Reviewer 1: Anonymous reviewer

In this preprint manuscript, the authors evaluate parasitism and social network metrics in Japanese macaques, finding that some metrics of network centrality are associated with increased parasite shedding. In addition, the authors evaluate the effect of subsampling the network on subsequent inference, finding that removal of individuals can result in the failure to detect a relationship between parasitism and network centrality metrics. This is my second time reviewing this manuscript for PCI Network Science.
This manuscript has a great deal of potential, but I'm afraid I have to confess to being disappointed by this updated version of the manuscript and the response to reviewers. I've outlined my major concerns below.

To Reviewers: We would like to answer the major comments of the first reviewer with the add-on comments from the recommender together, as they are focusing on the same topic.

Major comments

Point 1. In the initial review, multiple reviewers pointed out concerns with the authors' previous permutation approach to managing the non-independence inherent in network data. The authors have since removed the permutation approach, but no longer take any measures to account for network non-independence. Further, I am not convinced by their argument that since the network metrics are predictors, we do not need to be concerned about this non-independence. Rather, I would liken this to spatial models, in which both response and predictors are intrinsically affected by the autocorrelation induced by the nature of the data. As such, I would argue that the authors still need to account for the non-independence of network data in their modeling. Silk et al 2017 (MEE), for example, outlines an example of using a network autocorrelation model (NAM) to model infection1 status as a function of host characteristics and flow centrality—an example that seems particularly relevant for this study. Perhaps Matt Silk can chime in here, on the appropriateness/need for a NAM or similar approach.
Add-on comments: The non-independence of disease outcomes in the network is an interesting issue – individuals that are frequently found in proximity would be expected to be more likely to share similar outcomes for directly-transmitted pathogens/parasites in particular. Network autocorrelation models are certainly a good way to account for these types of pattern. However, given the focus of the study (and to some extent the likelihood that correlations between network position and infection intensity may not be related to transmission for these parasites), I think the current approach is reasonable (especially given the focus of the study is more on sampling).
Answer: We thank the reviewers for providing more discussion on this topic. We decided to keep our current approach but we make mention that this might be a subject of debate and explain why we do not carry out permutations (lines 332-340).

Point 2. Further relevant to the statistical approach, I am increasingly uncomfortable with the degree of pseudoreplication in the data, as analysed. The modeling approach relies very heavily on random effects to manage multiple layers of pseudoreplication with predictors that go unchanged. Because there is not extensive longitudinal sampling, I agree with the previous reviewers that it makes more sense to average parasite burden across samples within an individual (but not across parasites), and the authors' response on that point has not

satisfied this concern for me. Further, I think it would be important to actually interrogate the assumption that centrality effects are generalizable across all three parasites, as the current conception seems to me an oversimplification of disease ecology (e.g., intraindividual parasite competition, differing importance of social behaviors/interactions for transmission of different parasites, etc). Perhaps testing for interactions between parasite burdens or testing for different effects of network centralitymetrics for individual parasites.
Add-on comments: It is (inevitably) hard to judge the ability of the random effect structure to deal with clustering/non-independence, although the approach seems potentially reasonable it would be good to consider this carefully. I think the question about interactions really depends on the aims of the study, if the authors are interested in a similar relationship across all three pathogens then the current model structure seems fine, although more complex model structures could lead to more nuanced results.
Answer: We thank the reviewers for several interesting points.

After careful discussion and consultation with a statistician, we think that our approach is reasonable and that the models' random effect structure should be able to manage the potential issues.

Considering the differences between parasite species, we agree that there may be different relationships between network centrality and infection intensity, although carefully designed experiments and studies would still be needed to properly test this paradigm. Our team partly demonstrated this difference: we showed that gastro-intestinal parasite intensity was positively linked to the strength of social connections (in this study and in MacIntosh et al. 2012) whereas ectoparasite intensity was negatively linked to the number of social connections one has (Duboscq et al. 2016). Here, we are concerned parasites that have similar transmission modes to each other. Although we understand that each has its own ecology and biology, each can be transmitted between hosts that share environmental resources, so although our (commonly shared and tested) assumption might be oversimplistic from a disease ecology point of view, from a social transmission one, we believe there are enough common patterns to justify our approach. This study is focusing more on the commonly shared effects across parasites sharing a similar transmission pathway. We therefore think our current approach is sufficient for this study. We add statements in the ms to clarify this point (lines 301-305).

