

Estimating the robustness and uncertainty of animal social networks using different
observational methods

Grace H. Davis¹, Margaret C. Crofoot^{1,2,3}, Damien R. Farine^{4,5,6}

¹Department of Anthropology, University of California, Davis, USA

²Animal Behavior Graduate Group, University of California, Davis, USA

³Smithsonian Tropical Research Institute, Panama

⁴Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford,
Oxford, UK

⁵Department of Collective Behaviour, Max Planck Institute for Ornithology

⁶Department of Biology, University of Konstanz

Abstract

Social network analysis is quickly becoming an established framework to study the structure of animal social systems. To explore the social network of a population, observers must capture data on the interactions or associations between individuals. Sampling decisions significantly impact the outcome of data collection, notably the amount of data available from which to construct social networks. However, little is known about how different sampling methods, and more generally the extent of sampling effort, impact the robustness of social network analyses. Here, we generate proximity networks from data obtained via nearly continuous GPS tracking of members of a wild baboon troop (*Papio anubis*). These data allow us to produce networks based

on complete observations of inter-individual distances between group members. We then mimic several widely used focal animal sampling and group scanning methods by subsampling the complete dataset to simulate observational data comparable to that produced by human observers. We explore how sampling effort, sampling methods, network definitions, and levels and types of sampling error affect the correlation between the estimated and complete networks. Our results suggest that for some scenarios, even low levels of sampling effort (5-10 samples/individual) can provide the same information as high sampling effort (>64 samples/individual). However, we find that insufficient data collected across all potentially interacting individuals, certain network definitions (how edge weights and distance thresholds are calculated) and misidentifications of individuals in the network can generate spurious network structure with little or no correlation to the underlying or “real” social structure. Our results suggest that data collection methods should be designed to maximize the number of potential interactions (edges) recorded each observation. We discuss the relative trade-offs between maximizing the amount of data collected across as many individuals as possible and the potential for erroneous observations.

Keywords

Animal social networks, observational sampling methods, proximity networks, social organization, focal sampling, scan sampling

Introduction

To understand the complexity of group structure and how individual behaviors impact group-level dynamics, we need to consider all relationships linking group members. Social network analysis provides a powerful framework to analyze the variety and variability of inter-individual connections within groups, including the strength and extent of relationships between group members (Wey, Blumstein, Shen, & Jordan, 2008). This analytical social network approach has provided insight into cooperation (Croft et al., 2006; Ohtsuki, Hauert, Lieberman, & Nowak, 2006), mating success (Ryder, McDonald, Blake, Parker, & Loiselle, 2008; Schlicht, Valcu, & Kempenaers, 2015; Scott, 1991), information transfer (Allen, Weinrich, Hoppitt, & Rendell, 2013; Aplin et al., 2015; Aplin, Farine, Morand-Ferron, & Sheldon, 2012; Brown, 1986; Couzin, James, Mawdsley, Croft, & Krause, 2007; Farine, Aplin, Sheldon, & Hoppitt, 2015; Valente, 1995), disease transmission (Adelman, Moyers, Farine, & Hawley, 2015; Cross et al., 2004; Duboscq, Romano, Sueur, & MacIntosh, 2016; VanderWaal, Atwill, Isbell, & McCowan, 2014; D. J. Watts & Strogatz, 1998), and the selective consequences of the social environment (Brent et al., 2013; Farine & Sheldon, 2015; Formica et al., 2011; Fowler & Christakis, 2008; Oh & Badyaev, 2010; Wey, Burger, Ebensperger, & Hayes, 2013). However, although existing studies use a wide variety of approaches to capture social relationships, how different data collection methods impact the results in social networks studies is rarely investigated.

At its simplest, a social network represents a set of individual entities, represented as “nodes,” and the connections between them, represented as “edges” (Wasserman & Faust, 1994). The aim of social network analysis is to quantify the extent or strength of relationships between individuals, and explore the group- or population-level structure that emerges. Relationships can be defined based on particular interactions, such as grooming or dominance, or from associations

defined by spatial proximity (see Carter, Lee, & Marshall, 2015; Castles et al., 2014; Farine, 2015 for a discussion on their relative merit). Edges typically represent the rate or probability that two individuals interact or associate in time (Farine, 2015; Whitehead, 2008), with stronger relationships having larger edge values. The set of connections that the edges form can then be captured, or described, using different social network metrics. Metrics can be calculated at the individual level (e.g. degree— how many connections each individual has, or how well connected it is relative to others), or at the level of the entire network (e.g. edge density—the proportion of all possible edges that are present in the network). Individual-level and network-level metrics are influenced both by the biology of the study organism, and also the definitions used to record observations. For example, using a large distance threshold to capture associations will result in more associations per observation, and thus a higher individual average degree and higher network density. Essential factors to consider when designing social network studies include the definition of an association or interaction, the method used to calculate network edge weights, the potential for and impact of observation errors, and the resolution at which the data can be collected.

