

1 **Common datastream permutations of animal social network data are not**
2 **appropriate for hypothesis testing using regression models**

3

4 Michael N. Weiss^{1,4*}, Daniel W. Franks^{2†}, Lauren J. N. Brent¹, Samuel Ellis¹, Matthew J. Silk³, and
5 Darren P. Croft^{1†}

6

7 ¹ Centre for Research in Animal Behaviour, College of Life and Environmental Sciences, University of
8 Exeter, Exeter, U.K. EX4 4QG

9 ² Department of Biology and Department of Computer Science, The University of York, York, U.K.
10 YO10 5DD

11 ³ Centre for Ecology and Conservation, University of Exeter in Cornwall, Penryn, U.K. TR10 9FE

12 ⁴ Center for Whale Research, Friday Harbor, WA, U.S.A. 98250

13

14 * Corresponding author: mw607@exeter.ac.uk

15 † These authors contributed equally

16

17 **Running title:** Regression and permutations in social networks

18

19

20

21

22 **Abstract**

- 23 1. Social network methods have become a key tool for describing, modelling, and testing
24 hypotheses about the social structures of animals. However, due to the non-independence
25 of network data and the presence of confounds, specialized statistical techniques are often
26 needed to test hypotheses in these networks. Datastream permutations, originally
27 developed to test the null hypothesis of random social structure, have become a popular
28 tool for testing a wide array of null hypotheses. In particular, they have been used to test
29 whether exogenous factors are related to network structure by interfacing these
30 permutations with regression models.
- 31 2. Here, we show that these datastream permutations typically do not represent the null
32 hypothesis of interest to researchers interfacing animal social network analysis with
33 regression modelling, and use simulations to demonstrate the potential pitfalls of using this
34 methodology.
- 35 3. Our simulations show that utilizing common datastream permutations to test the
36 coefficients of regression models can lead to extremely high type I (false-positive) error rates
37 (> 30%) in the presence of non-random social structure. The magnitude of this problem is
38 primarily dependent on the degree of non-randomness within the social structure and the
39 intensity of sampling
- 40 4. We strongly recommend against utilizing datastream permutations to test regression models
41 in animal social networks. We suggest that a potential solution may be found in regarding
42 the problems of non-independence of network data and unreliability of observations as
43 separate problems with distinct solutions.

44

45 **Keywords:** group living; null hypothesis significance testing; null model; permutation test;
46 randomisations; regression; social networks

47

48 **Introduction**

49 Social structure, defined as the patterning of repeated interactions between individuals (Hinde
50 1976), represents a fundamental characteristic of many animal populations with far-reaching
51 consequences for ecology and evolution, including for gene-flow, social evolution, pathogen
52 transmission, and the emergence of culture (Kurvers et al., 2014). The last two decades have seen
53 widespread adoption of social network methods in animal behaviour research to quantify social
54 structure (Webber & vander Wal, 2019). The network framework is appealing because it explicitly
55 represents the relationships between social entities from which social structure emerges (Hinde,
56 1976), and thus allows tests of hypotheses about social structure at a variety of scales (individual,
57 dyadic, group, population). Social networks can be based on direct observations of interactions, or
58 inferred from other data types, such as groupings of identified individuals (Franks et al., 2010), GPS
59 tracks (Spiegel et al., 2016), proximity loggers (Ryder et al., 2012), or time-series of detections
60 (Psorakis et al., 2012).

61 The analysis of animal social network data presents a statistical challenge. Specifically, two separate
62 issues must be addressed. First, network data are inherently non-independent, thus violating the
63 assumptions of independent observations inherent to many commonly used statistical tests. Second,
64 factors outside of social structure, such as data structure and observation bias, may influence the
65 structure of observed animal social networks, potentially leading to both type I and type II errors in
66 statistical tests (Croft et al., 2011).

67 To address the problem of non-independence, a wide array of statistical tools have been developed,
68 primarily in the social sciences. These methods include permutation techniques that allow for
69 hypothesis testing in the presence of non-independence. These permutations normally test
70 relationships between exogenous variables and network properties, such as the presence and
71 strength of social ties, or the centrality of nodes in the network. These methods typically build

72 empirical null distributions by randomly assigning the location of nodes in the network, while
73 holding the network structure constant (“node-label permutations”). The resulting null distribution
74 maintains the non-independence inherent to the network while breaking any relationship that exists
75 between network structure and potential covariates (Dekker et al., 2007).

76 While these methods are useful for dealing with the issue of non-independence, they do not address
77 the second issue, from which studies of animal social systems in particular often suffer. Because the
78 methods developed in the social sciences only permute the final constructed network, they do not
79 inherently account for common biases in the collection of the raw observational data used to
80 construct the final network. These biases may be introduced by the method of data collection (e.g.
81 group-based observations), individual differences in identifiability, or demographic processes (James
82 et al., 2009). For example, consider a situation where researchers are interested in differences in
83 social position between sexes, but females are more cryptic and thus observed with a lower
84 probability. This would lead to incorrect inferences due to biases in the observed network structure
85 that are unrelated to the true social processes of interest (Farine, 2017). To deal with these
86 problems, a suite of alternative permutation procedures has been developed. Rather than
87 permuting the final network, these methods permute the raw data used to construct the network.
88 These methods are therefore sometimes referred to as “datastream permutations.” The goal is to
89 construct permuted datasets that maintain structures of the original data that may influence the
90 observed network structure (e.g. the number of times individuals were observed and the sizes of
91 observed groups), while removing the social preferences that underpin the social network (Farine &
92 Whitehead, 2015).

