

This is a repository copy of *Common datastream permutations of animal social network data are not appropriate for hypothesis testing using regression models*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/165210/>

Version: Accepted Version

Article:

Weiss, Michael N., Franks, Daniel Wayne orcid.org/0000-0002-4832-7470, Brent, Lauren J N et al. (3 more authors) (2020) Common datastream permutations of animal social network data are not appropriate for hypothesis testing using regression models. *Methods in ecology and evolution*. ISSN 2041-210X

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

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

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 animal social networks. In particular, they
29 have been used to test whether exogenous factors are related to network structure by
30 interfacing these 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, if used to indicate whether a relationship exists between
36 network structure and a covariate, datastream permutations can result in extremely high
37 type I error rates, in some cases approaching 50%. In the same set of simulations, traditional
38 node-label permutations produced appropriate type I error rates ($\sim 5\%$).
- 39 4. Our analysis shows that datastream permutations do not represent the appropriate null
40 hypothesis for these analyses. We suggest that potential alternatives to this procedure may
41 be found in regarding the problems of non-independence of network data and unreliability
42 of observations separately. If biases introduced during data collection can be corrected,
43 either prior to model fitting or within the model itself, node-label permutations then serve
44 as a useful test for interfacing animal social network analysis with regression modelling.

45

46 **Introduction**

47 Social structure, defined as the patterning of repeated interactions between individuals (Hinde
48 1976), represents a fundamental characteristic of many animal populations with far-reaching
49 consequences for ecology and evolution, including for gene-flow, social evolution, pathogen
50 transmission, and the emergence of culture (Kurvers et al., 2014). The last two decades have seen
51 widespread adoption of social network methods in animal behaviour research to quantify social
52 structure (Webber & vander Wal, 2019). The network framework is appealing because it explicitly
53 represents the relationships between social entities from which social structure emerges (Hinde,
54 1976), and thus allows tests of hypotheses about social structure at a variety of scales (individual,
55 dyadic, group, population). Social networks can be based on direct observations of interactions, or
56 inferred from other data types, such as groupings of identified individuals (Franks et al., 2010), GPS
57 tracks (Spiegel et al., 2016), proximity loggers (Ryder et al., 2012), or time-series of detections
58 (Psorakis et al., 2012).

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

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

71 holding the network structure constant (“node-label permutations”), therefore representing the null
72 hypothesis that the network measure serving as the response is unrelated to the predictor, while
73 controlling for network structure and non-independence. The resulting null distribution maintains
74 the non-independence inherent to the network while breaking any relationship that exists between
75 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 “pre-network permutations” or “datastream
89 permutations.” The goal is to construct permuted datasets that maintain structures of the original
90 data that may influence the observed network structure (e.g. the number of times individuals were
91 observed and the sizes of observed groups), while removing the social preferences that underpin the
92 social network (Farine & 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 data not constructed using group membership
107 (Whitehead, 1999). Controlling for gregariousness and sighting history is possible when each
108 sampling period is represented as a square matrix, where 1 indicates that individuals associated in
109 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 or node centralities, this null hypothesis can be equivalently stated as proposing that all
129 variance in a given value or network metric is due to data structure, confounds, and residual
130 variance. In network terminology, this null hypothesis is a random network, within a set of
131 constraints. This is precisely the null hypothesis that these permutations were designed to test, as
132 they were originally intended as a tool for detecting non-random social structure. However, we feel
133 there has been a lack of consideration about whether this null hypothesis is appropriate in other
134 contexts, such as regression modelling.

135

136 *Regression models in the context of social network analysis*

137 Most regression applications in social network analysis can be broadly considered in two broad
138 categories: nodal regression and dyadic regression. In the case of dyadic regression, researchers are
139 interested in determining if the strength or presence of social relationships themselves are predicted
140 by some dyadic variable, such as kinship or similarity in some trait. Nodal regression, on the other
141 hand, represents hypotheses linking individual level traits, such as age, sex, or personality, with the
142 position of nodes within the network, as summarized by any number of centrality measures. Here,
143 we will investigate whether datastream permutations specify the appropriate null hypothesis for the
144 typical inferences in these two regression contexts.

145 Consider the basic linear model:

146
$$Y = \beta X + \varepsilon \tag{1}$$

147 where Y is a response variable, X is a matrix of predictor variables, ε is the error term, and β is a
148 vector of estimated coefficients. The structure of Y , X , and ε differ between dyadic and nodal
149 regression contexts. In dyadic regression, Y is the $N \times N$ adjacency matrix (where N is the number of
150 individuals in the network), X is a $p \times N \times N$ array of predictors (where p is the number of predictors),
151 and ε is a square matrix. In nodal regression, Y is instead a vector of centrality measures of length N ,
152 X is a $p \times N$ matrix, and ε is a vector of length N .

153 We are typically interested in testing the null hypothesis $\beta = 0$, representing no relationship between
154 the response Y and the predictor(s) X . In permutation based hypothesis testing procedures, this null
155 hypothesis is tested by calculating a test statistic (such as the coefficient β or the t statistic) in the
156 observed data, and then repeatedly shuffling either X or Y to build a null distribution of this statistic.
157 These permutations maintain the distribution of both X and Y , but break the covariance between
158 them (Anderson & Robinson, 2001). This is the logic behind traditional permutation tests for
159 regression in social networks, such as node-label permutations and multiple regression quadratic
160 assignment procedures (MRQAP) (Croft et al., 2011).

161 Datastream permutations, however, do something very different, which is inappropriate for testing
162 the null hypothesis of no relationship between the response Y and the predictor(s) X . By permuting
163 the data underlying network measures and then re-calculating the response variable, these
164 procedures change the distribution of Y , instead of breaking relationships between the variables
165 (Figure 1). If the network has non-random social structure, even structure entirely unrelated to X ,
166 then we will typically see a reduction in the variance of Y as we permute the raw data. When Y has a
167 larger variance in the observed data than in the permutations, more extreme values of β are more
168 likely to occur in the observed data, even if the null hypothesis is true. This procedure is therefore
169 likely to result in much higher rates of false-positive (type I) error than is acceptable (Figure 1).