An important feature of animal social networks is that the number of connections (edges) between individuals often greatly exceeds the number of individuals in the study system or group. Further, because network metrics are constructed from many edges, the quality of a network can be only as good as its most uncertain edge. Thus, large quantities of data across all edges are required to build an accurate and meaningful social network because data must be collected on all pairs of individuals that could potentially interact. Without sufficient sampling, the resulting network model may have little basis in reality (James, Croft, & Krause, 2009).

93 Missing connections between individuals can have significant implications on social network
94 structure, even when the missing relationships are weak (see Figure 2 in Farine & Whitehead,
95 2015). In fact, weak edges form a critical aspect of social structure that has long been used to
96 justify the importance of studying social networks in human and animal systems (Granovetter,
97 1973). Simulation studies have shown that the number of observations of each pair of individuals
98 in the study is the most important factor determining how well a network represents the
99 underlying patterns of interactions. Franks, Ruxton, and James (2010) suggest that to construct a
100 robust representation of a real social network, a minimum of 20 potential observations (i.e. both
101 individuals are observed associating or not associating) for each pair of individuals is necessary
102 (see also Farine & Strandburg-Peshkin, 2015). Other analyses show that a minimum of 15
103 independent observations where two individuals could have interacted is essential to obtain 95%
104 confidence in the strength of their relationship (Whitehead, 2008). Thus, for large networks
105 containing many individuals and many edges, this could amount to thousands of necessary
106 observations. The results of these simulation studies hold equally true for both networks
107 constructed from associations or from interactions. Further, the importance of accurately
108 quantifying each edge in the network can far exceed the need to sample every individual in a
109 population, which can seem counter-intuitive. M. J. Silk, Jackson, Croft, Colhoun, and Bearhop
110 (2015) found that as few as 30% of individuals are needed to form an accurate representation of
111 those individuals' relative positions in the social network, suggesting that priority should be
112 given to collecting more samples per dyad (edges) rather than more individuals (nodes). Further,
113 some observational methods collect data on more pairwise associations per unit time than others.
114 At present, empiricists lack guidance as to the relative trade-offs between sampling strategies for
115 defining social networks.

116
117 The behavioral data that are central to social network analysis can be collected in a variety of
118 ways, with the sampling regime, level of sampling effort, and the definition of interactions
119 (edges) often differing from study to study (see Table S1). Focal sampling and scan sampling are
120 the two main approaches to collect observational data on animal behavior (Altmann, 1974).
121 Focal sampling centers on the actions of a single individual over a set period of time, while with
122 group scan sampling, observers take a ‘snapshot’ where they record the behaviors of all visible
123 members of the group at designated points in time. Studies of animal social networks also vary
124 substantially in their sampling effort, in terms of the number of individuals sampled, the length
125 of each sample, the number of repeated samples, and the overall duration of the study (see Table
126 S1). For example, based on our review of the literature, studies of primate social networks range
127 from 18 3-minute focal follows per individual over 2 weeks (brown capuchins and common
128 squirrel monkeys in Dufour, Sueur, Whiten, & Buchanan-Smith, 2011) to almost 10-hour focal
129 follows over a period of 9 years (chimpanzees in Lehmann & Boesch, 2009). Sampling effort for
130 scan sampling also varies widely across different animal species studied, from 20 total group
131 scans over a period of 10 days where all individuals were visible in a study of Kuhl’s pipistrelle
132 bats (Ancillotto, Serangeli, & Russo, 2012) to over 73,790 group scans over 27 years in
133 humpback whales (Allen et al., 2013). Studies also vary in how they define relationships or
134 associations. In the case of spatial proximity networks, many studies use a threshold, where all
135 individuals within a certain radius of the focal individual are noted (Crofoot, Rubenstein, Maiya,
136 & Berger-Wolf, 2011a; Fraser, Schino, & Aureli, 2008; Furuichi, 1983; Matthews, 2009;
137 Nakamichi & Koyama, 1997; Szykman et al., 2001). Other studies note the identity and distance
138 of the focal individual’s nearest neighbor (Henzi, Lusseau, Weingrill, van Schaik, & Barrett,

2009; King, Sueur, Huchard, & Cowlshaw, 2011; J. B. Silk, Altmann, & Alberts, 2006), or use a combination of nearest neighbor and threshold methods, recording the nearest neighbor up to a certain distance (Clark, 2011; Suzuki & Sugiura, 2011; D. P. Watts, 1992). A further alternative is the chain rule, or ambit (see Rimbach et al., 2015; Viscido & Shrestha, 2015). Here, individuals within a certain proximity of each other are connected, and the total set of individuals that can be linked together without any breaks are considered to be connected to one another (Ramos-Fernández, Boyer, Aureli, & Vick, 2009; Wolf, Mawdsley, Trillmich, & James, 2007).

