General Response

We thank the editor and reviewers for their feedback and patience as we revised this manuscript. In light of the comments of the reviewers we opted to analyze a number of additional complementary networks to add to the supplementary materials and provide additional context for our results. Due to the data processing requirements of analyzing an additional 2-4 networks we opted to move the entire data pipeline from using cytoscape plug-ins to calculate metrics and SAS to analyze data to using igraph and lme4 in R for social network and statistical analyses. Through this process we discovered a number of discrepancies in the resulting metrics (specifically the weighted centrality metrics) between these two programs and others offered in R (tnet, sna; see summary table of correlations below). While unweighted metrics were highly consistent across programs, weighted metrics showed a great degree of discrepancy. Because we believe that both the number and weight of connections are important when exploring the influence of social relationships, we attempted to track down the source of these discrepancies to attempt to understand how the underlying calculations might influence the inferences drawn when using these different metrics. Unfortunately, in the end we decided that including and explaining the different weighting algorithms would be beyond the scope of this paper and decided to only include centrality metrics from our original paper that showed high correlations (r > 0.7) between software. Therefore, we have removed most weighted metrics (with the exception of degree) and all information centrality metrics from the current manuscript. We appreciate your patience as we made the adjustments to our analyses to improve the rigor and reproducibility of our results. Given the addition of 4 new networks and the complexity of our analytical strategy we have added a figure to the paper to illustrate how the network filtering was done and added a figure to the supplementary materials with a flow-chart of analytical procedures.

	Multiplex	igraph	tnet (closeness only)	sna(info only)
Cytoscape	Degree	0.964		
	DegreeWeight	1		
	Betweenness	0.864		
	BetweennessWeight	0.65		
	Eigenvector	0.919		
	EigenvectorWeight	0.619		
	Closeness	1	0.417	
	ClosenessWeight	0.397	0.136	
	Information			0.201
	InformationWeight			-0.002
	Clustering/Transitivity	0.944		
tnet	Closeness	0.417		
	ClosenessWeight	0.592		

Correlation matrix for centrality metrics from the Multiplex networks calculated using cytoscape (CytoNCA and Network Analyzer; original manuscript) and R packages igraph, tnet, and sna.

	Uniplex	igraph	tnet (weighted closeness only)	sna(info only)
Cytoscape	Degree	0.974		
	DegreeWeight	1		
	Betweenness	0.834		
	BetweennessWeight	0.589		
	Eigenvector	0.961		
	EigenvectorWeight	0.898		
	Closeness	0.954	0.284	
	ClosenessWeight	0.472	0.295	
	Information			0.317
	InformationWeight			0.302
	Clustering/Transitivity	0.965		
tnet	Closeness	0.31		
	ClosenessWeight	0.182		

Correlation matrix for centrality metrics from the Uniplex networks calculated using cytoscape (CytoNCA and Network Analyzer; original manuscript) and R packages igraph, tnet, and sna.

Based on multiple questions from the reviewers regarding how the networks we present compare to other networks (e.g., all grooming, all contact sitting, multiplex and uniplex contact sitting/huddling) we have adjusted the paper to explicitly include the evaluation and characterization of these networks. Because grooming is known to have multiple functions we keep our discussion focused on comparing multiplex and uniplex grooming networks but we also analyze and present comparisons of grooming and contact sitting as well as multiplex vs. uniplex contact sitting networks. We calculated the network metrics the same way as the uniplex and multiplex grooming networks and used the same model building strategy. We also added this material to the supplementary materials as we agree with the reviewers that is important for contextualizing our results in the broader literature. Ultimately, our results remain the same with nearly all candidate models finding that high centrality in multiplex grooming networks is associated with lower inflammation while high centrality in uniplex grooming networks is associated with higher inflammation. In one candidate model, the grooming and contact sitting networks appear (i.e., TNF-a). We believe the lack of robust effects from these added networks supports the idea that multiplex networks are not simply a proxy for either grooming or contact sitting networks. We highlight key findings from these additional analyses, however, in our response to reviewers where relevant.

