Reviewer1: Matthew Silk

The preprint has now been assessed by three expert reviewers. As you will see, all three found the preprint interesting and well written, but had overlapping concerns that leave me to suggest a fairly substantial revision is necessary before I can recommend it. There are definitely some overall themes to the comments that are worth paying particular attention to.

Main comments:

Main Comment 1: Multiple reviewers suggest that the paper, in particular the introduction could be shortened and suggestions are provided as to some of the information that is somewhat redundant to research questions focused on. I very much agree with these comments and feel the paper would benefit from a more focused introduction (and that the discussion could be shortened also).
Answer: Following the reviewer's suggestions, we shortened the introduction by reducing parts that are not the main focus of this study (line 82-103).

Main Comment 2: There are also multiple suggestions related to the clarity of methods and model descriptions that suggest this is an area to focus on when revising the manuscript. I would encourage the authors to think carefully about the various thoughtful suggestions provided by the reviewers when deciding if and how to revise the analyses – it will not be possible to implement all of the suggestions made simultaneously but collectively they provide a helpful guide (including some useful links to different parts of the literature). Clearly assessing why different centrality measures were used (in terms of the hypotheses being tested) would also help work out the need to use multiple measures. I have provided a few additional comments about the analyses below as well which overlap with some of the points made by the reviewers.
Answer: We provide clarification of the methods and edited some parts of the analyses. Notably, we removed model averaging and randomisation as we decided they are less valuable in our study. We added a note in the introduction as well to show how the different network metrics are related to our hypotheses (line 144-147).

I have added some additional thoughts below from my own reading of the paper in case they are additionally helpful. First two general comments and then some more specific things.

General comments:

General comment 1: As picked up on by multiple of the reviewers some careful thought is needed about how permutations are used. First, it is important to think about the non-independence structure of the data and extent to which it is or is not controlled for by the statistical model constructed. As mentioned by one of the reviewers there is active debate on exactly when permutations are helpful for addressing questions such as this, and in this case with individual traits rather than individual social network measures alternative routes to controlling non-independence might be preferred. This is not saying the use of a permutation approach is necessarily invalid. However, if the authors are keen to use a permutation approach then it is important to stick with this approach (e.g. if using permutations then it implies you are avoiding using the confidence intervals of your model for statistical inference and so it is unclear to then show them in Figure 2). I would also like to see some justification for the permutation approach used – the rewire algorithm is a very general way to change the structure of the network and so has the potential to generate inappropriate reference models in some contexts – I would like to know why this approach was chosen and why constraints on the permutations were not used?

Answer: We thank the reviewers for their very helpful comments. We will respond here about several of the issues raised and refer to this response where necessary for the rest of the comments because we think the issues are related.

We would like to make several points clearer here and in the manuscript.

First and foremost, all the studies available (including the ones pointed towards by the reviewers) talk about the issue of non-independence when network measures are the response variable because it is important that residuals are independent and it relates to links between the response and the independent variable(s). In our study, we are testing the network centralities as predictor variables. We do not know of studies addressing issues of non-independence or appropriateness of randomisation/permutation in this case. We are also not concerned about whether the social structure we observed is robust or random. We are investigating links between individual centrality measures and their parasite burden (response variable, as proxied by EPG). Therefore, in explaining parasite burden, sociality was only included as much as for other individual traits such as rank or age.

Following talks and studies about controlling for non-independence of network data, we thought although network centralities are predictor variables, we still need to use permutations to test the probability of the sociality-parasitism relationship to give the hypothesized non-random connections between individuals, or in other words, what is the probability of this relationship to arise by chance. However, and most importantly, we do not use permutations to make statistical inference about the relationship between parasitism and sociality. The statistical inference is made through generalized linear mixed models.

For those reasons, we used a "post-network" permutation procedure (and not a "pre-network" procedure as was understood by one of the reviewers) as we were confident in our observations to build a representative social network. We also used the general "rewire" function of the package igraph with the option each_edge (which carries out edge rearrangements, i.e. takes one edge and attributes it to another pair of nodes without changing the node position or characteristics) and a probability of establishing a connection taken from a random distribution between 0 and 1 to take into account different levels of social preference (some connections are therefore highly probable, some others less so). We did not see the need to constrain the permutations further in this case.

That being said, after consulting with knowledgeable colleagues, arguing that the predictor also needs to show non-independence does not appear to make much sense: we would not be able to test anything in doing so, individuals in one group are not independent, a measure from one day and another involving the same individual is not independent, being either male or female creates dependencies, ect. Mixed models actually do account for dependencies in the data if they are well specified. So, we decided to remove randomisations from our analyses as we felt, in the end, that they do not bring much information anyway and are not therefore needed.

All this being said, we are nevertheless ready to implement changes in our analyses should the reviewers still think that it is necessary. Notably, we are open to suggestions and ideas regarding the analyses of centrality measures as a predictor and not a response variable.

General comment 2: Given the nature of the results, I am curious as to what extent the authors think the effects of sub-sampling detected are related to statistical power? The initial results were frequently marginal and there was no real difference between moving specific subsets of individuals and a random subset. One potential test that could be used to work out the importance of removing individuals from the network would be to subsample the same number of individuals from the dataframe (while including information

from the full network) to see whether the initial results were still detected. This could tell you whether the sub-sampling results are a generic power issue or specifically to do with missing interactions in the network. **Answer**: We added a test for the effect of subsampling on the observed relationship(line 346-354). As shown by the simulations (SP9, SP10), sub-sampling had a much weaker effect on the observed relationship compared with partial networks that were redrawn from remaining individuals, as would be predicted. This should confirm that the complete lack of a relationship we observed between networks and parasites in the subsampled networks (especially the random knockouts) cannot be explained by a smaller sample size.