93 The original datastream permutation technique for animal social data was proposed by Bejder et al.
94 (1998), based on the procedure outlined by Manly (1997) for ecological presence-absence data.
95 Bejder et al.’s procedure was designed to test whether a set of observed groupings of identified
96 animals showed signs of non-random social preferences. This procedure permutes a group-by-

97 individual matrix, where rows are groups and columns are individuals, with 1 representing presence
98 and 0 indicating absence. The algorithm finds 2 by 2 “checkerboard” submatrices, with 0s on one
99 diagonal and 1s on the other, that can be “flipped” (i.e. 0s replaced with 1s and vice versa). These
100 flips maintain row and column totals (the group size and observations per individual, respectively),
101 but permute group membership. In biological terms, matrices generated with this procedure
102 represent the null hypothesis that individuals associated completely at random, given the observed
103 distribution of group sizes and the number of sightings per individual.

104 Refinements of this method were later developed that constrained swaps within time periods,
105 classes of individual, or locations (Whitehead et al., 2005). One alteration also controls for
106 gregariousness, and allows for permutation of association data not constructed using group
107 membership (Whitehead, 1999). Controlling for gregariousness and sighting history is possible when
108 each sampling period is represented as a square matrix, where 1 indicates that individuals associated
109 in that period and 0 indicates no association. In this format, the data can be permuted in a way that
110 maintains the number of associates each individual had in each sampling period (Whitehead, 1999).

111 In recent years, datastream permutation methods have been developed that can handle more
112 complex data structures, such as GPS tracks (Spiegel et al., 2016), time-series of detections (Psorakis
113 et al., 2015), and focal follow data (Farine, 2017). All of these methods have in common that they
114 essentially randomise raw observations of social association (or interactions) data and thus remove
115 social structure while maintaining most other features of the data, including features potentially
116 causing biased measurements of social structure. They thus provide a robust null distribution to test
117 for non-random social structure in a dataset, which is a key step in understanding the behavioural
118 ecology of wild populations.

119 Many empirical studies and methodological guides have suggested interfacing these null models
120 with other statistical techniques, particularly regression models (including ordinary least squares,
121 generalized linear models, and mixed-effects models), to test hypotheses about network structure.

122 The logic of this recommendation is that permutation-based null models allow researchers to
123 account for sampling issues when testing hypotheses using these common statistical models.
124 However, it is important to recognize the limitations of this approach, and to think carefully about
125 the null hypothesis that these methods specify. In common datastream permutation null models,
126 the null hypothesis specified is that the population's social structure is random, once we control for
127 the structure of the data and other confounds. For a particular quantity of interest, such as edge
128 weights, node centralities, or differences between networks, this null hypothesis can be equivalently
129 stated as proposing that all variance in a given value or network metric is due to data structure,
130 confounds, and residual variance. In network terminology, this null hypothesis is a random network,
131 within a set of constraints. This null hypothesis is by design, because this form of permutation
132 procedure was specifically created to test for non-random social structure. However, we feel there
133 has been a lack of consideration about whether this null hypothesis is appropriate in other contexts,
134 such as regression modelling. We show here that these procedures do not provide an appropriate
135 null hypothesis for testing the null hypotheses of regression models.

136 Consider the basic linear model:

$$137 \qquad Y = \beta X + \varepsilon \qquad (1)$$

138 where Y is a response variable, X is a matrix of predictor variables, ε is the error term, and β is a
139 vector of estimated coefficients. We are typically interested in testing the null hypothesis $\beta = 0$,
140 representing no relationship between the response Y and the predictor(s) X . In permutation based
141 hypothesis testing procedures, researchers specify this null distribution by randomising either X or Y ,
142 often with constraints, thus maintaining the distribution of values but breaking any covariance
143 between the variables (Anderson & Robinson, 2001). This is the logic behind traditional node-label
144 permutation tests of regression in social networks (Croft et al., 2011).

145 Datastream permutations, however, do something very different, which is inappropriate for testing
146 this hypothesis. By permuting the data underlying network measures and then re-calculating the

147 response variable, these procedures change the distribution of Y , instead of breaking relationships
148 between the variables. If the network has non-random social structure, even structure entirely
149 unrelated to X , then we will typically see a reduction in the variance of Y as we permute the raw
150 data. When Y has a larger variance in the observed data than in the permutations, more extreme
151 values of β are more likely to occur in the observed data, even if the null hypothesis is true. This
152 procedure is therefore likely to result in much higher rates of false-positive type I error than is
153 acceptable (Figure 1).

154 Changes in variance between the observed and permuted data is more than just a technical issue.
155 There is a fundamental problem with this approach when it comes to testing hypotheses using
156 regression models. When researchers fit regression models to predict network properties from
157 exogenous variables, the null hypothesis they will be testing against can be stated as “the variation
158 in network structure is not related to the exogenous variable.” This, however, is not the null
159 hypothesis tested by the commonly used datastream permutation methods. Rather, the null
160 hypothesis that is proposed by these datastream permutations could be stated as “the degree of
161 variation in network structure and its relationship to the exogenous variable are both due to random
162 interactions of individuals within constraints.” The researcher cannot disentangle the null hypothesis
163 of no relationship between the network and the predictor from the null hypothesis of random social
164 structure. In other words, a significant result from this procedure could be due to a relationship
165 between the predictor and the network, or because individuals do not interact at random, whether
166 or not the true social structure is related to the predictor. This fundamental mismatch between the
167 null hypothesis of interest and that tested by the datastream permutation algorithm makes tests of
168 regression models using this procedure nearly uninterpretable.

