

Modelling human and non-human animal network data in R using STRAND

Submitted to Methods in Ecology & Evolution
XX(X):1–9
©The Author(s) 0000
DOI: To Be Assigned
DRAFT

Cody T. Ross¹, Richard McElreath¹, and Daniel Redhead¹

Abstract

1. There have been recent calls for wider application of generative modelling approaches in applied social network analysis. These calls have been motivated by the limitations of contemporary empirical frameworks, which have generally relied on *post hoc* permutation methods that do not actively account for interdependence in network data. At present, however, it remains difficult for typical end-users—e.g., field researchers—to apply generative network models, as there is a dearth of openly available software packages that make application of such methods as simple as other, permutation-based methods.
2. Here, we outline the STRAND R package, which provides a suite of generative models for Bayesian analysis of human and non-human animal social network data that can be implemented using simple, base R syntax.
3. To facilitate ease-of-use, we provide a tutorial demonstrating how STRAND can be used to model binary, count, or proportion data using stochastic blockmodels, social relations models, or a combination of the two modelling frameworks.

Keywords

social networks, animal networks, social relations, R software, generative models

Introduction

The application of theory and methods from network science (i.e., social network analysis) to ethological data has led to important advances in our understanding the structural features of animal societies. Similarly, the role that sociality—broadly conceived—plays in the differential success and survival of individuals and groups (e.g., Clutton-Brock 2009) has been a topic of perennial interest across the social, behavioural and biological sciences. By quantifying the social interactions (i.e., ties or edges) that are observed between individuals (i.e., nodes or vertices), researchers can more formally study how various dyadic phenomena are related to one another (e.g., Smith-Aguilar et al. 2019) and to key individual-level properties (e.g., Pike et al. 2008).

Recent network-based research has, for example, advanced theory on how social relationships—both positive (e.g., food sharing and grooming) and negative (e.g., agonistic behaviours)—guide the emergence and maintenance of social hierarchies (Kawakatsu et al. 2021; Redhead and Power 2022), influence the spread of disease (Read et al. 2008; Silk et al. 2017) and adaptive information (Waters and Fewell 2012; Hobaiter et al. 2014), and explain how individual actions culminate in group-wide movement patterns (Jacoby and Freeman 2016; Strandburg-Peshkin et al. 2015). To address these topics, and many others, network data are rapidly being compiled across a broad range of taxa (Sah et al. 2019). Given the flexibility of network-based frameworks for understanding behaviour, social network analysis has become one of the most popular areas of research in animal behaviour, behavioural ecology, and the quantitative evolutionary and social sciences more broadly.

Inferential concerns and topics of debate

While network analytical tools have great potential for advancing our understanding social behaviour across taxa, there are many statistical complexities inherent in such approaches. Network data are highly interdependent, and cannot be modelled using standard statistical approaches that assume uncorrelated residuals. Standard practices within the field often ignore such issues. For example, it is common for researchers to regress outgoing ties on incoming ties to estimate reciprocity (e.g., Carter and Wilkinson 2013), but such regressions are known to suffer from residual confounding (see Koster and Leckie 2014). Similarly, it is common for researchers to correct for sampling effort by creating a Simple Ratio Index (SRI; Cairns and Schwager 1987; Whitehead and James 2015; Farine and Whitehead 2015), but such indices are well-know to divide out sample size and give the weakest data-points disproportionate weight in downstream analyses (Hart et al. 2021b).

Permutation methods—such as the quadratic assignment procedure (QAP; Hubert and Schultz 1976; Krackardt 1987; Dekker et al. 2003)—have been used to “account for” the typical non-independence of network data (see Farine and Carter 2022; Farine and Whitehead 2015; Sosa et al. 2021). However, such *post hoc* methods do not actually permit unbiased estimation of generative model parameters (see

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

Corresponding author:

Cody T. Ross

Email: cody.ross@eva.mpg.de

Hart et al. 2021a), a point raised by a founder of the method thirty years ago (Krackhardt 1992). While there remains much lingering debate about the usefulness of permutation methods in applied network analysis, it is widely agreed upon that empirical data of all forms are best analysed using scientifically informed generative models (e.g., Reilly and Zeringue 2005). In the case of social network data, such models should *actively* account for non-independence of data points via correlated random effects at both the node and dyad level during the process of model fitting (see McElreath 2020, for a textbook example).

Permutation methods rose to prominence in large part for historical reasons, as correlated random-effects models were difficult to fit using early computer software. In fact, nuanced generative models of social networks have been available for decades, but Bayesian estimation of these models remained computationally intractable until only recently. As such, there is still a dearth of freely-available open-source software for implementing such generative modelling approaches as data analysis tools. To resolve this issue, we draw upon classic generative modelling approaches for network data, and integrate them with contemporary tools for Bayesian model fitting (Stan Development Team 2021b). In doing this, we have developed an R package, STRAND, that allows end-users to build complex network analysis models using simple `lm`-style syntax in base R, simplifying the process of modelling empirical network data.