Specific comments:

Specific comments 1: Reflecting comments from the reviewers I would like to see more justification for the reason to correlate parasite load and network centrality given the importance of indirect transmission in the study system. I don't think this makes the approach invalid but it warrants more explanation (and discussion) than is currently provided.

Answer: Even though it is true that the transmission pathway itself is generally indirect in the study system: 1) we currently lack well-tested methods to collect behavioural information that represents this indirect pathway, and the transmission pathway itself is not well explored in the wild (one of our current projects is aiming at exploring transmission pathways of intestinal parasites in Japanese macaques more directly, but such data are not yet available) (line 221-229). 2) One key feature of our work is that we have replicated – in a different population – the results of a previous study showing relationships between contact networks and parasite infection. We admit that we cannot know the mechanisms, yet, but these collective results do point to some connection between host social relationships and risk of parasite infection. We also argue that the main focus of our study does not require that we currently understand the exact mechanism at play, as our focus is on the question of (1) replication and (2) testing under what sampling conditions does the observed relationship hold, to inform future studies of potential pitfalls.

Specific comments 2: I would suggest using "eigenvector centrality" throughout in full to be clear that you are referring specifically to the centrality measure.
Answer: We applied the suggestion throughout.

I appreciate the authors not make directional predictions in cases where they had no a priori expectation.

Specific comments 3: Given the protocol described, it would be good to know what causes the conservable variability in the length of focal follow per individual? Presumably simply some individuals being difficult to observe or away from the group for periods of the study?
Answer: We made efforts to sample all individuals approximately equally, but indeed, some individuals were more difficult to find to start a focal observation, for example, solitary roaming males. We decided to include them anyway in the analysis as we had initially decided to sample all group members and they were interacting with other group members when around the group. We added this detail in the methods (line 204-205).

Specific comments 4: It would be good to be clear on how the social network is weighted as soon as it is initially described.
Answer: We had originally written "We constructed an undirected proximity network based on recorded 1m (close) proximity. We built a symmetrical matrix of undirected social proximities, in which dyadic proximity data were scaled by the number of observation scans collected from each individual of the corresponding dyad" (line 234-235), which provides the information needed.

Specific comments 5: Is the Elo-rating used in the analysis the score itself or a ranking based on the score?
Answer: It is the Elo-rating itself. We rephrased the sentences to make it clearer (line 303-304).

Specific comments 6: Throughout – it would be good to provide exact (rounded) p values rather than inequalities except when p<0.001 rather than providing them as inequalities.
Answer: We applied the suggestion throughout.

Specific comments 7: With only three parasites, it may be better to include them as a fixed rather than random effect to avoid potential biases in the estimation of model parameters (e.g. see http://dx.doi.org/10.1016/j.tree.2008.10.008 and http://dx.doi.org/10.1101/2021.05.03.442487)
Answer: We followed the reviewer's suggestion and now include parasite species as a fixed effect. (line 304-306)

Specific comments 8: I find Figure 1 very confusing – I don't entirely understand what it is showing and how?
Answer: We provided more information besides figure 1, notably more details in the legend and a "zoom in" window next to the figure to explicitly show variation in individuals' centrality according to network size and composition (line 355-360). We hope the reviewer feels the figure is improved, as we do. We still feel this is a useful representation of the study design and results, despite the figure's apparent complexity.

Specific comments 9: Please be careful and precise with language around the results (e.g. careful when using "significant" outside of the context of statistical significance, careful with the use of bias [c.f. accuracy and precision] or using it to described random sub-sampling, etc.)
Answer: we tried our best to adapt and correct any language that could be seen as ambiguous.

Specific comments 10: Talking about marginal results is always challenging, but it would be good to see this done with more care – e.g. a marginal result for eigenvector centrality is mnetiond towards the top of P19 and is associated with a p value of 0.904 when (I assume) statistical significance would be p<0.025 or p>0.975.
Answer: we tried our best to adapt and correct any language that could be seen as ambiguous or careless.

Specific comments 11: Could you use the proportion of model results equivalent to the observed dataset rather than a number for Table 5? It may help with interpretation.
Answer: we changed this accordingly.

Specific comments 12: The reference to Weber et al. 2013 at the start of the discussion is rather misleading – the study showed that individuals with more contacts outside of their group (or more important in connecting between groups) were more likely to be infected.
Answer: We changed the sentences accordingly (line 450-451)

Specific comments 13: The second argument about degree centrality on P25 Paragraph 2 related to dilution of infection risk could be equally relevant to other measures (especially strength) and so care should be taken in specifically singling it out here, the preceding argument seems much more convincing.
Answer: We changed the sentences accordingly (line 502-506).

Specific comments 14: P26 Paragraph 2: you mention lower model parameter estimates, it would be good to find a way to show this alongside statistical significance.
Answer: We provided reference to the table in the supplemental materials, which can show the difference between parameter estimates (line 532)

Specific comments 15: Page 27 Paragraph 2: It would be good to rethink the sentence starting "These differences might…" as it is correlations being discussed rather than the similarity in the numbers themselves

(degree in the full and partial network could still be highly correlated even if the average degree in the partial network is much lower). Extra information above correlation is needed to make this argument.