170 The problem here extends beyond the technical issue of reduced variance in the permuted datasets.
171 There is a fundamental problem with this approach when it comes to testing hypotheses using
172 regression models. When researchers fit regression models to predict network properties from
173 exogenous variables, the null hypothesis they will typically be testing against can be stated as “the
174 variation in network structure is not related to the exogenous variable.” This, however, is not the
175 null hypothesis tested by the commonly used datastream permutation methods. Rather, the null
176 hypothesis that is proposed by these datastream permutations could be stated as “the degree of
177 variation in network structure and its relationship to the exogenous variable are both due to random
178 interactions of individuals within constraints.” The researcher cannot disentangle the null hypothesis
179 of no relationship between the network and the predictor from the null hypothesis of random social
180 structure. In other words, a significant result from this procedure could be due to a relationship
181 between the predictor and the network, or because individuals do not interact at random, whether
182 or not the true social structure is related to the predictor. This fundamental mismatch between the
183 null hypothesis of interest and that tested by the datastream permutation algorithm makes tests of
184 regression models using this procedure nearly uninterpretable.

185 To further illustrate the problems that occur when combining datastream permutations of animal
186 social network data with regression we provide two simulated scenarios. In these scenarios, we
187 generate datasets with simple, but non-random social structure. We then introduce a random
188 exogenous variable that has no relationship to social structure, and test for a relationship between
189 network structure and this variable with linear models, using datastream permutations to determine
190 statistical significance. We show that even in the absence of any true relationship between
191 exogenous variables and social structure, datastream permutations are highly prone to producing
192 significant p -values when social structure is non-random. We caution against using these datastream
193 permutations to test the coefficients of regression models, and we discuss possible solutions and
194 alternative methods for regression analysis in social networks.

195

196 **Simulations**

197 *General framework*

198 We carried out simulations across two different scenarios, reflecting common research questions in
199 animal social network analysis. The first scenario simulates a case in which researchers are
200 interested in whether dyadic covariates (e.g. kinship or phenotypic similarity) influences the strength
201 of social bonds, which we will refer to as a case of “dyadic regression”. The second scenario
202 simulates a case when researchers are interested in how a quantitative individual trait (e.g. age or
203 personality) influences individual network position, which we refer to as “nodal regression.”

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

- 206 1. Generate observations of a network in which the quantity of interest (edge weight or node
207 centrality) has inherent variation.
- 208 2. Generate values for a trait that are unrelated to this variation.
- 209 3. Fit a linear model with the network property as the response variable and the trait as the
210 predictor.
- 211 4. Create permuted versions of the observed network via a common datastream permutation.
- 212 5. Compare the original model’s test statistics to those from the permuted data sets to
213 calculate a p -value.

214 For each simulation, we perform 200 runs, with varying parameter values (Table 1). For each run of
215 both simulations, we produce six outputs. The first two outputs are the p -values from the
216 datastream permutation test when using either the coefficient or t -value as the test statistic. We
217 additionally calculate the p -values for the same two test statistics using node-label permutations,
218 although further analysis showed that the t statistic and coefficient always produced identical results

219 in these cases. The final two outputs give information about the characteristics of the dataset not
220 given by the initial inputs. The first is the standard deviation of the response variable (either the
221 edge weights or strengths), indicating the degree of non-randomness in the social structure, and the
222 second is the average number of sightings per individual, a common measure of sampling effort in
223 social network studies.

224 All simulations and subsequent analyses were performed in R (R Core Team 2020), using the
225 packages *asnipe* (Farine 2019), *lhs* (Carnell 2019), and *truncnorm* (Mersmann et al. 2018). All code
226 necessary to reproduce our analysis is included in the online supplementary materials.

227

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

229 In our first simulation, we investigate the case in which the researcher is interested in the influence
230 of a dyadic predictor (such as similarity in phenotype or kinship) on the rates at which dyads
231 associate or interact. Our simulation framework is heavily inspired by those of Whitehead & James
232 (2015) and Farine & Whitehead (2015). We simulate a population of N individuals, and assign each
233 dyad an association probability p_{ij} from a beta distribution with mean μ and precision ϕ ($\alpha = \mu\phi$, $\beta =$
234 $(1-\mu)\phi$). By assigning association probabilities in this way, we create non-random social preferences
235 in the network, and thus larger variance in edge weights than would be expected given random
236 association (Whitehead et al., 2005).

237 We then simulate τ sampling periods. For simplicity, individuals are sighted in each sampling period
238 with a constant probability σ , and associations between dyads where both individuals are sighted
239 occur with probability p_{ij} . We then build the observed association network by calculating dyadic
240 simple ratio indices (*SRI*):

241
$$SRI_{ij} = \frac{X_{ij}}{D_{ij}} \quad (2)$$

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

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

250 We then fit the linear model:

$$251 \quad SRI_{ij} = \beta_0 + \beta_1 |trait_i - trait_j| + \varepsilon_{ij} \quad (3)$$

252 and save the estimate of β_1 and the associated t statistic. We compare this coefficient and t statistic
253 to a null model generated using the sampling period permutation method proposed by Whitehead
254 (1999). There are several algorithms available to perform these swaps. We use the “trial swap”
255 procedure described by Miklós & Podani (2004) and suggested for social network studies by Krause
256 et al. (2009). For each trial, this procedure chooses an arbitrary 2 by 2 submatrix of the lower
257 triangle within a random sampling period. If a swap is possible, it is performed (and symmetrized),
258 otherwise the matrix stays at its current state. These steps when the matrix is not changed are
259 referred to as “waiting steps.” This algorithm is ideal because it ensures that the Markov chain
260 samples the possible matrices uniformly, while other algorithms that do not include waiting steps
261 exhibit biases in their sampling of the possible matrices (Miklós & Podani, 2004). We generate
262 10,000 permuted datasets for each simulation, with 1,000 trial swaps between each permutation,
263 and re-fit our linear model to each permuted dataset, recording the coefficient and t statistic. We
264 then use these distributions to calculate p -values for the linear model’s coefficient. Across the 200
265 runs, we vary the parameters of the simulation by drawing μ , ϕ , N , σ , and τ randomly using Latin
266 hypercube sampling (Table 1).

