Accuracy and Power Analysis of Social Interaction Networks

Jordan D. A. Hart [*1], Daniel W. Franks[2], Lauren J. N. Brent[1], and Michael N. Weiss[1,3]

[1]Centre for Research in Animal Behaviour, University of Exeter, UK
[2]Departments of Biology and Computer Science, University of York, UK
[3]Center for Whale Research, Friday Harbour, WA, USA

Abstract

1. Power analysis is used to estimate the probability of correctly rejecting a null hypothesis for a given statistical model and dataset. Conventional power analyses assume complete information, but the stochastic nature of behavioural sampling can mean that true and estimated networks are poorly correlated. Power analyses do not currently take the effect of sampling into account. This could lead to inaccurate estimates of statistical power, potentially yielding misleading results.
2. Here we develop a method for computing *network correlation*: the correlation between an estimated social network and its true network, using a Gamma-Poisson model of interaction rates. We use simulations to assess how the level of network correlation affects the power of nodal regression analyses. We also develop a generic method of power analysis applicable to any statistical test, based on the concept of diminishing returns.
3. We demonstrate that our network correlation estimator is both accurate and moderately robust to its assumptions being broken. We show that social differentiation, mean interaction rate, and the harmonic mean of sampling times positively impacts the strength of network correlation. We also show that the required level of network correlation to achieve a given power level depends on many factors, but that 80% network correlation usually corresponds to around 80% power for nodal regression in ideal circumstances.
4. We provide guidelines for using our network correlation estimator to verify the accuracy of interaction networks, and to conduct power analysis. This can be used prior to data collection, in post hoc analyses, or even for subsetting networks in dynamic network analysis. The network correlation estimator and custom power analysis methods have been made available as an R package.

Introduction

Understanding the form and function of social systems is a core aim of behavioural ecology, and social network analysis has become a central tool for exploring these topics (Croft et al., 2008). However, social network analysis comes with challenges unique to the field, ranging from issues of missing or censored data, through to problems of non-independence in network measures (James et al., 2009; Farine and Whitehead, 2015). Significant advances have been made in many of these areas over the last two decades, but the problem of conducting power analysis has generally received little attention, with the exception of Whitehead (2008). Power analysis is used to determine the amount of data required to reject a null hypothesis assuming a given effect size, and is used to determine how a study should be conducted prior to collecting data, or in post-hoc analysis to determine if a study was sufficiently powered to reject a null hypothesis (Cohen, 1992, 2013). Animal social network studies rarely use power analysis prior to data collection but often do employ post hoc power analysis to indicate the reliability of results (Martin et al., 1993; Stadtfeld et al., 2020). In this study, we build on previous work to develop a method for estimating the accuracy of social networks, and show how this can be used to conduct power analysis on social networks constructed from interaction rates.

Whitehead (2008) introduced a method for estimating the accuracy of animal social networks constructed from association rate data by estimating the correlation of the observed association rates with the true underlying association rates. This method has often been used as a post hoc measure to verify the robustness of animal social networks before further analyses (Findlay et al., 2016; Ellis et al., 2017). However, this method is designed specifically for association data, but there are several other common types of data used to construct animal social networks. Networks are typically built from weighted edges between nodes that encode the strengths of relationships between individuals (Krause et al., 2015). Edge weight can be quantified in a number of different ways depending on the nature of the social system and the data available. There are two types of data often used for constructing social networks: association and interaction. Association data record instances where individuals have the opportunity to interact, generally using proximity thresholds or co-detections, and interaction count data record interactions such as grooming or aggression between individuals (Croft et al., 2011; Brent et al., 2013; Brask et al., 2021).

The correlation between true and estimated networks, which will refer to as *network correlation*, is a useful intuitive measure of the accuracy of a network, but even more importantly it is directly related to the power of statistical analyses. Whitehead (2008) made a series of recommendations based on the social preference test developed by Bejder et al. (1998), suggesting that a good general guide for maximal power is when the product of the squared social differentiation and mean number of observations per individual is greater than 5 (in their notation: $S^2 \times H > 5$). This recommendation is based only on the social preference test, and it isn't known how well this generalises to other statistical tests commonly used in social network analysis Bejder et al. (1998). In particular, one of the most common methods for testing hypotheses in social network analysis is nodal regression (Croft et al., 2011). Nodal regression uses a node-level social network metric such as node strength, eigenvector centrality, or closeness to quantify an individual's position in their social structure, and relates this social position to quantifiable traits such as age or sex (Farine and Whitehead, 2015). The relationship between network metric and trait is usually analysed using regression (Weiss et al., 2021b).

The statistical power of a conventional regression depends on sample size, effect size, and significance level. In a nodal regression, sample size is the number of individuals, the true effect size is fixed and unknown, and the significance level is set by convention, usually to 0.05 (Cohen, 1994). However, the true effect size is the effect size if the nodal regression was run on the true, unknown network. The estimated network will not perfectly match the true network, meaning that in turn the estimated effect size will deviate from the true effect size. Applying conventional power analysis to nodal regression could therefore lead to under- or over-estimates of power, potentially yielding misleading results.