The definition of edge weights (i.e. the strength of individuals' relationships) provides yet another layer of variation in research on animal social networks. While most studies calculate a rate over time (e.g., the proportion of observations of animal A in which it groomed animal B), some studies (e.g. Castles et al., 2014; Fraser et al., 2008) instead estimate the proportion of interactions directed towards each social partner (e.g. the proportion of animal A's grooming events directed toward animal B). While the latter may be useful when investigating the relative investment made by an individual towards all its potential interaction partners, Farine (2015) argues that this approach is not appropriate for estimating global social network metrics because all individuals have the same degree (i.e. their edge weights all sum to 1). Despite this extensive variation, how sampling strategy ultimately impacts the results drawn from the analysis of the network is rarely tested. This means that there is little guidance available to evaluate trade-offs between different methodological approaches (in terms of the quality of the data collected) for the purpose of generating robust social networks.

In addition to methodological choices, certain errors or biases made by the researcher can influence the robustness of social network metrics (Voelkl, Kasper, & Schwab, 2011). Information about how misidentifications and missing observations of individuals affect the ability of an estimated network to reconstruct the real underlying social structure is relatively scarce (but see M. J. Silk et al., 2015). Misidentifications are defined as the erroneous identification of individuals in an animal study group. Conceivable scenarios include field data acquisition and subsequent data transcription errors where the names of study animals might be recorded incorrectly, mistakenly swapped, or mis-transcribed. Missing observations occur when individuals that are present in the group cannot be identified or are not observed. While sources of error in observational data collection are often stochastic, they can also be biased, resulting in particular individuals being misidentified or missed during sampling more often than others. Random errors may arise if, for instance, an individual may be misidentified if it has its back to the observer, or some group-members may be missed when observations take place in dense vegetation. Biased errors may occur if, for example, some individuals are more likely to be mistaken for another individual within their same age/sex class (i.e.- an adult female is more likely to be misidentified as another adult female).

In this paper, we examine the accuracy of different sampling methodologies in replicating a social network. We use an unusual dataset that provides complete information on the spatial associations (i.e. proximity) among members of a troop of wild baboons. This data was collected by fitting high-resolution GPS collars, recording 1 GPS point per second, to nearly all (>80%) troop members (Crofoot, Kays, & Wikelski, 2015; Farine et al., 2016; Strandburg-Peshkin, Farine, Couzin, & Crofoot, 2015). These data provide a nearly complete set of inter-individual

distances, as they contain a continuous record of the relative position of troop members, over a two-week period. Using this GPS data, we create 6 social networks based on widely used edge definitions (referred to as the *complete networks*). These networks are based on the entire, second-by-second, information about the spatial associations of the troop members. We then subsample the complete GPS data to generate datasets comparable to those produced by human observers employing focal and scan sampling methods using these 6 edge definitions (referred to as the *estimated networks*). We compare these estimated networks to the matching complete networks to investigate which sampling techniques reliably reconstruct actual patterns of association. Different approaches used to collect data vary in how much data is collected per observation (i.e. one focal sample or one scan sample, see Table 1). We therefore evaluate how sampling effort for each sampling method affects the correlation of the estimated social networks relative to the complete networks generated by the full GPS dataset. We additionally investigate the potential effects that errors in data collection—namely misidentification of individuals and missing observations—have on network statistics by simulating the occurrence of such errors at different rates. Finally, because the aim of many studies is to evaluate how position within a social network (i.e. the relative connectedness of each individual defined by different social network metrics) is related to different individual characteristics (typically one or more phenotypic traits), we also simulate a relationship between social network position and a biological trait using the complete networks. We then evaluate whether this relationship can be inferred using each of the sampling methods used to construct the estimated social networks. Our approach thus evaluates robustness and reliability from both a methodological and an ‘hypothesis-testing’ perspective.

Methods

Data collection

We generate proximity networks using spatial data from high-resolution GPS collars fitted to members of a wild olive baboon troop (*Papio anubis*). All fieldwork was conducted at the Mpala Research Centre (MRC), a wildlife conservancy in central Kenya composed of approximately 200 km² of savannahs and dry woodlands. From July 21st-29th, 2012, we fit 26 baboons (14 adults, 10 subadults, and 2 large juveniles) with GPS collars (e-Obs, Gruenwald, Germany) (see Strandburg-Peshkin et al., 2015 for details). Collared adults and subadults represented approximately 80% (24/29) of all the adults and subadults in the troop. GPS collars were programmed to record location estimates continuously at 1 Hz during daylight hours (06-18h). To quantify the error in the location estimates of the GPS data, we conducted a test walk with a pair of GPS collars fixed 1 m apart. The average relative positional error was 0.26 m (95% CI: 0.03-0.69). For the purpose of this study, all analyses use data from the first 14 days of the study when the majority of collars were active. However, several collars failed early due to a programming bug, and so the total number of individuals tracked each day varied between 16 and 25. This enables us to partially explore the effect of changes in the underlying network, an important and realistic scenario as all biological networks undergo changes over time.

