social Network of a Wild Primate: doi:10.1371/journal.pone.0051144.

McCowan, B., K. Anderson, A. Heagarty, and A. Cameron, 2008, Utility of social network analysis for primate behavioral management and well-being: 396–405 p., doi:10.1016/j.applanim.2007.02.009.

McDonald, M., and S. Johnson, 2014, 'There's an app for that': a new program for the collection of behavioural field data: 81–87 p., doi:10.1016/j.anbehav.2014.06.009.

Modrý, D., B. Pafčo, K. J. Petrželková, and H. Hasegawa, 2018, Parasites of Apes: An Atlas of Coproscopic Diagnostics: Edition Chimaira.

Mohammed, K., Y. Abba, N. S. B. Ramli, M. Marimuthu, M. A. Omar, F. F. J. Abdullah, M. A. Sadiq, A. Tijjani, E. L. T. Chung, and M. A. M. Lila, 2016, The use of FAMACHA in estimation of gastrointestinal nematodes and total worm burden in Damara and Barbados Blackbelly cross sheep: Tropical Animal Health and Production, v. 48, no. 5, p. 1013–1020, doi:10.1007/s11250-016-1049-y.

Moller, A. P., R. Dufva, and K. Allander, 1993, PARASITES AND THE EVOLUTION OF HOST SOCIAL-BEHAVIOR: 65–102 p., doi:10.1016/s0065-3454(08)60405-2.

Müller Klein, N., Michael Heistermann, C. Strube, M. Franz, O. Schülke, and J. Ostner, 2019, Exposure and susceptibility drive reinfection with gastrointestinal parasites in a social primate: doi:10.1111/1365-2435.13313.

Müller Klein, N., M. Heistermann, C. Strube, Z. M. Morbach, N. Lilie, M. Franz, O. Schulke, and J. Ostner, 2019, Physiological and social consequences of gastrointestinal nematode infection in a nonhuman primate: 322–335 p., doi:10.1093/beheco/ary168.

Neumann, C., J. Duboscq, C. Dubuc, A. Ginting, A. M. Irwan, M. Agil, A. Widdig, and A. Engelhardt, 2011, Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating: 911–921 p., doi:10.1016/j.anbehav.2011.07.016.

Neumann, C., and L. Kulik, 2014, EloRating: Animal Dominance Hierarchies by Elo Rating.

Nunn, C. L., S. Altizer, K. E. Jones, and W. Sechrest, 2003, Comparative tests of parasite species richness in primates: 597–614 p., doi:10.1086/378721.

Nunn, C. L., F. Jordan, C. M. Mc-Cabe, J. L. Verdolin, and J. H. Fewell, 2015, Infectious disease and group size: More than just a numbers game: doi:10.1098/rstb.2014.0111.

Poulin, R., 2007, Are there general laws in parasite ecology? 763–776 p., doi:10.1017/s0031182006002150.

R. Development Core Team, 2021, R: A Language and Environment for Statistical Computing.

Rimbach, R., D. Bisanzio, N. Galvis, A. Link, A. Di Fiore, and T. R. Gillespie, 2015, Brown spider monkeys (Ateles hybridus): A model for differentiating the role of social networks and physical contact on parasite transmission dynamics: doi:10.1098/rstb.2014.0110.

Roberts, S. G. B., and A. I. Roberts, 2020, Social and ecological complexity is associated with gestural repertoire size of wild chimpanzees: 276–292 p., doi:10.1111/1749-4877.12423.

Roberts, C. W., W. Walker, and J. Alexander, 2001, Sex-associated hormones and immunity to protozoan parasites: 476-+ p., doi:10.1128/cmr.14.3.476-488.2001.

Romano, V., J. Duboscq, C. Sarabian, E. Thomas, C. Sueur, and A. J. J. MacIntosh, 2016, Modeling infection transmission in primate networks to predict centrality-based risk: 767–779 p., doi:10.1002/ajp.22542.

Rushmore, J., D. Bisanzio, and T. R. Gillespie, 2017, Making New Connections: Insights from Primate-Parasite Networks: 547–560 p., doi:10.1016/j.pt.2017.01.013.

Rushmore, J., D. Caillaud, R. J. Hall, R. M. Stumpf, L. A. Meyers, and S. Altizer, 2014, Network-based vaccination improves prospects for disease control in wild chimpanzees: 10 p., doi:10.1098/rsif.2014.0349.

Sandel, A. A., J. Rushmore, J. D. Negrey, J. C. Mitani, D. M. Lyons, and D. Caillaud, 2021, Social Network Predicts Exposure to Respiratory Infection in a Wild Chimpanzee Group: 12 p., doi:10.1007/s10393-020-01507-7.