In this study we extend the method developed by Whitehead (2008) for estimating network correlation from networks constructed from association data to networks constructed from interaction count data. Interaction count data are usually collected by recording instances of pairs of individuals interacting over many sampling periods using either scan or focal sampling (Martin et al., 1993). The result is an integer count of interactions X_{ij} between each pair of individuals i and j, and a positive real-valued sampling time for each pair d_{ij}, which is simply the sum of time spent sampling at least one individual of the pair. The rate of interactions is then computed by dividing the interaction count by the sampling time, giving the interaction rate:

$$\hat{\lambda}_{ij} = \frac{X_{ij}}{d_{ij}}. \tag{1}$$

As with any empirical measure, sampling has the potential to influence the estimated interaction rate significantly (Franks et al., 2010). Consider the case when X_{ij} is 10 interactions and d_{ij} is 5 minutes, if interaction counts are Poisson-distributed, the estimate will be 2 interactions per minute, but the 95% confidence interval will be $(0.8, 3.4)$, with a range of 2.6. Compare this to the case where X_{ij} and d_{ij} are 100 interactions and 50 minutes respectively; the point estimate is the same, but the 95% confidence interval is $(1.6, 2.4)$, with a range of 0.8, less than one third of the range of the previous case, but for around 10 times the sampling effort. This shows that we should have less confidence in the estimate of the first case compared to the second case, and demonstrates how the estimated network could deviate significantly from the true network because of low sampling time. The magnitude of this deviation will impact the reliability of an analysis, and ideally should be taken into account when conducting social network analyses. This problem was recognised by Whitehead (2008), who developed a method based on the binomial distribution for estimating network correlation for networks constructed from association data.

We demonstrate that our method for interaction count data provides accurate estimates of network correlation in realistic scenarios, and we use simulations to suggest guidelines for the minimum level of sampling required for nodal regression depending on the desired level of statistical power. We also develop a generic alternative approach for guiding data collection based on the principle of diminishing returns that can be used for any type of statistical analysis, including but not limited to nodal regression. We contrast our guidelines to those from Whitehead (2008) to show that the amount of sampling required depends on the type of data, and that data-specific guidelines may be highly useful when designing and conducting social network analysis. We have made the methods available as an R package: *pwrIRGP*, which is available at `www.github.com/JHart96/pwrIRGP`.

Methods

By modelling interaction counts as being distributed according to a Gamma-Poisson process, we are able to analytically derive an equation for network correlation. We verify our network correlation equation using simulations which either follow the assumptions of the model, or break them to varying degrees, to test the robustness of the method. Following this, we use simulations to determine the level of network correlation required to obtain a desired level of power when performing nodal regression on interaction networks.

Gamma-Poisson model of dyadic interactions

As sampling time increases, we expect that the estimated interaction rate $\hat{\lambda}_{ij}$ will get closer to the true interaction rate λ_{ij}. However for lower sampling times there may be a considerable error between the estimated and true interaction rate. This error can be modelled by treating the interaction counts X_{ij} as draws from a Poisson distribution. Since the underlying true interaction rates λ_{ij} are unknown, we assume the true interaction rates of the dyads are drawn from a gamma distribution: $\lambda_{ij} \sim \text{Gamma}(\alpha, \beta)$. A Poisson-distributed random variable with rates drawn from the gamma distribution is equivalent to a random variable following the negative binomial distribution, therefore the number of observed interactions X_{ij} is given by:

$$X_{ij} \sim \text{NegBinomial}\left(\alpha, \frac{\beta}{\beta + d_{ij}}\right). \tag{2}$$

Using this, we can estimate the variance of both the true and estimated interaction rates, which allows us to estimate the Pearson's correlation coefficient, ρ, between them:

$$\rho(\lambda, \hat{\lambda}) = \frac{S\sqrt{\mu H(d)}}{\sqrt{1 + \mu S^2 H(d)}} = \frac{S\sqrt{I}}{\sqrt{1 + S^2 I}} \tag{3}$$

where $S = 1/\sqrt{\alpha}$ is the social differentiation, $\mu = \alpha/\beta$ is the dyadic mean interaction rate, $H(d)$ is the harmonic mean of the sampling times d_{ij}, and $I = \mu H(d)$ reflects the sampling effort. The harmonic mean $H(d)$ of the m dyads is defined as $H(d) = m/\sum_{i,j} d_{ij}^{-1}$, and is equal to the arithmetic mean only in the case where all d_{ij} are equal. Note that when sampling time d_{ij} is even across all dyads, sampling effort I is the number of interactions observed per dyad. When sampling time d_{ij} is uneven, sampling effort I will be lower, and more sampling time will be required to reach the same sampling effort as the equivalent network with evenly-sampled dyads. The network correlation is computed only over dyads that have non-zero sampling times.

The parameters α, β of the underlying gamma distribution can be estimated numerically using maximum likelihood or an MCMC algorithm such as Gibbs sampling (Hammersley, 2013). We use point estimates from maximum likelihood in this study to reduce computation time, and to avoid model fitting problems, but we have included both a quadratic approximation and MCMC version in the code to allow the confidence and credible intervals of the network correlation to be estimated.

Simulations 1: Verification of the Gamma-Poisson model

To confirm that Equation 3 is appropriate for interaction rate data, and to determine how robust it is to the Gamma-Poisson assumptions being broken, we ran simulations under three different scenarios. These scenarios follow or break the assumptions of Equation 3 to varying degrees (see Table 1 for more details). The parameter space of these models was explored using a random search for $S \in (0, 2]$ and $\mu \in (0, 10]$. For scenario 2, an additional constraint was applied such that interaction rates are stronger between members of the same group than between members of different groups. The simulations proceeded as follows:

1. Data were generated according to one of the scenarios outlined in Table 1.
2. The true correlation between λ and $\hat{\lambda}$ was computed.
3. The parameters μ and S were estimated using maximum likelihood and used to compute the estimated correlation $\rho(\lambda, \hat{\lambda})$ between λ and $\hat{\lambda}$.
4. The network correlations were recorded and the process was repeated 200 times.

The relationship between the true and estimated network correlations was quantified by computing the Pearson correlation coefficient between the estimated network correlations and true network correlations, the mean absolute error, and the standard deviation of the error.

Table 1: Simulation models used for verifying the Gamma-Poisson model, where simulation 1 follows the assumptions perfectly, but simulations 2 and 3 break the assumptions in different ways. The amount of sampling time d_{ij} for each dyad is drawn from a Poisson distribution. In model 2 the probability of a dyad belonging to the same group is given by p.