169 Here, we demonstrate the problems that occur when combining datastream permutations of animal
170 social network data with regression using two simulated scenarios. In these scenarios, we generate
171 datasets with simple non-random social structure. We then introduce a random exogenous variable

172 that has no relationship to social structure, and test for a relationship between network structure
173 and this variable with linear models, using datastream permutations to determine statistical
174 significance. We show that even in the absence of any true relationship between exogenous
175 variables and social structure, datastream permutations are highly prone to producing significant p -
176 values when social structure is non-random. We caution against using these datastream
177 permutations to test the coefficients of regression models, and we discuss possible solutions and
178 alternative methods for regression analysis in social networks.

179

180 **Methods**

181 *General framework*

182 To illustrate the problems with using datastream permutations to test the coefficients of regression
183 models, we carried out simulations across two different scenarios, reflecting common research
184 questions in animal social network analysis. The first scenario simulates a case in which researchers
185 are interested in whether dyadic covariates (e.g. kinship or phenotypic similarity) influences the
186 strength of social bonds, which we will refer to as a case of “dyadic regression”. The second scenario
187 simulates a case when researchers are interested in how a quantitative individual trait (e.g. age or
188 personality) influences individual network position, which we refer to as “nodal regression.”

189 While the methods of network generation differ slightly for each scenario, the general steps are the
190 same:

- 191 1. Generate observations of a network in which the quantity of interest (edge weight or node
192 centrality) has inherent variation.
- 193 2. Generate values for a trait that are unrelated to this variation.
- 194 3. Fit a linear model with the network property as the response variable and the trait as the
195 predictor

- 196 4. Create permuted versions of the observed network via a common datastream permutation
- 197 5. Compare the original model's coefficient to those fit to the permuted data to calculate a p -
- 198 value

199 For each simulation, we perform 200 runs, with varying parameter values (Table 1). For each run of

200 both simulations, we produce four outputs. The first two outputs are the coefficient of the fitted

201 linear model and the p -value from the permutation test. The other two outputs give information

202 about the characteristics of the dataset. The first is the standard deviation of the response variable

203 (either the edge weights or weighted degrees), indicating the degree of non-randomness in the

204 social structure, and the second is the average number of sightings per individual, a common

205 measure of sampling effort in social network studies.

206

207 *Dyadic regression: Does similarity in a trait predict the strength of social relationships?*

208 In our first simulation, we investigate the case in which the researcher is interested in the influence

209 of a dyadic predictor (such as similarity in phenotype or kinship) on the rates at which dyads

210 associate or interact. Our simulation framework is heavily inspired by those of Whitehead & James

211 (2015) and Farine & Whitehead (2015). We simulate a population of N individuals, and assign each

212 dyad an association probability p_{ij} from a beta distribution with mean μ and precision ϕ ($\alpha = \mu\phi$, $\beta =$

213 $(1-\mu)\phi$). By assigning association probabilities in this way, we create non-random social preferences

214 in the network, and thus larger variance in edge weights than would be expected given random

215 association (Whitehead et al., 2005).

216 We then simulate t sampling periods. For simplicity, individuals are sighted in each sampling period

217 with a constant probability σ , and associations between dyads where both individuals are sighted

218 occur with probability p_{ij} . We then build the observed association network by calculating dyadic

219 simple-ratio indices:

220
$$SRI_{ij} = \frac{X_{ij}}{D_{ij}} \quad (2)$$

221 Where X_{ij} is the total number of sampling periods in which i and j were observed associating, and D_{ij}
222 is the total number of periods in which either i or j was observed (including periods where they were
223 observed, but did not associate with any individuals).

224 We then assign each individual a trait value from a uniform distribution (0,1). We do not need to
225 specify what this trait represents for our simulation, but it could represent any quantitative trait
226 used as a predictor in social network studies (age, personality, cognitive ability, dominance rank,
227 parasite load, etc.). Note that the trait value is generated after the observations of association and
228 has no influence on any network property.

229 We then fit the linear model:

230
$$SRI_{ij} = \beta_0 + \beta_1 |trait_i - trait_j| + \varepsilon \quad (3)$$

231 and save the estimate of β_1 . We compare this coefficient to a null model generated using the
232 sampling period permutation method proposed by Whitehead (1999). There are several algorithms
233 available to perform these swaps. We use the “trial swap” procedure described by Miklós & Podani
234 (2004) and suggested for social network studies by Krause et al. (2009). For each trial, this procedure
235 chooses an arbitrary 2 by 2 submatrix of the lower triangle within a random sampling period. If a
236 swap is possible, it is performed (and symmetrized), otherwise the matrix stays at its current state.
237 These steps when the matrix is not changed are referred to as “waiting steps.” This algorithm is ideal
238 because it ensures that the Markov chain samples the possible matrices uniformly, while other
239 algorithms that do not include waiting steps exhibit biases in their sampling of the possible matrices
240 (Miklós & Podani, 2004). We generate 10,000 permuted datasets for each simulation, with 1,000
241 trial swaps between each permutation, and re-fit our linear model to each permuted dataset,
242 recording the coefficient. We then use this distribution of coefficients to calculate the p -value of the

243 linear model's coefficient. Across the 200 runs, we vary the parameters of the simulation by drawing
244 μ , ϕ , N , σ , and t randomly using Latin hypercube sampling (Table 1).

245

246 *Nodal regression: Do individual traits influence network centrality?*

247 We next investigate the same concept in the context of nodal regression. This form of analysis tests
248 whether some individual attribute is related to variation in network position. This is perhaps the
249 most common use of datastream permutation null models for testing the significance of linear
250 regression coefficients in animal social networks (e.g. Cowl et al., 2020; Poirier & Festa-Bianchet,
251 2018; Zeus et al., 2018). For simplicity, we focus on weighted degree, which is simply the sum of an
252 individual's edge weights.

253 In this simulation, we consider the case where networks are derived from patterns of shared group
254 membership ("gambit of the group"). This form of data collection is extremely common in animal
255 social network studies, and was the basis for the original datastream null model developed by Bejder
256 et al. (1998).

