Peer Community In
Network Science

Centrality and parasite loads in sampled networks

Matthew Silk *based on peer reviews by* ***Krishna Balasubramaniam***, ***Quinn Webber*** *and 1 anonymous reviewer*

Zhihong Xu, Andrew J. J. MacIntosh, Alba Castellano-Navarro, Emilio Macanás-Martínez, Takafumi Suzumura, Julie Duboscq (2022) Linking parasitism to network centrality and the impact of sampling bias in its interpretation. bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Network Science.
https://doi.org/10.1101/2021.06.07.447302

Submitted: 20 June 2021, Recommended: 05 July 2022

Cite this recommendation as:
Silk, M. (2022) Centrality and parasite loads in sampled networks. *Peer Community in Network Science*, 100005.
https://doi.org/10.24072/pci.networksci.100005

Published: 05 July 2022
Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Networks provide an ideal tool to link social behaviour and infection in animal societies (White et al. 2017). A major focus of previous research has been on the links between social centrality and infection (Briard & Ezenwa 2021). But what happens when conclusions are drawn from sampled networks in which some individuals are not observed, or when studies focus on some individuals at the expense of others (e.g. adults versus juveniles or females versus males)? Xu et al. (2022) examine how focusing on different samples of individuals in a network analysis relating centrality to parasite load in Japanese macaques *Macaca fuscata* influence the conclusions drawn.

Xu et al. (2022) use faecal egg counts to estimate parasite loads of three environmentally transmitted parasites *Oesophagostomum aculeatum*, *Strongyloides fuelleborni* and *Trichuris trichiura* in a group of macaques on Koshima Island, Japan. After showing positive associations between parasite load and strength (the sum of an individual's connections) and eigenvector centrality (accounting for second-order connections) in a 1-metre proximity network, the authors explore how this result is impacted by focusing on only adult females, only juveniles or random sub-samples of the population. Their results indicate that the positive association persists more strongly in the adult female networks albeit with reduced statistical power to detect it. It is largely absent in juveniles (either based on their centrality in the full network or centrality in the juvenile-only network). Random removal of individuals from the network led to a rapid reduction in the ability to detect the same positive association between centrality and parasite load due to a combination of changes in individual centrality in re-sampled networks and reduced statistical power.

The timescale of network data collection and proximity networks studied are (likely) not fully relevant for the transmission of these parasites and social transmission of the parasites studied here is likely to be limited, there remain other reasons that we may expect correlations between sociality and infection (Ezenwa et al. 2016). Nevertheless, this is a useful contribution to the literature on sampling effects in animal networks, complementing existing work (Franks et al. 2010, Silk et al. 2015, Davis et al. 2018, Silk 2018). The results from considering different sub-samples of the group show the potential importance of carefully considering whether social network effects will be equivalently important for the whole population and which interactions will contribute either to promoting health or increasing the risk of infection. The results of random sub-sampling show how in small (within-group) networks such as these even small numbers of missing individuals could have substantial impacts on testing how traits are associated with the social network position of individuals.

The findings set up some interesting questions about how best to develop effective sampling designs in single-group studies such as these, or in how best to extend these types of projects across multiple groups (see also Silk 2018). Testing the generality of these findings across taxa with different social systems and infection prevalences or loads will also be a valuable next step for behavioural disease ecology.

References:

Briard L, Ezenwa VO. 2021. Parasitism and host social behaviour: a meta-analysis of insights derived from social network analysis. Anim. Behav. 172, 171-182. https://doi.org/10.1016/j.anbehav.2020.11.010

Davis GH, Crofoot MC, Farine DR. 2018. Estimating the robustness and uncertainty of animal social networks using different observational methods. Anim. Behav. 141, 29-44. https://doi.org/10.1016/j.anbehav.2018.04.012

Ezenwa VO, Ghai RR, McKay AF, Williams AE. 2016. Group living and pathogen infection revisited. Curr. Opin. Behav. Sci. 12, 66-72. https://doi.org/10.1016/j.cobeha.2016.09.006

Franks DW, Ruxton GD, James R. 2010. Sampling animal association networks with the gambit of the group. Behav. Ecol. Sociobiol. 64, 493-503. https://doi.org/10.1007/s00265-009-0865-8

Silk MJ, Jackson AL, Croft DP, Colhoun K, Bearhop S. 2015. The consequences of unidentifiable individuals for the analysis of an animal social network. Anim. Behav. 104, 1-11. https://doi.org/10.1016/j.anbehav.2015.03.005

Silk MJ. 2018. The next steps in the study of missing individuals in networks: a comment on Smith et al. (2017). Soc. Net. 52, 37-41. https://doi.org/10.1016/j.socnet.2017.05.002

White LA, Forester JD, Craft ME. 2017. Using contact networks to explore mechanisms of parasite transmission in wildlife. Biol. Rev. 92, 389-409. https://doi.org/10.1111/brv.12236

Xu Z, MacIntosh AJJ, Castellano-Navarro A, Macanás-Martinez E, Suzumura T, Dubosq J. 2022. Linking parasitism to network centrality and the impact of sampling bias in its interpretation. bioRxiv 2021.06.07.447302, ver. 6 peer-reviewed and recommended by Peer Community in Network Science. https://doi.org/10.1101/2021.06.07.447302

Reviews

Evaluation round #3

DOI or URL of the preprint: https://www.biorxiv.org/content/10.1101/2021.06.07.447302v4.article-info

Version of the preprint: V4

Authors' reply, 29 June 2022

Download author's reply

Decision by Matthew Silk, posted 28 March 2022

Some very minor final suggestions

The authors have done an excellent job of responding to the previous round of reviewer comments and the article now reads very well overall. I have a few very minor comments, suggestions and queries before I write my recommendation.