267

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

269 We next investigate the same concept in the context of nodal regression. This form of analysis tests
270 whether some individual attribute is related to variation in network position. This is perhaps the
271 most common use of datastream permutation null models for testing the significance of linear
272 regression coefficients in animal social networks (e.g. Cowl et al., 2020; Poirier & Festa-Bianchet,
273 2018; Zeus et al., 2018). For simplicity, we focus on strength, which is simply the sum of an
274 individual's edge weights.

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

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

287
$$P(i \text{ in } g) \propto \frac{1}{(S_g - P_i)^2} \quad (4)$$

288 This gives the simulation the property that individuals with higher assigned gregariousness scores
289 tend to occur in larger groups, and vice versa. This leads to non-random differences in

290 gregariousness (and thus strength centrality) between individuals. We then calculate the association
291 network, again using the SRI:

$$292 \quad SRI_{ij} = \frac{X_{ij}}{X_{ij} + Y_i + Y_j} \quad (5)$$

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

$$297 \quad \sum_j SRI_{ij} = \beta_0 + \beta_1 trait_i + \varepsilon_i \quad (6)$$

298 and again save the estimate of β_1 , along with the associated t statistic. We compare these statistics
299 to those derived from networks generated using the group-based permutation procedure proposed
300 by Bejder et al. (1998). This procedure again sequentially permuted the observed dataset, while
301 maintaining the size of each group and the number of groups per individual. We again use the trial
302 swap method to perform these permutations, generating 10,000 permuted datasets with 1,000 trials
303 per permutation, and derived p -values in the same way as above. We vary the parameters of this
304 simulation by using Latin hypercube sampling to draw values of N , M , G , and V (see Table 1 for
305 ranges).

306

307 *Analysis*

308 We use the outputs of the simulations primarily to derive overall type I error rates (calculated as the
309 portion of runs in which a p -value less than 0.05 was obtained) when using either regression
310 coefficient or t -value as the test statistic. We further investigated the sensitivity of these results to
311 non-random social structure, sampling effort, and population size. Previous work suggests that the
312 sensitivity of datastream permutation techniques are highly dependent on variation in social
313 structure and sampling intensity (Whitehead, 2008). We use binomial generalized linear models to

314 summarize how population size, response variance, and sampling intensity influence the probability
315 of false positives. We further analyse these relationships qualitatively using conditional probability
316 plots. We compare these results to those derived from node-label permutation tests on the same
317 simulated datasets.

318

319 **Simulation results**

320 *Dyadic regression*

321 The overall type I error rate for the dyadic regression case was high, with 41% (81/200) of runs giving
322 false positives when using the coefficient as the test statistic, and 21% (42/200) when using the t -
323 value. When using the regression coefficient as the test statistic, the false positive rate increased
324 with greater sampling effort ($\beta = 0.012 \pm 0.004$, $z = 2.82$, $p = 0.005$) and variance in SRI values ($\beta =$
325 6.35 ± 3.04 , $z = 2.09$, $p = 0.03$), but was not strongly influenced by the network size ($\beta = -0.007 \pm$
326 0.006 , $z = -1.085$, $p = 0.278$). When the t -value was used as the test statistic, only the sampling effort
327 significantly influenced the false positive rate ($\beta = 0.014 \pm 0.004$, $z = 3.00$, $p = 0.003$), while neither
328 the number of individuals ($\beta = 0.0007 \pm 0.008$, $z = 0.091$, $p = 0.927$) or variance in edge weights ($\beta = -$
329 0.59 ± 3.72 , $z = -0.177$, $p = 0.859$) were significantly correlated with the false positive rate. In
330 contrast, the node-label permutation method had a much lower false positive rate of 6% (12/200)
331 and was unaffected by sampling effort ($\beta = -0.004 \pm 0.008$, $z = -0.443$, $p = 0.658$), network size ($\beta =$
332 0.001 ± 0.013 , $z = 0.086$, $p = 0.931$), or edge weight variance ($\beta = 3.574 \pm 5.438$, $z = 0.657$, $p = 0.511$).

333

334 *Nodal regression*

335 In the case of nodal regression, type I errors were once again high when using datastream
336 permutations. When using the regression coefficient as the test statistic, our simulation resulted in a
337 type I error rate of 43.5% (87/200), and when using the t -value the type I error rate was 28%

338 (56/200). When using the regression coefficient as the test statistic, both sampling effort ($\beta = 0.029$
339 ± 0.012 , $z = 2.434$, $p = 0.015$) and variance in centrality ($\beta = 1.444 \pm 0.479$, $z = 3.017$, $p = 0.003$) were
340 positively correlated with type I errors, while the number of individuals was not related to type I
341 errors ($\beta = -0.005 \pm 0.007$, $z = -0.732$, $p = 0.464$). When using the t -statistic, sampling effort was still
342 positively related to type I error rate ($\beta = 0.042 \pm 0.013$, $z = 3.265$, $p = 0.001$), however the variance
343 in centrality was not ($\beta = -0.287 \pm 0.498$, $z = -0.577$, $p = 0.564$), and, interestingly, the size of the
344 network appears to be positively correlated with type I error ($\beta = 0.017 \pm 0.009$, $z = 1.990$, $p = 0.047$).
345 As in the case of dyadic regression, the node-label permutations produced an acceptable false
346 positive rate of 7% (14/200), which was unaffected by sampling ($\beta = 0.014 \pm 0.021$, $z = 0.663$, $p =$
347 0.507) network size ($\beta = 0.008 \pm 0.015$, $z = 0.579$, $p = 0.562$) or variance in centrality ($\beta = 0.194 \pm$
348 0.868 , $z = 0.224$, $p = 0.823$).

349

350 **Discussion**

351 These two simple simulated scenarios show that the commonly used datastream permutation
352 procedures for animal social network data produce extremely high and thus unacceptable false-
353 positive rates when used as a test of regression models. This is because datastream permutations
354 represent a null hypothesis that is different from the typical null hypothesis that researchers are
355 interested in testing when fitting regression models (i.e. that the model coefficients are 0).