257 The framework for this simulation is based on that used by Firth et al. (2017). We simulate G
258 observations of groupings in a population of N individuals. Each group is assigned a group size S from
259 a discrete uniform distribution on $[1, M]$. We assign each individual a preference for a particular
260 group size P from a truncated normal distribution with mean $(1+M)/2$, standard deviation σ , lower
261 bound 0, and upper bound M . Higher values of σ will therefore lead to higher variation in
262 gregariousness in the population. For each group g , membership is determined by sampling S_g
263 individuals without replacement, with individual sampling probability determined by the size of
264 group g and each individual's group size preference:

265
$$P(i \text{ in } g) \propto \frac{1}{(S_g - P_i)^2} \quad (4)$$

266 This gives the simulation the property that individuals with higher assigned gregariousness scores
267 tend to be seen in larger groups, and vice versa. This leads to non-random differences in
268 gregariousness (and thus weighted degree) between individuals. We then calculate the association
269 network, again using the SRI:

$$270 \quad SRI_{ij} = \frac{X_{ij}}{X_{ij} + Y_i + Y_j} \quad (5)$$

271 Where X_{ij} is the number of groups in which the dyad was seen together, and Y_i and Y_j are the number
272 of groups in which only i or only j were seen, respectively. After calculating the network, we
273 determine each individual's weighted degree. We again generate a trait value for each individual at
274 random from a uniform distribution on (0,1) and fit the linear model

$$275 \quad \sum_j SRI_{ij} = \beta_0 + \beta_1 \text{trait}_i + \varepsilon \quad (6)$$

276 and again save the estimate of β_1 . We compare this coefficient to random coefficients fit to
277 networks generated using the group-based permutation procedure proposed by (Bejder et al.,
278 1998). This procedure again sequentially permuted the observed dataset, while maintaining the size
279 of each group and the number of groups per individual. We again use the trial swap method to
280 perform these permutations, generating 10,000 permuted datasets with 1,000 trials per
281 permutation, and derived p -values in the same way as above. We vary the parameters of this
282 simulation by using Latin hypercube sampling to draw values of N , M , G , and σ (see Table 1 for
283 ranges).

284

285 *Analysis*

286 We use the outputs of the simulations primarily to derive overall type I error rates for both
287 scenarios, calculated as the portion of runs in which a p -value less than 0.05 was obtained. We
288 further investigated the sensitivity of these results to non-random social structure, sampling effort,
289 and population size. Previous work suggests that the sensitivity of datastream permutation

290 techniques are highly dependent on variation in social structure and sampling intensity (Whitehead,
291 2008). We use binomial generalized linear models to summarize how population size, response
292 variance, and sampling intensity influence the probability of false positives. We further analyse these
293 relationships qualitatively using conditional probability plots.

294

295 **Results**

296 *Dyadic regression*

297 The overall type I error rate for the dyadic regression case was high, with 35% of runs giving false
298 positive results (70 out of 200 runs). Sensitivity analysis suggested that the most important factors
299 influencing type I error rate in our simulations were the average number of sightings per individuals
300 and the variance of association probabilities. As the average number of sightings increased, so did
301 the false positive rate ($\beta = 0.012 \pm 0.004$, $z = 3.149$, $p = 0.002$, Figure 2a). Similarly, networks with
302 higher variance in edge weights experienced higher type I error rates ($\beta = 8.35 \pm 8.93$, $z = 2.37$, $p =$
303 0.02 , Figure 2b). There was a less clear, but statistically significant relationship between network size
304 and type I error rates, with networks of larger size typically having lower type I error rates ($\beta = -$
305 0.014 ± 0.007 , $z = -2.02$, $p = 0.04$, Figure 2c).

306

307 *Nodal regression*

308 The nodal regression case resulted in even higher type I error rates than the dyadic case, with almost
309 half of runs giving false positive results (95 out of 200 runs; 47.5%). The rate of type I errors was
310 strongly influenced by the variance in weighted degree; as the standard deviation of the response
311 increased, so too did the false positive rate ($\beta = 1.18 \pm 0.50$, $z = 2.34$, $p = 0.019$, Figure 3a). In
312 contrast, as the size of the network increased, the false positive rate decreased, although never
313 approaching the target false positive rate of 0.05 in our simulations ($\beta = -0.02 \pm 0.01$, $z = 2.89$, $p =$

314 0.004, Figure 3c). In this simulation, the number of sightings per individual did not appear to
315 significantly influence the type I error rate ($\beta = 0.018 \pm 0.013$, $z = 1.43$, $p = 0.153$, Figure 3b). This
316 may be because, in networks with few groupings but high sightings per individual, there were fewer
317 possible permutations of the observed network, and therefore the permuted networks were more
318 similar to the original network.

319

320 **Discussion**

321 These two simple simulated scenarios show that the commonly used datastream permutation
322 procedures for animal social network data produce extremely high and thus unacceptable false-
323 positive rates when applied to regression models. This is because datastream permutations do not
324 generate appropriate null distributions for testing the significance of model coefficients. We
325 therefore strongly warn against using this procedure.

326 We now turn to some potential solutions to this problem that may still facilitate inference in these
327 situations. This is not intended to be a comprehensive guide to hypothesis testing in social networks,
328 and other solutions are certainly possible. We encourage other researchers to consider these and
329 other possible solutions.

330

331 *Transforming the response variable*

332 If variations in social behaviour are present in the network, datastream permutations undesirably
333 eliminate social influence and reduce the variance in the response of a regression model. A potential
334 fix for this problem is to simply standardize the variable of interest, perhaps to have a mean 0 and
335 standard deviation 1 (Z-scores), and to repeat this process for all permutations. While this is likely to
336 reduce type I error rates, we caution that this is a quick fix of the symptom and does not address the

337 cause: that the null model generated does not test the desired hypothesis. We therefore do not
338 consider this to be an adequate solution in itself.