Generative modelling approaches

There are a plethora of generative network models that reflect complex data generating procedures (see Newman 2018). Through Bayesian inversion (Allmaras et al. 2013), these models can be used as analytic tools that support statistical inference on the basis of empirical data. One such model that holds particular promise for research on animal social networks is the *social relations model* (Kenny and La Voie 1984; Snijders and Kenny 1999), which examines and accounts for correlations in node-level and dyad-level random effects. Across many contexts, animal social networks may further be partitioned into observable sub-groups—such as coalitions (Kajokaite et al. 2019), matriline (Ilany et al. 2021), and groupings based on identity or physical location (De Dreu and Triki 2022). These groupings may create gross community structure in networks, whereby individuals preferentially interact with those in their own sub-groups (to a greater or lesser extent across contexts and relationship types; e.g., Pisor and Ross 2021). Given this, *stochastic blockmodels* show further promise for animal social network analysis (Pearl and Schulman 1983), as they capture these higher order structures (Peixoto 2019). Together, these approaches provide a framework for a more direct analysis of the mechanisms involved in the data generating process of social relationships.

Our contribution

In order to address the concerns outlined above, and to make data analysis using complex network models easier for end-users, we have introduced an R package for Bayesian social network analysis, STRAND, that facilitates applications

of generative network modelling approaches. The STRAND package supports stochastic blockmodelling, social relations modelling, and more complex latent-network modelling approaches (see Redhead et al. 2021).

In this paper, we outline how stochastic blockmodels and social relations models can be fit to both human and non-human animal network data to answer key research questions. By presenting a tutorial for fitting these models to each of the three most commonly collected types of outcome data used in studies of animal behaviour and behavioural ecology—i.e., binary tie data (e.g., via self-reports), count data (e.g., through behavioural observations over a standardised time window), and proportion data (e.g., counts of behavioural observations where the sampling rate is variable across possible dyads)—we hope to inspire more wide-spread application of principled generative network models in empirical research.

Our approach to network modeling here complements that of the BISO team (Hart et al. 2021b), who have also developed a suite of network analysis models using Bayesian methods. Our R package provides additional functionality to typical end-users, however, in that it integrates new Bayesian network analysis models on the ‘back-end’ with a user-friendly ‘front-end’ interface, that allows even casual R users to specify complex network models (which may include a variety of block-, individual-, and dyad-level covariate data) on-the-fly, using nothing more than base R syntax.

Using STRAND

Much of the functionality of STRAND is made possible by Stan (Stan Development Team 2021b) and CmdStanR (Stan Development Team 2021a). Users must install these programs prior to installing STRAND. Installation and loading of STRAND is then simple: just run three lines of code from R:

```
library(devtools)
install_github("ctross/STRAND")
library(STRAND)
```

All of the tutorial code elaborated on below can also be found online at: <https://github.com/ctross/STRAND>, where the package will be maintained.

Building data objects

The first step in building any STRAND model is to organise the data. Social network data are normally complex, with some variables being reported at the level of the individual and others being reported at the level of the dyad. The `make_strand_data` function serves to organise all of these data into a unified format that can be read by later functions. After data are compiled, they can then be analysed with simple, `lm`-style function calls, as we discuss below.

We will illustrate how STRAND data objects are built, using human friendship network data from Dalla Ragione et al. (2022). First, outcome data and dyad-level predictors (both structured as adjacency matrices) are stored as labeled lists:

```
# Organize dyadic measures
outcome = list(Friends = friends)

dyad = list(Relatedness = pedigree,
            Distance = distance)
```

The outcome data must be either binary data or integers. The dyadic covariate data can include numeric variables, indicator variables, or even categorical variables. These dyadic covariate data can be used to estimate associations between dyad-level characteristics—such as genetic relatedness or physical proximity—and the likelihood of a tie in the outcome network.

Next, the individual-level covariates are stored in a data-frame:

```
# Organize individual measures
indiv = data.frame(
  Age = age,
  BMI = bmi,
  PA = pa
)
```

The individual-level covariate data can also include numeric variables, indicator variables, or categorical variables. Individual-level covariate data can be used to estimate associations between individual-level characteristics and the likelihood of either sending or receiving a tie.

Finally, individual-level covariates that govern group/block structure are stored in a separate data-frame:

```
# Organize blocking measures
groups = data.frame(
  Ethnicity = as.factor(ethnicity),
  Sex = as.factor(sex)
)
```

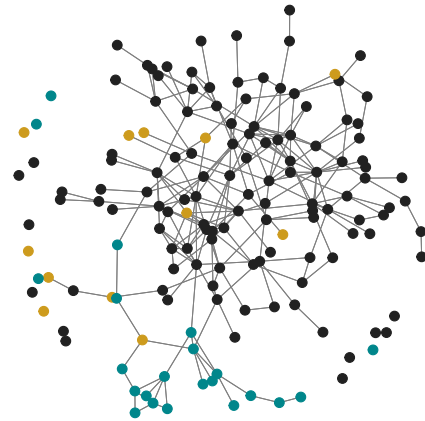
Although these block-structuring variables are also individual-level data, they are treated differently than other variables by STRAND; these variables must be factors, and are used to create random intercept offsets unique to the interaction of focal/sender and alter/receiver block IDs.