356 It is important here to stress that the permutation procedure is not doing anything “wrong” in these
357 examples. The permutations are in fact generating a distribution of statistics that is correct for the
358 null hypothesis that the algorithm is designed to test, which is that the social structure is random.

359 The “type I errors” that we discuss here are introduced when the rejection of this null hypothesis is
360 taken as evidence that a relationship exists between the non-random structure of the network and
361 an exogenous variable, when in fact these rejections in our simulations are simply indicating that
362 social structure is not in fact random. For this reason, we recommend against datastream

363 permutations as a test for regression models with social network data. Datastream permutations,
364 however, will continue to play an important role in animal social network analysis; the results of
365 datastream permutations can tell us whether a given dataset shows signs of non-random social
366 structure. This is key, not just for social analyses generally but for regression analyses in particular. If
367 a dataset does not show signs of non-random social structure, it likely does not make sense to
368 continue with regression analyses that attempt to uncover the correlates of social network
369 structure.

370 In this study, we focused on the case where network measures are the response variable in a linear
371 model. A different, but related scenario is when we try to predict individual attributes (such as
372 measures of fitness or personality) using network measures as a predictor. The statistical problems
373 presented by this scenario are slightly different than those of the network response case. Here, the
374 non-independence of the network data are not a problem, as linear models do not make any
375 assumptions about the distribution or covariance structure of the predictors (n.b. there can still be
376 covariance in the attribute used as a response variable related to network position that, if present,
377 would need accounting for in the statistical model). The issue of data unreliability, however, may still
378 be present. As in the simulations used here, datastream permutations alone would not serve as an
379 adequate test. These models would test the null hypothesis that the relationship between the
380 response and the network arose due to random social structure, when in fact the researcher is likely
381 interested in whether the non-random social structure influences the individual attributes. A
382 significant result from the datastream permutation method could simply indicate that the social
383 structure is not random, rather than serving as an indicator that a relationship exists between the
384 network and the response.

385 The high false-positive rate we describe here is the result of decreased variance in the response
386 variable after permuting the raw data, as the variation due to social processes has been removed. A
387 potential “quick fix” that might be mooted is to simply standardize the response variable in the

388 observed network so that all subsequent permutations to have a constant variance, i.e. using Z-
389 scores. This may reduce the type I error rate. However, we strongly recommend against this as a
390 solution to the problem. Standardizing the variance does not address the inconsistency at the heart
391 of the problem. The null hypothesis being specified by the null model, that the social structure is
392 random, is still not the same as the null hypothesis of interest in the regression.

393 In the following sections, we highlight some potential ways forward for the application of regression
394 in animal social network analyses, and give some general recommendations for researchers. We
395 hope that this discussion will encourage further work that may provide an extended toolkit for
396 ecologists interested in these kinds of problems.

397

398 *Carrying out regression in social networks by separating non-independence and bias*

399 If datastream permutations alone cannot be used to test regression models in animal social network
400 analyses, how should we conduct these analyses? While there are numerous potential solutions, and
401 a full accounting of them is beyond the scope of this paper, we suggest that a general way forward is
402 to recognize that the two issues of non-independence and unreliability of the data are separate
403 problems requiring distinct statistical solutions.

404 Not all animal network data will be subject to the issue of unreliability (e.g., in cases where sampling
405 is balanced across subjects and relevant contexts) and in some instances data may be complete and
406 unbiased. In these cases, node permutations or other statistical network models will be appropriate
407 (Croft et al. 2011). When structure or bias in the observations need to be controlled for, we propose
408 two general approaches that may be useful; other solutions are certainly possible, and we
409 encourage further work on this matter.

410 The first method (Figure 4A) would first attempt to remove the bias from the network using

411 generalised affiliation indices (GAIs; Whitehead & James, 2015) or similar corrections to account for

412 confounding variables that may influence observed edge weights. GAls fit the observed associations
413 or interactions as the response in a binomial or Poisson generalized linear model, with confounding
414 factors such as space use, sightings frequency, or joint gregariousness as predictors. The residuals of
415 this model are then used as measures of affiliation, as they reflect the difference between observed
416 and expected association rates given the confounding factors. While a flexible and appealing
417 approach, GAls require that potential confounds be properly specified in terms of dyadic covariates,
418 and that the relationship between confounds and edge weights be linear. This second issue could be
419 solved by deriving affiliations from generalized additive models (GAMs), where the relationship
420 between covariates and the response can be represented by smooth functions. While GAls
421 represent the most well developed method for correcting social network edge weights, other
422 methods are certainly possible. Once corrections are made, researchers can use the corrected social
423 network to derive responses to use in the statistical model. A potential drawback of GAls is that
424 avoidance between individuals is represented as negative edge weights. While this is not a problem
425 for dyadic regression (in fact it better conforms to the assumptions of traditional linear models), this
426 complicates the calculation of some centrality measures, requiring that negative edge weights be
427 ignored or set to zero (Whitehead & James 2015). Inference would be carried out using post-
428 network permutation methods, such as node-label permutations or MRQAP.

429 A second, different approach (Figure 4B) would be to incorporate confounds in the inferential model
430 itself. If researchers identify likely confounds and summarize them quantitatively at the same level
431 as the hypothesis being tested (e.g. dyadic or nodal), these could be used directly in the statistical
432 model. Where potential non-linearity between confounds and responses exist, data transformations,
433 polynomials, and smooth functions may present a possible solution. Again, post-network
434 permutation methods would be employed for inference to correct for the non-independence of the
435 data. Franks et al. (2020) explore this method in detail.

436 We feel that these approaches have the potential to address the current issue that we have
437 identified and we strongly encourage new work to explore and validate these approaches. These
438 suggestions are general, identifying the ways in which we might approach separately address non-
439 independence and bias. It is important to note that the methods we propose are only useful if the
440 question of interest is about the structure of social affinity, rather than the empirical pattern of
441 encounters between individuals. If, instead, researchers are interested in the actual rates of contact
442 (as is the case in disease research and studies of social learning), this approach may not be
443 appropriate. Extensions of recent work using hidden state modelling may be more appropriate for
444 disentangling true association patterns when detections are potentially biased or imperfect
445 (Gimenez et al., 2019).