339

340 *Alternative test statistics*

341 Another potential solution could be found in using a test statistic other than the coefficients from
342 the model. In the context of node-label permutations such as MRQAP, pivotal statistics such as the t
343 statistic for ordinary least squares regression typically perform better than raw coefficient values
344 (Dekker et al., 2007). While previous authors have recommended against using the t or Z statistic,
345 because they represent deviations from a parametric distribution rather than direct features of the
346 data (Farine, 2017), such statistics could experience a lower type I error rate than those reported
347 here. However, as in the case of transforming the response, this does not address the larger issue of
348 an incorrectly specified null hypothesis. We therefore do not view the adoption of alternative test
349 statistics from regression models compared to datastream permutations as an appropriate solution.

350

351 *Separating the issue of non-independence from biases in the data*

352 We suggest that the way forward for hypothesis testing in animal social networks is to recognize that
353 the problems of non-independence of network measures and the influence of data structure
354 underlying networks are separate issues, requiring separate solutions. Not all animal network data
355 will be subject to the issue of unreliability (e.g., in cases where sampling is balanced across subjects
356 and relevant contexts) and in some instances the data may be complete and unbiased. In these
357 cases, node permutations or other statistical network models are appropriate (Croft et al. 2011). In
358 instances where structure in the data needs to be controlled we propose two potential methods;
359 other solutions are certainly possible, and we encourage further work on this matter.

360 The first method would utilize generalized affiliation indices (GAIs; Whitehead & James, 2015) or
361 similar corrections to account for confounding variables that may influence observed edge weights.
362 GAIs fit the observed data associations or interactions as the response in a binomial or Poisson
363 generalized linear model, with confounding factors such as space use, sightings frequency, or joint
364 gregariousness as predictors. The residuals of this model are then used as measures of affiliation, as
365 they reflect the difference between observed and expected association rates given the confounding
366 factors. While a flexible and appealing approach, GAIs require that potential confounds be properly
367 specified in terms of dyadic covariates, and that the relationship between confounds and edge
368 weights be linear. This second issue, however, may be addressed by fitting generalized additive
369 models (GAMs), where relationships are represented by smooth functions.

370 A related, but slightly different approach would be to incorporate confounds in the inferential model
371 itself. Rather than deriving new edge weights via GAIs, if researchers identify likely confounds and
372 summarize them quantitatively for each dyad or individual, these could be used directly in the final
373 model. Where potential non-linearity between confounds and responses exist, data transformations,
374 polynomials, and smooth functions may present a possible solution.

375 We feel that these methods have the potential to address the current issue that we have identified
376 and we strongly encourage new work to explore and validate these approaches. It is important to
377 note that the methods we propose are only useful if the question of interest is about the structure
378 of social affinity, rather than the empirical pattern of encounters between individuals. If, instead,
379 researchers are interested in the actual rates of contact (as is the case in disease research and
380 studies of social learning), this approach may not be appropriate. Extensions of recent work using
381 hidden state modelling may be more appropriate for disentangling true association patterns when
382 detections are potentially biased or imperfect (Gimenez et al., 2019).

383

384 *Building better null models*

385 The problems we have identified here arise because the commonly used null models for animal
386 societies do not generate datasets representing the null hypothesis of interest in a regression
387 setting. These models were specifically designed to test the null hypothesis of random social
388 structure, not the null hypothesis that aspects of social structure are unrelated to exogenous factors.
389 An obvious way forward would be the development of permutation procedures that generate
390 datasets that correctly represent the relevant null hypothesis. In the case of dyadic regression, these
391 datasets would maintain the structure of the data (e.g. sightings per individual, associations per
392 sampling period, spatial patterns of observations), randomise identities of associated individuals,
393 and simultaneously preserve the variance in edge weights. In the case of nodal regression, permuted
394 datasets would maintain the same (or at least a similar) distribution of individual centrality within
395 the network, in addition to structural confounds such as the size of groups, sightings per individual,
396 and timing of sightings. The design of such procedures is far from trivial, and is beyond the scope of
397 this paper, but we suspect that the development of algorithms that simultaneously maintain aspects
398 of data structure and features of the social system will be an important area of methodological
399 research going forward.

400

401 **Conclusion**

402 The development of permutation techniques that control for sampling biases while maintaining
403 temporal, spatial, and structural aspects of the raw data is an important development in the study of
404 animal social systems, and we suspect that these procedures will remain a key tool for hypothesis
405 testing in ecology and evolution. However, a lack of consideration regarding the matching up of the
406 null hypothesis being tested with the null model being generated using datastream permutations
407 has led to unwarranted application of these techniques, particularly in the context of hypothesis
408 testing using regression models.

409 We recommend that researchers think critically and carefully about the null hypothesis they wish to
410 test using social network data, and ensure that the null model they specify does in fact represent
411 that hypothesis. We suspect that in most cases, the null hypothesis of random social structure will
412 clearly not be appropriate, and therefore traditional datastream permutations will not be a viable
413 approach. We hope that our discussion of this issue and the results of our simulations will result in
414 reconsideration of how researchers employ null models when analysing animal social networks,
415 promote further research and discussion in this area, and lead to the development of procedures
416 that correctly specify null hypotheses and allow robust inference in animal social network studies.

417

418 **Acknowledgements**

419 We would like to thank colleagues at the Centre for Research in Animal Behaviour, particularly the
420 members of the CRAB Social Network Club, for extremely valuable discussion that greatly improved
421 this manuscript. DPC, DWF and SE acknowledge funding from NERC (NE/S010327/1). LJB
422 acknowledges funding from the NIH (R01AG060931; R01MH118203).

