

# Robust Bayesian modeling of animal networks subject to censoring and sampling biases

Sebastian Sosa<sup>1</sup> ...

## Abstract

Animal Social Network Analysis (ASNA) has emerged as a crucial approach, leading to significant progress in theoretical and empirical studies of animal social behavior, social development, genetic factors, fitness impact, epidemiology, animal culture, and social structures. This progress owes much to the development of cutting-edge analytical techniques.

## Keywords

social networks, animal networks, social interactions, generative models

## Introduction

Over the past 50 years, graph theory has become a vital tool in studying natural and artificial systems in various fields like anthropology (Apicella et al. 2012), sociology (Milgram 1967), economics (Ter Wal and Boschma 2009), ecology (Sosa et al. 2021a). It applies to both small-scale (e.g., proteomics (Ravasz et al. 2002)) and large-scale (e.g., ecosystems (Ulanowicz et al. 2014)) systems. In the realm of animal sociality research, innovative techniques, such as association indices and pre-network permutations (Whitehead 1999; Farine 2013; Farine and Whitehead 2015), have expanded upon traditional graph theory. These methods help address specific challenges in this field, like variations in sampling efforts. However, recent studies have highlighted significant reliability concerns associated with hypothesis testing protocols developed in Animal Social Network Analysis (ASNA) (Weiss et al. 2021; Puga-Gonzalez et al. 2021; Hart et al. 2022; Farine and Carter 2022).

These issues involve elevated rates of false negatives (meaning failing to reject a false null hypothesis) and false positives (accepting a false null hypothesis). For instance, in a simulation study by Puga-Gonzalez et al. (2021), they considered a scenario with data biases stemming from the data collection process (like oversampling certain individual categories). They identified false positive rates as high as 60.8% and false negatives rates ranging up to 36.6%. Since very few biological data collected in natural settings are entirely free from biases related to the studied system or limitations in sampling, these findings emphasize a common challenge that urgently needs addressing to ensure the reliability of hypothesis testing in ASNA. Several approaches have been proposed since the issues with rates of false positives and negatives were raised. These new methods can be categorized into two types based on whether they control for bias within the steps of the ASNA hypothesis testing protocol. In ASNA, there are typically two hypothesis testing protocol steps: 1) Estimating social interaction patterns among individuals, which are quantified within

a social network and 2) Testing statistical hypotheses about these patterns.

To achieve the first step, researchers calculate a measure of the tendency for individuals to associate (in undirected behavior) or interact (in directed behavior). This measure is referred to as a social index and is computed for each pair of individuals (dyad). These values are used to create a social network, where each individual is a node, and the social index values represent the strength of the connection (edge weight) between dyad members. In addition, social indices have also the goal to control for sampling effort (i.e. heterogeneity in sampling between individuals). Currently two main types of social index are used: association indices (Hubalek 1982; Sailer and Gaulin 1984) and interaction index. Association indices have been used by behavioral ecologists to estimate the proportion of time that a pair of individuals spends together (Whitehead 2008). The higher the index value, the stronger the association within the dyad. The most used association index so far is the simple ratio index (SRI; Eq. 1), designed for data collected in discrete sampling periods (e.g. gambit of the group, scan sampling). The Interaction Index (Eq. 2) is primarily utilized by primatologists for data collected during continuous sampling periods, such as focal sampling. This index estimates the rate of social interactions per unit of time (e.g. the total time of focus), as opposed to association indices that are proportion-based. A higher index value indicates a stronger rate of interaction within the dyad. Finally, new approaches based on the Bayesian generative network have recently been proposed to estimate interaction patterns among individuals (Hart et al. 2023; Ross et al.). Also, there hasn't been any testing conducted to assess the reliability

<sup>1</sup>Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

## Corresponding author:

Sebastian Sosa

Email: s.sosa@live.fr

of results based on biased data. The lingering question is then whether Bayesian generative networks can outperform sociality indices or not.

$$SRI_{ij} = \frac{x_{ij}}{y_i + y_j + y_{ij} + x_{ij}} \quad (1)$$

Where  $x_{ij}$  is the number of sampling periods with  $i$  and  $j$  observed associated,  $y_i$  is the number of sampling periods with only  $i$  identified,  $y_j$  is the number of sampling periods with only  $j$  identified and  $y_{ij}$  is the number of sampling periods with  $i$  and  $j$  identified but not associated.

$$II_{ij} = \frac{x_{ij}}{y_i + y_j} \quad (2)$$

Where  $x_{ij}$  represents the overall frequencies or total interaction time,  $y_i$  and  $y_j$  denote the number of observations or the total observation time for individuals  $i$  and  $j$ , respectively.