446

447 *Building better null models*

448 The problems we have identified here arise because the commonly used null models for animal
449 societies do not generate datasets representing the null hypothesis of interest in a regression
450 setting. These models were specifically designed to test the null hypothesis of random social
451 structure, not the null hypothesis that aspects of social structure are unrelated to exogenous factors.
452 An obvious way forward would be the development of permutation procedures that generate
453 datasets that correctly represent the relevant null hypothesis. In the case of dyadic regression, these
454 datasets would maintain the structure of the data (e.g. sightings per individual, associations per
455 sampling period, spatial patterns of observations), randomise identities of associated individuals,
456 and simultaneously preserve the variance in edge weights. In the case of nodal regression, permuted
457 datasets would maintain the same (or at least a similar) distribution of individual centrality within
458 the network, in addition to structural confounds such as the size of groups, sightings per individual,
459 and timing of sightings. The design of such procedures is far from trivial, and is beyond the scope of
460 this paper, but we suspect that the development of algorithms that simultaneously maintain aspects

461 of data structure and features of the social system will be an important area of methodological
462 research going forward. This area of research is still in its early days, although there has been some
463 potentially applicable work in other sub-fields of network science (e.g. Chodrow 2019).

464

465 **Conclusion**

466 The development of permutation techniques that control for sampling biases while maintaining
467 temporal, spatial, and structural aspects of the raw data is an important development in the study of
468 animal social systems, and we suspect that these procedures will remain a key tool for hypothesis
469 testing in ecology and evolution. These techniques are particularly crucial when it is not clear
470 whether a dataset shows signs of non-random social structure. However, a lack of consideration
471 regarding the matching up of the null hypothesis being tested with the null model being generated
472 using datastream permutations has led to unwarranted application of these techniques, particularly
473 in the context of hypothesis testing using regression models. Here, we have shown that significant p -
474 values from applying datastream permutations to regression models cannot be used as evidence of a
475 relationship between the social network and exogenous predictors.

476 We recommend that researchers think critically and carefully about the null hypothesis they wish to
477 test using social network data, and ensure that the null model they specify does in fact represent
478 that hypothesis (Table 2). We suspect that in most cases, the null hypothesis of random social
479 structure will clearly not be appropriate in regression analysis, and therefore traditional datastream
480 permutations will not be a viable approach. We hope that our discussion of this issue and the results
481 of our simulations will result in reconsideration of how researchers employ null models when
482 analysing animal social networks, promote further research and discussion in this area, and lead to
483 the development of procedures that correctly specify null hypotheses and allow robust inference in
484 animal social network studies.

485

486 **Acknowledgements**

487 We would like to thank colleagues at the Centre for Research in Animal Behaviour, particularly the
488 members of the CRAB Social Network Club, for extremely valuable discussion that greatly improved
489 this manuscript. We would like to thank Josh Firth and Josefine Bohr Brask for very useful comments
490 and discussion on an earlier version of this manuscript. DPC, DWF and SE acknowledge funding from
491 NERC (NE/S010327/1). LJNB acknowledges funding from the NIH (R01AG060931; R01MH118203).

492

493 **Author contributions**

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

496

497 **Data availability**

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

500

501 **References**

- 502 Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian and New*
503 *Zealand Journal of Statistics*, 43(1), 75–88. <https://doi.org/10.1111/1467-842X.00156>
- 504 Beijder, L., Fletcher, D., & Bräger, S. (1998). A method for testing association patterns of social
505 animals. *Animal Behaviour*, 56(3), 719–725. <https://doi.org/10.1006/anbe.1998.0802>

506 Carnell, R. (2019) lhs: Latin hypercube samples. R package version 1.0.1. [https://CRAN.R-](https://CRAN.R-project.org/package=lhs)
507 [project.org/package=lhs](https://CRAN.R-project.org/package=lhs)

508 Chodrow, P. S. (2019) Configuration models of random hypergraphs. [arXiv:1902.09302](https://arxiv.org/abs/1902.09302)

509 Cowl, V. B., Jensen, K., Lea, J. M. D., Walker, S. L., & Shultz, S. (2020). Sulawesi Crested Macaque
510 (Macaca nigra) Grooming Networks Are Robust to Perturbation While Individual Associations
511 Are More Labile. *International Journal of Primatology*, 41(1), 105–128.
512 <https://doi.org/10.1007/s10764-020-00139-6>

513 Croft, D. P., Madden, J. R., Franks, D. W., & James, R. (2011). Hypothesis testing in animal social
514 networks. In *Trends in Ecology and Evolution*, 26(10), 502-207.
515 <https://doi.org/10.1016/j.tree.2011.05.012>

516 Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and
517 autocorrelation conditions. *Psychometrika*, 72(4), 563–581. [https://doi.org/10.1007/s11336-](https://doi.org/10.1007/s11336-007-9016-1)
518 [007-9016-1](https://doi.org/10.1007/s11336-007-9016-1)

519 Farine, D. R. (2017). A guide to null models for animal social network analysis. *Methods in Ecology*
520 *and Evolution*, 8(10), 1309–1320. <https://doi.org/10.1111/2041-210X.12772>

521 Farine, D. R. (2019). asnipe: Animal social network inference and permutation for ecologists. R
522 package version 1.1.12. <https://CRAN.R-project.org/package=asnipe>

523 Farine, D. R., & Whitehead, H. (2015). Constructing, conducting and interpreting animal social
524 network analysis. *Journal of Animal Ecology*, 84(5), 1144–1163. [https://doi.org/10.1111/1365-](https://doi.org/10.1111/1365-2656.12418)
525 [2656.12418](https://doi.org/10.1111/1365-2656.12418)