423

424 **Author contributions**

425 MNW conceived of the project and designed the simulations, with input from all authors. All authors
426 contributed to drafting the manuscript.

427

428 **Data availability**

429 This study used no empirical data. All code necessary to reproduce our analysis is included in the
430 online supplementary material.

431

432 **References**

- 433 Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian and New*
434 *Zealand Journal of Statistics*, 43(1), 75–88. <https://doi.org/10.1111/1467-842X.00156>
- 435 Bejder, L., Fletcher, D., & Bräger, S. (1998). A method for testing association patterns of social
436 animals. *Animal Behaviour*, 56(3), 719–725. <https://doi.org/10.1006/anbe.1998.0802>
- 437 Cowl, V. B., Jensen, K., Lea, J. M. D., Walker, S. L., & Shultz, S. (2020). Sulawesi Crested Macaque
438 (*Macaca nigra*) Grooming Networks Are Robust to Perturbation While Individual Associations
439 Are More Labile. *International Journal of Primatology*, 41(1), 105–128.
440 <https://doi.org/10.1007/s10764-020-00139-6>
- 441 Croft, D. P., Madden, J. R., Franks, D. W., & James, R. (2011). Hypothesis testing in animal social
442 networks. In *Trends in Ecology and Evolution* (Vol. 26, Issue 10, pp. 502–507). Elsevier Current
443 Trends. <https://doi.org/10.1016/j.tree.2011.05.012>
- 444 Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and
445 autocorrelation conditions. *Psychometrika*, 72(4), 563–581. [https://doi.org/10.1007/s11336-](https://doi.org/10.1007/s11336-007-9016-1)
446 [007-9016-1](https://doi.org/10.1007/s11336-007-9016-1)
- 447 Farine, D. R. (2017). A guide to null models for animal social network analysis. *Methods in Ecology*
448 *and Evolution*, 8(10), 1309–1320. <https://doi.org/10.1111/2041-210X.12772>
- 449 Farine, D. R., & Whitehead, H. (2015). Constructing, conducting and interpreting animal social
450 network analysis. *Journal of Animal Ecology*, 84(5), 1144–1163. [https://doi.org/10.1111/1365-](https://doi.org/10.1111/1365-2656.12418)
451 [2656.12418](https://doi.org/10.1111/1365-2656.12418)
- 452 Firth, J. A., Sheldon, B. C., & Brent, L. J. N. (2017). Indirectly connected: simple social differences can
453 explain the causes and apparent consequences of complex social network positions.
454 *Proceedings of the Royal Society B: Biological Sciences*, 284(1867), 20171939.
455 <https://doi.org/10.1098/rspb.2017.1939>

- 456 Franks, D. W., Ruxton, G. D., & James, R. (2010). Sampling animal association networks with the
457 gambit of the group. *Behavioral Ecology and Sociobiology*, *64*(3), 493–503.
458 <https://doi.org/10.1007/s00265-009-0865-8>
- 459 Gimenez, O., Mansilla, L., Klaich, M. J., Coscarella, M. A., Pedraza, S. N., & Crespo, E. A. (2019).
460 Inferring animal social networks with imperfect detection. *Ecological Modelling*, *401*, 69–74.
461 <https://doi.org/10.1016/j.ecolmodel.2019.04.001>
- 462 Hinde, R. A. (1976). Interactions , Relationships and Social Structure. *Man*, *11*(1), 1–17.
- 463 James, R., Croft, D. P., & Krause, J. (2009). Potential banana skins in animal social network analysis. In
464 *Behavioral Ecology and Sociobiology* (Vol. 63, Issue 7, pp. 989–997). Springer.
465 <https://doi.org/10.1007/s00265-009-0742-5>
- 466 Krause, S., Mattner, L., James, R., Guttridge, T., Corcoran, M. J., Gruber, S. H., & Krause, J. (2009).
467 Social network analysis and valid Markov chain Monte Carlo tests of null models. *Behavioral*
468 *Ecology and Sociobiology*, *63*(7), 1089–1096. <https://doi.org/10.1007/s00265-009-0746-1>
- 469 Kurvers, R. H. J. M., Krause, J., Croft, D. P., Wilson, A. D. M., & Wolf, M. (2014). The evolutionary and
470 ecological consequences of animal social networks: Emerging issues. In *Trends in Ecology and*
471 *Evolution* (Vol. 29, Issue 6, pp. 326–335). Elsevier Ltd.
472 <https://doi.org/10.1016/j.tree.2014.04.002>
- 473 Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.).
474 Chapman & Hall.
- 475 Miklós, I., & Podani, J. (2004). Randomization of presence-absence matrices: Comments and new
476 algorithms. *Ecology*, *85*(1), 86–92. <https://doi.org/10.1890/03-0101>
- 477 Poirier, M. A., & Festa-Bianchet, M. (2018). Social integration and acclimation of translocated
478 bighorn sheep (*Ovis canadensis*). *Biological Conservation*, *218*, 1–9.
479 <https://doi.org/10.1016/j.biocon.2017.11.031>