To achieve the second step, which involves testing statistical hypotheses about social interaction patterns among individuals, researchers often use node-based measures computed on the network of association. These measures are derived from the network and help assess various aspects of an individual's position within the network. However, using node-based measures straightforwardly to test hypotheses about individual associations isn't ideal because each association/interaction is counted twice, once for each individual in the dyad, which violates the assumptions of parametric tests. Therefore, much of the methodological work in ASNA has focused on developing techniques to enable valid hypothesis testing. Since 2000, in ASNA permutations have been the main standard coming in two forms: network permutations, pre-network permutations (Farine 2013). For more detailed information on the principles of permutation approaches, readers can refer to (Farine 2013; Sosa et al. 2021b). Issues related to false positive and negative rates were originally demonstrated in network permutations and pre-network permutations. To address these issues, double permutations, which combine both, have been developed and tested on biased data to estimate rates of false positives and false negatives. Double permutations have shown substantial improvement in false positives and false negatives rates compared to other permutation approaches. However, permutations have recently faced criticism for assuming exchangeability of residuals (Hart et al. 2022). Moreover, studies that highlighted problems with permutation approaches or propose new methods have been employing the same simulation protocol and as demonstrated in Appendix 1, this simulation has some issues. Firstly, these simulations were designed to generate "bias of interaction". This bias pertains to missing associations or interactions while observing individuals. It differs from observation bias, as the heterogeneity of observation among individuals is known by the observer(s) and is used by indices of sociality to control for it. In contrast, bias of interaction is unknown to the observer. However, during simulation, generating bias of interaction inadvertently leads to observation bias. This complicates the separation of the effects of both biases and the estimation of which aspects are controlled by indices and/or permutation approaches.

To tackle these concerns, we developed a new simulation that allows to independently specify sampling bias and interaction bias. Using this simulation, we created various scenarios to assess the reliability of results obtained through different methods: association index, interaction index, Bayesian generative network, network permutation, double permutation and two new methods that will be discussed in the methodology section. Through this research, we have multiple goals. Firstly, we aim to assess whether issues related to false positives and false negatives in ASNA methods are associated with observation or interaction biases. We are expecting to observe accurate results for all approaches regarding bias of observation but we are also expecting that once interaction bias will be introduced in the simulation all methods will be inaccurate as none of them are accounting for such bias type. Our secondary objective is to determine if Bayesian generative networks reliability in terms of providing accurate estimates and/or differences in hypothesis testing. This is particularly important because the reliability of Bayesian generative networks hasn't been addressed yet. Additionally, having good reliability tests for them could enable us to account for interaction bias as they enable consideration of various types of confounding factors, as we will demonstrate through the presentation of one of your approaches. Our final aim is to identify analytical protocols that can yield more reliable results for hypothesis testing in ASNA. Over the years, multiple approaches have been proposed and faced with such a plethora of methods, researchers may encounter difficulties in assessing the strengths and weaknesses of each.

## Methods

### Modeling animal networks

In the vast majority of cases, researchers in animal behavior collect network data in the form of numerical outcomes—i.e., the number of times that ties *were observed*, conditional on the number of times that ties *could have been observed*. For example, researchers may conduct a specific number of observations (scans or focals), and the outcome data  $Y_{[i,j]}$  might reflect the number of observations in which directed (e.g., grooming or aggression events) or undirected (e.g., spatial associations) ties from individual  $i$  to individual  $j$  were observed. In this case, the number of observations is the sampling effort, an *exposure* variable which strongly—indeed, *proportionally*—influences the outcome variable. Thus, we let the variable  $E_{[i,j]}$  be the number of observations in which individuals  $i$  and  $j$  could have observed engaging in the dyadic behavior of interest. As such, a Bayesian model of the network data will generally take the form:

$$Y_{[i,j]} \sim \text{Binomial}(E_{[i,j]}, \phi_{[i,j]}) \quad (3)$$

where  $\phi_{[i,j]}$  is the *true* latent directed connection strength between  $i$  and  $j$ , represented as the probability that a directed tie from  $i$  to  $j$  occurs in a given observation.  $Y_{[i,j]}$  is typically a count of events or duration summed over all scans in the study.  $E_{[i,j]}$  is the number of observations in which ties between  $i$  and  $j$  could have been detected (e.g., the total number of scans or the sum of focals for individuals  $i$  and  $j$ . Farine and Whitehead 2015). Note that in the generative

context,  $\phi$  represents a true *biological* phenomenon, whereas the outcome variable,  $Y$ , and the exposure variable,  $E$ , may be strongly affected by the sampling protocol, researcher behavior, and features of animals  $i$  and  $j$  other than sociality (e.g., cryptic coloration might cause the researcher to fail to observe true ties).

The goal of modeling, is to measure  $Y$  and  $E$ , and then recover  $\phi$ . This can be complicated, however, if there are structural biases in  $Y$  and/or  $E$ . Here, we will consider two potential causes of bias: (1) sampling bias, where features of  $i$  and  $j$  influence  $E_{[i,j]}$ , and (2) detection bias, where features of  $i$  and  $j$  influence measurement of  $Y_{[i,j]}$  given  $E_{[i,j]}$  independent of  $\phi_{[i,j]}$ .

We will now build up a Bayesian model of the data generating process. We will start by describing the sub-model underlying the true weighted network of ties, and then integrate the sub-models underlying the measurement process.