Answer: Even though it is true that correlation is different from the numbers themselves, the network metrics are normalized so that comparisons of exact numbers are equally valued as correlation.

Specific comments 16:Page 28: It is not clear how a meaningful comparison has been the random removals versus the removals of specific types of individual.

Answer: We removed the sentence for clarification.

I hope the authors find these comments helpful and I look forward to seeing the revised version of the manuscript.

Reviewer 2: Quinn Webber

This article quantifies the relationship between parasitism and social network centrality and dominance in Japanese macaques. Overall, the paper is well written and interesting. I do have some comments on the structure of the introduction and the statistical analysis that will hopefully help improve the paper for publication.

Over all comment 1: Please note, there are no line numbers – so in some cases the line numbers I provide are
approximate.

Answer: We apologise for the inconvenience and added line numbers in the updated version.

Comments

Introduction

Major comment: I think the introduction is overall well written, but I think a bit more detail and precision earlier on in the introduction would benefit the paper and provide the paper with more of a focus (specific comments below). I also think the last paragraph of the introduction could be a bit more explicit about the hypotheses being tested and the associated predictions. Currently, there is no hypothesis – so I would suggest starting there and then introducing the predictions in the context of the hypotheses.

Answer: We rephrased the paragraph accordingly (line 143-158).

Comments 1: Page 3, line 1: this is the same sentence as the first sentence in the abstract. Perhaps change up the language a little bit between these sections.

Answer: We followed the suggestion. (line16)

Comments 2: Page 3, lines ~4-7: this sentence is kind of long and unwieldly – is there any way it can be separated into fewer sentences that provide more specifics about each of the factors (i.e. density, more social partners etc) that might influence increase transmission?

Answer: We followed the suggestion. (Line48-53)

Comments 3: Page 3, lines ~14-18: the sentences about how social interactions are structured by kinship, dominance etc as well as the subsequent sentence suggesting that social structure should influence exposure to parasites should both have a references.
Answer: We rephrased the paragraph and added references (line 54-58).

Comments 4: Page 3, lines ~19: does "accounting for social structure" help make sense of parasite transmission? I would argue that we don't use social networks to account for social structure, we use social networks as the tool to directly test the relationship between social processes and other ecological processes (in this case parasitism). I would suggest revising this sentence accordingly.
Answer: We rephrased the sentence accordingly (line 60-61)

Comments 5: Page 3, lines ~20-21: what can provide more information and what type of information? I would suggest this last sentence on page 3 needs to be more precise and answer the question of "why" the relationship between social behaviour and parasitism is important in the evolution of animal behaviour.
Answer: We rewrote the sentence accordingly. (Line 61-64)

Comments 6: Page 4, lines ~21-22; Page 5 lines 6, 13, 15, and 16: at all of these liens the authors discuss immune responses. This is clearly a very important aspect of parasitism, however, the present study does not quantify immune responses and so I found the repeated mention of immune response and systems to be somewhat misleading. I would suggest limiting the introduction of immune responses in the introduction and focussing instead of the variables that are actually measured in the paper (i.e. dominance, centrality, age, and sex).
Answer: We shortened those parts (Line 84-95)

Statistical analysis

Major comment: overall, I think the authors did the right things, I just have some clarifying comments and some suggestions to ensure the analyses are as robust as possible. See below for specific comments on the stats.
Answer:

Comments 1: Page 13-14: I think the description of the model set up is confusing given the number of fecal samples and number of parasite species differs per individual. It is unclear to me if for each fecal sample per individual there is a new row with the same information for all other variables? If so, why not average the parasite numbers across samples – especially given that the social network represents the sum of social encounters over the full study period, presumably the average number of parasites per fecal sample should correspond? My concern specifically is that if the same information is included multiple times in the model, there is non-independence and the coefficients will be inflated. For example, if the dataset were to look like this:

Animal ID	Parasite score	Degree	Age	Sex	
1	0.5	10	2	M	
1	0.7	10	2	M	
2	0.1	6	4	F	
2	0.3	6	4	F	

My concern with this type of data structure is that the data for degree, age, and sex is replicated for each individual and included in the models multiple times. As I understand it, even accounting for ID as a random

effect will still inflate coefficients because the only data that are changing is the response variable (i.e. parasitism). The alternative option is to have shorter time frames to build social networks so that if a fecal sample was collected on day 10 of the study, the degree of that individual in the days around the time of that sample would be different from the degree around day 50 (or something like that). I think this would be an elegant solution to the problem, but would require making many mini-focal networks for each individual around the time a fecal sample was collected. A good example of a study that does something like this is Turner et al. 2021 (https://doi.org/10.1111/1365-2656.13282). I would say that it isn't necessarily easy, but I would argue that perhaps it is the right way to do it.

Answer: We thoroughly discussed this comment amongst ourselves and with statistician colleagues, and then to be sure conducted a small simulation to test whether including similar data multiple times in the model would produce inflated results. The simulation showed that coefficients were not inflated, and we thus decided to keep our original data structure. In general, we have never encountered this criticism before, and a brief look at the literature will show that our approach is used predominantly in studies of behavior and ecology. With the right model specifications, and as our simulations show, we don't believe maximizing the data available is problematic, and can have numerous benefits in our case such as accounting for sample-to-sample variance, which an average would preclude. But we thank the reviewer for allowing us to consider this point.