Once all covariate data are organised as above, they can be compiled into a single STRAND object:

```
# Create the STRAND data object
dat = make_strand_data(
  self_report = outcome,
  block_covariates = groups,
  individual_covariates = indiv,
  dyadic_covariates = dyad,
  outcome_mode = "bernoulli",
  exposure = NULL
)
```

At this point, the user must define which outcome model to use. The STRAND package supports three outcome modes for each model type: “bernoulli” for binary tie data (e.g., for human self-report/name-generator data, or similar binary tie data from non-human animals, like whether or not there was any observed conflict between each dyad on a given day), “poisson” for raw count data (e.g., the number of times GPS

Figure 1. Friendship network data from a rural Colombian town published in Dalla Ragione et al. (2022). Nodes are coloured according to ethnic group, with blue nodes representing indigenous Emberá individuals, goldenrod nodes representing Mestizo individuals, and dark-grey nodes representing Afrocolombian individuals. Group structure is modelled by including ethnicity as an observed block variable. Variance in node degree is estimated using random effects on the probability of sending and receiving friendship nominations.



trackers were within 5 meters of each other over a fixed 1-week period), or finally “binomial” for proportion data (e.g., if the outcome variable is a matrix containing a count of grooming events between each dyad, and the exposure variable is matrix containing a count of the number of scans in which grooming events between each dyad could have been observed). If the outcome mode is set to “binomial”, then the exposure variable (a labeled list containing a matrix of sample size values) must be provided.

Example model with Binary data

Classic social network data, especially as collected using self-report survey designs in human research, is frequently represented as a matrix of binary ties (i.e., zeros indicating the absence of ties, and ones indicating the presence of ties). For our first example, we will model binary human friendship network data (see Figure 1) published in Dalla Ragione et al. (2022). These data include individual-level measures (e.g., sex, ethnicity, age, BMI, and physical attractiveness) and dyad-level measures (e.g., relatedness and inter-household distance) that are thought to be associated with the structure of friendships.

To model the data, we use a hybrid of the stochastic block-model and social relations model (see Supporting Information for full mathematical details). The STRAND syntax is based on standard `lm` syntax from base R. To model the data, we write out equations for block effects, focal/sender effects, alter/receiver effects, and dyadic effects:

```

fit =
fit_block_plus_social_relations_model(
  data=dat,
  block_regression = ~ Sex + Ethnicity,
  focal_regression = ~ Age + BMI + PA,
  target_regression = ~ Age + BMI + PA,
  dyad_regression = ~ Relatedness +
    Distance,
  mode="mcmc",
  stan_mcmc_parameters = list(
    chains = 2,
    iter_warmup = 1500,
    iter_sampling = 1500,
    max_treedepth = 12,
    adapt_delta = 0.98)
)

```

In this model, we estimate block-level effects for sex and ethnicity using the argument: `block_regression = ~ Sex + Ethnicity`. These effects are indicative of how the likelihood of friendship nomination varies as a function of block categories—i.e., we can measure if male-to-male ties are more or less likely than male-to-female ties, female-to-male ties, or female-to-female ties (and likewise for ties within and between ethnic groups).

Next, the focal regression model, `focal_regression = ~ Age + BMI + PA`, explores how the age, body mass index, and physical attractiveness rating for a given individual is related to that individual's propensity to nominate others as friends (i.e., it measures the effects of individual-level covariates on *out-degree*). Similarly, the target regression model, `target_regression = ~ Age + BMI + PA`, explores how the age, body mass index, and physical attractiveness rating for a given individual is related to that individual's propensity to be nominated by others as a friend (i.e., it measures the effects of individual-level covariates on *in-degree*). Finally, the dyad regression model, `dyad_regression = ~ Relatedness + Distance`, explores how the likelihood of friendship ties is associated with the genetic relatedness of each dyad, as well as with the spatial distance between the homes of each dyad.

When the model has finished running, the MCMC samples can be processed and summarised using a convenience function:

```
res = summarize_strand_results(fit)
```

The results can then be visualised as a table (e.g., see Table 1) or as a forest plot (e.g., see Figure 2).

As in Dalla Ragione et al. (2022), we find that relatedness is positively associated with the probability of observing friendship ties, while inter-household distance is negatively associated the probability of friendship ties. Older individuals have a higher probability of making friendship nominations, while individuals who were rated by the community as more physically attractive are less likely to nominate others as friends. On the other hand, individuals rated as more physically attractive were more likely to be nominated by others as friends. Age and BMI are also associated with the likelihood on being nominated as a friend. Next, when considering block variables, such as ethnicity

and sex, we find evidence that individuals preferentially assort with others of their same group. For example, the intercept offset for Afrocolombian-to-Afrocolombian ties is -3.37 (90%HPDI; -4.83, -1.62), is reliably greater than the intercept offset for Afrocolombian-to-Emberá ties, -9.30 (90%HPDI; -11.94, -6.73). Finally, we find evidence of both dyadic reciprocity (dyadic ρ) and generalised reciprocity (focal-target ρ) in friendship nominations (see Table 1).