## Social relations

**Model definition** We begin the process of building up the model described in Eq. 3 by providing a generative model for the *true* latent directed connection strength between individuals  $i$  and  $j$ ,  $\phi_{[i,j]}$ . In order to generate networks with empirically plausible typologies, it is generally necessary to define a model for  $\phi_{[i,j]}$  that includes correlated random effects for the propensity to send and receive ties (nodal random effects), and correlated dyad-level random effects for the propensity of  $i$  to send to  $j$  and  $j$  to send to  $i$ . Additionally, such models should permit inclusion of covariate effects that influence block/group structure, node-level tie propensity, and dyad-level tie propensity. See Fig. 1 for an example network generated under such a model.

Following prior work of Redhead et al. (2021); Ross et al. (2023), we recommend the use of the social relations model (Kenny and La Voie 1984; Snijders and Kenny 1999; Back and Kenny 2010) with an additional set of stochastic block-model parameters (Holland et al. 1983; Karrer and Newman 2011; Peixoto 2019) to account for group structure when necessary. As such,  $\phi_{[i,j]}$  can be modeled as:

$$\text{logit}(\phi_{[i,j]}) = \alpha + \lambda_{[i]} + \pi_{[j]} + \delta_{[i,j]} + \Gamma(i, j, V) + \dots \quad (4)$$

where  $\alpha$  is an intercept term,  $\lambda$  is a vector of individual-specific focal/sender/nominator effects,  $\pi$  is a vector of individual-specific target/receiver/nominee effects,  $\delta$  is a matrix of dyadic effects,  $\Gamma(i, j, V)$  is a function giving an intercept offset as a function of group/block structuring variables, and the ellipsis signifies any linear model of coefficients and focal, recipient, or dyadic covariates.

For example, if  $S$  is an animal-specific measure, like body size, and  $Q$  is a dyad-specific measure, like a matrix of relatedness ties, then the ellipsis may be replaced with:  $\kappa_{[1]}S_{[i]} + \kappa_{[2]}S_{[j]} + \kappa_{[3]}Q_{[i,j]}$ , to give the effects of body size on the probability of sending ties to any target and receiving ties from any target, and the effects of kinship on the probability of dyadic ties.

To model block structure, we can consider a list of  $V$  categorical variables describing individuals  $i$  and  $j$ . Let  $B_{[v,b_1,b_2]}$  be a three-dimensional parameter array,

$$\Gamma(i, j, V) = \left( \sum_{v=1}^V B_{[v,b(i,v),b(j,v)]} \right) \quad (5)$$

where the probability of a tie from individual  $i$  in block  $b(i, v)$  to individual  $j$  in block  $b(j, v)$  for variable  $v$  is controlled by the corresponding entry in the array of For discrete block modeling, we can utilize the assortativity coefficient: <https://arxiv.org/pdf/cond-mat/0209450.pdf>

**Priors** To complete the model definition, we define vague priors. We model the sender and receiver effects jointly using a multivariate normal distribution. This allows for generalized correlations at the individual level to be detected—i.e., we can detect if individuals who groom others are also more likely to be groomed by others. For computational efficiency (Stan Development Team 2021; Lewandowski et al. 2009), it is best to write the multivariate normal as:

$$\begin{pmatrix} \lambda_{[i]} \\ \pi_{[i]} \end{pmatrix} = \begin{pmatrix} \sigma_\lambda \\ \sigma_\pi \end{pmatrix} \circ \left( L * \begin{pmatrix} \hat{\lambda}_{[i]} \\ \hat{\pi}_{[i]} \end{pmatrix} \right) \quad (6)$$

where  $L$  is a Cholesky factor from the decomposition of the  $2 \times 2$  correlation matrix with  $\rho$  on the off-diagonal, and  $\hat{\lambda}_{[i]} \sim \text{Normal}(0, 1)$  and  $\hat{\pi}_{[i]} \sim \text{Normal}(0, 1)$  are unit-normal random effects. Weakly informative priors may then be independently specified on the variance and correlation terms (Lewandowski et al. 2009):

$$\sigma_\lambda \sim \text{Exponential}(1.0) \quad (7)$$

$$\sigma_\pi \sim \text{Exponential}(1.0) \quad (8)$$

$$L \sim \text{LKJ Cholesky}(2.5) \quad (9)$$

We use the above approach to define the dyad-level random effects as well:

$$\begin{pmatrix} \delta_{[i,j]} \\ \delta_{[j,i]} \end{pmatrix} = \begin{pmatrix} \sigma_\delta \\ \sigma_\delta \end{pmatrix} \circ \left( L_\delta * \begin{pmatrix} \hat{\delta}_{[i,j]} \\ \hat{\delta}_{[j,i]} \end{pmatrix} \right) \quad (10)$$

where  $\hat{\delta}_{[i,j]} \sim \text{Normal}(0, 1)$  have unit-normal priors, and the variance and correlation terms have weakly informative priors:

$$\sigma_\delta \sim \text{Exponential}(1.0) \quad (11)$$

$$L_\delta \sim \text{LKJ Cholesky}(2.5) \quad (12)$$

Under this model,  $\rho_\delta$  provides a measure of dyadic reciprocity—i.e., whether the probability of focal  $i$  giving to alter  $j$ , increases with the probability that focal  $j$  gives to alter  $i$ .