Scenario	Description	Model
1	Follow the assumptions of the model exactly, with gamma distributed interaction rates and Poisson-distributed interaction counts.	$X_{ij} \sim \text{Poisson}(\lambda_{ij} d_{ij})$ $\lambda_{ij} \sim \text{Gamma}(\alpha, \beta)$
2	Introduce community structure with two classes of dyads: same or different group. The two classes have two different parameterisations of the gamma distribution. Interaction counts are Poisson-distributed.	$X_{ij} \sim \text{Poisson}(\lambda_{ij} d_{ij})$ $\lambda_{ij} \sim \text{Gamma}(\alpha_{k_{ij}}, \beta_{k_{ij}})$ $k_{ij} \sim \text{Bernoulli}(p)$
3	Simulating sparse networks by modelling interaction counts as a zero-inflated Poisson with a sparsity $(1 - \pi)$.	$X_{ij} = Z_{ij} Y_{ij}$ $Y_{ij} \sim \text{Poisson}(\lambda_{ij} d_{ij})$ $\lambda_{ij} \sim \text{Gamma}(\alpha, \beta)$ $Z_{ij} \sim \text{Bernoulli}(\pi)$

Simulations 2: Statistical power of nodal regression

We ran simulations to test how statistical power relates to sampling effort in the case where the true effect size is the minimum required to achieve 100% power. This made it possible to see the effect of sampling effort without the effect being hidden behind under- or over-powered tests. To simulate data collection we followed the core assumptions of the model: that dyad-level interaction rates are drawn from a gamma distribution, and observations are made such that interaction counts follow a Poisson distribution. Sampling time for each dyad was modelled as the number of sampling periods, which was drawn from a Poisson distribution with mean $D \in [10, 10000]$. The dyad-level interaction rates were converted to an edge list from which the true network was built. Node strength s_i was used as the node-level network metric for these simulations. A linear relationship between network metric and individual trait was created by assigning traits t_i to each individual i according to a linear equation $t_i = a + bs_i + \epsilon_i$, where a is the equivalent of an intercept term, b encodes the relationship between metric and trait, and ϵ_i is a normally-distributed noise term: $\epsilon_i \sim N(0, 1)$. The value of the effect b was set depending on the number of nodes n. This was determined using a preliminary simulation in such a way that the value of b for a given number of individuals was equal to the lowest value for which a power $\geq 99.9\%$ can be achieved.

Networks were estimated using Equation 1 and node strength was estimated from the corresponding estimated networks. To estimate the relationship b, a simple linear regression was fitted to the simulated data, and node-label permutations were used to calculate the p-values (R Core Team, 2013; Croft et al., 2011). The power of the nodal regression was computed for each set of parameters by repeatedly assigning traits t_i according to the true network, fitting a linear model to an estimated network, and computing the p-value of the estimated relationship. The proportion of p-values less than 0.05 gave the power of the test for the current set of parameters. We searched the parameter space using a random search to assess the relationship between network correlation, statistical power, and the number of individuals.

Finally, to provide guidelines on the level of network correlation required in a nodal regression, LOESS curves approximating the relationship between correlation and power for different numbers of individuals were fitted to the simulated data (Cleveland, 1979; R Core Team, 2013). The resulting curves were used to predict the level of network correlation required to achieve 80% power, assuming the underlying relationship has power $\approx 100\%$.

Simulations 3: Optimal network correlation estimator for generic tests

The relationship between network correlation (Equation 3) and statistical power is affected by several factors, many of which will depend on the type of analysis being conducted. The relationship between power and network correlation was explored in detail for nodal regression in the previous section, but simulation-based studies like this are limited to the analyses they focus on, and cannot generalise to other methods. However we can expect that as network correlation increases, the power of any statistical tests should also improve (or at least stay the same). Assuming that increases in network correlation positively affect statistical power, and that increases in sampling effort come at a cost to researchers, the problem of finding the optimal sampling effort can be seen as finding the point at which increases in sampling effort lead to diminishing returns.

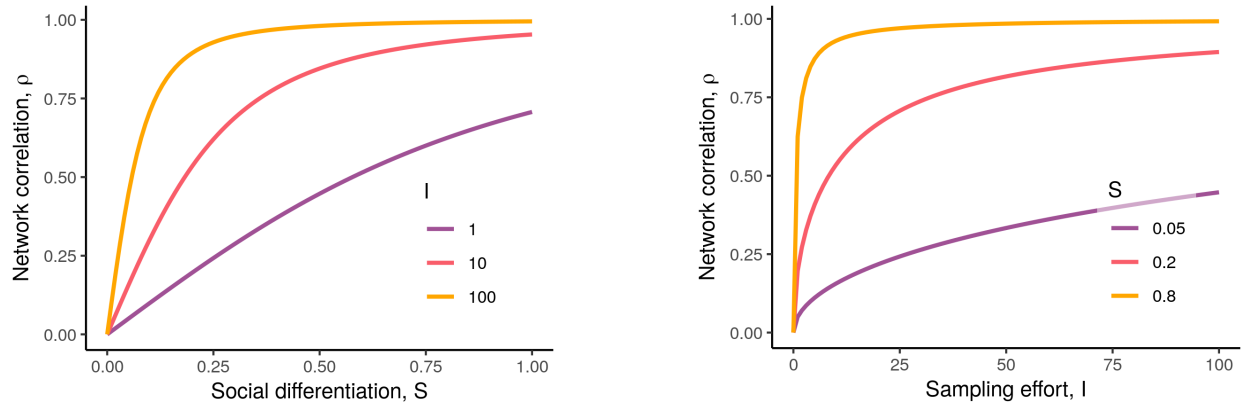
Diminishing returns describes how the rate of increase of one variable decreases as another variable increases (Shephard and Färe, 1974). In our case the aim is to find the point at which increases in sampling effort lead to diminishing increases in network correlation. This is the same problem as finding the "elbow point" of the relationship between network correlation and sampling effort. There is no guarantee of the power of the analysis at the elbow point, but we do know that additional sampling would provide increasingly small gains in network correlation. Since collecting behavioural data is generally time-consuming, financially expensive, and may even have ethical implications, assuming a cost to increases in sampling effort allows us to use the elbow point as an estimate for the optimal level of network correlation and corresponding amount of sampling (Martin et al., 1993). Though this will not guarantee the power of an analysis, if social differentiation and sampling effort can be estimated well, it will indicate the point at which additional sampling will yield increasingly small increases in network correlation.