Point 3. I do not feel that the authors have provided adequate additions regarding the appropriateness of proximity-based network metrics as predictors of parasite load for environmentally transmitted parasites. In their response to the reviewers and revisions, the authors emphasize a lack of methods and a publication in which they used this approach previously, and underemphasize a biological rationale for this choice. I strongly recommend rewriting the material in lines 221-229 to something with the flow of "spatial proximity is representative of direct interactions; while we lacked data for spatially-explicit or time-lagged indirect interactions, we assumed that direct interactions in this highly social species would be representative of the probability of indirect transmission-relevant resource sharing between individuals; our previous work with this species supports this assumption." With this kind of flow, I am arguing that the authors should emphasize a biological underpinning for their assumption, and de-emphasize methodological constraints or reference to a single prior publication within this research group, as a biological rationale is more generalizable, justifiable, and comprehensible.
Add-on comments: This remains an important point and I am not sure it has been adequately dealt with in the way the manuscript has been revised. In addition to the reviewer's comments about the lack of clarity in the biological rationale for the approach, makes a variety of statements related to transmission in relation to

differences in the association between centrality measures and infection intensity when, given the fact that parasite is indirectly transmitted, it is not clear whether this is from social connections being a proxy for spatial proximity or because social centrality is related to health in some other way. It is fairly misleading to focus the discussion around ideas related to direct transmission (the introduction provide a much more holistic/balanced view of the relationship between social centrality and infection).
Answer: We rephrased the sections in methods (lines 221–228) according to the reviewer comments, and provided more clarification about our choice of the index for constructing networks. We also simplified our discussion about how or why different centrality metrics might be differentially associated with parasite infection phenotypes or inferred transmission (lines 501-514）.

Point 4. I think the authors misunderstood my point about providing a justification for the time over which the network observations were aggregated. This is not a question of the time between infection and detection. This is an issue of "are the interactions aggregated at a time scale representative of transmission processes?" This question has still not been addressed in the text. Are associations aggregated across 3 months representative of transmission-relevant interactions for the parasites being studied? From page 403 of White et al 2017 (Biological Reviews): "If researchers opt for association data, it is very important that they define contacts and collect data at intervals that are relevant to the epidemiology of the pathogen of interest (Cross et al., 2004; Keeling & Eames, 2005; Grear et al., 2013). Data collection should take into account host life history and the relevant infectious and exposure periods for pathogens in order for the contact network to be meaningful. For instance, when evaluating a population's vulnerability to epidemics, combining months and years of data into a single contact network can suggest an extremely interconnected population, when in fact, the infectious period is much shorter in duration than the time between when individuals come into contact (Cross et al., 2005)." Please justify the choice of 3 months as the aggregation period in the context of the parasites under study.
Add-on comments: This remains an important point to consider, but perhaps becomes less important if infection intensity is treated more as a variable trait of each node associated with network position and less as a directly-transmitted parasite (as discussed in relation to Point 3). This treatment seems feasible given the focus of the study (analyses at least) is predominantly on sampling methodology.
Answer: We agree with the reviewer's points. This study does not focus on transmission patterns or processes but rather on cumulative mid-to-long term effects of sociality (in the broad sense of individuals being near one another or not) on parasitism and how social structure correlates with parasite load. We recognise the limitations of using aggregated data over either point-sample transmission event data or arbitrary time scale, but we do not investigate transmission processes per se; we lack adequate data anyway. The two main foci of this study were (1) the replication of a previous phenomenological study of the same species at a different location (MacIntosh et al 2012, PLoS ONE) and (2) testing the effects of sampling methodology; infection intensity was treated more as a trait of each node (each individual's infection phenotype) though we admit that this trait assumes that a relevant transmission process has occurred. We provided more clarification about this issue in the methods (lines 229-244) and made sure to address more that we do not investigate transmission events per se in this study.