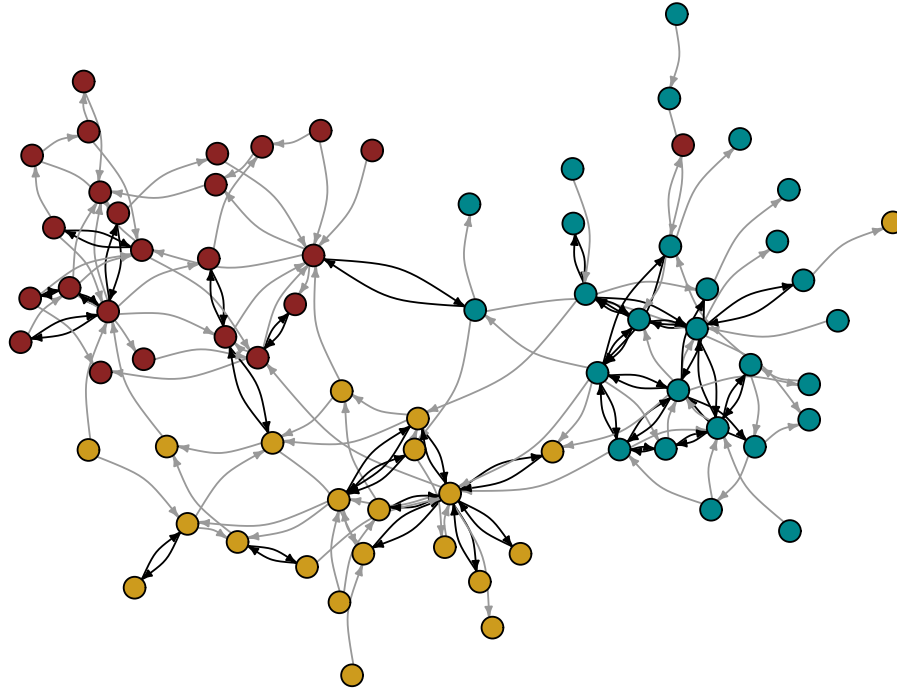
We recommend standardizing predictor variables. We can then use weakly regularizing priors on the  $\kappa$  terms:

$$\kappa_{[p]} \sim \text{Normal}(0, 1) \quad (13)$$

## Sampling bias

The next step in building the model described in Eq. 3, is to provide a generative model for the sampling effort between individuals  $i$  and  $j$ ,  $E_{[i,j]}$ . Most data collection protocol designs can be separated in two categories: discrete time sampling rule and continuous recording sampling rule (Bateson and Martin 2021). Discrete time sampling rule are instantaneous or one-zero sampling such as “gambit of the group” or “scan” sampling protocol (See Sosa et al. (2021b) for further details) which involves behavioral events without duration collected through “snapshots”. In such data collection protocol, the sampling effort of an individual

**Figure 1.** Simulated animal network data under the generative model described in [Ross et al. \(2023\)](#). Nodes are colored by group. The stochastic block model introduces gross substructure, where ties within-groups are more likely than ties between groups. These rates can be varied continuously, producing networks with no group structure on one extreme, to networks fully segmented by group identity on the other. Random effects for sending and receiving ties lead to variable degree. Here, some individuals have small degree and are connected to the network by only a single tie, and others have a high degree and are connected to many other individuals. As with the parameters controlling group structure, the parameters controlling variance (and covariance) in individual-level propensity to send and receive ties can be varied continuously, producing networks with approximately uniform degree on one extreme, to networks with highly unequal degree distributions on the other extreme. Finally, we note that some dyads form reciprocal ties (note the black, bidirectional arrows), while other dyads are linked only by unidirectional ties (note the grey, unidirectional arrows). The parameters controlling variance (and covariance) in dyadic random effects, are continuously variable, and can produce networks with close to no dyadic tie reciprocation on one extreme, and networks with high rates of tie reciprocation on the other.



$i$  would be the count of sampling periods during which it was observed. Continuous recording sampling rule are behavioral state with (e.g., time of grooming) or without (e.g., frequencies of grooming) duration collected with focal sampling protocol. The sampling effort of an individual  $i$  would be the time of all focal bouts made on  $i$ . Whether for discrete or continuous recording sampling a simple model of sampling effort can be written as:

$$\hat{E}_{[i]} \sim \text{Binomial}(\bar{E}, \epsilon_{[i]}) \quad (14)$$

where  $\bar{E}$  is maximum sampling effort (i.e., the count of observations for individual  $i$ , either as the number of scans or the number of focals),  $\hat{E}_{[i]}$  is the realized number of scans or focals of individual  $i$ , and  $\epsilon_{[i]}$  is the probability that individual  $i$  is observed from the researcher. The parameter  $\epsilon_{[i]}$  might vary as a function of individual-level characteristics, that also influence network structure. For example, an animal with a cryptic phenotype may be more likely to be unobserved—even as the focal in a focal-follow—if their coloration allows them to escape the attention of the researcher performing the behavioral observation. Similarly, individuals with a cryptic phenotype may be under-sampled, because they are less likely to be found on a given day, and are thus subject to fewer focal follows than desired overall.

$$\text{logit}(\epsilon_{[i]}) = \nu_{[0]} + \dots \quad (15)$$

where  $\nu_{[0]}$  is an intercept, and the ellipsis signifies any linear model of coefficients and individual-level covariates. For example, if  $S$  is an animal-specific measure, like a binary variable for cryptic coloration, then the ellipsis may be replaced with:  $\nu_{[1]}S_{[i]}$ , to give the effects of coloration on number of scans.