Quantifying the complete networks

To quantify the *complete* proximity networks for our baboon troop, we calculate the distances between all pairs of individuals using the entire second-by-second dataset (excluding data within 100m of the sleeping site—that is, using exactly the same dataset as what we use to generate the estimated networks). We then construct 6 social networks using different edge definitions (listed in Table 1). For example, for the threshold1 network, we record the conspecifics within 10m of each individual for each second over the 14 days. These networks are *real* in the sense that they are based on complete GPS information recorded about the spatial associations of the troop members, as opposed to *estimated* networks that are based on sampling from the complete set of GPS locations (see below). It does not necessarily mean that they are biologically real (i.e. the real patterns of spatial proximity may or may not be an accurate reflection of the social relationships between individuals), but they provide a useful test case for evaluating the relative strengths and weaknesses of different choices of sampling methods.

Simulating observational sampling

We construct 6 *estimated* networks using the same edge definitions as the complete networks (listed in Table 1). We subsample our GPS proximity dataset to simulate the process of a group of observers collecting focal and scan samples by following a troop of baboons. Our literature search suggests that, on average, studies use 52 focal samples per individual for analyses, with a minimum of 13 samples (Suzuki & Sugiura, 2011) and a maximum of over 100 (Fraser et al., 2008, see Table S1). To collect this many samples, many observational studies employ multiple observers collecting data simultaneously (e.g. by sampling animals separately from opposite ends of the group). This generates more focal follow data than a single observer is able to collect.

We simulate this practice by generating all social network datasets as if three “observers” were sampling focal individuals concurrently. This allows us to produce a range of 1-64 focal samples per individual baboon, incorporating the average sampling effort seen in the literature. Due to increasing numbers of failing collars beyond day 14 of our study, we were limited in our ability to collect additional data without compromising data quality. Because the GPS noise was highest when the collars first switched on in the morning and when baboons were stationary, we start sampling each day when the troop first moved outside a 100 m radius around their sleeping site. All analyses and simulations were done using R version 3.2.2 (R Development Core Team, 2010).

Simulating focal-follow sampling

For each sample collected by one of our three ‘observers’, we simulate a focal follow by selecting one baboon at random, and recording information about its spatial neighborhood every 60 seconds for 10 minutes (see Table 1 for details), resulting in 10 neighborhood samples per focal follow. In line with the practices of human observers, we introduce a 5-minute gap between focal follows. Data are collected continuously from the time the troop leaves their sleeping site in the morning, until they return in the evening. For each follow, the ‘observers’ sample different group members, and ensure that the individuals they select had not been sampled in the preceding round of data collection. Randomizing the order in which individuals are sampled enables us to repeat the entire process 100 times to create 100 replicated sets of simulated observation data, where each dataset is unique. Over 14 days of GPS data, this results in a total of 1,593 focal follows, or an average of 64 focal follows per GPS-tracked baboon, in each

replicate. For each sample in a focal follow, we record the individual's nearest neighbor and the neighborhood data using one of 5 methods of defining spatial proximity (see Table 1). Our choice of distance thresholds is based on the most commonly used definitions for proximity networks in field studies of wild baboons (see Table S1 for literature review).

Simulating scan sampling

To make group scan data directly comparable to the data collected using focal sampling, we sample the group once every 15 minutes (i.e., a 10-minute interval between samples, plus the 5-minute 'rest period'). This sampling procedure is in line with many published studies where group scans are performed at similar rates (see Table S1). We note that certain focal sampling methods are unrealistic and infeasible when collecting observational data using instantaneous scans (namely chain rule and nearest neighbor), and thus we do not implement these as scan analyses. During each scan sample, our simulated observers record pairs of individuals that are within 10 meters of one another (threshold 1 in Table 1). Because collecting data on all troop members at once is unrealistic, we limit the percent of the group identified in each scan sample to 80%.

Constructing estimated social networks

We investigate the effects of sampling effort by constructing social networks for each method as the number of samples per individual increases. We call these our *estimated* networks. Our aim was to both determine the relationship between sampling effort and network robustness, and

298 create some general guidelines for field studies. Therefore, we generate an estimated network
 299 after 1 focal follow per individual on average, 2 focal follows per individual on average, and so
 300 on until we reached the end of our available data (64 focal follows per individual on average). In
 301 our simulations, we control for the drop-out in some GPS collars over the 14 days by calculating
 302 dyadic association rates as a function of the time the collars collected data on both individuals
 303 simultaneously. Additionally, by sequentially adding samples starting from the first day of
 304 observation onwards, we are able to explore the sensitivity of different observational methods to
 305 changing conditions in the network (in this case, individuals disappearing due to collar drop out).
 306 To disentangle the potential effects of increasing sample sizes versus temporal changes in the
 307 network structure due to collar drop out, we also simulated the same results across the 14-day
 308 study period for a single “observer” collecting data (this resulted in an average of 1-22 samples
 309 per individual baboon). We create estimated networks for the scan sampling data at the same
 310 point in time that we construct the estimated networks for focal follows. We repeat this process
 311 for each of the 100 unique replicates of simulated observation data. We compare these estimated
 312 networks to their matched *complete networks* (see below). To ensure that there was no
 313 systematic sampling bias in our datasets between focal and scan sampling (i.e. that the average
 314 number of samples per individual was different for focal versus scan sampling methods), we
 315 calculated the coefficient of variation in the number of observations across individuals and dyads
 316 under both sampling regimes across the 1-64 samples.
 317