Point 5. I was incredibly disappointed by the response to my previous concerns about Figure 2. There are many options to present the results in a way that is clear and accessible to the reader. If regression lines for

different models are nearly identical, the authors can: (1) have different plots for each model; (2) have a single representative example in a plot in the main text with a statement as such, and put other results plots in a supplement; (3) simply add a clause to the figure legend that regression lines overlap due to their close similarity. At the very least, please increase the size of the figures and/or the resolution to make the lines more distinguishable. Further, the authors should follow Reviewer 2's recommendation to jitter the points in the plot for sex effect; if the authors feel this makes the plot too "busy," a violin plot can be used to display the distribution of the data.
Add-on comments: I would certainly agree with the reviewer that the plot for sex (finding an alternative way to display the distribution of the data that is clearer) could be more informative. There is also a mismatch (I think) between the legend and the number of lines/confidence intervals on the plots for the non-network metric variables. It would be good to use the comments provided to tidy up the figures a little more
Answer: We are thankful for the suggestions for improving our figures and have made some amendments in this revised version.

Point 6. Relevant to my previous comment, while the updates to Figure 1 aid in its comprehension, I still find the figure to be overly complicated and difficult to parse. Based on the final line of the figure legend, the point of the figure appears to be to show that removing individuals from the network affects Eigenvector centrality. However, visual assessment of the change in size of nodes is not an effective way to convey this result. The authors should focus on conveying this specific result; for example, by plotting change in rank from the complete network per individual as sampling effort changes.
Add-on comments: I continue to agree with the reviewer here. Figure 1 looks very nice but I honestly don't feel like I learn any useful information from it. As it doesn't detract from the manuscript this is up to the authors really, but the reviewer provides some suggestions for alternatives that it may be worth considering.
Answer: We replaced the network plot with a Trellis plot that illustrates the change in eigenvector across partial networks and hope this provides better visualization.

Point 7. I previously inquired why the authors chose to use Pearson's correlation instead of Spearman's, as was done in other similar studies. Perhaps the authors misunderstood my question, as they responded about GLMM results, and I meant specifically when comparing centrality metrics between complete and subsampled networks (L332-333). Indeed, Figure 1 specifically uses Eigenvector rank to compare network metrics, supporting my suggestion that Spearman's would be a more appropriate comparison. Please clarify the choice for Pearson's.
Add-on comments: While I think I would probably also have used Spearman Rank, I think the choice of Pearson's is probably reasonable especially if the metrics have been standardized (hard to tell definitively without the data). One thing that needs to be made clearer in the methods is for which analyses metrics have been standardized – this was something I only learned from the figure legends.
Answer: We thank the reviewer for insisting on this. We now use Spearman's correlation instead of Pearson's (lines 313-315; lines 434-436). The results changed slightly but do not change our conclusions. We also made related changes in the discussion (line 547-564), and added information about standardizing metrics in the methods (lines 298-299).

Reviewer 2: Matthew Silk

The authors have done a good job overall in addressing the previous round of reviewer comments, especially in terms of the introduction, methods/analyses and results. However, there still remain issues that need to be addressed before I can recommend the paper. In particular, the revision of the discussion is inconsistent with some suggested changes not made and some text/ideas that seem to remain from the previous version that does not fit well with the results from the new analysis. I sent the paper back to two of the original reviewers. While one of the reviewers was happy with the changes made, the other still had concerns over various aspects of the manuscript. I've provided some additional thoughts on these comments below as well as some more specific points on my own.

In the abstract it would be good to distinguish between the two types of subsampling (that of the dataframe of metrics versus of the network data) given this finding is important.
Answer: We added the corresponding part to the abstract (lines 34-36).

