*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.

# INTRODUCTION

Over the past 50 years, graph theory has become a vital tool in studying natural and artificial systems in various fields like anthropology, sociology, economics, ecology, ethology, and animal societies. It applies to both small-scale (e.g., proteomics) and large-scale (e.g., ecosystems) systems. In the realm of animal sociality research, innovative techniques, such as association indices and pre-network permutations, 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).

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., they considered a scenario with data biases stemming from the data collection process (like oversampling certain individual categories). They found false positive rates ranging from 35.6% to 60.8%, and false positive rates ranging from 36.6% to XXX. Variations in error rates are connected to the different protocols used and will be discussed further. 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 were 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. 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 observations between individuals). Currently two main types of social index are used: association indices and interaction indices. Association indices have been by behavioral ecologist. Although, many of the existing association indices estimate the proportion of time that a pair of individuals spends together. 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. Currently, there is only one approach that has been suggested for addressing the social interaction patterns among individuals. XXX employs a Bayesian generative network approach to estimate the uncertainty of links related to observations. Also, there hasn't been any testing conducted to assess the reliability of results based on biased data. This is particularly important because it doesn't estimate the rates of false positives and false negatives. The lingering questions is then whether Bayesian generative networks can have good rates of false positives and false negatives and whether it can outperform sociality indices approaches or not.

(Eq.1)

Where is the number of sampling periods with *i* and *j* observed associated, is the number of sampling periods with only *i* identified, is the number of sampling periods with only j identified and is the number of sampling periods with i and j identified but not associated.

To achieve the second step, which involves testing statistical hypotheses about sociality, 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 three forms: network permutations, pre-network permutations, and double permutations. For more detailed information on the principles of permutation approaches, readers can refer to XXXX. 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. Permutations have recently faced criticism for producing permutations that are essentially similar. Moreover, studies that highlighted problems with permutation approaches or propose new methods have been employing the same simulation protocol. As demonstrated in Appendix 1, this simulation has some issues. Firstly, these simulations were designed to generate a specific bias known as the "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, pre-network permutation, network permutation, and double permutation. Through this analytical approach, 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. Secondly, we want to determine if Bayesian generative networks outperform traditional ASNA methods in terms of providing accurate estimates and/or differences in hypothesis testing. Lastly, this will help us identify analytical protocols that can produce more reliable results for hypothesis testing in ASNA.

# METHODS

For a given population *N* we generated ties probability using the following mass function:

Where is the probability of getting exactly *x* successes, *n* is the number of trials and *p* the probability of success on each trial. We define *p* using the following multivariate normal distribution:

Where is XXX, is XXX, is XXX, is XXX, is a parametrizable predictors' effect, is a parametrizable individual characteristic. Thus, by specifying , we can accurately define a particular tie probability based on individuals' characteristics of a given dyad (*i, j*).

*n* is define as fellow:

Where is the number of observations for individual *i,*  is themean population number of observations and the given variance for individual *i.* is a given parametrizable observation predictors' effect. is individual *i* characteristic.is individual *i* variance draw from a parametrizable gaussian distribution. Thus, by specifying , we can accurately define a particular observation bias based on individuals' characteristics.

RESULTS

# DISCUSSION

For observation bias, SRI seems equivalent to Bayesian and rates is the less accurate approach. However, SRI is usually computed on undirected networks and this undirected version is found in all software's. People will thus need some coding experience to adapt the SRI (code can be found in R script XXX and ANTs will have a version of it)