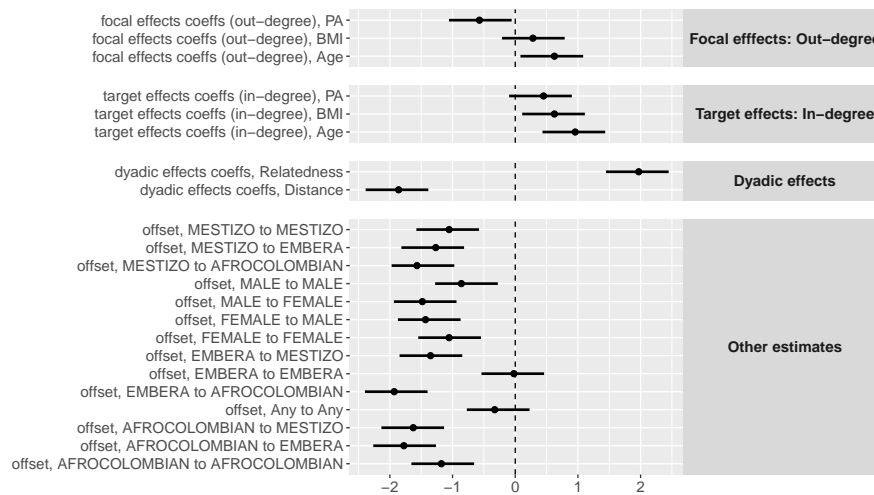
Example model with Poisson data

Throughout the non-human literature, network data are often recorded using numerical measurements (e.g., the number of times two animals are observed fighting, or the number of minutes two animals spend grooming one another). As an example of such numerical measures, we draw on data investigating blood-sharing among vampire bats (Carter and Wilkinson 2013). These data include individual-level predictors (i.e., sex) and dyad-level predictors (i.e., genetic relatedness, and whether each dyad had the opportunity to be observed sharing blood).

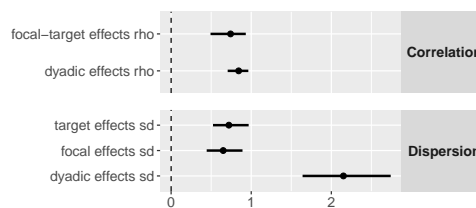
As before, we start by organising the data:

Table 1. Parameter estimates for the effects of various covariates on the structure of friendships in a rural Colombian town. *Focal*, *target*, and *dyadic* effects are interpreted as slopes. *Block* effects are interpreted as intercept offsets. *Random* effects include terms that control the variance of random effects, σ , and terms that control the correlation of random effects, ρ .

Type	Variable	Median	HPDI:L	HPDI:H	Mean	SD
random	focal σ	0.669	0.434	0.87	0.666	0.139
random	target σ	0.75	0.527	0.976	0.758	0.139
random	dyadic σ	2.16	1.706	2.726	2.213	0.341
random	focal-target ρ	0.741	0.514	0.957	0.724	0.148
random	dyadic ρ	0.836	0.689	0.959	0.826	0.089
focal	Age	0.016	0.003	0.027	0.016	0.007
focal	BMI	0.02	-0.016	0.056	0.02	0.023
focal	PA	-0.02	-0.037	-0.003	-0.02	0.011
target	Age	0.023	0.011	0.034	0.023	0.007
target	BMI	0.045	0.011	0.085	0.046	0.023
target	PA	0.015	0	0.029	0.015	0.009
dyadic	Distance	-3.593	-4.591	-2.691	-3.602	0.609
dyadic	Relatedness	4.484	3.256	5.605	4.468	0.718
block	Any to Any	-1.889	-4.408	1.036	-1.912	1.648
block	Female to Female	-4.306	-6.123	-2.322	-4.284	1.168
block	Female to Male	-5.898	-7.662	-4.065	-5.885	1.168
block	Male to Female	-6.082	-8.025	-4.239	-6.097	1.174
block	Male to Male	-3.362	-5.174	-1.567	-3.376	1.145
block	Afro. to Afro.	-3.379	-4.831	-1.62	-3.328	0.967
block	Afro. to Emberá	-9.309	-11.943	-6.735	-9.36	1.615
block	Afro. to Mestizo	-5.221	-6.786	-3.407	-5.176	1.018
block	Emberá to Afro.	-6.611	-8.413	-4.655	-6.637	1.148
block	Emberá to Emberá	0.071	-1.513	1.827	0.094	1.011
block	Emberá to Mestizo	-4.976	-6.797	-2.938	-4.946	1.192
block	Mestizo to Afro.	-4.852	-6.585	-3.296	-4.844	1.029
block	Mestizo to Emberá	-7.464	-10.845	-4.741	-7.609	1.881
block	Mestizo to Mestizo	-5.956	-8.703	-2.873	-6.027	1.806



(a) Effects of covariates.



(b) Variance and correlation of random effects.

Figure 2. Parameter estimates for the effects of various covariates on the structure of friendship in a rural Colombian town. Points represent the posterior median, and bars represent 90% highest posterior density intervals.

```

# Number of minutes of blood licking
net = list(Lick = round(Lick/60, 0))

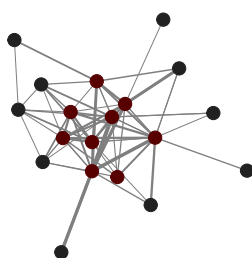
# Dyadic variables
dyad = list(Relatedness = Relatedness,
           NoOppertunity = NoOppertunity
           )

# Block variables
group_ids = data.frame(
  Sex = as.factor(Sex)
)
    
```

290

Then, the data can be compiled into a STRAND data object. This time, we must include the argument: `outcome_mode = "poisson"`, so that STRAND treats the outcome data as

Figure 3. Blood-sharing network data from vampire bats published by Carter and Wilkinson (2013). Red nodes represent females and dark-grey nodes represent males. Group structure is modelled by including sex as a block variable. Variance in node degree is estimated using random effects.



integers. Also, since there are no individual-level covariates other than sex, which is used as a blocking variable, we can set: `individual_covariates = NULL`.