526 Firth, J. A., Sheldon, B. C., & Brent, L. J. N. (2017). Indirectly connected: simple social differences can
527 explain the causes and apparent consequences of complex social network positions.
528 *Proceedings of the Royal Society B: Biological Sciences*, 284(1867), 20171939.
529 <https://doi.org/10.1098/rspb.2017.1939>

530 Franks, D. W., Ruxton, G. D., & James, R. (2010). Sampling animal association networks with the
531 gambit of the group. *Behavioral Ecology and Sociobiology*, 64(3), 493–503.
532 <https://doi.org/10.1007/s00265-009-0865-8>

533 Franks, D. W., Weiss, M. N., Silk, M. J., Perryman, R. J. Y., & Croft, D. P. (2020). Calculating effect sizes
534 in animal social network analysis. *Methods in Ecology and Evolution*, 00: 1-9.
535 <https://doi.org/10.1111/2041-210X.13429>

536 Gimenez, O., Mansilla, L., Klaich, M. J., Coscarella, M. A., Pedraza, S. N., & Crespo, E. A. (2019).
537 Inferring animal social networks with imperfect detection. *Ecological Modelling*, 401, 69–74.
538 <https://doi.org/10.1016/j.ecolmodel.2019.04.001>

539 Hinde, R. A. (1976). Interactions , Relationships and Social Structure. *Man*, 11(1), 1–17.

540 James, R., Croft, D. P., & Krause, J. (2009). Potential banana skins in animal social network analysis. In
541 *Behavioral Ecology and Sociobiology* (Vol. 63, Issue 7, pp. 989–997). Springer.
542 <https://doi.org/10.1007/s00265-009-0742-5>

543 Krause, S., Mattner, L., James, R., Guttridge, T., Corcoran, M. J., Gruber, S. H., & Krause, J. (2009).
544 Social network analysis and valid Markov chain Monte Carlo tests of null models. *Behavioral*
545 *Ecology and Sociobiology*, 63(7), 1089–1096. <https://doi.org/10.1007/s00265-009-0746-1>

546 Kurvers, R. H. J. M., Krause, J., Croft, D. P., Wilson, A. D. M., & Wolf, M. (2014). The evolutionary and
547 ecological consequences of animal social networks: Emerging issues. *Trends in Ecology and*
548 *Evolution*, 29(6). 326-335. <https://doi.org/10.1016/j.tree.2014.04.002>

549 Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.).
550 Chapman & Hall.

551 Miklós, I., & Podani, J. (2004). Randomization of presence-absence matrices: Comments and new
552 algorithms. *Ecology*, 85(1), 86–92. <https://doi.org/10.1890/03-0101>

553 Poirier, M. A., & Festa-Bianchet, M. (2018). Social integration and acclimation of translocated
554 bighorn sheep (*Ovis canadensis*). *Biological Conservation*, *218*, 1–9.
555 <https://doi.org/10.1016/j.biocon.2017.11.031>

556 Psorakis, I., Roberts, S. J., Rezek, I., & Sheldon, B. C. (2012). Inferring social network structure in
557 ecological systems from spatiotemporal data streams. *Journal of the Royal Society Interface*,
558 *9*(76), 3055–3066. <https://doi.org/10.1098/rsif.2012.0223>

559 Psorakis, I., Voelkl, B., Garroway, C. J., Radersma, R., Aplin, L. M., Crates, R. A., Culina, A., Farine, D.
560 R., Firth, J. A., Hinde, C. A., Kidd, L. R., Milligan, N. D., Roberts, S. J., Verhelst, B., & Sheldon, B. C.
561 (2015). Inferring social structure from temporal data. *Behavioral Ecology and Sociobiology*,
562 *69*(5), 857–866. <https://doi.org/10.1007/s00265-015-1906-0>

563 R Core Team (2020). R: A language and environment for statistical computing. R Foundation for
564 Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

565 Ryder, T. B., Horton, B. M., van den Tillaart, M., Morales, J. D. D., & Moore, I. T. (2012). Proximity
566 data-loggers increase the quantity and quality of social network data. *Biology Letters*, *8*(6),
567 917–920. <https://doi.org/10.1098/rsbl.2012.0536>

568 Spiegel, O., Leu, S. T., Sih, A., & Bull, C. M. (2016). Socially interacting or indifferent neighbours?
569 Randomization of movement paths to tease apart social preference and spatial constraints.
570 *Methods in Ecology and Evolution*, *7*(8), 971–979. <https://doi.org/10.1111/2041-210X.12553>

571 Webber, Q. M. R., & vander Wal, E. (2019). Trends and perspectives on the use of animal social
572 network analysis in behavioural ecology: a bibliometric approach. *Animal Behaviour*, *149*, 77–
573 87. <https://doi.org/10.1016/j.anbehav.2019.01.010>

574 Whitehead, H. (1999). Testing association patterns of social animals. *Animal Behaviour*, *57*(6), F26–
575 F29. <https://doi.org/10.1006/anbe.1999.1099>

576 Whitehead, H. (2008). Precision and power in the analysis of social structure using associations.
577 *Animal Behaviour*, 75(3), 1093–1099. <https://doi.org/10.1016/j.anbehav.2007.08.022>

578 Whitehead, H., Bejder, L., & Ottensmeyer, C. A. (2005). Testing association patterns: Issues arising
579 and extensions. *Animal Behaviour*, 69(5). <https://doi.org/10.1016/j.anbehav.2004.11.004>

580 Whitehead, H., & James, R. (2015). Generalized affiliation indices extract affiliations from social
581 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)
582 [210X.12383](https://doi.org/10.1111/2041-210X.12383)