- 480 Psorakis, I., Roberts, S. J., Rezek, I., & Sheldon, B. C. (2012). Inferring social network structure in
481 ecological systems from spatiotemporal data streams. In *Journal of the Royal Society Interface*
482 (Vol. 9, Issue 76, pp. 3055–3066). Royal Society. <https://doi.org/10.1098/rsif.2012.0223>
- 483 Psorakis, I., Voelkl, B., Garroway, C. J., Radersma, R., Aplin, L. M., Crates, R. A., Culina, A., Farine, D.
484 R., Firth, J. A., Hinde, C. A., Kidd, L. R., Milligan, N. D., Roberts, S. J., Verhelst, B., & Sheldon, B. C.
485 (2015). Inferring social structure from temporal data. *Behavioral Ecology and Sociobiology*,
486 69(5), 857–866. <https://doi.org/10.1007/s00265-015-1906-0>
- 487 Ryder, T. B., Horton, B. M., van den Tillaart, M., Morales, J. D. D., & Moore, I. T. (2012). Proximity
488 data-loggers increase the quantity and quality of social network data. *Biology Letters*, 8(6),
489 917–920. <https://doi.org/10.1098/rsbl.2012.0536>
- 490 Spiegel, O., Leu, S. T., Sih, A., & Bull, C. M. (2016). Socially interacting or indifferent neighbours?
491 Randomization of movement paths to tease apart social preference and spatial constraints.
492 *Methods in Ecology and Evolution*, 7(8), 971–979. <https://doi.org/10.1111/2041-210X.12553>
- 493 Webber, Q. M. R., & vander Wal, E. (2019). Trends and perspectives on the use of animal social
494 network analysis in behavioural ecology: a bibliometric approach. *Animal Behaviour*, 149, 77–
495 87. <https://doi.org/10.1016/j.anbehav.2019.01.010>
- 496 Whitehead, H. (1999). Testing association patterns of social animals. *Animal Behaviour*, 57(6), F26–
497 F29. <https://doi.org/10.1006/anbe.1999.1099>
- 498 Whitehead, H. (2008). Precision and power in the analysis of social structure using associations.
499 *Animal Behaviour*, 75(3), 1093–1099. <https://doi.org/10.1016/j.anbehav.2007.08.022>
- 500 Whitehead, H., Bejder, L., & Ottensmeyer, C. A. (2005). Testing association patterns: Issues arising
501 and extensions. *Animal Behaviour*, 69(5). <https://doi.org/10.1016/j.anbehav.2004.11.004>

502 Whitehead, H., & James, R. (2015). Generalized affiliation indices extract affiliations from social
503 network data. *Methods in Ecology and Evolution*, 6(7), 836–844. [https://doi.org/10.1111/2041-](https://doi.org/10.1111/2041-210X.12383)
504 210X.12383

505 Zeus, V. M., Reusch, C., & Kerth, G. (2018). Long-term roosting data reveal a unimodular social
506 network in large fission-fusion society of the colony-living Natterer’s bat (*Myotis nattereri*).
507 *Behavioral Ecology and Sociobiology*, 72(6), 1–13. <https://doi.org/10.1007/s00265-018-2516-4>

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523 **Figure captions**

524

525 **Figure 1.** Example of the mechanism by which datastream permutations may lead to false positives
526 in linear regression. In the original network, there is variation in strength among individuals driven
527 by differences in gregariousness (represented by node size in the social networks). Individuals are
528 assigned a trait value (represented by colour in the social network) unrelated to their network
529 position. By chance, there is a slight negative relationship between network strength and trait value
530 in the observed network. After several permutations, there is a reduction in the variance in the
531 strength of individuals in the permuted network, and thus the magnitude of the relationship is
532 reduced. The bottom histogram shows the distribution of null coefficients after 10,000 permutations
533 (black), and the coefficient from the original linear model (red).