Comments 2: Page 14, lines 10-17: this is an interesting interpretation of the Webber et al. (2020) paper discussing the use of multiple social network metrics. As one of the authors of this paper, I would argue that the authors missed the point a little bit and the spirit of the "less is more" approach is really about a priori selecting a single network metric. I would also point out that model averaging has been highly criticized and may not be appropriate (see Cade 2015 https://doi.org/10.1890/14-1639.1). My recommendation would be to remove the model averaging and select a single metric and run a single model.

Answer: We thank the debate that the reviewer introduced us to about using model averaging. After careful consideration, we have decided to remove model averaging from our analysis as it is not an essential component of our study anyway. With only 3 network metrics being considered this is not a major burden on the reader in any case. However, considering the point about selecting a single metric, we do have specific reasons for selecting the three metrics we did, and we do not feel that reducing our analysis to a single one is appropriate. Each of degree, strength and eigenvector have the potential to be linked with infection in different ways, and we cannot decide a priori which is more appropriate for our study. Moreover, we have already finished our statistical analyses, so that going back and omitting only one of them would produce a "HACKing" problem. So, we have kept the three separate models in our study.

Comments 3: Page 15, line 16: I have never used, or heard of, the rewire function. Perhaps some more detail is required here? I also would suggest the authors consider explaining the randomizations in a little more detail. Does the rewire function randomly re-assign interactions between individuals throughout the dataset? What if animals A and B didn't interact until day 50 of the study – do the authors account for temporal (or spatial) heterogeneity in interactions? Typically, network randomizations hold everything constant (i.e. space, time, group size) and shuffle the identities of individuals interacting/associating within a given sampling period. The recent review by Hobson et al. 2021 (https://doi.org/10.1111/brv.12775) is a great over-view of the theory and application of network randomizations, which I would suggest the authors read and adopt the terminology used in this paper.

Answer: We removed the randomisation part after careful consideration of reviewers' comments (see the response to general comment 1 at the beginning).

Comments 3: Figure 2: I would suggest the panel with sex should be a boxplot with the points jittered behind the boxes. I would also suggest indicating in the caption that the values on the x-axis are scaled (as opposed to having "scaled value" written next to the top left panel). Also, the panels are missing y-axis labels and units. I assume it's EPG, but this should be included.

Answer: We changed the panel with sex into a semi-box plot, in which the box represents the effect of this factor, instead of the distribution of data. Considering the y-axis labels, we added the y-axis labels.

Reviewer 3: Krishna Balasubramaniam

Comments:

Comments 1: P2:This sentence is too long, and needs to be broken up
Answer: We followed the suggestion. (Line32-37)

Comments 2: P3: I understand what you're trying to get at, but feel that these lines can be written more clearly. For instance --Living in larger groups may mean a generally greater exposure to parasites. However, the fact that animals within groups may interact at different rates and/or with different numbers of partners thereby leading to heterogeneity in contact patterns means that the expected positive relationship between increased group size and parasitism is not always observed.
Answer: We rephrased the sentence accordingly (line54-56).

Comments 3: P4: From these lines, I take it that male susceptibility to parasitism is linked almost entirely to intrinsic factors. Would you not also say that their explorstory tendencies, wider ranging, and space-use would also, and as such, contribute to increased parasitism?
Answer: We now give an example for extrinsic factors and an example for intrinsic factors. . (Line 82-103)

Comments 4: P5: Here I think you do a more thorough job of reviewing both intrinsic (immunity) and extrinsic (environmental interactions) factors that may lead to increased parasitism in Juveniles. Use a similar framework for males above.
Answer: As mentioned in the response to the previous comment, we streamlined this paragraph by giving only one example for both intrinsic and extrinsic factors. We hope this reduces confusion. (Line 82-103)

Comments 5: P7-1: In my opinion, there is no such thing as an "ideal" model system. I would say "well-suited" instead
Answer: We followed the suggestion. (Line 122)

Comments 6: P7-2: Again, this sentence is too long. Probably break it up as --"... social structure. Furthermore, they are mainly terrestrial..."
Answer: We followed the suggestion. (line 123-126)

Comments 7: P8: This seems to be a fair argument. Out of curiosity, though, assuming that your proximity network was correlated to a social contact or grooming network, did you try disassociating the effects of space-use sharing and social contact patterns, to see if one versus the other type of interaction better explains transmission? See the below recent reference for this:
Albery, G. F., Kirkpatrick, L., Firth, J. A. & Bansal, S. Unifying spatial and social network analysis in disease ecology. Journal of Animal Ecology 90, 45-61 (2021).
Answer: This is certainly an interesting question and one we hope to be able to address in the future. However, since our study is largely about the impact of sampling bias, we decided to focus on the social data set that we felt best represented opportunities for parasites to transmit considering their life cycles. In our ongoing project, though, we will be exploring this to the degree that it is possible for geohelminths, and we hope to have a better answer in the future.

Comments 8: P15: This describes a pre-network randomization approach, but are the authors aware of recent methodological concerns raised regarding the implementation of this approach in GLMM-based null-hypothesis testing? In particular, its susceptibility to Type-I errors?