The introduction is much improved. It's on the long side but nicely written. However, as highlighted by the reviewer the justification for studying non-directly transmitted parasites still weak in the current version.
Answer: We provided more justification for our choice of study subjects (lines 128-133).

Methods section in general is nice and clear, some minor suggestions below:
Comments 1: L172-173: Suggest "macaques FROM DIFFERENT GROUPS were never" for clarity
Answer: We applied this comment to the manuscript (lines 174-175).

Comments 2: L222-223: There are typos in this sentence
Answer: We thank the reviewer for noticing the typos. The mentioned words were removed during revising.

Comments 3: L223-229: As the reviewer did, I found this rationale rather weak. I would suggest (in part at least) highlighting that the focus of the analysis is as much on the sampling effects as on examining why infection intensity is associated with network position.
Answer: We changed this part substantially (lines 234-244). We would like to add here that we do not investigate nor do we have the data for investigating "why" infection intensity is associated with network position. The study is correlational and determines whether and maybe in what social context (with the different social network measures) infection intensity is associated with network position.

Comments 4: L293: From the equation provided above it seems unlikely that EPGs were whole numbers – how did this factor into using negative binomial models?
Answer: We provided more information for clarification (lines 298-299). Basically, EPG are always rounded.

Comments 5: L301-302: Is the problem with including them in the same model not more related to covariance between them making the interpretability of coefficients challenging (given they are typically highly correlated).
Answer: We added this comment to the manuscript (lines 315-317).

Comments 6: L309: One question related to statistical model design is if individuals with more samples were more likely to have higher EPGs? That could lead to potentially problematic informative cluster sizes in the mixed model.
Answer: EPG is randomly distributed around a certain range related to the actual infection of the individual. More samples will lead to a more robust measurement of infection intensity, but it wouldn't lead to higher/lower value EPG for a specific individual. To ensure this is true, we did run a model that included the number of faecal samples collected from each individual, and it did not show a significant relationship with EPG, nor did it increase model fit.

Comments 7: L314-316: This text is odd here as you have already told us that you used zero-inflated models previously.
Answer: Thanks for pointing this out. It was a miss in the writing process that has now been rectified.

Some minor comments on the results:
Comments 1: L364: I would advocate for a bit more text descriptive text in this currently very short section!
Answer: We provided a more detailed description in this section (lines 381-387).

Comments 2: Tables 4-6: Given the null models are the same model minus the effect of the centrality measure it would perhaps be informative to provide the model estimates for each term included with just the dash/slash for the centrality measure that's not in the model?
Answer: Considering the focus on this study, we think it is not as informative to provide null model results in the main text but we provided the information in the supplementary material now (Table SI-4 ~ Table SI-8).

Comments 3: Tables 4-6: The information provided for random effects seems like it may be given incorrectly (this is worked out from the tables so may/may not be correct). It seems the authors have taken the standard deviation and variance directly from the model summary table, in which cas these are providing the same information (if you square the SD then this calculates the variance). You can calculate the standard error around random effect estimates but this is done in a different ways. Personally, I think it would be sufficient to provide the point estimate of the random effect variance, but the authors could provide the standard error also if preferred, Improve the clarity of references to the supplementary material (SP1 etc. is more confusing than it needs to be)
Answer: We thank the reviewer for pointing out this mistake. We now provide point estimates of the random effect variance instead of the previous information. We also provided more information when referencing to the supplementary materials (lines 206-207, 420-421; 455-456, 465-466).

Comments 4: L416-417: Looking at the supplementary tables it seems that this is written incorrectly. It looks like sub-sampling from the metrics dataframe produces results that are more similar to the original full analysis? This is a key finding and its important it is written clearly.
Answer: We thank the reviewer for noticing this unclearly written section. We have revised the section to avoid misunderstandings (lines 442-450).

