Re-review of the preprint, "Linking parasitism to network centrality and the impact of sampling bias in its interpretation."

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.

Major comments

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 infection 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.

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 centrality metrics for individual parasites.

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.

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.

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.

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.

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.