Puga-gonzalez, I. Sueur, C. & Sosa, S. Null models for animal social network analysis and data collected via focal sampling : Pre-network or node network permutation? Methods Ecol. Evol. 12, 22-32 (2020)

Weiss, M. N., Franks, D. W., Brent, L. J. N., Ellis, S., Silk, M. J., & Croft, D. P. Common datastream permutations of animals social network data are not appropriate for hypothesis testing using regression models. Methods Ecol. Evol. 12, 255-265 (2020)

More generally, the topic of null models in network analysis is receiving a lot of scrutiny now, following recent debates:

Hart, J. D. A., Weiss, M. N., Brent, L. J. N., & Franks, D. W. Common permutation methods in animal social network analysis do not control for non-independence. https://www.biorxiv.org/content/10.1101/2021.06.04.447124v1 (2021)

Farine, D. R. & Carter, G. G. Permutation tests for hypothesis testing with animal social data: problems and potential solutions. bioRxiv 2020.08.02.232710 (2020).

Given these recent arguments, I suggest that the authors either refrain from using null models entirely, or make a stronger case for why they used pre-network randomizations and preferred these approaches to post-network node-swapping procedures.

Answer: Please refer to our answer to General comment 1 from Matthew Silk. We have decided to revise our approach and do away with network randomizations entirely. If the reviewer has any further comments on this we would be happy to entertain them as well.

Comments 9: P20: Could you convert the plot for sex into a box/violin plot, but still retain the raw data-points?

Answer: We now present the effect plot for sex in a semi-box plot which shows the range of effect in a box with raw data-points.

Comments 10: P21: See my earlier comment (and references) related to recent debates regarding the value of null models in hypothesis-testing, and the appropriate choice(s) of randomization tests. Related to that, did your permuted p values differ from the original p values from the GLMMs? Did the latter reach significance? Finally, it is my understanding that permutation-based null models have been used only in situations where network measures are the outcome (rather than predictor) variables, to account for their non-independence (which itself is now being heavily debated!).

So I'm left wondering whether your results would be different if they did not implement these null model approaches.

Answer: Please refer to our answer to General comment 1 from Matthew Silk. We agree regarding the response versus predictor quandary, and have adjusted our methods accordingly.

Reviewers 4: Anonymous reviewer

In this preprint manuscript, the authors assess geohelminth parasitism in a population of free-ranging Japanese macaques. The authors first test for a relationship between network measures of macaque centrality and the intensity of helminth parasitism (quantified by eggs per gram of fecal sediment; EPG), finding that, when considering the whole population of macaques, strength and eigenvector centrality are positively associated with parasitism, and macaque age is negatively associated. The authors then mimic sampling conditions in which only adult females, only juvenile macaques, or a random subset of the population are considered in analysis, finding that the relationship between centrality and parasitism is obscured or lost under these conditions. The effect of inadequate and/or biased subsampling of a population on subsequent network structure is a significant concern in

network studies, with a number of publications devoted to understanding this relationship. This study is an important extension of such previous work, focusing on how subsampling a population affects network inference, a largely understudied issue in disease ecology. That being said, I have a few major comments about this preprint, in addition to more minor comments.

Major comments:

Major comments 1: My biggest comment here is concerning the appropriateness of the bounds used in the network construction.

   a. Firstly, the authors use proximity to construct their macaque social network in a study of parasites which are environmentally transmitted. Individuals within one meter of each other at one time are considered connected in the network, but there is not assessment of potential network connectivity between individuals in the same location with a time lag, despite the authors' acknowledgement that "there is a time-lag between infection of one host and transmission to the next host through the environment" (pg 25). This calls into question the appropriateness of proximity as the edge definition without considering the role of specific locations/environmental exposures. Presumably, many individuals should be connected in the network through time-lagged proximity events, which would likely greatly affect results. The authors further assume that close proximity events are associated with resources, yet there may be a mismatch between locations in which individuals aggregate/are in close proximity and locations in which they perform behaviors most conducive to environmental parasite transmission. The authors should provide a justification for their choice of edge definition for this study of environmentally transmitted parasites and discuss any limitations to this choice.

Answer: First, please refer to our answer to Specific question 1 from Matthew Silk where we discuss part of this question. Now, we completely agree with this criticism of our work. There are various challenges working with such data sets involving natural populations, and matching relevant behavior to transmission events is one of them, particularly when the prepatent period (time between infection acquisition and onset of worm reproduction and egg shedding) is not known (see more comments below on this). In the present study, we decided to use an aggregate approach, assuming that individuals that spent time in spatiotemporal proximity by extension shared space to a greater degree than did others. However, we have no data on true space sharing and this is indeed a limitation of our study. We also did not compartmentalize our proximity data into activities that may be more or less likely to lead to transmission events. Such events would be hard to identify in practice, but perhaps with more data we could begin to create more robust networks based on both spatial behavior and specific activity patterns to better understand parasite transmission. In fact, we are currently working on such a study now, but unfortunately data will not be available in a time frame relevant to the current manuscript. We are also trying to address uncertainty in terms of both networks and infection dynamics (including accounting for time lags), but again we are not able to provide data on this at present. We have added some text to better address this issue (line 221-233).

   b. Secondly, the network is constructed over a period of 3 months of behavioral observations, but there is no justification provided for this choice. Indeed, the authors later discuss potential usefulness of a much longer period of observations (i.e., 16 months). However, in the context of parasite transmission, networks should generally be constructed over a time period which aligns with transmission of the parasite(s) in question. How long do the parasites under study survive and transmit in the environment? If only for a few days, for example, then edges should be aggregated over a similarly short time period to represent the interactions occurring during the "transmission period" of the parasite (e.g., see White et al 2017, Biological Reviews). The authors should provide a justification for their choice of aggregating interactions over a 3 month period and use caution in suggesting that aggregating over longer time periods would be more appropriate without further epidemiological justification.