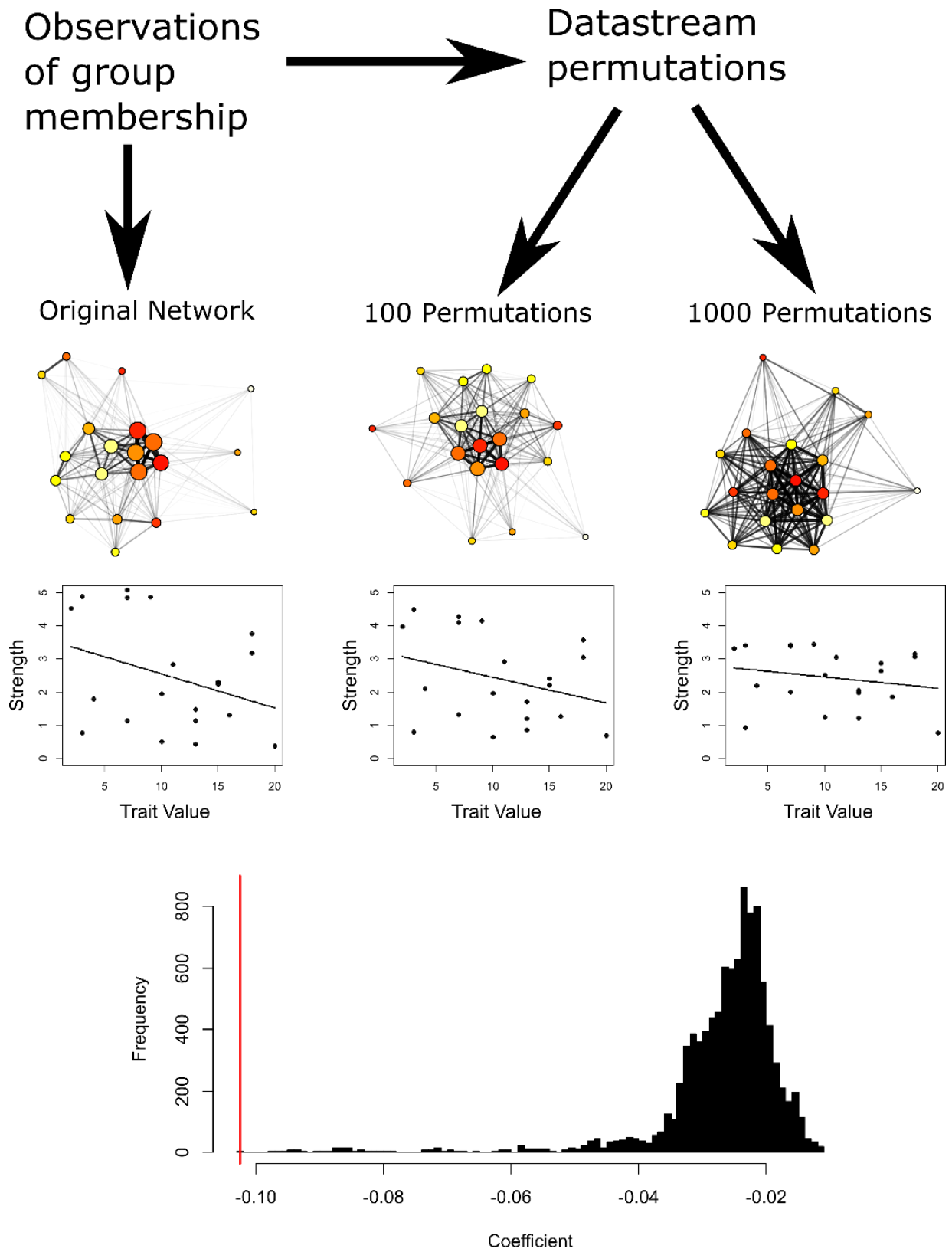
534

535 **Figure 2.** Conditional probability plots from dyadic regression simulation. Points indicate results of
536 individual simulation runs (1 = significant p -value, 0 = non-significant p -value). Red lines are
537 smoothed condition probabilities of a significant p -value. Dotted line indicates target type I error
538 rate of 0.05.

539

540 **Figure 3.** Conditional probability plots of type I error rates for the nodal regression simulation. Points
541 indicate the outcome of individual runs (1 = significant p -value at 0.05, 0 = non-significant p -value).
542 Red lines are smoothed conditional probabilities of a significant p -value. Dotted lines indicate the
543 target error rate of 0.05.

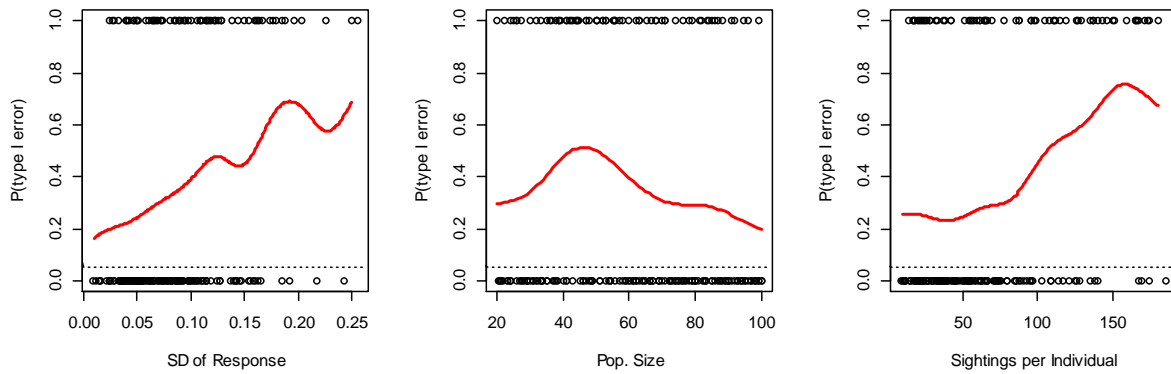
544 **Figure 1**



545

546

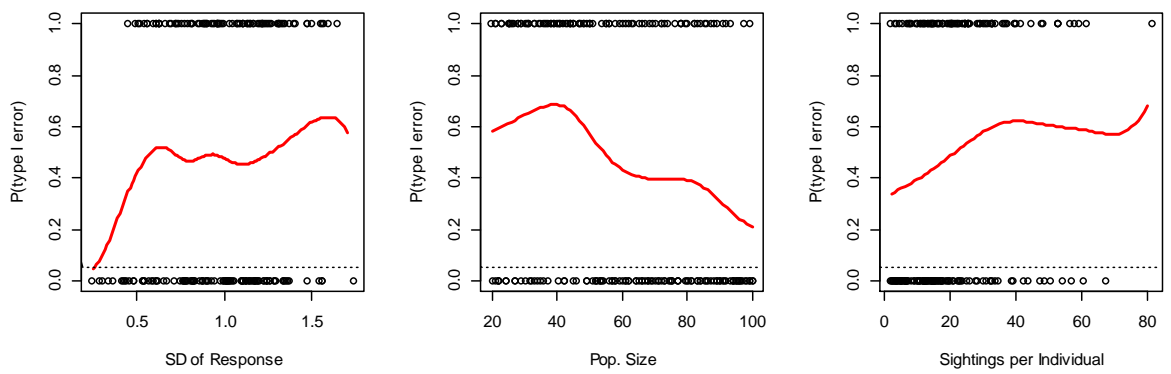
547 **Figure 2**



548

549

550 **Figure 3**



551

552

553

554

555

556

557

558

559 **Table 1.** Ranges for varied parameters used in simulations

Parameter	Meaning	Dyadic	Nodal	Range
N	Number of individuals in population	✓	✓	20 – 100
μ	Mean association probability	✓		0.01 – 0.5
t	Number of sampling periods	✓		20 – 200
ϕ	Precision of beta distribution for association probabilities	✓		1 – 10
o	Observation probability per sampling period	✓		0.1 – 1
G	Number of observed groupings		✓	20 – 500
M	Maximum grouping size		✓	5 – 10
σ	Standard deviation of group size preference		✓	0.1 – 2.0

560