Sarabian, C., R. Belais, and A. J. J. MacIntosh, 2018, Feeding decisions under contamination risk in bonobos: 13 p., doi:10.1098/rstb.2017.0195.

Sarabian, C., V. Curtis, and R. McMullan, 2018, Evolution of pathogen and parasite avoidance behaviours: 7 p., doi:10.1098/rstb.2017.0256.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson, 1998, Patterns of macroparasite aggregation in wildlife host populations: 597–610 p., doi:10.1017/s0031182098003448.

Silk, M. J., 2018, The next steps in the study of missing individuals in networks: a comment on Smith et al. (2017): 37–41 p., doi:10.1016/j.socnet.2017.05.002.

Silk, M. J., D. P. Croft, R. J. Delahay, D. J. Hodgson, N. Weber, M. Boots, and R. A. McDonald, 2017, The application of statistical network models in disease research: Methods in Ecology and Evolution, v. 8, no. 9, p. 1026–1041, doi:10.1111/2041-210X.12770.

Silk, M. J., A. L. Jackson, D. P. Croft, K. Colhoun, and S. Bearhop, 2015, The consequences of unidentifiable individuals for the analysis of an animal social network: 1–11 p., doi:10.1016/j.anbehav.2015.03.005.

Smith, J. A., and J. Moody, 2013, Structural effects of network sampling coverage I: Nodes missing at random: 652–668 p., doi:10.1016/j.socnet.2013.09.003.

Smith, J. A., J. Moody, and J. H. Morgan, 2017, Network sampling coverage II: The effect of non-random missing data on network measurement: 78–99 p., doi:10.1016/j.socnet.2016.04.005.

Stephens, P. R. et al., 2017, Global Mammal Parasite Database version 2.0: 1476-+ p., doi:10.1002/ecy.1799.

Teichroeb, J. A., E. C. Wikberg, and P. Sicotte, 2011, Dispersal in male ursine colobus monkeys (Colobus vellerosus): influence of age, rank and contact with other groups on dispersal decisions: Behaviour, v. 148, no. 7, p. 765–793.

Tiddi, B., P. Romina, and A. Ilaria, 2019, The impact of food provisioning on parasite infection in wild black capuchin monkeys: a network approach: 297–306 p., doi:10.1007/s10329-018-00711-y.

VanderWaal, K. L., E. R. Atwill, S. Hooper, K. Buckle, and B. McCowan, 2013, Network structure and prevalence of Cryptosporidium in Belding's ground squirrels: 1951–1959 p., doi:10.1007/s00265-013-1602-x.

Venables, W. N., and B. D. Ripley, 2002, Modern Applied Statistics with S: Springer.

Watanabe, K., 2008, A Review of 50 Years of Research on the Japanese Monkeys of Koshima: Status and Dominance: 405–417 p.

Webber, Q. M. R., R. M. Brigham, A. D. Park, E. H. Gillam, T. J. O'Shea, and C. K. R. Willis, 2016, Social network characteristics and predicted pathogen transmission in summer colonies of female big brown bats (Eptesicus fuscus): 701–712 p., doi:10.1007/s00265-016-2093-3.

Weiss, M. N., D. W. Franks, L. J. N. Brent, S. Ellis, M. J. Silk, and D. P. Croft, 2021, Common datastream permutations of animal social network data are not appropriate for hypothesis testing using regression models: Methods in Ecology and Evolution, v. 12, no. 2, p. 255–265, doi:10.1111/2041-210X.13508.

White, L. A., J. D. Forester, and M. E. Craft, 2017, Using contact networks to explore mechanisms of parasite transmission in wildlife: 389–409 p., doi:10.1111/brv.12236.

Wood, E. L. D., J. B. Matthews, S. Stephenson, M. Slote, and D. H. Nussey, 2013, Variation in fecal egg counts in horses managed for conservation purposes: individual egg shedding consistency, age effects and seasonal variation: 115–128 p., doi:10.1017/S003118201200128X.

Wooddell, L. J., S. S. K. Kaburu, and A. M. Dettmer, 2020, Dominance rank predicts social network position across developmental stages in rhesus monkeys: American Journal of Primatology, v. 82, no. 11, p. e23024, doi:10.1002/ajp.23024.

Wren, B. T., M. J. Remis, J. W. Camp, and T. R. Gillespie, 2016, Number of Grooming Partners Is Associated with Hookworm Infection in Wild Vervet Monkeys (Chlorocebus aethiops): 168–179 p., doi:10.1159/000448709.

Yamagiwa, J., and D. Hill, 1998, Intraspecific variation in the social organization of Japanese macaques: Past and present scope of field studies in natural habitats: 257–273 p., doi:10.1007/bf02573076.