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

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604 **Figure Captions**

605

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

615

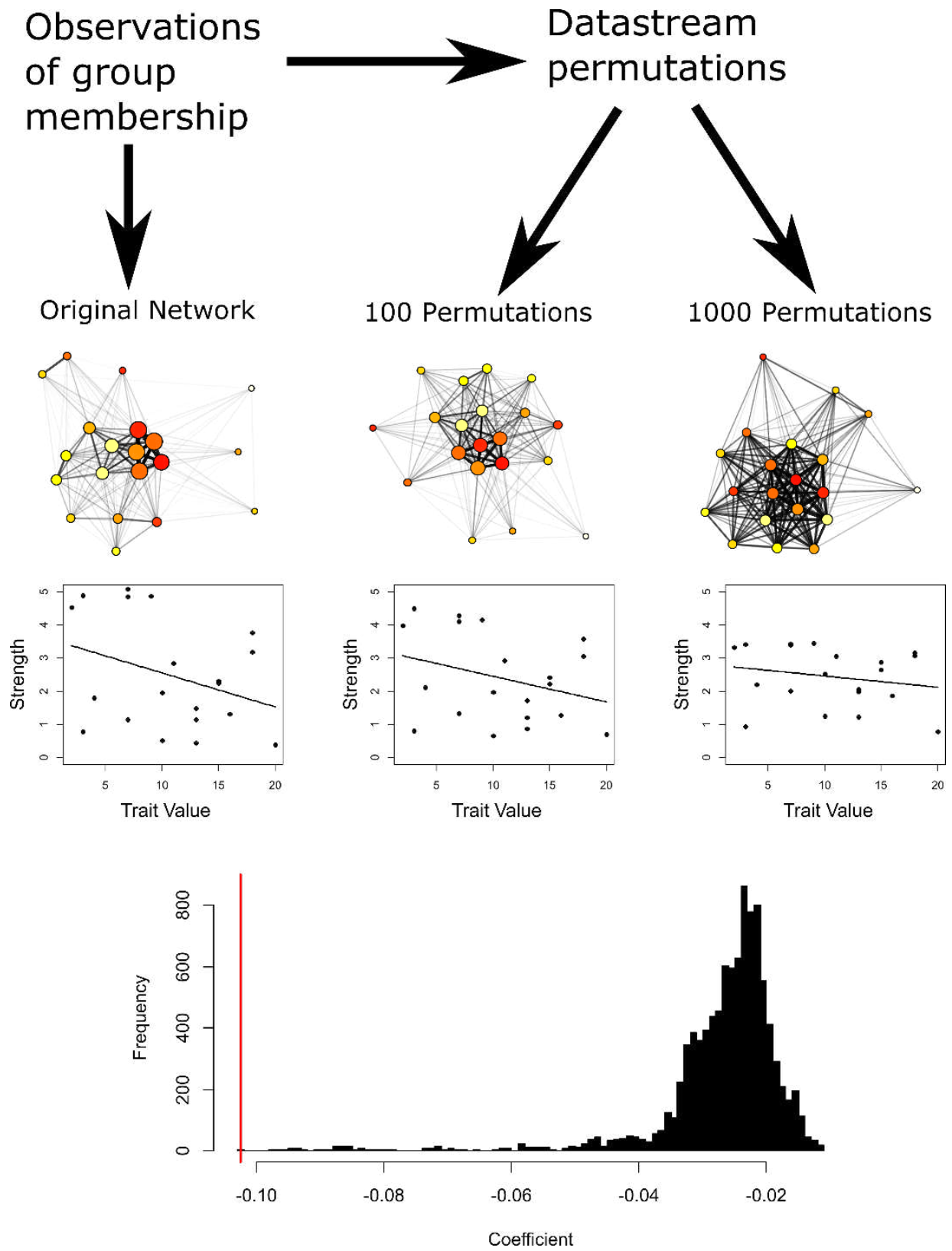
616 **Figure 2.** Conditional probability plots from dyadic regression simulation. Lines indicate smoothed
617 conditional probabilities of a type I error (a p -value less than 0.05) for datastream permutations
618 using the coefficient (red) or t -value (orange), and node-label permutations (blue) in relation to
619 three covariates. Dotted line indicates target type I error rate of 0.05.

620

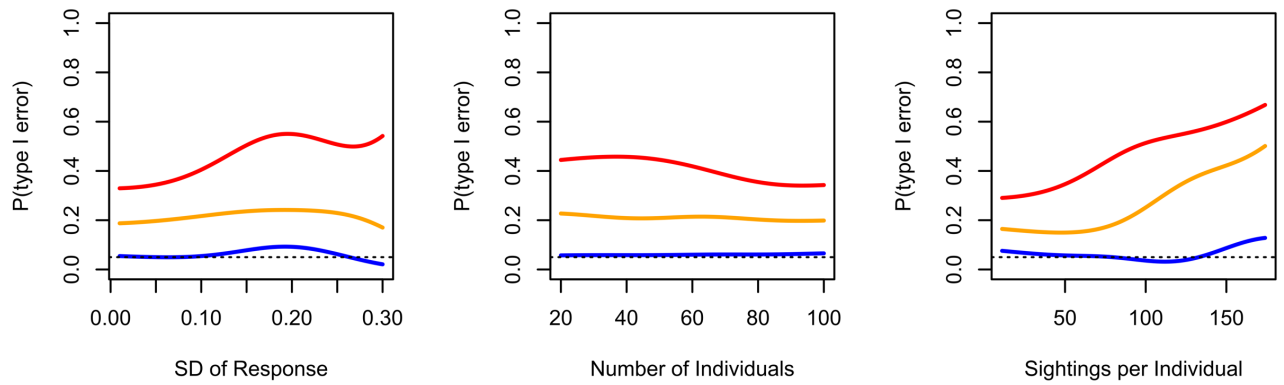
621 **Figure 3.** Conditional probability plots from nodal regression simulation. Lines indicate smoothed
622 conditional probabilities of a type I error (a p -value less than 0.05) for datastream permutations
623 using the coefficient (red) or t -value (orange), and node-label permutations (blue) in relation to
624 three covariates. Dotted line indicates target type I error rate of 0.05.

625

626 **Figure 4.** Flowcharts of two approaches for regression analysis in animal social networks. In the first
627 (A), a network is generated that attempts to adjust for confounding effects (through e.g. GAI) which
628 is then used to derive the response. In the second (B), the original network is used to derive the
629 response variable, with confounds instead being incorporated as covariates in the inferential model.
630 In both methods, inference is based on post-network permutations (such as MRQAP or node-label
631 permutations).