Network name	Network type	Data recorded	Edge definition	Number of edges sampled per observation
Threshold1	10m threshold	All individuals within 10m of the focal individual	Number of times A and B were within 10m divided by number of times A and B were both present	$N - 1$

Threshold2	20m threshold	All individuals within 20m of the focal individual	Number of times A and B were within 20m divided by number of times A and B were both present	$N - 1$
Chain rule	Neighborhood	All individuals separated by a gap of no more than 5m starting from the focal individual	Number of times A and B were in the same group divided by number of times A and B were both present	0 to $\frac{N \times (N-1)}{2}$
NN1	Nearest neighbor	Focal individual's nearest neighbor	Number of times A and B were nearest neighbors by number of times A and B were both present (asymmetric directed network)	$N - 1$
NN2	Nearest neighbor	Focal individual's nearest neighbor	Proportion of observations of A that B was its nearest neighbor (asymmetric directed network)	$N - 1$
Group Scan	10m threshold	All pairs of individuals within 10m of each other	Number of times A and B were within 10m divided by number of times A and B were both present	$\frac{N \times (N-1) \times x^2}{2}$, where x is the observation probability

318

319 **Table 1.** Summary of the sampling methods and the edge definitions used to construct each of
320 the 6 different types of social network. Each method captures data on different number of edges
321 per observation. Threshold methods capture whether the focal individual is associated or not with
322 all other (N-1) group members. The 10m threshold was chosen as this represents 20% of group
323 spread in our study baboon group and is similar to common used cut-offs for defining proximity
324 in baboon research (see Table S1), while a 20m threshold was selected because it increased the
325 number of edges sampled in each period. Nearest neighbor observations capture 1 nearest
326 neighbor edge, but also the fact that the other N group members are not the nearest neighbor.
327 Nearest neighbor edges are asymmetric (i.e. directed). For the chain rule, individuals forming a
328 cohesive group are sampled, with the group size ranging from 1 (the focal individual only) to N.
329 However, because the status of other subgroups (whose members are not directly or indirectly
330 connected to the focal individual) is not recorded, the number of edges sampled is not consistent
331 in each observation (i.e. two individuals that are not found in the subgroup are known not to be

associated with members of the subgroup, but their status relative to each other is unknown). For scan sampling, all edges between visible individuals are captured. In this study, we introduce a scan sampling observation probability (x) of 0.8 to simulate missed observations in each sample, and also to introduce variation among the 100 replicated data sets.

Introducing biases and errors in observational sampling

To evaluate the consequences of sampling errors, we explore one focal sampling method (focal sampling with threshold1; all individuals within 10 meters of the focal animal are recorded) in greater detail. For comparison, we implement the same analyses for group scan sampling where all pairs of individuals within 10 meters are recorded as associating. We investigate the effects of misidentification and missing observations, under both random and non-random scenarios.

Misidentifications involve erroneously assigning the unique identifier (animal ID number) of one study animal to another. We generate misidentifications by swapping the identity of individuals in pairs (i.e. individual A is incorrectly identified as individual B and vice versa). This swapping method allows us to compare data across all methods by shuffling the identities of individuals, as sometimes can occur in observational studies. We do this at varying levels of misidentification, where 1 pair, 2 pairs, 3 pairs, 4 pairs, 5 pairs, and 6 pairs of individuals are swapped during each sample (each focal has 10 samples, each scan has 1 sample). To replicate missing observations, we randomly remove individuals from the data generated by focal and scan sampling. For focal sampling this means that a certain percentage of a given focal's spatial associates are not recorded (e.g. if focal individual A is within 10m of individuals B and C, and the observer does

not record individual C as nearby, then individual C is a missing observation and the edge connection between A and C is not recorded). For scan sampling this means an individual who is present in the group is missed, and thus their spatial associations are not recorded. We mimic observers missing 5%, 10%, 20%, 30%, 40%, and 50% of individuals each sample (1, 3, 5, 8, 10, and 13 missed observations at each sampling time step). For both misidentifications and missing observations, we implement this randomly where all pairs were equally likely to be misidentified, or all individuals were equally likely to be missed. We further investigate group scan sampling where all pairs of individuals within 20 meters are recorded as associating for random misidentifications, to assess if distance threshold impacts the effect of errors. To simulate the effects of biased, or non-random, errors in data collection, we introduce a bias such that the probability of an individual being missed is proportional with its distance from the center of the group. Because group members show consistent spatial positioning within the group (Farine, Strandburg-Peshkin, Couzin, Berger-Wolf, & Crofoot, 2017), such spatially-biased errors lead to consistent patterns in which individuals are missed.