The elbow of the curve can be computed numerically by rotating the curve of ρ against I by θ radians about the origin $(0.0, 0.0)$ and finding the maximum point of the rotated curve. The angle θ is usually determined by the maximum point of the curve, but in our case ρ is asymptotic to 1.0, so no maximum exists. We can get around this problem by introducing a free parameter ρ_{MAX} to describe the effective maximum network correlation to be used for computing the elbow. The choice of ρ_{MAX} encodes the minimum acceptable trade-off between increases in network correlation and increases in sampling effort, and therefore will affect the estimated elbow point. However, using a value of ρ_{MAX} sufficiently high that we would consider a sampled network to be negligibly different to the true network (for biological purposes) represents a meaningful choice since increases in network correlation beyond ρ_{MAX} would add no further value for our purposes. We use a value of $\rho_{\text{MAX}} = 0.99$ for our analysis, but a brief exploration of the impact of this choice is included in the supplementary material.

To assess the levels of power obtained by using the optimal network correlations, we simulated nodal regression analyses for different levels of social differentiation $S \in (0.0, 0.5]$ and sampling effort $I \in (0, 500]$ using a similar setup as the previous section. The power of the analysis for each simulated "true" network was computed, and the proportion of networks with lower than the optimal level of network correlation with power greater than 80% were calculated. This was repeated for the proportion of networks with the same or higher than the optimal level of network correlation. This provides a descriptive measure of the performance of the method for estimating the sampling required to achieve 80% power. This level of power was chosen from convention, but there is no reason that the optimal network correlation estimator should obey this convention as it is based on a different concept to conventional power analysis. We also estimated the level of power that most closely matched the estimator for nodal regression analysis using a grid search.

Case study: Southern resident killer whale contacts

To demonstrate our method, we applied it to a publicly-available dataset of near-surface physical contact interactions between 22 southern resident killer whales *(Orcinus orca)*. Interactions were observed using an unoccupied aerial vehicle over a total of 11 hours during the summer of 2019 (Weiss et al., 2021a). The mean number of observed interactions per dyad was 3.43, with the average sampling time per dyad being 210 minutes. We used numerical



(a) Social differentiation

(b) Total interactions

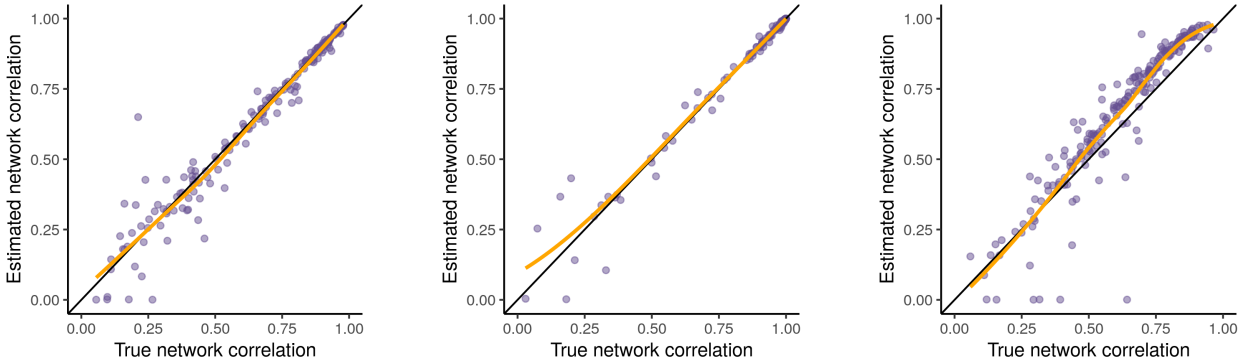
Figure 1: Plots of network correlation against social differentiation S and sampling effort $I = \mu H(d)$ for different levels of sampling and social differentiation respectively using Equation 3. This shows that for relatively low levels of social differentiation around $S = 0.25$, a high level of sampling effort around $I = 100$ is required to achieve a network correlation of $\rho = 0.90$. It also shows that for low social differentiation of $S = 0.05$, even high levels of sampling effort of $I = 100$ will not yield a network correlation of $\rho = 0.90$.

MLE to estimate the confidence intervals of social differentiation S and network correlation ρ. This is implemented in the *pwrIRGP* R package in the function `net_cor`. Additionally, we conducted a nodal regression power analysis on the data for different levels of true effect size ($r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$) using our function `pwr_nodereg`. We also applied the diminishing returns sampling effort estimator using the function `pwr_elbow` to estimate the optimal level of network correlation and its corresponding sampling effort.

Results

Gamma-Poisson model of dyadic interactions

The analytical equation for estimated network correlation given by Equation 3 was used to produce the plots in Figure 1 of network correlation against social differentiation and sampling effort. This shows that, as expected, network correlation increases towards an asymptote at 1.0 with both social differentiation S and sampling effort I. Lower numbers of interactions ($I = 1$) did not reach a network correlation of 1.0 even for high levels of social differentiation ($S = 1.0$). Higher sampling efforts of $I = 100$ reached close to a network correlation of 1.0 even at low levels of social differentiation, $S < 0.25$. This shows that social differentiation is an important factor in estimating the network correlation of a sampled network, which is in line with the findings of (Whitehead, 2008). The results show that relatively low values of $S \approx 0.25$ are required to make achieving a high level of network correlation feasible, but that lower values of social differentiation ($S = 0.05$) require sampling effort $I > 100$ to achieve even 50% network correlation.