Then, following standard approaches (e.g., [Farine and Whitehead 2015](#)), the dyad-level count of scans is given by the formula:

$$E_{[i,j]} = \hat{E}_{[i]} + \hat{E}_{[j]} \quad (16)$$

as ties between  $i$  and  $j$  can be detected either when  $i$  is subject to observation or when  $j$  is subject to observation.

From our formalization, it is obvious that variation in sampling effort should have little to no effect on our ability to identify accurate estimates of  $\phi_{[i,j]}$  from Eq. 3, as long as we apply statistical methods that propagate uncertainty according to the rules of probability theory. A binomial model of  $Y_{[i,j]}$  given  $E_{[i,j]}$  provides information about expected value of  $\phi_{[i,j]}$  that does not depend on  $E_{[i,j]}$ ; however, the narrowness of the posterior distribution of  $\phi_{[i,j]}$  does depend on  $E_{[i,j]}$ , as we can be more confident in estimates of  $\phi_{[i,j]}$  that are based on more observations.



### Interaction bias, or censoring

Censoring, or “interaction bias” is a more severe problem. In this case, a researcher performed  $E_{[i,j]}$  scans and detected  $\hat{Y}_{[i,j]}$  ties. However, if he or she is following animal  $i$ , and animal  $j$  has a cryptic phenotype, then it is possible that ties from  $i$  to  $j$ , or  $j$  to  $i$ , did occur, but the researcher did not detect them. For a single observation, the indicator of a true tie be  $Q_{[i,j]} \in \{0, 1\}$ , the indicator of a detected tie be  $\hat{Q}_{[i,j]} \in \{0, 1\}$ , and the indicator for individual  $i$  being detectable be  $D_{[i]} \in \{0, 1\}$ . Then, we find that:

$$Pr[\hat{Q}_{[i,j]} = 0] = Pr[(1 - Q_{[i,j]})D_{[i]}D_{[j]}] + Pr[1 - D_{[i]}D_{[j]}] \quad (17)$$

$$Pr[\hat{Q}_{[i,j]} = 1] = Pr[Q_{[i,j]}D_{[i]}D_{[j]}] \quad (18)$$

where zeros arise as a mixture of two process: (1) a process where both  $i$  and  $j$  were detectable, but no tie occurred, and (2) a process where either  $i$  or  $j$  was censored; ones, however, can only occur when both  $i$  and  $j$  were detectable and a true tie occurred.

Let  $\eta_{[i]} = Pr[D_{[i]} = 1]$  describe the detectability of individual  $i$  and  $\phi_{[i,j]} = Pr[(1 - Q_{[i,j]})]$ . We note that Eqs. 17 and 20 define the probability mass function of a Bernoulli random variable, and so we can aggregate over scans,  $E_{[i,j]}$ , to yield a Binomial model. Thus, we can rewrite Eq. 3 as:

$$\hat{Y}_{[i,j]} \sim \text{Binomial}(E_{[i,j]}, \phi_{[i,j]}\eta_{[i]}\eta_{[j]}) \quad (19)$$

We can let  $\eta_{[i]}$  depend on individual-specific covariates:

$$\text{logit}(\eta_{[i]}) = \psi_{[0]} + \dots \quad (20)$$

where  $\psi_{[0]}$  is an intercept, and the ellipsis signifies any linear model of coefficients and individual-level covariates. For example, if  $S$  is an animal-specific measure, like a binary variable for cryptic coloration, then the ellipsis may be replaced with:  $\psi_{[1]}S_{[i]}$ , to give the effects of coloration on detectability.

Here, however, we run up against the limits of inference. Since  $\phi_{[i,j]}$  and  $\eta_{[i]}$  multiply, the likelihood cannot distinguish between effects, of  $S_{[i]}$  for example, that increase the odds of a tie and effects that increase the odds of detectability.

The only way to permit accurate inference, is to use an independent source of data to anchor estimation of  $\eta_{[i]}$ . If, for example, a researcher conducts  $\bar{Z}_{[i]}$  trials to encounter individual  $i$  in a setting where individual  $i$  is known to be present, and observes individual  $i$  in  $Z_{[i]}$  of those trials, this permits inclusion of a model like:

$$Z_{[i]} \sim \text{Binomial}(\bar{Z}_{[i]}, \eta_{[i]}) \quad (21)$$

allowing estimation of social-network parameters that are robust to censoring, or ‘interaction bias’.