Metrics for quantifying individual network position

We measure three node-based metrics to evaluate the robustness of social networks derived using different sampling procedures in each of the estimated networks (see next section). These represent both local (direct) and global (indirect) measures. First, we use the weighted degree (or strength) metric, a local network measure often used to estimate individual gregariousness, with more social individuals having a higher degree. We believe this is an important direct measure in social networks as many studies of social differences in animals focus on simple indices of

sociality, like degree, to capture an individual's number of social partners or the frequency/rate they interact with others (J. B. Silk, Alberts, & Altmann, 2003; J. B. Silk et al., 2010). Second, we calculate eigenvector centrality and betweenness—two commonly-used, global network metrics (Aplin et al., 2015; Boogert, Farine, & Spencer, 2014; Oh & Badyaev, 2010). Eigenvector centrality is a global measure of an individual's direct and indirect contacts, where individuals with high values have either many associates or are connected to other individuals with many associates. Such measures of indirect connections (i.e. connections beyond the dyadic level) have been shown to have biological importance. For example, eigenvector centrality has been shown to be a better predictor of offspring survival in female baboons than dyadic-level connections (Cheney, Silk, & Seyfarth, 2016). Betweenness centrality is a global measure of how important a node is at providing the shortest path between other nodes in the network, and is thought to be linked to dispersal tendency and social movements between different groups/sub-groups (Farine & Milburn, 2013; Oh & Badyaev, 2010). Betweenness centrality also plays an important role in the spread of information and disease in social groups (Adelman et al., 2015; Aplin et al., 2015).

Estimating the robustness of the estimated social networks

We compare metrics applied to the estimated network with the same metrics applied to the matched complete network. For example, the degree of an individual estimated using 10 m threshold focal samples is compared to degree in the network constructed using a 10 m threshold applied to all of the data available for focal sampling (i.e. discarding data within 100m of the sleep site). We use the rank correlation on the values of each metric derived from the estimated

vs. complete networks to determine how well each observation method performs. We calculate the mean and 95% range of correlation values for each level of sampling effort (average number of focal follows ranging from 1 to 64) using the 100 replicated datasets. This correlation measures the ability of different sampling methods and levels of sampling effort to estimate the structure of the network generated using the complete data, and using the full range of samples gives us a good indication of how the performance of the estimate increases (or not) with additional sampling effort.

Estimating robustness from a ‘question-driven’ perspective

Although the first step for evaluating the quality of different sampling strategies is to compare the similarity in network structure (see above), many studies use social networks to explore relationships between animals’ positions in their social network and one or more biological traits (Aplin et al., 2013; Boogert et al., 2014; Brent et al., 2013). For example, one might test if degree centrality is correlated with dominance. Thus, in our second test, we simulate an individual level trait that is strongly correlated with an individual-level social network metric (Figure 1, top row). We do this separately for all three network metrics: degree, eigenvector centrality, and betweenness. For each complete network, we first create a trait value for each individual by ranking individuals according to the given network metric and normalize the rank values (ranging from 0 to 1). This simulates a relationship, such as a body-size index, that is perfectly correlated to an individual’s social network position. For each of the estimated networks generated (i.e. 100 for each sampling method and level of sampling effort), we then regress the individual-level social network positions (e.g. rank in eigenvector centrality score) from the

424 estimated network against the trait value generated from the equivalent complete network
 425 (Figure 1, bottom row). Because many biological relationships are often imperfect, we repeat this
 426 process, adding noise when simulating the trait values (which otherwise perfectly represented
 427 each individual's rank in a given social network metric, Figure 1a). Noise is generated by
 428 drawing random values from a normal distribution with a mean of 0 and a standard deviation
 429 increasing from 0 to 0.3 in steps of 0.05 (see Figure 1b & 1c for examples). We create 100 such
 430 relationships from each of the complete networks (one per replicate), and recorded the proportion
 431 of the regressions between the trait and the *real* network data in which the effect was significant
 432 at $\alpha=0.05$ (to calculate the expected rate of false negatives for that network at that level of noise).
 433 We then regress the same trait values against each of the estimated networks (after 1 to 64
 434 average focal follows) to determine if the relationship between the estimated network metric and
 435 the individual trait value can be detected.