(a) Scenario 1: Gamma interaction rates

(b) Scenario 2: Community structure

(c) Scenario 3: Zero inflated interaction rates

Figure 2: Comparisons of the true network correlations and the network correlations estimated using Equation 3. The black line is diagonal and shows the ideal relationship between the true and estimated network correlations. In these visualisations the three models were run with different maximum social differentiation parameters to ensure the full distribution of network correlations was visible, these were 0.20, 0.01, and 1.00 for models 1, 2, and 3 respectively. In each of these models the true and estimated network correlations are closely related, with few major deviations. The yellow lines show the general relationship using LOESS curves fitted to the points.

Simulations 1: Verification of the Gamma-Poisson model

The results of our simulations (see Figure 2) show that true and estimated network correlation match closely across the parameter space. Over the full parameter space outlined in Table 1, the mean absolute error between the true and estimated network correlations was 5.5%, 0.047%, and 3.5% for scenarios 1, 2, and 3 respectively. The standard deviations of the errors for the three scenarios were 0.0084, 0.0038, and 0.047 respectively, and the correlations between the true and estimated network correlations were 99%, 99%, and 96% respectively. The relationship between true and estimated network correlation is shown in Figure 2 over a limited part of the parameter space, where the social differentiation was adjusted manually to visualise the full width of the distribution of network correlations. Social differentiation was used to adjust the mean level of network correlation purely for visualisation purposes, and was not used when computing the quantitative statistics.

Simulations 2: Statistical power of nodal regression

The results of the nodal regression simulation are shown in Figure 3 for four different network sizes (10, 20, 50, 100). The relationship between network correlation and power is approximately logistic, with a higher power for networks with larger numbers of nodes across the range of network correlations. Large gains in power could be seen for networks with 20 nodes against those with 10 nodes (37% versus 27% respectively, at a network correlation of 50%), whereas networks with 100 nodes against 50 nodes had a much smaller gain in power (55% versus 51% respectively, again at a network correlation of 50%). To attain a statistical power

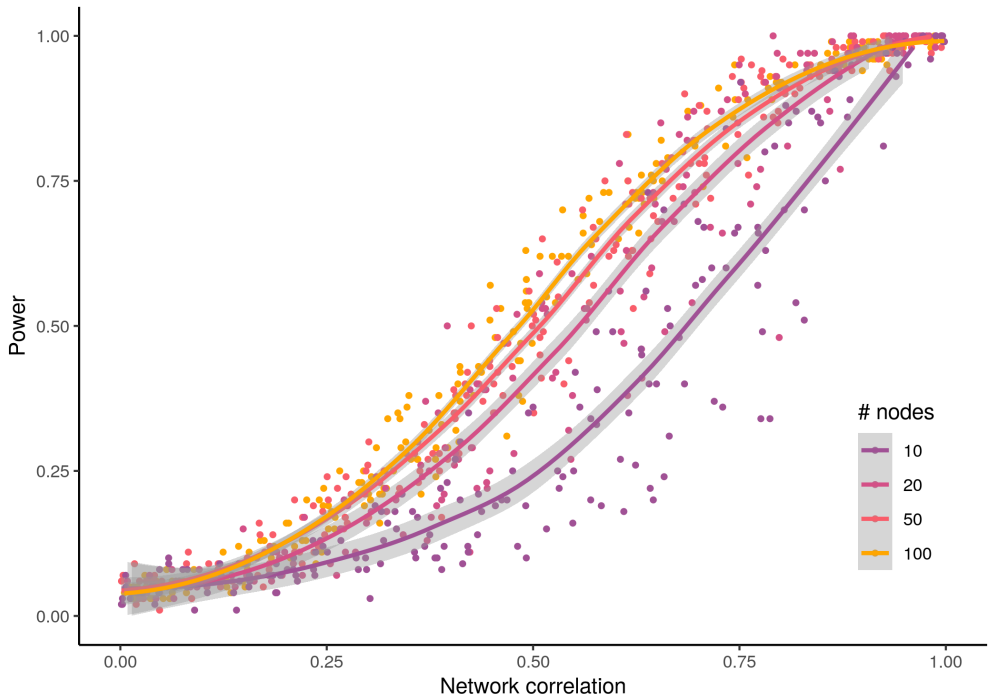


Figure 3: The relationship between estimated network correlation and statistical power for various numbers of nodes over 1000 simulations, where the minimal effect size was chosen such that power was $\geq 99.9\%$. Power and network correlation are monotonically related in a relationship that resembles a logistic curve. Larger numbers of nodes achieved a higher power for a lower network correlation than smaller numbers of nodes, as is expected in a nodal regression where sample size is equal to the number of nodes. A statistical power of 80% is common by convention, and was generally achieved for all simulated networks with at least 10 nodes for network correlations greater than 80%.

of the conventional 80%, network correlations of 81%, 78%, 69%, and 65% were required for network sizes 10, 20, 50, and 100 respectively. The maximum required level of network correlation to achieve 80% statistical power was 81% (for $n = 10$).

Table 2 shows the sampling effort I required to achieve 80% power, depending on the social differentiation S and size of a network n. For systems with low levels of social differentiation and low network sizes, high sampling effort ($I = 530$ for $n = 10, S = 0.05$) is required. However, as social differentiation increases, much lower sampling effort is required ($I = 5.3$ for $n = 10, S = 0.5$). This reflects the findings shown in Figure 1. Larger network sizes also have an effect on the required amount of sampling, with $n = 100$ requiring less than half the sampling effort as $n = 10$.