Round #1

Reviews

Reviewed by anonymous reviewer, 24 Nov 2022 09:02

This manuscript presents a study linking social structure to measures of individual health in captive primates. The authors show that individuals who were more connected when considering “multiplex” networks were healthier, while the opposite effect was found in “uniplex” networks. I found the results potentially interesting, but I have serious reservations regarding the network definitions and the analysis.

The most major issue is that multiplex networks are a type of multilayer networks, where there are several types of interactions between the same set of nodes/individuals. Here, both multiplex and uniplex interactions were analyzed as single-layer networks, with both types of networks using only one interaction type (lines 204-206). Basically, instead of analyzing one grooming network where some dyads also huddled, the authors split the same grooming interaction into two different networks. The authors did not offer any biological reasoning for defining the two networks, and I have never seen such analysis. This type of analysis makes the interpretation of the results very difficult. The networks were divided based on one interaction that was not included in the analysis (huddling). I suggest the authors rethink the analytical approach. The differences between the two networks can come from individual traits, such as sex, kinship, or rank. In that case these factors should be included in the analysis.

This method was chosen to ask the question of whether the diversity of affiliative behaviors used in a relationship contains information about the quality or function of a dyad’s relationship. Hinde (1976a) highlights in his foundational paper on social relationships and social structure that a key component determining the features of a relationships is the co-occurrence of behaviors within a relationship (p. 6). Additionally, Hinde further highlights in a second paper (1976b) that the diversity or breadth of interactions contained within a relationship (including use of the distinction between one type of interaction, uniplex, or multi-stranded/multiplex interactions) contains crucial information regarding the nature of a relationship. While acknowledged to be important very few methods have been developed to explore the importance of the diversity of behaviors in assessing relationships quality or function (Silk, Cheney, Seyfarth, 2013, except see Fischer 2017). We have attempted to clarify this perspective and justification for our analysis in the introduction. The general approach we take, in which we take a single behavioral network and filter it into two networks, and then examine the networks separately is not novel. For example, Croft (2009) separated association networks into strong and weak network and used both networks in the further analysis. The novelty in our approach is in using a second behavior (contact sitting) as a metric of relationship “strength” rather than a cutoff of dyadic strength of association as was done by Croft. To aid in the interpretation of our filtered networks compared to more traditional approaches of examining all interactions of a single behavior as separate networks we have added results from that approach to the manuscript. Finally, we provide data on how kinship and rank are associated with overall network structure in the first section of our results. Indeed, the multiplex network is more likely to include kin grooming, and although both networks tend to have grooming directed up the hierarchy, the uniplex network contains dyads of more disparate ranks. However, in our

analyses of the predictors of inflammation, we do include status related predictors in our models (e.g. rank and dominance certainty) and they do not explain the observed patterns.

Hinde 1976a: https://www.jstor.org/stable/2800384

Hinde 1976b: J. Child Psychol. Psychiat., Vol. 1

Other comments:

Line 73 and on: At some point I asked myself if this discussion relates only to non-human animals. It is worthy to clarify that, because the beginning of the introduction referenced both humans and other animals.

The general idea that the quality of the relationship (including whether relationships are multidimensional) does pertain to humans and nonhuman animals. Measurement of these concepts in animals, however, presents an additional challenge, however as we cannot ask them about how they perceive their relationships or what function each relationship serves in their life. We present this method as an additional way to use their behavior to attempt to tease apart the potential functions and meaning of different relationships. We have edited the introduction in an attempt to more clearly delineate where we are referring to humans vs. nonhuman animals or nonhuman primates.

Line 100: Multiplex networks should be defined.

This is now defined.

Line 129: Were there dyads that only huddled? How were these cases treated?