295

```

dat = make_strand_data(
  self_report = net,
  block_covariates = group_ids,
  individual_covariates = NULL,
  dyadic_covariates = dyad,
  outcome_mode = "poisson"
)
    
```

Then, a model can be fit to the data:

```

fit =
fit_block_plus_social_relations_model(
  data=dat,
  block_regression = ~ Sex,
  focal_regression = ~ 1,
  target_regression = ~ 1,
  dyad_regression = ~ NoOppertunity +
    Relatedness,
  mode="mcmc",
  stan_mcmc_parameters = list(
    chains = 1,
    iter_warmup = 1500,
    iter_sampling = 1500,
    max_treedepth = NULL,
    adapt_delta = .98)
)
    
```

Here, we set the focal and target regression models to be intercept only (as we have no individual-level covariates)

300

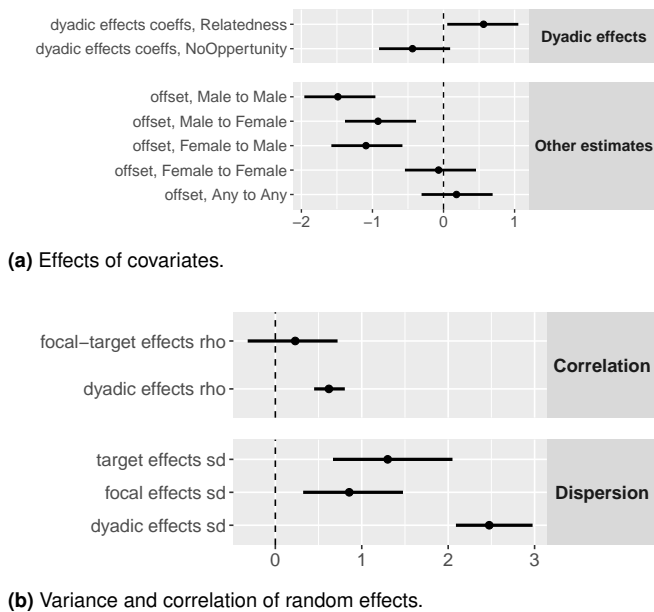


Figure 4. Blood sharing among of vampire bats. Points represent the posterior median, and bars represent 90% highest posterior density intervals. Vampire bats are more likely to share blood with relatives than less-related conspecifics. Transfers are also more likely to flow from females to other females than from females to males, as indicated by the credible regions for these offsets not overlapping. Blood sharing is also reciprocal, as indicated by the strong correlation in dyadic random effects.

using the `lm`-style syntax: `focal_regression = ~ 1` and `target_regression = ~ 1`.

Finally, the results can be summarised and plotted:

```
# Summarize results
res = summarize_strand_results(fit)

# Plot slopes
vis = strand_caterpillar_plot(res,
  normalized=TRUE, only_slopes=TRUE)
```

Table 2 and Figure 4 present the effects of relatedness and sex on the rate of blood-sharing transfers. We recover the primary results of Carter and Wilkinson (2013), finding that genetic relatedness is a reliable predictor of blood-sharing, and that transfers are reliably more likely between female dyads than between male or mixed-sex dyads.

Table 2. Predictors of vampire bat blood-sharing relationships.

Type	Variable	Median	HPDI:L	HPDI:H	Mean	SD
random	focal σ	0.855	0.322	1.477	0.872	0.348
random	target σ	1.3	0.667	2.05	1.332	0.437
random	dyadic σ	2.475	2.09	2.977	2.496	0.274
random	focal-target ρ	0.23	-0.32	0.72	0.196	0.322
random	dyadic ρ	0.62	0.449	0.804	0.61	0.113
dyadic	No Opportunity	-0.971	-2.016	0.207	-0.976	0.685
dyadic	Relatedness	1.299	0.121	2.426	1.285	0.726
block	Any to Any	0.716	-1.213	2.713	0.694	1.216
block	Female to Female	-0.275	-2.164	1.819	-0.29	1.248
block	Female to Male	-4.203	-6.08	-2.228	-4.209	1.226
block	Male to Female	-3.607	-5.427	-1.513	-3.563	1.226
block	Male to Male	-6.439	-8.481	-4.151	-6.421	1.331

Example model with Binomial data

Another type of data frequently encountered in studies of animal sociality represents ties strength as a weighted social association matrix (e.g., Brask et al. 2019). These measures are typically created by applying a simple ratio association index (SRI; Cairns and Schwager 1987; Farine and Whitehead 2015), where, for example, the edge weight of each dyad is calculated as the ratio of the number of scans/observations in which the dyad is observed together divided by the number of scans/observations in which at least one of them was observed.

While this approach of weighting counts by an exposure variable is significantly better than ignoring variation in risk of observation (Farine and Whitehead 2015), construction of a simple ratio divides out sample size information, leading dyadic observations based on little data to carry disproportionate weight in downstream analyses (see McElreath 2020; Hart et al. 2021b, for a review of this issue). Moreover, zeros arising from censoring (i.e., due to members of a dyad being unavailable; “denominator zeros”) are often confounded with true zeros (i.e., members of a dyad being present but not interacting; “numerator zeros”). A better approach involves modelling the actual count of the number of scans/observations in which each dyad is observed together using a Binomial model, in which the sample size parameter is—for example—the number of scans/observations in which at least one member of the dyad was observed.