Simulations 3: Optimal network correlation estimator for generic tests

The powers of the nodal regression analyses for this simulation were split into five equally sized groups between 0% and 100%, with the top group of 80 - 100% being the desired power, by convention. The results of the simulations are shown in Figure 4 with the optimal sampling

Table 2: Required sampling effort $I = \mu H(d)$ to achieve a statistical power of 80% for networks with social differentiation S and number of nodes n. Relatively low numbers of interaction observations are required to achieve 80% power for levels of social differentiation of 0.5 and higher.

	n			
S	10	20	50	100
0.05	530	410	230	200
0.2	33	26	14	12
0.5	5.3	4.1	2.3	2
0.8	2.1	1.6	0.89	0.77
2	0.33	0.26	0.14	0.12
10	0.013	0.01	0.0057	0.005

effort overlaid. The optimal sampling effort for low levels of social differentiation ($S < 0.1$) was in excess of a sampling effort of $I > 500$. For slightly larger values of social differentiation ($0.1 \leq S \leq 0.2$), the required sampling effort drops quickly from approximately $I = 500$ to $I = 100$. The sampling effort for larger values of social differentiation quickly asymptotes towards zero.

The optimal sampling effort generally fell around the boundary between power levels of 60 - 80% and 80 - 100% for the nodal regression analyses. Considering the curve as a classifier of under- and over-powered analyses, using a power of 80% as the boundary between the two, 50.3% analyses below the curve were under-powered, and 99.9% of the analyses above the curve were adequately- or over-powered. In the context of our simulated nodal regressions, the curve was relatively conservative, with the power level that most closely corresponded to the power of the optimal network correlations being 90%.

Case study: Southern resident killer whale contacts

The southern resident killer whale contact network had an estimated network correlation of 97.7%, with a 95% confidence interval of $[96.6\%, 98.5\%]$. Social differentiation was estimated to be 2.54, with a confidence interval of $[2.23, 2.89]$. According to our power analysis, these levels of social differentiation and network correlation would yield power levels of at least 6.6%, 26.5%, 66.2%, 98.2%, and 100.0% for true effect sizes 0.1, 0.3, 0.5, 0.7, and 0.9 respectively. The diminishing returns estimator suggested that the optimal amount of network correlation to sampling effort is approximately 90%, corresponding to sampling efforts between 0.51 and 0.86. The true sampling effort was 3.30, several times larger than the maximum recommended sampling effort of 0.86.

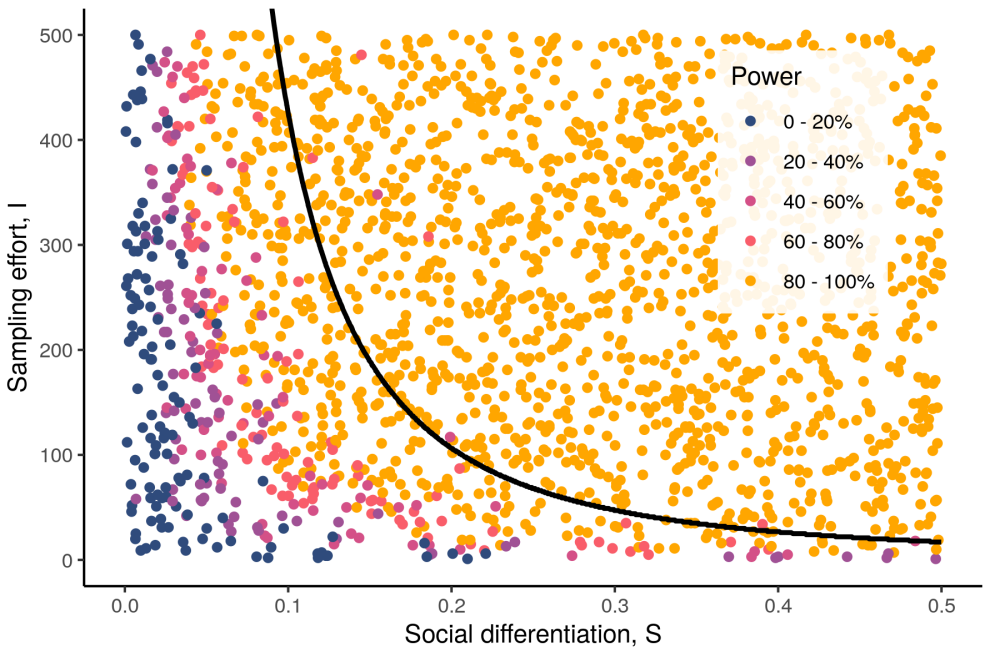


Figure 4: Power of nodal regression analyses for networks with varying levels of social differentiation S and sampling effort I. The black line shows the optimal sampling effort estimated using the diminishing returns method, corresponding to the optimal network correlation as social differentiation increases. This can be viewed as a classifier where analyses below the line are under-powered, and analyses above the line are adequately- or over-powered.

Discussion

In this study we showed that sampling effort can have a major impact on the accuracy of social interaction networks. We showed that this can severely affect the power of statistical analyses, and demonstrated how to carry out power analysis by accounting for sampling effort in both linear regression and generic statistical analysis. We derived an equation to describe how well a sampled network correlates with the true underlying network using a Gamma-Poisson model of interaction counts. We showed that the equation is a good estimator of the true network correlation, and is robust to the assumptions of the model being moderately broken. We used simulations of nodal regression analyses to find the relationship between network correlation and statistical power. Additionally we developed an estimate for the optimal level of network correlation required for generic analyses by using the concept of diminishing returns, where increases in network correlation come at the cost of increasing levels of sampling effort. We showed that this can be used as an alternative to conventional power analysis by verifying the results against the nodal regression simulations.