### Testing ASNA methods

**Scenarios:** Using the previously explained simulation model, we generated six types of scenarios. Scenario 1, we simulated data where individuals’ characteristics have no effect on their social ties and without sampling or censoring biases. This will allow to assess rates of false positives for each approach tested. Scenario 2, we simulated data where

individuals’ characteristics have an effect on their social ties and without sampling or censoring biases. This will allow to assess rates of false negatives for each approach tested. In Scenario 3, we simulated data where individuals’ characteristics have no effect on their social ties, introducing sampling bias but without censoring biases. This helps assess rates of false positives for each approach tested in the presence of sampling biases. In Scenario 4, data simulation includes individuals’ characteristics affecting social ties with sampling biases and no censoring biases, allowing us to assess rates of false negatives for each approach tested under sampling biases conditions. In scenario 5, we simulated data where individuals’ characteristics have no effect on their social ties, introducing censoring biases without sampling biases. This helps assess rates of false positives for each approach tested in the presence of censoring biases. In scenario 6, data simulation includes individuals’ characteristics affecting social ties without sampling biases and with censoring biases, allowing us to assess rates of false negatives for each approach tested under censoring biases conditions.

For each scenario, we generated 100 simulations with random variations in the simulation parameters related to the effect size of individuals’ characteristics on social ties, sampling biases, and censoring bias. Details on the range of parameter variations are presented in Table 1.

**Tested approaches:** For each simulation, we assessed 5 approaches. Initially, we tested the association index without permutations, along with network and double permutations. Next, we examined the interaction index with the same permutation variations. Additionally, Bayesian generative networks modeling interaction observations (e.g., Bison) and link weight probabilities (e.g., STRAND) have been tested. Finally, we introduced a new method using weighted regression model for association and interaction indices.

Indeed, apply association (Eq. 1) or interaction (Eq. 2) indices to a point-estimates  $\hat{\phi}_{[i,j]}$  models can violate homoscedasticity linear regression assumptions, especially in small samples. This occurs because Eq. 1 and Eq. 2 divide out sample size, and provide samples based on sparse evidence disproportionate weight in downstream analysis Fig. 2.a. Unlike Bayesian approaches that automatically adjust the influence of each point in proportion to the sample size of observations Fig 2.b, regression models assume homoscedasticity. Homoscedasticity implies equal error distribution between data points, leading to similar values in the diagonal covariance matrix (Eq. 22)

$$C = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix} \quad (22)$$

Where  $C$  is the covariance matrix, Diagonal elements  $\sigma^2$  represent the variance of each observation error and they are all the same because errors are identically distributed.

We propose using weighted linear regression that incorporates the sampling effort of each data point into the covariance matrix of errors, providing advantages when handling heteroscedastic data (Eq. 23).

$$C = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix} \quad (23)$$

Where  $C$  is the covariance matrix and where diagonal elements  $\sigma_1^2$  to  $\sigma_m^2$  are not identical and each observation has its own variance, unlike in Eq.23.

For each approach in every simulation, we calculated the estimated effect size along with the permuted p-value. This was done to assess the accuracy of approach estimations and the correctness of hypothesis testing conclusions, respectively.

## Results

### Evaluating the accuracy of the simulation model.

We assessed the simulation method reliability based on variations in three terms. Initially, we simulated data without any influence of individuals' characteristics on their social interaction patterns, observations, or censoring (Fig. 3.A). Next, we simulated data with only an effect of individuals' characteristics on their social interaction patterns (Fig. 3.B). Additionally, we conducted a simulation with only an effect of individuals' characteristics on their observations (Fig. 3.C). Finally, we performed a simulation with only an effect of individuals' characteristics on their censoring (Fig. 3.D). For each simulation type, we used linear regressions to examine the relationship between individuals' characteristics and their social interaction patterns, observations, and censoring, ensuring the reliability of the simulation results. Regression outcomes are summarized in Table 2.