To demonstrate how to fit such a model in STRAND, we draw on grooming data from captive Guinea baboons published by Gelardi et al. (2020). We investigate three questions here: 1) do individuals who groom others more also receive more grooming in general (i.e., generalised reciprocity)?, 2) accounting for individual-level differences in the probability of grooming, does the probability of individual i grooming individual j increase with the probability that individual j grooms individual i (i.e., dyadic reciprocity)?, and 3) is the probability of individual i grooming individual j associated with whether individual j “presents” to individual i ? (Presenting is defined as: “approaching another individual gently with or without lipsmacks and grunts and presenting the rear”; Gelardi et al. 2020).

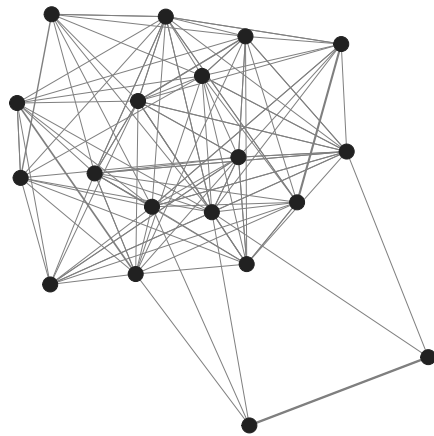
As before, we start by organising the data:

```
# Number of grooming event and a
# sample-size measure
nets = list(Grooming = Grooming)
exposure = list(Exposure = Exposure)

# Dyadic variable: transpose of
# Presenting
dyad = list(Presenting = t(Presenting))
```

And then we compile the data into a STRAND object:

Figure 5. Grooming data in captive Guinea baboons published by Gelardi et al. (2020). Group structure is not modelled as no individual-level data were provided. Variance in node degree is estimated using random effects on the probability of providing and receiving grooming.



```
dat = make_strand_data(
  self_report = nets,
  block_covariates = NULL,
  individual_covariates = NULL,
  dyadic_covariates = dyad,
  outcome_mode = "binomial",
  exposure = exposure
)
```

This time, the argument: `outcome_mode = "binomial"` is included. In order for the Binomial model to run, we also need to store the sample size data, using: `exposure = exposure`.

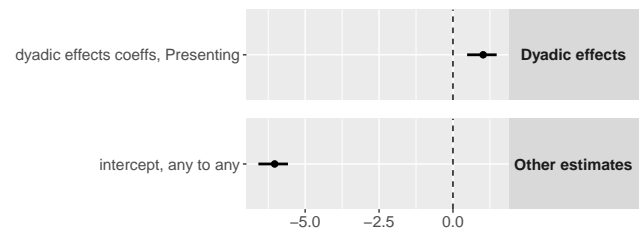
This time, because we have no blocking variables, we use the basic social relations model with no block-level effects. We run this model using the function:

```
fit =
fit_social_relations_model(
  data=dat,
  focal_regression = ~ 1,
  target_regression = ~ 1,
  dyad_regression = ~ Presenting,
  mode="mcmc",
  stan_mcmc_parameters = list(
    chains = 1,
    iter_warmup = 1500,
    iter_sampling = 1500,
    max_treedepth = NULL,
    adapt_delta = .98)
)
```

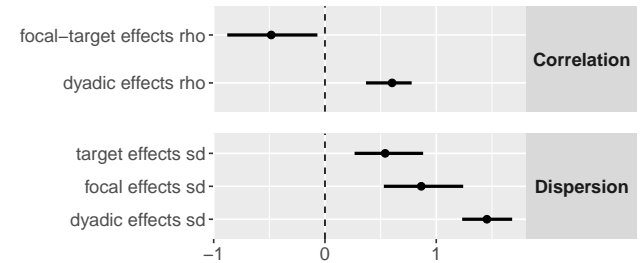
Finally, the results can be summarised and plotted:

```
# Summarize results
res = summarize_strand_results(fit)
```

The results are presented in Table 3 and Figure 6. We note that the correlation, ρ , of focal and target effects is negative. This indicates that baboons who frequently groom others



(a) Effects of covariates.



(b) Variance and correlation of random effects.

Figure 6. Grooming data in captive Guinea baboons. Points represent the posterior median, and bars represent 90% highest posterior density intervals. Guinea baboons appear more likely to groom conspecifics who regularly “present” to them. Interestingly, the correlation, ρ , of focal and target effects is negative. This indicates that baboons who frequently groom others are less likely to be groomed themselves by others. However, after accounting for this individual-level variation in grooming propensity, there is evidence of dyadic reciprocation, as indicated by the strong correlation, ρ in dyadic random effects.

are actually less likely to be groomed themselves by others. Such unidirectional behavioural propensities are consistent with dominance hierarchies (e.g., Gullstrand et al. 2021) in which unbalanced benefits are tolerated. However, after accounting for this individual-level variation in grooming propensity, there is evidence of dyadic reciprocation, as indicated by the strong correlation, ρ , in dyadic random effects. Finally, in this sample of Guinea baboons, individuals appear more likely to groom conspecifics who regularly “present” to them.