Answer: The reviewer is right that the best case scenario would be to synchronize network data with transmission data (inclusive of a time lag). Our decision to aggregate was based on a few things. Limitations in data collection was one issue, because constructing reliable networks requires considerable amounts of data, which we could not achieve in shorter time periods with a single observer (we would argue this could

not be done without observing all individuals simultaneously). Note that the 16-month study referenced (MacIntosh et al 2012) was not in fact an aggregation over 16 months but instead examined changes in network structure across seasons (<2-month blocks). Perhaps this was just a misunderstanding. Now, the geohelminths in our study have estimated prepatent periods ranging between 2 and 6 weeks (as observed in livestock studies). But we have no way of creating robust linkages here, so instead we settled on an aggregate approach. Aggregating smooths the variance in both networks and infections, but we admit that it also affects the biological plausibility in an unavoidable - we think - way. Our results should thus be interpreted as being smoothed and encompassing general, rather than specific, relationships. (line 229-233)

Major comments 2: The modeling approach needs additional explanation and clarification in the methods.
   a. I was never certain exactly which models were run with which variables. I would recommend at least a table outlining the models which were run. Alternatively, the authors might consider a methods flow chart-style figure to show the different modeling approaches in the context of the different networks constructed. **Answer:** We added a table outline for the model (line310-311).

   b. I think there is perhaps an error in the description of the randomization procedure (pg 16): as written, it sounds as though there were many observations compared to many randomizations. Shouldn't there be one observation compared to the randomization-based null distribution? Or did the authors fit 1000 "observed" models?
Answer: We now removed the randomization procedure as we decided that it's not required for our analysis. Please refer to the answer to General comment 1 of Matthew Silk for more information.

Major comments 3: I have several questions about EPG as the outcome variable.
   a. The authors state that EPG is "not always indicative of true worm infection intensity," but do not later discuss how this limitation might affect their results.
Answer: Positive relationships as well as no relationships have previously been found, depending on the study system, but EPG is still the most standardized way to present infection intensity in wild animals that are not destructively sampled. (line 273-276) We currently lack such data for primates, unfortunately, so we are not able to discuss in depth about its influence on our results. We accept this as a caveat of our study.

b. Furthermore, the methods and first entries in the dataset (which are presented in the supplementary materials) suggest that EPG is extremely variable within the same individual over time. How might this variability affect results?
Answer: We provided clarification in the methods part. (line 306-309)

c. It appears that EPG was pooled across parasite species for modeling. The authors provide a statistical justification, but not a biological/epidemiological one. Do all parasites examined have identical epidemiological characteristics? If not, wouldn't they be expected to have different relationships between social structure and transmission?
Answer: We provided clarification in the methods part. (line 289-292)
We can understand the comment here. There are certainly differences in the life cycle of each parasite, but it is not currently possible to determine which is more or less likely to be associated with sociality as they are all encountered through interaction with the environment and contaminated substrates, so we decided to keep our current approach to reflect a kind of generalized phenomenon.

Major comments 4: I'm a little confused about the interpretation of the results when focusing on just adult females or juveniles. The authors state that age was significantly associated with parasitism (juveniles having higher EPG than older individuals); does it not then follow that we would struggle to find a relationship between centrality and parasitism when focusing on only juveniles? I imagine there is less variability in the EPG measures for juveniles, if they tend to have higher EPG, which would reduce the power for detecting an

effect in just this subset of the population. In contrast, because sex was not significantly associated with parasitism, we might expect to have higher power for detecting a relationship between centrality and parasitism with this subset. Indeed, the results seem to support this to some extent (e.g., strength was still significantly associated with EPG for females only). I would suggest the authors could discuss how changes in statistical power with subsampling may affect inference.

Answer: As shown in SP 6 & SP 7, age was associated with infection in juveniles but not in adult females: juveniles are found to have large age-related differences in infection in a previous data set, in which younger juveniles displayed significantly higher EPG and parasite richness than older juveniles (MacIntosh et.al, 2010). We also provide a discussion about changes in statistical power related to this in the discussion (line 582-585). Finally, note also that we included a new subsampling element in which we did not redraw the networks but instead kept all original relationships between individuals and their associated network metrics but remove comparable numbers of individuals. We describe the results above and in the supplementary material, but you can see that statistical power may not be the driving factor here (SP 6, SP7, SP9, SP10).

Major comments 5: Overall, I think the manuscript—and especially the introduction—could be streamlined. The real target of the manuscript is only arrived at by the end of the 7th paragraph and I think much of the introductory material, while interesting, is perhaps beyond the scope of the objectives of the study. To tighten up the manuscript and improve readability, I'd recommend streamlining and reducing the length of the introduction.

Answer: We thank the reviewer for the comment, and agree. We shortened the introduction part accordingly.