For the sections on sub-sampling:

It is not immediately clear why the intuitive explanation for results from female and juvenile networks is a combination of power and sampling (L446-448) – this has not been set out very well. The results would also seem to suggest that the relationship between social centrality and EPG is different for females and juveniles. For females there is a statistically significant association with both their position in the full network and that in the female only network, while for juveniles it is neither suggesting a biological difference here (less variation in juvenile EPGs? Juvenile EPGs depending on status of mother? Something else unrelated to the network?). Another curiosity here is that the results for females persist in both networks despite them being uncorrelated (at least according to Spearman rank). It could be this means that both overall connections and connections to other adult females are important? [or is it that the centrality measures are correlated when ranks aren't used – I can't remember the results before this was changed?]

I am also intrigued by the results on random sub-sampling and strength. It seems very striking that there is no (apparent) decline with an increasing number of removed individuals in Table 6 (one is apparent for eigenvector centrality albeit only just). Can I just check how the comparisons were made – was it a% of statistically significant effects in the same direction as the observed effect or any statistically significant effects? If only the latter can I suggest that you provide both options in the table? It might help explain this slightly odd pattern a little more clearly if reductions in statistically significant effects in the same direction are being compensated for by increases in statistically significant effects in the opposite direction (the results level out at the ~5% error rate expected). It's really striking that just removing ~2 individuals from the analysis has such a big effect and it would be good to see this explored even just a little more. Following this suggestion might also help point out that with smaller sub-samples it is possible (albeit rare) to have results that are statistically significant effects in the wrong direction.

Now just a few quick suggestions on writing/typos that I noted when carefully reading the article.

L35-36: Suggest "partly an effect of sampling the incomplete network"

L324: Would be good to clarify/clearly state that the zero-inflated part of the model was intercept only

Figure 2: is it worth considering a log scale for the y axis given the distribution of the data. It may not be clearer but it may be worth checking?

L444: Suggest changing "than" to "to"

L530-531: These results are not clear from the main text because of the unusual choices for the presentation of the Tables – might be worth stating explicitly in the relevant section as is done for adult females already

L543-544: This statement currently has the potential to be misleading. I would suggest changing "in general" to "can"

SI Tables (where relevant): In the captions it would be good to integrate the two separate sentences about which entries are highlighted in bold as this is quite confusing currently.

Evaluation round #2

DOI or URL of the preprint: https://doi.org/10.1101/2021.06.07.447302
Version of the preprint: V3

Authors' reply, 18 March 2022

Download author's reply

Decision by Matthew Silk, posted 23 November 2021

Revisions required

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.

Please see the attached file for details **Download recommender's annotations**

Reviewed by Quinn Webber, 10 November 2021

This is the second time I have been asked to review this manuscript. I appreciate the authors effort to address the comments from the previous round of revisions and I don't really have anything else to add. I would suggest the authors would be well primed to submit their paper and likely have a positive outcome. Good luck!

Reviewed by anonymous reviewer 1, 17 November 2021

Download the review

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.1101/2021.06.07.447302

Authors' reply, 22 October 2021

We posted the final formatted version on the preprint server.
Download author's reply

Decision by Matthew Silk, posted 29 July 2021

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.

- 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

focussed on. I very much agree with these comments and feel the paper would benefit from a more focussed introduction (and that the discussion could be shortened also).

- 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.

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?

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.

Specific comments:

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.

I would suggest using "eigenvector centrality" throughout in full to be clear that you are referring specifically to the centrality measure.

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

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?

It would be good to be clear on how the social network is weighted as soon as it is initially described.

Is the Elo-rating used in the analysis the score itself or a ranking based on the score?

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.

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 https://doi.org/10.1016/j.tree.2008.10.008 and https://doi.org/10.1101/2021.05.03.442487)

I find Figure 1 very confusing – I don't entirely understand what it is showing and how?

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.)

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.

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.

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.

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.

P26 Paragraph 2: you mention lower model parameter estimates, it would be good to find a way to show this alongside statistical significance.

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.

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

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

Reviewed by Krishna Balasubramaniam, 17 July 2021

In this study, the authors address a critical gap in our understanding of how animals' social network connectedness with all other members of their group irrespective of demographic attributes (e.g. age or sex class) influences parasite risk. They do so by systematically implementing knock-out simulations *in silica*, to remove sub-sets of individuals of specific age-sex classes to determine their effects on parasite risk. I found the study interesting, well written, and easy to follow. I nevertheless have a number of comments, mostly minor but one or two more major considerations (more so than comments) that the authors would need to address. I provide an attachment of the PDF of the manuscript with these comments added as notes adjacent to the relevant, highlighted parts of the manuscript.

I look forward to reading a revised version that has taken these comments into consideration.

Download the review

Reviewed by Quinn Webber, 22 July 2021

Download the review

Reviewed by anonymous reviewer 1, 20 July 2021

Download the review