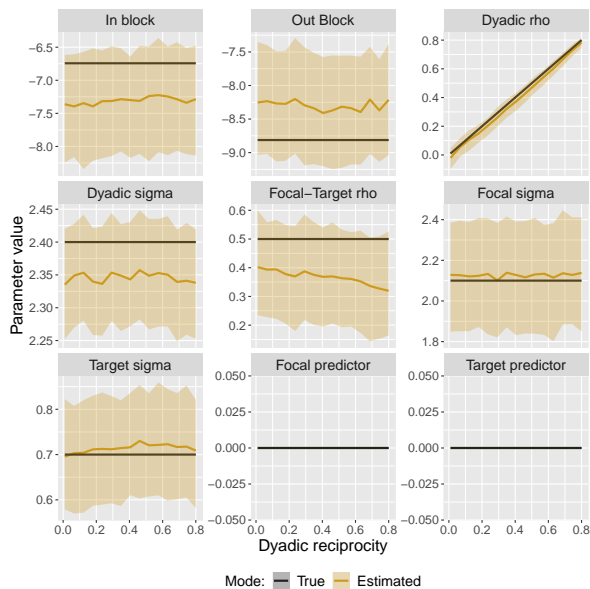
Table 3. Results of captive Guinea baboon grooming.

Type	Variable	Median	HPDI:L	HPDI:H	Mean	SD
random	focal σ	0.865	0.529	1.242	0.896	0.234
random	target σ	0.541	0.266	0.882	0.559	0.195
random	dyadic σ	1.455	1.233	1.682	1.46	0.138
random	focal-target ρ	-0.483	-0.878	-0.068	-0.444	0.266
random	dyadic ρ	0.603	0.369	0.779	0.59	0.128
dyadic	Any to Any	-5.645	-6.161	-5.225	-5.66	0.294
dyadic	Presenting	0.339	0.159	0.491	0.341	0.103

Validating the models with simulated data

For interested readers, we include a detailed mathematical description of our statistical models in the Supporting Information. There, we walk readers through model specification, parameter interpretation, and model validation procedures. We test each statistical model, for each outcome mode (Bernoulli, Poisson, and Binomial) using a suite of unit-tests (see also Redhead et al. 2021). Specifically, we first generate network data using forward simulations

Figure 7. Example of model validation results. In this case, we vary the dyadic reciprocity parameter of a generative network model and simulate test data-sets for a range of values, $\rho_\delta \in (0.01, 0.8)$. We then use STRAND to estimate the model. Black lines represent the generative parameter values. Yellow regions represent estimated posterior distributions of the same parameters. We find that our model accurately recovers all generative parameters. We repeat this process for all combinations of network models and outcome modes, varying all key model parameters for each test case. See supporting information for details.



of—and full documentations for—all functions included in STRAND by visiting: <https://github.com/ctross/STRAND>. Additional R code examples are provided there as well.

Acknowledgements

C.T.R., R.M.E., and D.R. were supported by the Department of Human Behavior, Ecology, and Culture at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. D.R. was supported by ESRC Grant number ES/V006495/1. Authors declare no conflict of interest.

References

Allmaras M, Bangerth W, Linhart JM, Polanco J, Wang F, Wang K, Webster J and Zedler S (2013) Estimating parameters in physical models through bayesian inversion: A complete example. *siam REVIEW* 55(1): 149–167.

Brask JB, Croft DP, Edenbrow M, James R, Bleakley BH, Ramnarine IW, Heathcote RJ, Tyler CR, Hamilton PB, Dabelsteen T et al. (2019) Evolution of non-kin cooperation: social assortment by cooperative phenotype in guppies. *Royal Society open science* 6(1): 181493.

Cairns SJ and Schwager SJ (1987) A comparison of association indices. *Animal Behaviour* 35(5): 1454–1469.

Carter GG and Wilkinson GS (2013) Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment. *Proceedings of the Royal Society B: Biological Sciences* 280(1753): 20122573.

Clutton-Brock T (2009) Structure and function in mammalian societies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1533): 3229–3242.

Dalla Ragione A, Ross CT and Redhead D (2022) Friendship and partner choice considerations in rural colombia .

De Dreu CK and Triki Z (2022) Intergroup conflict: origins, dynamics and consequences across taxa.

Dekker D, Krackhardt D and Snijders T (2003) Multicollinearity robust gap for multiple regression. In: *1st annual conference of the North American Association for Computational Social and Organizational Science*. Citeseer, pp. 22–25.

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(1): 144–156.

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

Gelardi V, Godard J, Paleressompouille D, Claidière N and Barrat A (2020) Measuring social networks in primates: wearable sensors versus direct observations. *Proceedings of the Royal Society A* 476(2236): 20190737.

Gullstrand J, Claidière N and Fagot J (2021) Computerized assessment of dominance hierarchy in baboons (*papio papio*). *Behavior Research Methods* 53(5): 1923–1934.

Hart JD, Weiss MN, Brent LJ and Franks DW (2021a) Common permutation methods in animal social network analysis do not control for non-independence. *bioRxiv* .

Hart JDA, Weiss MN, Franks DW and Brent LJN (2021b) Bison: A bayesian framework for inference of social networks. *bioRxiv* DOI:10.1101/2021.12.20.473541. URL

390 from a stochastic blockmodel, a social relations model, or the combined model, which includes both stochastic blockmodel and social relations model parameters. We then use the corresponding inferential statistical models to analyse the simulated data sets and ensure that we can recover the generative parameter values. In each simulation experiment, we generally vary only a single generative parameter (e.g., the dyadic reciprocity coefficient) across a broad parameter space that contains realistic values, while fixing all other parameters in the model to empirically plausible values. See Figure 7 for an example, and the Supporting Information for the full suite of unit-tests, which all indicate that our models accurately recover generative parameters.