Minor comments:

Throughout comment: Just a quick note to perhaps add line numbers to future submissions. I've had some difficulty in describing the precise locations of minor comments without line numbers, so I apologize for any confusion or lack of clarity in the following comments. I have chosen to give the page and paragraph number for the following comments, with "paragraph 1" (abbreviated "P") always the first text on the page, regardless of if this is a new paragraph on that page.

Answer: We apologize for this omission! Line numbers were added in this version.

Minor comments 1: Pg 3, P 2: The flow of logic in the paragraph was difficult to follow and citations were missing. Is the objective of this paragraph to introduce the idea that social behavior affects transmission? This is a paragraph that could be drastically altered to streamline the introduction (see major comments).

Answer: we rephrased this paragraph accordingly. (line 54-64)

Minor comments 2: Pg 4, P 2: I did not follow the mention of young males dispersing. Recommend taking this point further and explaining the relevance if the authors choose to keep this sentence/paragraph.

Answer: we shortened this part according to comments about the introduction.

Minor comments 3: Pg 5, P 1: Use caution with vague/non-specific language such as "other factors" here. It decreases the readability of the manuscript. Other examples would be on page 26, paragraph 2: "For various reasons, numerous studies…"; page 29, conclusions paragraph: repeated use of "different" is vague and nonspecific to the direction of changes.

Answer: we applied the suggestion throughout except for the conclusions: In the conclusion part, we are not able to fully specify the direction of changes. Although we received false negative results in our simulations, it can only represent one kind of situation, while we still lack information of other analyses involving incomplete networks.

Minor comments 4: Pg 6, P 2: Recommend rephrasing the sentence with the very long parenthetical statement for readability.

Answer: we applied the suggestion. (line 104-111)

Minor comments 5: Pg 6, P 2: The authors repeatedly cite the same two papers here and in the discussion (Smith et al 2017 and Silk et al 2015). These are excellent choices, but additional relevant papers include Davis et al 2018, Animal Behaviour and Gilbertson et al 2021, Methods in Ecology and Evolution. In the interest of full disclosure: this reviewer is an author on one of these papers, but I do feel they are both particularly relevant to this manuscript (perhaps especially Davis et al 2018).
Answer: we applied the suggestions. (line 105)

Minor comments 6: Pg 9, P 2: Recommend using "macaques" throughout for consistent terminology; I think there were two instances where the authors chose to use "monkey" instead.
Answer: we applied the change throughout.

Minor comments 7: Pg 9, P 2: Does the focus group of macaques ever use spaces previously occupied by the other group on the island? I'm wondering if there could be any environmentally-mediated transmission relationships between the two groups.
Answer: we provided clarification in the methods part. (line 171-173)

Minor comments 8: Pg 10, P 1: Should the Kappa index results perhaps be given in the results rather than the methods?
Answer: as the index is part of our methods evaluation, we think it is better to present it in the methods.

Minor comments 9: Pg 12, P 1 and 2: Paragraph one states that samples were sent to Kyoto University; paragraph 2 then states that "we" processed the samples. Did the authors perform parasitological testing themselves, or was this work performed separately by the Primate Research Institute? I see that two of the authors are affiliated with the Primate Research Institute, but it was unclear if testing was performed by the authors or not, as phrased.
Answer: we rephrased the sentence to avoid confusion (line 249-250)

Minor comments 10: Pg 13, P 3: The authors do not need to provide information about how the data was laid out. However, please define "SP" in "(SP 1)."
Answer: As treating EPG of different parasite species from a sample separately is less common, we think that it is better to clarify our data layout to avoid confusion. We now provide the definition of SP. (line 205)

Minor comments 11: Pg 14, P 1: The Git repository does not need to be given in the main text as it is listed in a data availability statement later. However, the authors should provide information about how/where this repository will be versioned and archived in the data availability statement. For example, BES journals require Github code to be archived and versioned with Zenodo upon acceptance for publication.
Answer: we removed the sentences. As our data is archived on Git already, we would like to wait until our decision of where to publish for other data/code repository.

Minor comments 12: Pg 14, P 2: The different models for the different centrality measures and model averaging descriptions were very unclear. Recommend additional explanation and a table or figure to show the different models (see major comments). In addition, in GLMMs, were individuals and dates included as random intercepts or also as random slopes?
Answer: we provided extra information about the models.

Minor comments 13: Pg 15, P 1: AIC results should perhaps be in results, not methods?
Answer: Here AIC is to explain why we decided to use the Zi-model instead of normal glmm. So we decided to put it into methods.

Minor comments 14: Pg 15, P 2 heading: It is unclear, as written, how this heading and therefore section is different from the previous section. I think this is where a methods flow chart could be particularly helpful (see major comments).
Answer: the randomisation part was removed after thoughtful consideration.

Minor comments 15: Pg 15, P 2: I did not follow the "because we were confident in the observed edges" explanation. Could you explain further?
Answer: the randomisation part was removed after thoughtful consideration.

Minor comments 16: Pg 16, P 2: I would recommend stating the size of the subsampled networks, especially for the adult female and juvenile only networks.
Answer: We applied the suggestion. (line 330-332)

Minor comments 17: Pg 16, P 2: Why Pearson's and not a non-parametric correlation measure (i.e., Spearman's)? Both Davis et al 2018 and Gilbertson et al 2021 (mentioned in minor comments for page 6) use ranked correlations rather than Pearson's.
Answer: Considering the change of network metrics, we think that the change of values might be more influential compared with change of ranks, as we are focusing on the estimates calculated by glmms.