The nodal regression simulations suggested that a reasonable level of network correlation to achieve at least 80% power is also around 80%. This value of network correlation has precedent in the behavioural sciences, and would be categorised as "very strong" by Evans (1996). However we arrived at this value through the use of simulations, and would advise careful application of this guideline for different nodal regression-based analyses, and discourage its use

on other types of analysis. We believe it is particularly important to note that in our simulations we used the smallest possible effect sizes required to achieve approximately 100% power with infinite sampling. Effect sizes that satisfy this property are highly unlikely in the real world, with the true effect size either being lower, in which case no amount of sampling will be able to obtain full power, or the effect size being higher, in which case lower sampling efforts will be able to obtain full power. For this reason, we suggest using the code we have made available to run a custom power analysis for any specific study.

An alternative method for calculating the required level of network correlation is to use the diminishing returns-based optimal network correlation estimator. In the nodal regression simulations it proved to be a conservative classifier of analysis power, often suggesting that more samples were required than would have been necessary to obtain a power of 80%, and was more in line with a desired power level of 90% in this specific simulation. Since the estimator is based on diminishing returns, it implicitly models the trade-off between lower sampling variance and the cost of sampling, which is a fundamental part of behavioural data collection. For this reason, we believe this generic estimator to be a useful tool for generating estimates of the amount of sampling required, and in making few assumptions, it is flexible, allowing it to be used for any type of social network analysis. Another strength of the method is that knowledge of the desired statistical power is not required. Power can often be difficult to define, and is therefore usually treated as a nuisance parameter when conducting power analyses, with researchers reverting to conventions such as 80%. The free parameter ρ_{MAX} could be considered a similar theoretical limitation to this method, but we view it as the maximum value of network correlation a researcher is interested in. We chose $\rho_{\text{MAX}} = 0.99$ because the final increase in network correlation by 0.01 from 0.99 to 1.00 is unlikely to be of biological or practical interest, though other values could be used depending on the specific context. We suggest that for nodal regression, the simulations we have provided will be a more precise tool, but for other types of analyses, or where the desired power level is difficult to define, the generic estimator will give a good guide to the amount of sampling needed. The two methods can also be used in conjunction with each other to gain additional understanding about the power of an analysis.

The inspiration for this work, Whitehead (2008), suggested a guideline sampling effort for the social association test of $S^2 \times H > 5$, where H is the mean number of observed interactions per individual (Bejder et al., 1998). This is a classifier analogous to the classifier shown in Figure 4 for determining the required level of sampling effort for a given value of social differentiation to achieve a sufficiently-powered analysis. An obvious difference between the shapes of the two classifiers is that Whitehead's is linear, and ours is highly non-linear. The original was determined "by eye" whereas ours was calculated using a LOESS curve, so sampling and fitting differences could be responsible for some of the discrepancies. However, we found that when it came to sampling effort, I – the number of interactions observed per dyad – was the important factor for the classifier, whereas Whitehead (2008)'s guidance suggested that H – the number of interactions per individual – was the important factor. This is a key point because unless a strict sampling regime is used, sampling is unlikely be even across dyads and nodes, and a subset of well-sampled individuals that share connections might be over-sampled, while other dyads remain severely undersampled. This is taken into account by the definition of sampling effort, $I = \mu H(d)$, since the harmonic mean is lower than the arithmetic mean for unevenly sampled dyads. The previous study showed that this was unimportant in the case of association data on a specific test, but our equation suggests that it should be taken into

account for interaction count data.

We suggest that in addition to conducting power analyses when developing studies or assessing the feasibility of a study, our methods could also be used in post hoc or dynamic social network analyses where subsetting is required. It is common for long-term interaction data to be split up into multiple networks, often for the purposes of studying changes in social behaviour over time (Pinter-Wollman et al., 2014). Our methods could be used as a data-driven means to determine the amount of data required for each network, and therefore how many networks should be constructed. In this case a network correlation of 0.99 could be considered to be a representative network by the arguments above, but again this may depend on the context.

Our method uses a Gamma-Poisson model of estimated interaction rates. This introduces several assumptions that may not be met by empirical data, namely 1) interactions occur due to a Poisson point process, 2) true underlying interaction rates do not change over the course of data collection, and 3) true underlying interaction rates follow a gamma distribution. Assumptions are necessary for any statistical analysis, and assumptions 1 and 2 are usually assumed implicitly when constructing social networks from interaction count data. Modelling interaction count data as a Poisson point process means that the method will not work with interaction duration data. These assumptions could be broken by sampling biases such as increased observation of gregarious individuals or sampling based on number of interactions observed. Assumption 3 is necessary for the method, but may not be met in systems where interaction rates follow a multimodal distribution. The validity of the assumptions can be tested using the diagnostic tools included in the R package we have provided.

Correlation is a familiar statistical concept, so using network correlation to quantify the impact of sampling on the accuracy of social networks is an attractive method. However, network correlation on its own can only offer limited information about how useful a sampled network will be in further analyses. We used simulations in this study to estimate how network correlation relates to statistical power in nodal regression, but these simulations are limited to specific circumstances, under a number of assumptions. This will always be a limitation of simulation studies, and makes the generality of the results somewhat restricted. We have shown that the required level of network correlation is context dependent. This contrasts with some studies that have used Whitehead (2008)'s method to calculate network correlation, and suggest a minimum acceptable level of network correlation at 40% (Chabanne et al., 2017; Hawkins et al., 2020; Frau et al., 2021). Our results suggest a threshold value of 40% is not generally optimal when conducting nodal regression on interaction networks, and that the true optimal will depend on the level of social differentiation and sampling effort.