Comments 5: L434-442: It would be good to do more (either here, the discussion or both) to integrate these findings with the slightly conflicting results obtained with targeted sub-sampling (e.g. the very rapid loss of statistically significant results with random sampling the networks vs. the result for strength for females persisting even though they make up a minority of the network).
Answer: We provided more description in the targeted sub-sampling result part (lines 442-450) and also more discussion (lines 594-613)

Overall the discussion was disappointing, it is very long and repetitive in places. There appeared to be left over text from the previous version that did not match well with the current results and other places where the response said changes had been made but they had not.

Comments 1: L450-455: These first two sentences are rather vague and also introductory in tone. The first sentence also seems at odds with both the introduction and literature in the field in not presenting that other studies have not found such a clear relationship
Answer: we removed the redundant sentences.

Comments 2: L487-488: This statement is not relevant given the stated mode of transmission of the parasites.
Answer: The sentence was removed while streamlining this section.

Comments 3: L489-491: As highlighted previously this is equally true of other social centrality measures and not specific to degree, and so is misleading to include as an argument to include specifically here.
Answer: We removed this argument.

Comments 4: L472-517: These sections are very focused on the association between infection intensity and network position being driven by transmission which is somewhat disingenuous with the mode of transmission of the parasites and ignores a more holistic view of the factors that may drive this relationship. It also ends up being rather repetitive and restricts space for discussion of the findings related to sampling (despite these being prominent in the title and abstract)
Answer: We streamlined this section, and discuss more substantially the issues of sampling. (lines 501-514)

Comments 5: L524: Would suggest “non-significant” rather than “insignificant”
Answer: The mentioned words were removed during revising.
Comments 6: L531-533: I don’t follow the logic of this argument
Answer: This argument was removed while revising the discussion.

Comments 7: L534-547: Some of this discussion is at odds with there being no statistically significant sex effect in the current models
Answer: We now provided the null models of each of the models in the supplementary materials, in which we can find statistically significant sex effects in the model of the juvenile-only network (Table SI-6).

Comments 8: L551-557: This argument is incorrect as currently presented. While it is possible that sub-sampling may produce noise that impacts the correlations, simply being in a smaller group doesn't impact the correlation strengths. Imagine a degree distribution in a larger group of [10,8,6,4,2]. The same individuals are measured at another time step and still have the degree distribution [10,8,6,4,2]. However, at a third time step they are in a smaller group might have the degree distribution [5,4,3,2,1]. While the values for degree are much lower in the third group the strength of the correlation is identical. While there are reasons that sub-samples of different sizes might change the strength of correlations, the magnitude of the numbers is not one of them.
Answer: This argument was removed.

Comments 9: L558-560: This argument can work but needs to be adjusted to take into account the previous point.
Answer: We adjusted this argument accordingly (lines 554-560).

Comments 10: L561-563: The topic sentence of this paragraph seems odd in the context of the new findings where strength and eigenvector centrality were more similar than strength and degree
Answer: We rephrased this section and removed the odd sentence (lines 547-564).

Comments 11: L569-571: Seems odd to focus on females when it was zero models in juveniles?
Answer: This sentence was removed while revising the discussion section (lines 547-564).

Comments 12: L576-585: This paragraph is rather repetitive of the results and lacks additional discussion
Answer: We removed the redundant sentences and added more discussion (lines 565-580).

Comments 13: L588-595: This direct comparison is a little misleading as targeted node removal was based on network position in the studies cited but a non-network trait in the current study (which could be close to random removal of nodes with respect to network position depending on the relationships between age, sex and social centrality). Another potentially relevant and interesting point related to this discussion about sub-sampling networks is that the impact on studies investigating infectious disease epidemiology may depend on the transmission mode itself (sometimes removing certain subsets may reduce noise).
Answer: we provided a further discussion on this section, which we hope reduces the confusion. We were discussing the potential effects of random removal, while one of these effects would be removing central nodes (lines 575-580).

In the supplementary material it would be ensure clear Figure/Table captions and check formatting to make sure it is correct (seemed to be some bolding missing at least). I found some of the current Table captions very similar and hard to distinguish from each other.

Answer: We checked and edited the supplementary information. We also now provided clearer captions in the supplementary information, and added bolded text to make it easier to distinguish each table/graph.