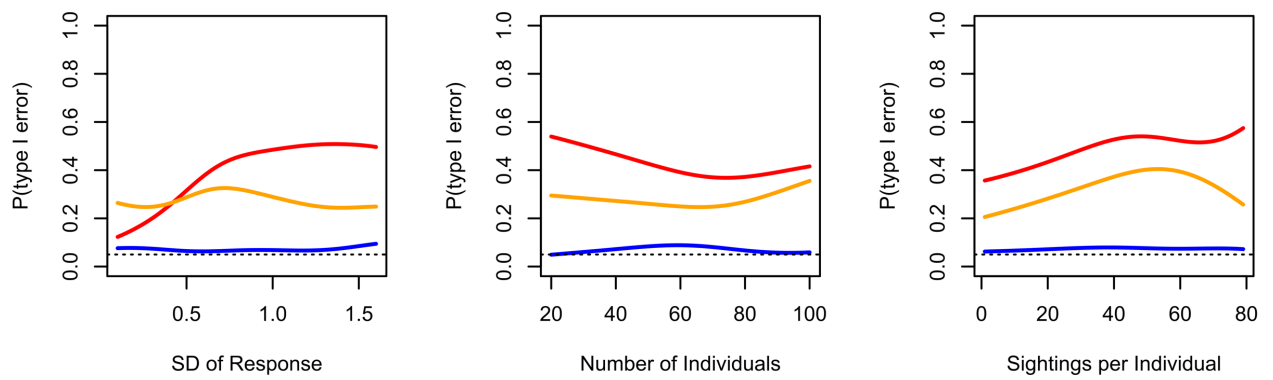
635 **Figure 2**



636

637

638 **Figure 3**



639

640

641

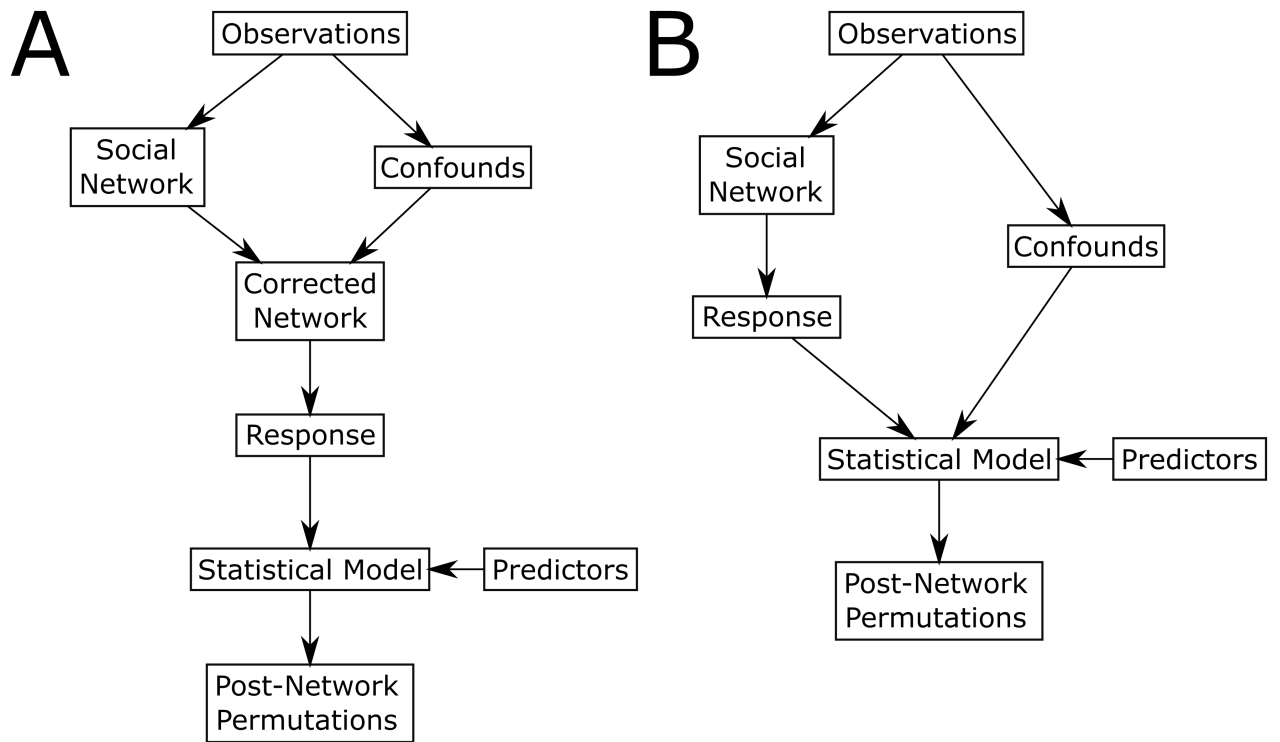
642

643

644

645

646



648

649

650

651

652

653

654

655

656

657 **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
σ	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

658

659

660 **Table 2.** Comparison of datastream and node-label permutations

	Datastream permutations	Node-label permutations
Dyadic H_0	There is no variation in the strength of social ties once data structure, sampling noise, and constraints (time, location, etc.) are accounted for.	The observed variation in the strength of social ties is unrelated to dyadic covariates (e.g. kinship, trait similarity)
Nodal H_0	There is no variation in centrality once data structure, sampling noise, and constraints are accounted for.	Observed variation in centrality is unrelated to node characteristics (e.g. age, sex, personality)
Applications	Testing for the presence of social preferences	Testing relationships between observed social ties and dyadic predictors
	Testing for non-random variation in social position	Testing relationships between centrality and node attributes
Benefits	Corrects for bias in data collection from differences in detection probability and demographic processes	Corrects for the structure of the observed network
	Accounts for complex data structures such as focal follows and gambit of the group	Specifies the null distribution of interest for most regression applications
Drawbacks	Results in a decrease in variance in network measures compared to observed data when social structure is non-random	Does not account for data collection method or complex data structures
	Cannot be used to test regression models against the null hypothesis of zero effect	Does not correct for bias or uncertainty due to sampling

661

662