Minor comments 18: Pg 17, Figure 1: I found figure 1 to be extremely "busy" and it was difficult to decipher what the authors intended the reader to understand from this figure. If limited for space, I would recommend swapping this figure for a methods flow-chart style figure.
Answer: We provided more information besides figure 1, notably more details in the legend and a "zoom in" window next to the figure to explicitly show variation in individuals' centrality according to network size and composition (line 355-360).

Minor comments 19: Pg 19, P 1: "Full model" is never really defined. This is an example of how the methods are currently unclear, which has downstream impact on the interpretability of the results (see major comments).
Answer: we applied the suggestions throughout.

Minor comments 20: Pg 20, Figure 2: The figure legend states that the authors constructed three different models for each network metric, but this is unclear in the main text. Further, the different colored lines are not readily visible in the figures, as currently presented.
Answer: We rephrased the sentences for clarification.
About the different coloured lines, even though we are aware that they are not visible enough, it is what shows. The results are very close to each other, and we decided to represent the data and results we got accordingly instead of manipulating the figure. That being said, we are open to suggestions for further improvements.

Minor comments 21: Pg 20, Table 2: Symbols being defined (e.g., p) should generally be given in parentheses after the definition on first use. Additionally, the "name" for the summed output is excessively long and complicated. Please use a shorter abbreviation or name and define it. I should also add that the null model for LRT is also never defined, again adding to confusion about the methods. These comments for table 2 are relevant for other similar tables in the manuscript.
Answer: we applied the suggestions throughout.

Minor comments 22: Pg 23, P 1: It would appear that the authors are using the partial networks to test for "false negative" results (Type II error), but I would think that false positives (Type I error) would also be

interesting and informative. I would therefore think that including results for degree would be useful, rather than just strength and eigenvector centrality.

Answer: We agree that it is not completely appropriate to speak only to one type of statistical error here, and in fact it is not entirely accurate to claim that female or juvenile subsets, given divergent results, actually reflect Type I or Type II errors; whole and partial networks reflect different data sets. But we do feel the random knockouts better allow us to test for true errors. We have tried to clarify this distinction in the text. Also, we now also include results for degree.

Minor comments 23: Pg 23, Table 5: "Equivalency" appears to be qualitatively defined (are the randomization results also showing a positive estimate?), without information about the effect size of the detected relationship. Why? Furthermore, I think giving the "equivalent" results as proportions/percents would be more understandable than the current numbers. Importantly: it appears that the centrality measures in table 2 are incorrect. The legend states these should be strength and eigenvector, but the column in the table shows degree and strength.

Answer: we did not consider effect size in the simulations as it is less representative than the direction or significance itself. We now give the results as percentages.

Minor comments 24: Pg 24, P 1: Recommend defining "social integration" as I found this language unclear.
Answer: We rephrase the sentence. (line 509-510)

Minor comments 25: Pg 25, P 2: Regarding individuals with "high degree but low frequencies of interaction" – how common are these individuals? You should be able to comment on this with the data you have.
Answer: we applied the suggestion. (lines 486-487)

Minor comments 26: Pg 25, P 2: Regarding the example of lice - the direct transmission mentioned here would seem to reinforce my concerns about the appropriateness of proximity as the edge definition in this study (see major comments).
Answer: we refer here to the response to the major comment

Minor comments 27: Pg 26, P 3: Much of this paragraph is repetitive with the introduction. I would suggest cutting/trimming from one location to help streamline (see major comments).
Answer: We streamlined the introduction, and hope the current version is less redundant.

Minor comments 28: Pg 27, P 2: Did the authors not normalize degree for network size? This is standard functionality in igraph and would be valuable for controlling for the effect of network size.
Answer: Even though network size itself can be normalized, the change of values can not be normalized. When fewer individuals are removed, the values will be more likely to remain similar to the original ones.

Minor comments 29: Pg 28, P 2: "Unlike with targeted removal" - this phrasing was confusing and seemed to contradict the previous paragraph stating that targeted removal was a problem for inference.
Answer: we removed the sentence for clarification.

Minor comments 30: Pg 28, P 2: I am very surprised that random removals would have a stronger impact on inference than targeted/biased removals. I would suggest further interrogation of this result as it has major implications to sampling strategies. Furthermore, this finding contradicts the statement in the next paragraph that incomplete networks should have "little imbalance as to who gets excluded."
Answer: We rephrased the sentence for clarification. incomplete networks still provide robust network metrics. What changed dramatically was the result calculated when they are included in a glmm.

Minor comments 31: Supplement, SP 2 and 5: I found these figures very difficult to interpret. I would recommend providing additional information for readers. For example, the legend states there should be vertical lines for the original result and the 95% confidence intervals, but there appear to be 4 vertical lines in each figure. Also, what does "statistically significant results" mean in the context of these figures? I needed a little more "hand-holding" here.
Answer: we provided more information in the figure legends.

Throughout comment: I noticed a number of instances of grammatical errors or informal language throughout. I'd recommend perhaps having a native English speaker read and correct the manuscript. In addition, some in-text citations had doubled close parentheses; recommend fixing these.
Answer: Our in-house native English speaking author and co-writer apologizes for this inconvenience and vows it will not happen again!