Yes, some individuals huddled (now termed contact sit) but did not groom. We now include explicit analyses of these networks in the paper. Overall, neither multiplex nor uniplex contact sit based centralities are predictors of inflammation.

Line 133: Modularity and clustering are not the same. A network may be highly modular but not clustered, if in each module, or sub-group, clustering is low.

We thank the reviewer for this comment and have edited our text to be more accurate.

Lines 164-176: This paragraph should be part of the introduction. It provide general information on the species and why it was chosen.

We have chosen to leave the details of model choice in the methods. We attempted to include it in the introduction but given our recent revisions we felt that it fit best at the beginning of the methods in a new section describing our study system. We feel this is more appropriate because while many of our predictions and behaviors are dependent on our choice of study system, the more general concept that integrating information from multiple behaviors to understand qualities of relationships is not limited to our study system. We believe the wealth of knowledge about our study system just makes is a good system in which to test this approach and method.

Line 188: How were individuals identified?

Information on how animals were identified and criteria for observer ID reliability on lines 215-216.

Figure 1: Is there any text in each node? It's impossible to read so either enlarge it or delete it.

We have replaced this figure.

Statistics: Was there any treatment of the fact that network measures are not independent? Usually, some sort of null model based on permutations of the data is being used to compare to the observed patterns.

Generally, we have seen network permutation approaches suggested when network measures are outcomes of generalized linear modeling due to the violation of the independence assumption. Given that our network measures are predictors in these models, these techniques do not apply. Also, there is recent evidence to suggest that even in the case where network measures are outcomes, network permutation approaches may not be superior to traditional regression approaches (Hart et al., 2022, Weiss et al., 2020) We initially attempted a network permutation approach to contextualize the difference between our multiplex and uniplex networks, however, methods in this application are not well developed and we chose to use simple standard parametric and non-parametric approaches.

Weiss, Franks, Brent, Ellis, Silk, Croft 2020. https://doi.org/10.1111/2041-210X.13508

Hart, Weiss, Brent, Franks 2022. https://doi.org/10.1007/s00265-022-03254-x

Dominance: Perhaps I have missed it, but was there a correlation between dominance and measures of centrality in the multiplex network? I am wondering if the main result could be, at least to some extent, be explained by dominance. Readers would benefit from data addressing this question, for example by adding rank to the selected models and testing its effect.

Our second step in running the glmms was to evaluate whether age, dominance rank, dominance certainty, or physiological sampling order predicted the outcomes. Based on AIC scores these demographic factors were not significant predictors of inflammation and cannot explain the results we

present. The correlation between rank and eigenvector centrality in uniplex and multiplex networks is variable and there do not appear to be consistent differences across groups in this metric which makes it unlikely this would be an explanation for our results.

Line 261: Multidimensional vs unidimensional network metrics?

This should read multiplex and uniplex and has been corrected.

Line 271: Why undirected?

We have added some clarification on lines 334-337.

Figure S1: What is the number of dyads without interactions?

This was added to the histograms in supplementary figure 1.

Line 265 Table S2 or table S3?

We have double checked our table and figure references.

Reviewed by anonymous reviewer, 06 Dec 2022 12:44

I really like the idea of the paper and also the aim to explore social relationships in more complex way. The paper is very well written, the text is clear and comprehensive. The authors cite relevant literature sources. The specific aim of the study is clearly presented. The sample size in terms of groups and individual females is impressive. The paper is overall very informative in terms of discrepancies in previous studies and methods used.

My major concern stems if the presented results of the comparison of uni vs multiplex networks is really representative of the two types of networks or if it is more related to the way how data were filtered for each network. In the methods section Affiliative network analysis the authors stated that "…and a network containing edges for dyads that groomed but were never seen huddling (uniplex affiliative relationships)." However if one would do an independent study using grooming for uniplex network one will not filter out edges based on the huddling behavior thus the resulting uniplex network based on grooming will be different of that presented by the authors. So it might not be completely fair to say that the uniplex network is "grooming only" because is grooming only and never huddling", thus grooming ties between individuals who also huddled were not included in this grooming network. This approach in my opinion leads to bigger differences in the two networks which is not necessarily connected to multiplex vs uniplex networks but to the way how the edges were filtered.