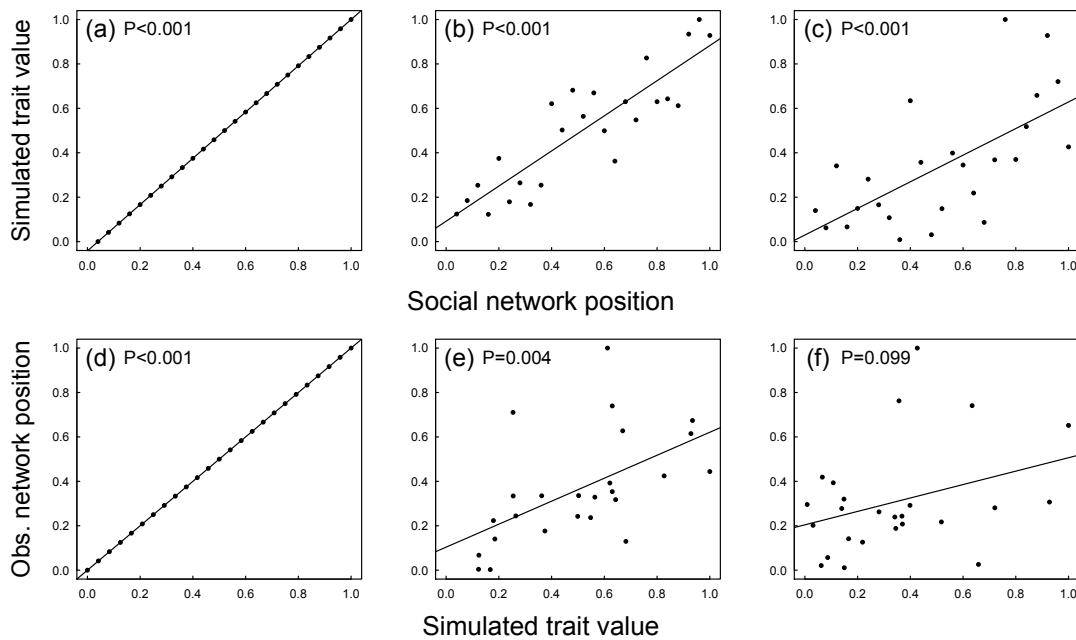


Figure 1. Example of the procedure for simulating a relationship between social network position in any given complete network and a biological trait, and then testing whether that relationship can be recovered from the social network position inferred using observational data. Trait values are simulated by ranking individuals according to their social network position in the complete network (top row). Noise is added in panels a to c with standard deviation values of 0, 0.15 and 0.3 respectively. We then calculate each individual's social network position in an estimated network for a given level of sampling effort (bottom row). This toy example demonstrates a decrease in the accuracy of the estimated network in panels (d) to (f).

Ethical note

All research activities described in this paper were approved by the Republic of Kenya's National Council for Science and Technology (Permit # NCTS/RCD/12B/012/26B), Kenya Wildlife Services (KWS/BRM/5001) and the Smithsonian Tropical Research Institute's Animal Care and Use Committee (IACUC # 2012.0601.2015).

Results

The effect of sampling type and effort on network robustness

To assess the impact of observational methodology and sampling effort on the robustness of estimated proximity networks, we compare our complete networks to networks estimated using a

variety of different observational methods. Specifically, we correlate three individual-level, weighted social network metrics: degree, eigenvector centrality, and betweenness, measured on complete networks (based on 1Hz GPS data) and on estimated networks. The amount of sampling effort required to produce an accurate representation of the complete proximity network varies depending on the sampling method used (Figures 2, 3, and 4). The performance of each method in re-creating the complete networks is relatively consistent when using either degree (Figure 2) or eigenvector centrality (Figure 3), although notably not for betweenness. For all sampling methods except nearest neighbor (NN1 and NN2), the correlation value between the estimated and complete networks generally increases rapidly in the first 10 samples, after which only marginal increases in correlation are gained from further sampling. Thus more sampling effort (>10 samples per individual) does not considerably increase the accuracy of these estimated networks in replicating the complete networks. In contrast, increased sampling is necessary for the nearest neighbor NN1 and NN2 methods: the correlation increases up to approximately 30 samples per individual. NN2 performs very poorly overall for eigenvector centrality, producing low correlation between estimated and complete networks across all levels of sampling effort. In general, all correlations between estimated networks and complete networks are lower when using betweenness (Figure 4). Our results simulating varying sampling intensities (one versus three observers) for each time point in the dataset also indicate that these same correlation patterns are robust to finer details of sampling design and hold across the sample study period (Figures S1-S3). No sampling method produces a network that correlates 100% with the equivalent complete network even after 1,593 sets of focal observations. However, scan sampling consistently results in high correlations across all network metrics (>0.8) after as few as 10 observations. Additionally, when calculating the coefficient of variation

in the number of observations per individuals and per dyads under both focal and scan sampling methods, we found that when the network is largely unchanged (in terms of the presence/absence of individuals), both methods achieve accurate sampling with no variance in the number of samples per individual. Once some network structure dynamics are introduced (in this case, GPS collar drop out), the focal method sampling cannot perfectly track the underlying variance in amount of time individuals are present, and thus produces marginally higher variance levels. Conversely, scan sampling accurately tracks underlying changes in the network (Figure S28). Overall, the similarity in variance values for the number of observations per individual for focal versus scan sampling indicates there is no systematic bias resulting in our sampling methods.