We have made the code from this study publicly available, and our method of power analysis can be used for any prospective studies where interaction count data will be collected, in post hoc analysis when reporting results, to indicate whether a null result may be due to insufficient sampling, and even when subsetting data to ensure sufficient data are used in each network. The diminishing returns method presented here for computing optimal network correlation naturally encodes the cost of collecting behavioural data and provides a simple method for estimating required sampling effort without specifying additional parameters such as desired power. We believe its flexibility could prove it to be useful for a wide variety of social network analyses.

Author contributions

MNW conceived of the negative binomial network correlation method and JDAH conceived of the power analysis methods. JDAH derived the equations and implemented the simulations with input from MNW, DWF, and LJNB. The manuscript was written by JDAH with input from MNW, DWF, and LJNB.

Acknowledgments

We would like to thank members of the CRAB Social Network Club and Alexander Mielke for useful discussions on this topic. JDAH acknowledges funding from the Engineering and Physical Sciences Research Council [grant number EP/R513210/1]. LJNB acknowledges funding from a European Research Council Consolidator Grant (FriendOrigins - 864461). DWF and MNW acknowledge funding from the Natural Environment Research Council [grant number NE/S010327/1]. The authors declare no conflict of interest.

References

Bejder, L., Fletcher, D., and Bräger, S. (1998). A method for testing association patterns of social animals. *Animal Behaviour*, 56(3):719–725.

Brask, J. B., Ellis, S., and Croft, D. P. (2021). Animal social networks: an introduction for complex systems scientists. *arXiv:2005.09598*. arXiv: 2005.09598.

Brent, L. J. N., MacLarnon, A., Platt, M. L., and Semple, S. (2013). Seasonal changes in the structure of rhesus macaque social networks. *Behavioral Ecology and Sociobiology*, 67(3):349–359.

Chabanne, D. B. H., Finn, H., and Bejder, L. (2017). Identifying the Relevant Local Population for Environmental Impact Assessments of Mobile Marine Fauna. *Frontiers in Marine Science*, 4. Publisher: Frontiers.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3):98–101.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist*, 49(12):997–1003. Place: US Publisher: American Psychological Association.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

Croft, D. P., James, R., and Krause, J. (2008). *Exploring Animal Social Networks*. Princeton University Press.

Croft, D. P., Madden, J. R., Franks, D. W., and James, R. (2011). Hypothesis testing in animal social networks. *Trends in Ecology & Evolution*, 26(10):502–507.

Ellis, S., Franks, D. W., Nattrass, S., Cant, M. A., Weiss, M. N., Giles, D., Balcomb, K. C., and Croft, D. P. (2017). Mortality risk and social network position in resident killer whales: sex differences and the importance of resource abundance. *Proceedings of the Royal Society B: Biological Sciences*, 284(1865):20171313. Publisher: Royal Society.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co, Belmont, CA, US.

Farine, D. R. and Whitehead, H. (2015). Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84(5):1144–1163.

Findlay, R., Gennari, E., Cantor, M., and Tittensor, D. P. (2016). How solitary are white sharks: social interactions or just spatial proximity? *Behavioral Ecology and Sociobiology*, 70(10):1735–1744.

Franks, D. W., Ruxton, G. D., and James, R. (2010). Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology*, 64(3):493–503.

Frau, S., Ronchetti, F., Perretti, F., Addis, A., Ceccherelli, G., and Manna, G. L. (2021). The influence of fish farm activity on the social structure of the common bottlenose dolphin in Sardinia (Italy). *PeerJ*, 9:e10960. Publisher: PeerJ Inc.

Hammersley, J. (2013). *Monte Carlo Methods*. Springer Science & Business Media.

Hawkins, E. R., Pogson-Manning, L., Jaehnichen, C., and Meager, J. J. (2020). Social dynamics and sexual segregation of Australian humpback dolphins (Sousa sahulensis) in Moreton Bay, Queensland. *Marine Mammal Science*, 36(2):500–521. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mms.12657.

James, R., Croft, D. P., and Krause, J. (2009). Potential banana skins in animal social network analysis. *Behavioral Ecology and Sociobiology*, 63(7):989–997.

Krause, J., James, R., Franks, D. W., and Croft, D. P. (2015). *Animal Social Networks*. Oxford University Press.

Martin, P., Bateson, P. P. G., and Bateson, P. (1993). *Measuring Behaviour: An Introductory Guide*. Cambridge University Press.

Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., de Silva, S., Waters, J. S., Prager, S. D., Sasaki, T., Wittemyer, G., Fewell, J., and McDonald, D. B. (2014). The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology*, 25(2):242–255.

R Core Team (2013). R: A language and environment for statistical computing.

Shephard, R. W. and Färe, R. (1974). The Law of Diminishing Returns. *Zeitschrift für Nationalökonomie/Journal of Economics*, (H. 1/2):69–90. Publisher: JSTOR.

Stadtfeld, C., Snijders, T. A. B., Steglich, C., and van Duijn, M. (2020). Statistical Power in Longitudinal Network Studies. *Sociological Methods & Research*, 49(4):1103–1132.

Weiss, M. N., Franks, D., Giles, D. A., Youngstrom, S., Wasser, S. K., Balcomb, K. C., Ellifrit, D. K., Domenici, P., Cant, M. A., Ellis, S., Nielsen, M. L. K., Grimes, C., and Croft, D. P. (2021a). Age and sex influence social interactions, but not associations, within a killer whale pod. *Proceedings of the Royal Society B: Biological Sciences*. (In press).

Weiss, M. N., Franks, D. W., Brent, L. J. N., Ellis, S., Silk, M. J., and Croft, D. P. (2021b). Common datastream permutations of animal social network data are not appropriate for hypothesis testing using regression models. *Methods in Ecology and Evolution*, 12(2):255–265. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13508.

Whitehead, H. (2008). Precision and power in the analysis of social structure using associations. *Animal Behaviour*, 75(3):1093–1099.