In other words if the uniplex network would be based on all grooming edges (without filtering out the dyads that huddle as well) would there be same differences between the two networks? I understand

that using the approach the authors made the two networks rather independent in terms the data used but then I think you are comparing rather two different types of relationships one based on grooming and huddling and one based on grooming only and never huddling but not necessarily uniplex and multiplex networks.

This also has consequences when comparing results of previous studies which in many cases would use "full" grooming uniplex networks which is not completely comparable with the uniplex grooming network in this study.

We appreciate both reviewers questions regarding how our filtered networks compare with other approaches that have been more commonly employed. Therefore, we have constructed a number of new networks and explicitly include analysis using more traditional single behavior networks (e.g. all grooming, all huddling (now termed contact sitting) and uniplex vs multiplex contact sitting). Because grooming behavior is the most commonly used behavior to infer social relationships in this species we keep the focus of the manuscript the same but add information and analysis of these other networks to provide greater context of how to place our results in the greater literature. We agree that uniplex is an imperfect term to describe a network which is determined by the presence of one behavior (i.e. grooming) and absence of a second behavior (i.e., contact sitting) we use this terminology based on its use in research on sociality (Hinde 1976, Berkman 2000). Overall, the separate behavioral networks partially reproduce the results from our filtered networks (i.e., results in the grooming vs. contact sit networks for TNF-a analysis roughly parallel with grooming effects similar to uniplex affiliation and contact sit effects similar to multiplex affiliation), the resulting model is not a better fit to the data compared to the analysis of the multiple and uniplex affiliative networks. We believe this supports that these effects are not an artifact of the model filtering technique but instead are real effects of the types of relationships these networks likely represent.

Hinde 1976a: https://www.jstor.org/stable/2800384

Hinde 1976b: J. Child Psychol. Psychiat., Vol. 1

Berkman et al. 2000. Social Science & Medicine.

Additional comments:

Although authors stated in conclusion that is not clear if using different behaviors like proximity would provide similar results, still I would like to know why huddling was selected for this study? Why not the more traditional proximity or body contact? It makes me think that the authors in fact did expect different results with huddling behavior.

We have adjusted the language we use in the paper to term our behavior contact sitting since that more accurately represents what is described in our ethogram. Huddling is included in contact sitting but other forms of contact sitting (i.e. stationary body contact) that may not be described as huddling are included in our data collection (see lines 217-218). We chose this behavior because it is a commonly seen affiliative behavior in our captive groups. We did not choose proximity in our groups because the meaning of proximity in the context of captive groups with relatively higher population density than in the wild would be more difficult to interpret. However, we do not think that this approach is specific to

the behaviors we used and that this approach could be applied to whatever affiliative behaviors are relevant to a species given the environment they live in. We have tried to clarify this perspective throughout the introduction.

Is it possible from this study to say if huddling might be also beneficial just as a thermoregulatory behavior? In other words, it is not beneficial for one's health because of the social relationship it represents but because one is not cold?

I think this is a reasonable hypothesis for a uniplex contact sitting network. We did create these networks however this might not be the best dataset to answer that question. While we are able to calculate a uniplex contact sitting network, I would expect that the association of this network with thermoregulation would be dependent on the season (i.e., primarily seen during cold months). Since our groups were observed in spring and fall we may not have the best data to answer this question.

Just a nitpicky question which is not specific to this paper but Is the word "multiplex" warranted if there are two types of interactions only?

We have struggled to identify a clear term to describe our networks. We settled on the terms uniplex and multiplex largely because they are the terms used to describe the theoretical role that the diversity of interaction types might have on a relationship by Hinde (1976b) and Berkman et al (2000).