Conclusions

405 The tools included in STRAND provide easy-to-use and efficient methods for generative modelling of human and animal social networks. Here, we have outlined the functionality of STRAND, defined the suite of models that are included, and provided detailed unit-tests to show that the software performs correctly. Using openly available example data-sets, we have provided tutorials for end-users interested in running network analysis models in R using STRAND. We hope that this software will help end-users with limited programming experience easily
410
415 deploy otherwise complex statistical models, and thus effortlessly investigate fundamental research questions in their fields of interest. End-users can find a complete index

- 475 <https://www.biorxiv.org/content/early/2021/12/21/2021.12.20.473541>.
- Hobaiter C, Poisot T, Zuberbühler K, Hoppitt W and Gruber T (2014) Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS biology* 12(9): e1001960.
- 480 Hubert L and Schultz J (1976) Quadratic assignment as a general data analysis strategy. *British journal of mathematical and statistical psychology* 29(2): 190–241.
- Ilany A, Holekamp KE and Akçay E (2021) Rank-dependent social inheritance determines social network structure in spotted hyenas. *Science* 373(6552): 348–352.
- 485 Jacoby DM and Freeman R (2016) Emerging network-based tools in movement ecology. *Trends in Ecology & Evolution* 31(4): 301–314.
- 490 Kajokaite K, Whalen A, Panchanathan K and Perry S (2019) White-faced capuchin monkeys use both rank and relationship quality to recruit allies. *Animal Behaviour* 154: 161–169.
- Kawakatsu M, Chodrow PS, Eikmeier N and Larremore DB (2021) Emergence of hierarchy in networked endorsement dynamics. *Proceedings of the National Academy of Sciences* 118(16).
- 495 Kenny DA and La Voie L (1984) The social relations model. *Advances in experimental social psychology* 18: 141–182.
- Koster JM and Leckie G (2014) Food sharing networks in lowland nicaragua: an application of the social relations model to count data. *Social Networks* 38: 100–110.
- 500 Krackardt D (1987) Qap partialling as a test of spuriousness. *Social networks* 9(2): 171–186.
- Krackhardt D (1992) A caveat on the use of the quadratic assignment procedure. *Journal of quantitative anthropology* 3(4): 279–296.
- 505 McElreath R (2020) *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
- Newman M (2018) *Networks*. Oxford university press.
- Pearl MC and Schulman SR (1983) Techniques for the analysis of social structure in animal societies. In: *Advances in the Study of Behavior*, volume 13. Elsevier, pp. 107–146.
- 510 Peixoto TP (2019) Bayesian stochastic blockmodeling. *Advances in network clustering and blockmodeling* : 289–332.
- Pike TW, Samanta M, Lindström J and Royle NJ (2008) Behavioural phenotype affects social interactions in an animal network. *Proceedings of the Royal Society B: Biological Sciences* 275(1650): 2515–2520.
- 515 Pisor A and Ross CT (2021) How generalizable are patterns of parochial altruism in humans? .
- 520 Read JM, Eames KT and Edmunds WJ (2008) Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface* 5(26): 1001–1007.
- Redhead D, McElreath R and Ross CT (2021) Reliable network inference from unreliable data: A tutorial on latent network modeling using strand .
- 525 Redhead D and Power EA (2022) Social hierarchies and social networks in humans. *Philosophical Transactions of the Royal Society B* 377(1845): 20200440.
- 530 Reilly CS and Zeringue A (2005) Improved predictions of lynx trappings using a biological model. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*. Wiley-Blackwell, pp. 297–308.
- Sah P, Méndez JD and Bansal S (2019) A multi-species repository of social networks. *Scientific data* 6(1): 1–6.
- 535 Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Weber N, Boots M and McDonald RA (2017) The application of statistical network models in disease research. *Methods in Ecology and Evolution* 8(9): 1026–1041.
- 540 Smith-Aguilar SE, Aureli F, Busia L, Schaffner C and Ramos-Fernández G (2019) Using multiplex networks to capture the multidimensional nature of social structure. *Primates* 60(3): 277–295.
- 545 Snijders TA and Kenny DA (1999) The social relations model for family data: A multilevel approach. *Personal Relationships* 6(4): 471–486.
- Sosa S, Pasquaretta C, Puga-Gonzalez I, Dobson FS, Viblanc VA and Hoppitt W (2021) Reliable hypotheses testing in animal social network analyses: global index, index of interactions and residual regression. *bioRxiv* .
- 550 Stan Development Team (2021a) Cmdstanr: A lightweight interface to stan for r users. URL <http://mc-stan.org/>.
- Stan Development Team (2021b) Stan modeling language users guide and reference manual, version 2.21. URL <http://mc-stan.org/>.
- 555 Strandburg-Peshkin A, Farine DR, Couzin ID and Crofoot MC (2015) Shared decision-making drives collective movement in wild baboons. *Science* 348(6241): 1358–1361.
- 560 Waters JS and Fewell JH (2012) Information processing in social insect networks. *PloS one* 7(7): e40337.
- Whitehead H and James R (2015) Generalized affiliation indices extract affiliations from social network data. *Methods in Ecology and Evolution* 6(7): 836–844.