### Scenario 1

### Scenario 2

### Scenario 3

### Scenario 4

### Scenario 5

### Scenario 6

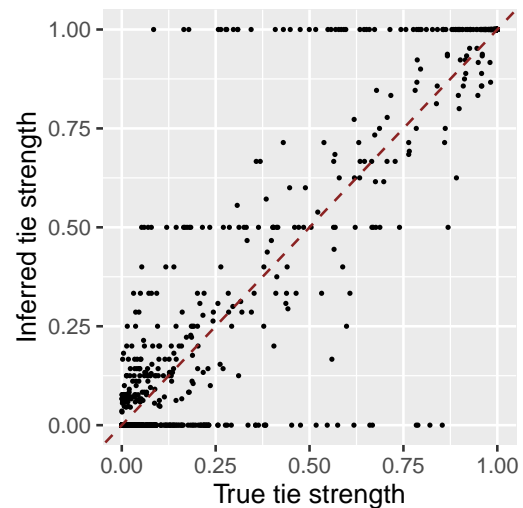
## Discussion

## References

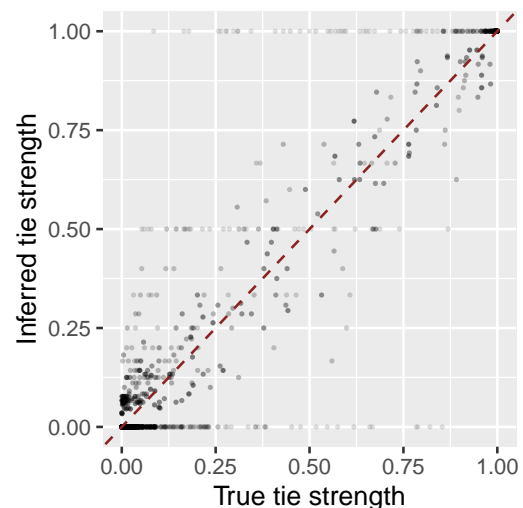
- Apicella CL, Marlowe FW, Fowler JH and Christakis NA (2012) Social networks and cooperation in hunter-gatherers. *Nature* 481(7382): 497–501.
- Back MD and Kenny DA (2010) The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass* 4(10): 855–870.
- Bateson M and Martin P (2021) *Measuring behaviour: an introductory guide*. Cambridge university press.
- Farine DR (2013) Animal social network inference and permutations for ecologists in r using asnipe. *Methods in Ecology and Evolution* 4: 1187–1194.

**Figure 2.** Example data on tie strength generated under a model where a covariate  $Z$  has a strong positive effect on tie probability,  $\phi_{[i,j]}$ , but a negative effect on sampling effort,  $E_{[i,j]}$ . We plot the true dyadic tie strength on the x axis, and the simple ratio index on the y axis. In frame (a), we see that many points lie along the diagonal (as expected), but there are also strips of points along the horizontal lines at  $y = 0.0$ ,  $y = 0.5$ , and  $y = 1.0$ , where the inferred tie strength is not all that reflective of true tie strength. These points correspond to estimates from small samples. Simply plugging in SRI estimates into a downstream regression leads to poor inference, because samples based on few data-points obscure the underlying predictors of tie strength.

(a) Simple ratio index



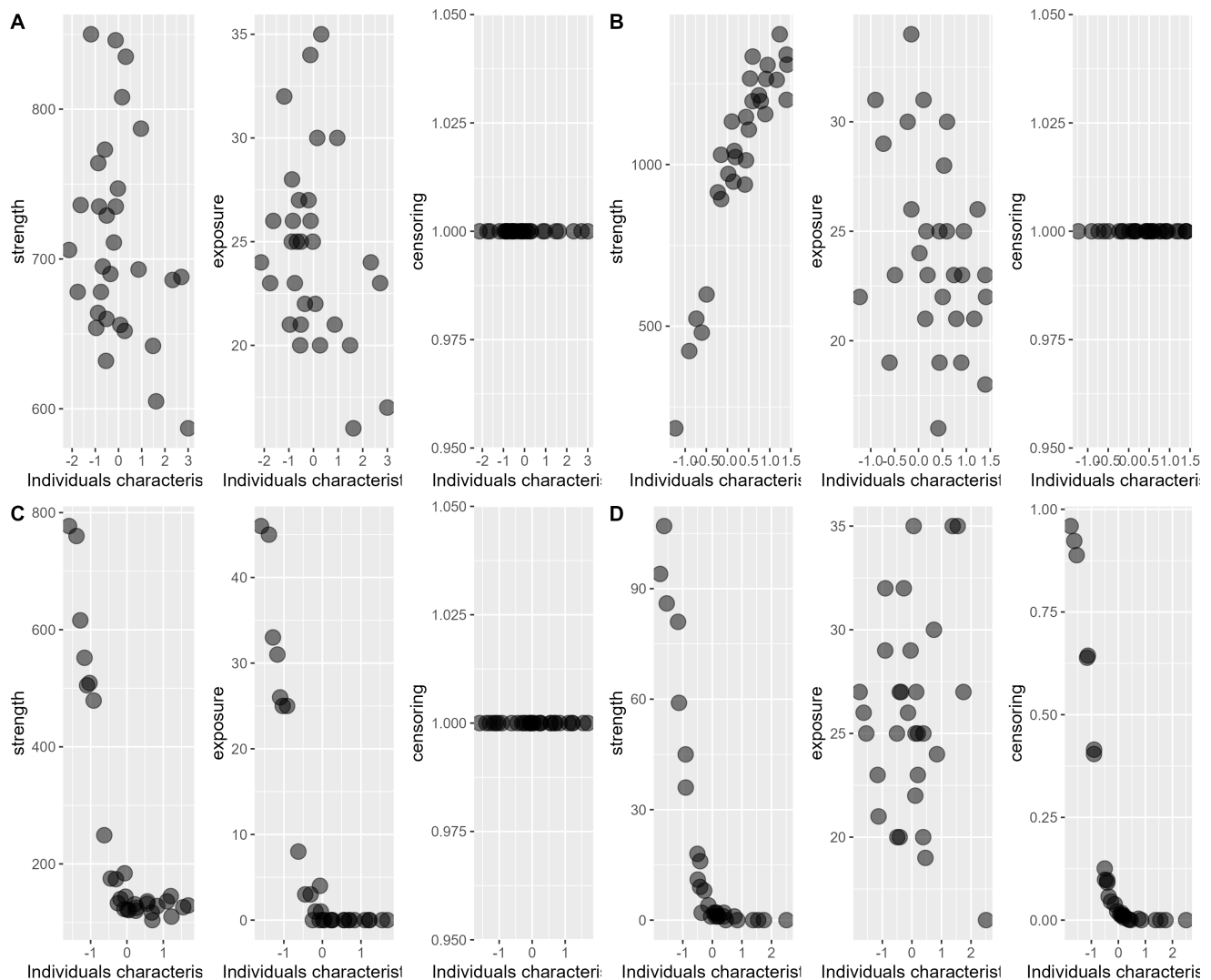
(b) A Bayesian model would scale the influence of each point in proportion to the sample size of observations, as we show here by scaling the alpha values of each point by  $E_{[i,j]}$ . The fact that most of the weight of evidence lies along the diagonal is now much more apparent.



