Spurious inference when comparing networks

Damien R. Farine^{a,b,c,d,1} and Lucy M. Aplin^{c,d,e}

Comparing networks is challenging (1–3). van Leeuwen et al. (4) use variation in networks among 4 chimpanzee groups as evidence for cultural differences in social behavior. However, social networks can be shaped by many factors. The authors control for group size, an obvious correlate of many network metrics, but do not adequately consider alternative processes that can generate differences in social networks among groups. Many are unrelated to social decisions. For example, habitat geometry can maintain highly repeatable between-generation network community structure in wild populations of songbirds (5).

We illustrate why van Leeuwen et al.'s study cannot be used as evidence for culturally driven differences among chimpanzee groups using simulations (6). Briefly, we randomly placed N group members (with group sizes matching van Leeuwen et al.'s data) in an arena by drawing X and Y coordinates from a uniform distribution (forming an area matching each group's enclosure). We then calculated interindividual distances, using 1 unit (relative to side length) as our measure of "being connected" and a chain rule to define parties (repeating this process 10 times to generate 10 observations per group). We applied the statistical testing approaches described by van Leeuwen et al.—constructing models with and without group identity as random effects, and using postnetwork permutation tests. We repeated this procedure 100 times to estimate type I error rates. Our aim was not to replicate exactly the data produced by van Leeuwen et al.'s observations, but instead to show that simple nonsocial processes can generate highly significant and spurious results when using incorrect hypothesis-testing methods. Our results confirm this; every test following the method used by van Leeuwen et al. found significant differences between groups, with their methods generating anticonservative P values and extremely high rates of false positives (~94.3%; Table 1, van Leeuwen et al. method).

While comparison across networks remains challenging, one solution is to use permutation tests (3). Specifically, prenetwork permutation tests that swap observations among individuals within groups can be used to generate realistic "randomized" networks and thus produce a distribution of expected test statistics in which the social process of interest has been removed (7–9). Postnetwork (or node) permutations are not useful for comparing networks because they do not change the network-level distribution of metrics, making them only useful for within-network contrasts (8). Applying prenetwork permutation tests on exactly the same simulated data produced reliable type I errors rates (~3.6%; Table 1, Prenetwork permutation method), because prenetwork permutations effectively controlled for a simple nonsocial causal effect in our simulated data: density. We note, however, that density is unlikely to be the sole driver of between-group difference found by van Leeuwen et al.

We conclude that, as it stands, the study by van Leeuwen et al. does not provide robust evidence for significant cultural differences in social behavior. Using appropriate hypothesis-testing methods has the potential to completely change the conclusions of van Leeuwen et al.'s study. More broadly, even if significant between-group differences remain, attributing them to culture requires following groups over longer timescales to rule out alternative mechanisms arising from interactions between particular individuals. Individual differences in personality are well known to shape group-level collective behaviors (10), but this does not necessarily lead to cultures that outlast the individuals involved.

Acknowledgments

D.R.F. and L.M.A. were funded by the Max Planck Society and by the Deutsche Forschungsgemeinschaft Centre of Excellence 2117 "Centre for the Advanced Study of Collective Behaviour" (ID 422037984). We thank two anonymous reviewers for constructive comments that have improved our Letter.

^aDepartment of Collective Behavior, Max Planck Institute of Animal Behavior, 78547 Konstanz, Germany; ^bDepartment of Biology, University of Konstanz, 78464 Konstanz, Germany; ^cCentre for the Advanced Study of Collective Behaviour, University of Konstanz, 78547 Konstanz, Germany; ^dEdward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; and ^eCognitive and Cultural Ecology Research Group, Max Planck Institute of Animal Behavior, 78315 Radolfzell, Germany

Author contributions: D.R.F. and L.M.A. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Published under the [PNAS license](#).

Data deposition: Code and supplementary information for this paper have been deposited in the Edmond Repository (<https://dx.doi.org/10.17617/3.2b>).

¹To whom correspondence may be addressed. Email: dfarine@ab.mpg.de.

Published online August 13, 2019.

	van Leeuwen et al. method				Prenetwork permutation method			
Test	Lower P	Upper P	% False positives ($P < 0.05$)	% False positives ($P < 0.001$)	Lower P	Upper P	% False positives ($P < 0.05$)	% False positives ($P < 0.001$)
Party size	4.37×10^{-13}	0.017	98	91	0.232	0.99	0	0
Proximity	2.44×10^{-29}	1.58×10^{-6}	100	100	0.0143	0.995	6	3
Strength	0	0.001	100	99	0	0.990	11	4
Eigenvector centrality	0	0.009	100	52	0.015	0.990	7	2
Reach	0	0	100	100	0.005	1	9	3
Clustering	0	0.03	98	83	0.049	0.985	3	2
Affinity	0	0	100	100	0.020	1	10	1

^a“Lower *P*” and “Upper *P*” represent the 2.5% and 97.5% quantiles of the *P* value distribution, respectively. “% False positives” is the proportion of simulations that generated *P* values that were significant at different significance thresholds.

- 1 D. P. Croft, R. James, J. Krause, *Exploring Animal Social Networks* (Princeton University Press, Princeton, 2008).
- 2 B. C. M. van Wijk, C. J. Stam, A. Daffertshofer, Comparing brain networks of different size and connectivity density using graph theory. *PLoS One* **5**, e13701 (2010).
- 3 B. S. Anderson, C. Butts, K. Carley, The interaction of size and density with graph-level indices. *Soc. Networks* **21**, 239–267 (1999).
- 4 E. J. C. van Leeuwen, K. A. Cronin, D. B. M. Haun, Population-specific social dynamics in chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11393–11400 (2018).
- 5 D. R. Farine, B. C. Sheldon, Stable multi-level social structure is maintained by habitat geometry in a wild bird population. *bioRxiv:085944* (8 February 2019).
- 6 D. R. Farine, L. M. Aplin, Code and supplementary information for “Spurious inference when comparing networks.” Edmond Repository, 10.17617/3.2b. Deposited 23 March 2019.
- 7 L. Bejder, D. Fletcher, S. Bräger, A method for testing association patterns of social animals. *Anim. Behav.* **56**, 719–725 (1998).
- 8 D. R. Farine, A guide to null models for animal social network analysis. *Methods Ecol. Evol.* **8**, 1309–1320 (2017).
- 9 H. Whitehead, *Analyzing Animal Societies* (University of Chicago Press, Chicago, 2008).
- 10 L. M. Aplin, D. R. Farine, R. P. Mann, B. C. Sheldon, Individual-level personality influences social foraging and collective behaviour in wild birds. *Proc. Biol. Sci.* **281**, 20141016 (2014).