Hinde 1976b: J. Child Psychol. Psychiat., Vol. 1

Reviewed by Tamao Maeda, 06 Dec 2022 07:40

The paper proposed very interesting approach in order to assess how the quality of affiliative relationships affects the health status of Rhesus macaques. Overall, I found the study well-designed and nicely presented. Although the results seem robust and convincing, you will find several questions and comments below requesting additional input from the authors to justify and clarify some analytical approaches and expressions in the manuscript.

Comments and questions:

Would you please explain why you did not examine the opposite version of the network, i.e. creating huddling network and using grooming to filter it? The authors used huddling degree in the model to exclude the possibility that huddling behavior itself has an effect on the inflammation, but to be more precise, I thought it is better that you also compare only huddling and, huddling and grooming network.

Due to this question, and other raised by the other reviewers we have examined these networks as well as networks including all grooming and all contact sitting interactions. As stated elsewhere in our responses we find that using grooming and contact sitting networks either do not show the same patterns (e.g., IL-6) or produce the same results but with poorer model fit (e.g., TNF-a) as the models using multiplex and uniplex grooming networks. Therefore, we believe these analysis supports the idea

that multiplex networks are not simply a proxy for either grooming or contact sitting networks. We did some exploration of multiplex and uniplex contact sitting networks, but again these did not show strong associations with inflammation.

Line 80: I suggest you to change "assess the quality of affiliative social relationships" to something like "assess the overall amount/frequency of multiple behaviors". To my understanding, DSI is just a method of averaging all the correlated dyadic interactions, thus it is indicating the overall frequency of affiliative behavior and not actually measuring the quality (and I guess it is the point of your discussion).

Thank you for this comment. It is a measure of frequency of affiliative interaction but is often used to infer relationship quality. We have adjusted the language to clarify this. We are, however, trying to make the point that relationship quality may not only be how often a dyad affiliates, but also the types or diversity of affiliative behaviors they engage in.

Line 109: It is better to specify what is "huddling" behavior here, as it is not a very common term (especially for non-primatologist).

We changed the terminology to call it contact sitting since that more accurately reflects the breadth of behaviors recorded and have added clarification to this point on lines 217-218.

Line 113-115: "Therefore to understand the impact of social relationships on individual health and fitness, we must examine the cumulative impact of an individual's diversity and breadth of social relationships to really understand their potential influence on health."

I felt this part is duplicating the meaning. Please revise.

This section has been revised.

Line 187: Did you also observe other types of direct affiliative behavior? As you mentioned "Affiliation in primates takes many forms, including grooming, huddling, proximity, embracing, and less commonly coalitionary support" in Introduction, I presume that huddling and grooming are the main forms of direct affiliations, but I wondered if there is a possibility that other affiliative behavior potentially affects the inflammation.

Unfortunately for this dataset we only had access to these two affiliative behaviors, however, we believe that examination of other behaviors in the future would further shed light on what characterizes different types of affiliative relationships (e.g., what is the impact of an affiliative relationship that does or does not also contain agonism). We are not arguing contact sitting and grooming are special behaviors, but that affiliating via multiple species typical affiliative behaviors might carry meaning regarding the quality of that relationship. We also believe that the advancements in multilayer networks may also provide another tool to further answer these questions. We have added a section discussion multilayer network techniques to the introduction to explicitly address this.

Line 213: To my understanding, information centrality is based on concept of efficient propagation of information in a group. I rarely see it is used in the context similar to this study. Please justify the reason of using it here. In addition, the authors explain that information centrality indicates the social cohesion, but you should put the citation or more detailed explanations.

Due to the issues raised above relating to consistency in metric measurements across different software programs we have opted to remove information centrality-based measures from the paper entirely.

Table 2: As I mentioned above, I thought it is better to put citation for each interpretation of the network metrics.

We have included citations.

Figure 1. I cannot read the labels of each node. Also, the image resolution seems not good enough. Please make it clearer.

We have replaced this figure.