Farine DR and Carter GG (2022) Permutation tests for hypothesis testing with animal social network data: Problems and potential solutions. *Methods in Ecology and Evolution* 13: 144–156. DOI:10.1111/2041-210X.13741.

Farine DR and Whitehead H (2015) Constructing, conducting and interpreting animal social network analysis. *Journal of Animal*

Figure 3. XXXX



*Ecology* 84: 1144–1163. DOI:10.1111/1365-2656.12418.

475 Hart J, Weiss MN, Franks D and Brent L (2023) Bison: A bayesian framework for inference of social networks. *Methods in Ecology and Evolution* 14(9): 2411–2420.

Hart JD, Weiss MN, Brent LJ and Franks DW (2022) Common permutation methods in animal social network analysis do not control for non-independence. *Behavioral Ecology and Sociobiology* 76: 1–10. DOI: 10.1007/S00265-022-03254-X/FIGURES/3. URL <https://link.springer.com/article/10.1007/s00265-022-03254-x>.

485 Holland PW, Laskey KB and Leinhardt S (1983) Stochastic blockmodels: First steps. *Social networks* 5(2): 109–137.

Hubalek Z (1982) Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biological Reviews* 57(4): 669–689.

490 Karrer B and Newman ME (2011) Stochastic blockmodels and community structure in networks. *Physical review E* 83(1): 016107.

Kenny DA and La Voie L (1984) The social relations model. *Advances in Experimental Social Psychology* 18: 141–182.

495 Lewandowski D, Kurowicka D and Joe H (2009) Generating random correlation matrices based on vines and extended

onion method. *Journal of multivariate analysis* 100(9): 1989–2001.

Milgram S (1967) The small world problem 2: 60–67.

Peixoto TP (2019) Bayesian stochastic blockmodeling. *Advances in network clustering and blockmodeling* : 289–332. 500

Puga-Gonzalez I, Sueur C, Sosa S, Puga-Gonzalez I, Sueur C and Sosa S (2021) Null models for animal social network analysis and data collected via focal sampling: Pre-network or node network permutation? *Methods in Ecology and Evolution* 12: 22–32. DOI:10.1111/2041-210X.13400. URL <https://doi.org/10.1111/2041-210X.13400>. 505

Ravasz E, Somera AL, Mongru DA, Oltvai ZN and Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. *science* 297(5586): 1551–1555.

Redhead D, McElreath R and Ross CT (2021) Reliable network inference from unreliable data: A tutorial on latent network modeling using strand . 515

Ross CT, McElreath R and Redhead D (????) Modelling animal network data in r using strand. *Journal of Animal*

- 520 *Ecology* n/a(n/a). DOI:<https://doi.org/10.1111/1365-2656.14021>. URL <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.14021>.
- Ross CT, McElreath R and Redhead D (2023) Modelling animal network data in r using strand. *Journal of Animal Ecology* .
- 525 Sailer LD and Gaulin SJ (1984) Proximity, sociality, and observation: the definition of social groups. *American Anthropologist* : 91–98.
- Snijders TA and Kenny DA (1999) The social relations model for family data: A multilevel approach. *Personal Relationships* 6(4): 471–486.
- 530 Sosa S, Jacoby DMP, Lihoreau M and Sueur C (2021a) Animal social networks: Towards an integrative framework embedding social interactions, space and time.
- Sosa S, Pasquaretta C, Puga-Gonzalez I, Dobson FS, Viblanc VA and Hoppitt W (2021b) Reliable hypotheses testing in animal 535 social network analyses: Global index, index of interactions, and residual regression. *bioRxiv* .
- Stan Development Team (2021) Stan modeling language users guide and reference manual, version 2.21. URL <http://mc-stan.org/>.
- 540 Ter Wal AL and Boschma RA (2009) Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science* 43: 739–756.
- Ulanowicz RE, Holt RD and Barfield M (2014) Limits on ecosystem trophic complexity: insights from ecological network 545 analysis. *Ecology letters* 17(2): 127–136.
- Weiss MN, Franks DW, Brent LJ, Ellis S, Silk MJ and Croft DP (2021) Common datastream permutations of animal social network data are not appropriate for hypothesis testing using 550 regression models. *Methods in Ecology and Evolution* 12(2): 255–265.
- Whitehead H (2008) Analyzing animal societies. In: *Analyzing Animal Societies*. University of Chicago Press.
- Whitehead HAL (1999) Testing association patterns of social 555 animals. *Animal Behaviour* 57: F26–F29. DOI:10.1006/anbe.1999.1099.