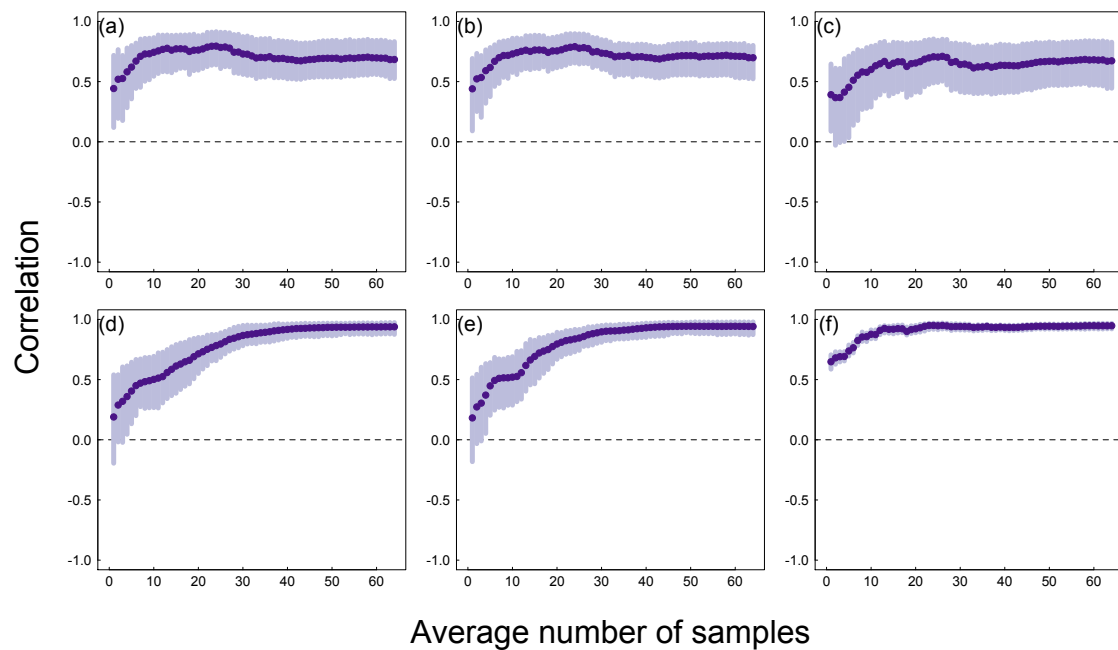
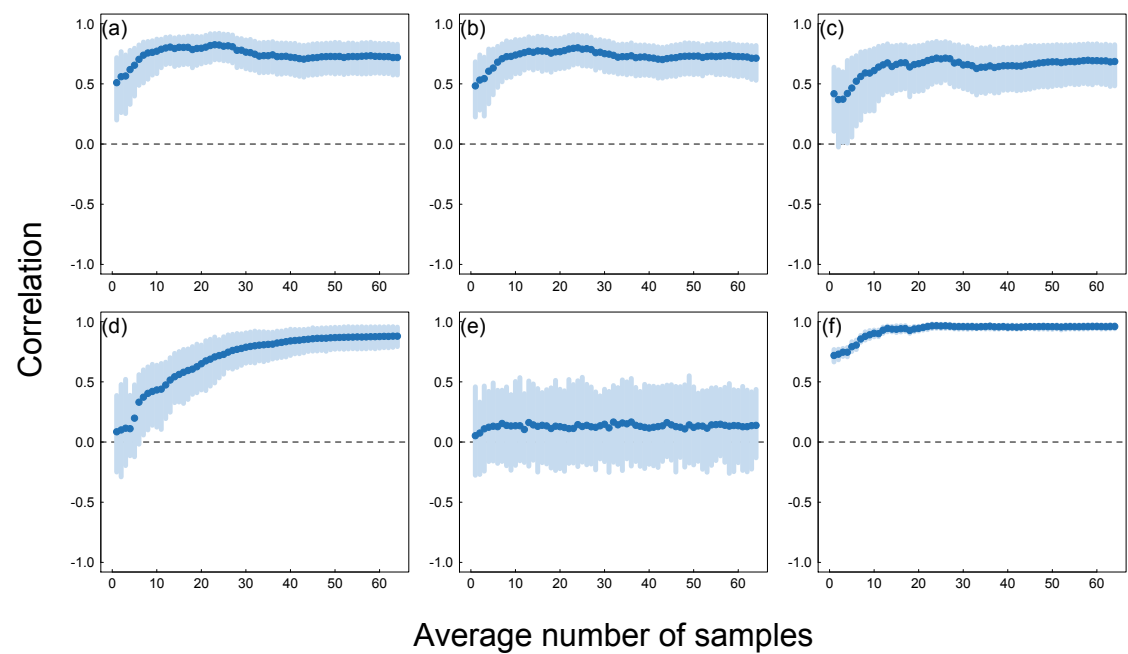


Figure 2. Mean and 95% confidence intervals of the rank correlation of degree centrality in the estimated vs. complete networks. Plots show how the rank correlation varies as a function of sampling intensity (average of 1-64 focal samples per individual) for: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) group-level scan sampling methods (see Table 1 for details).



500

501

502

503

504

505

506

Figure 3. Mean and 95% confidence intervals of the rank correlation of eigenvector centrality in the estimated vs. complete networks. Plots show how the rank correlation varies as a function of sampling intensity (average of 1-64 focal samples per individual) for: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) group-level scan sampling methods (see Table 1 for details).

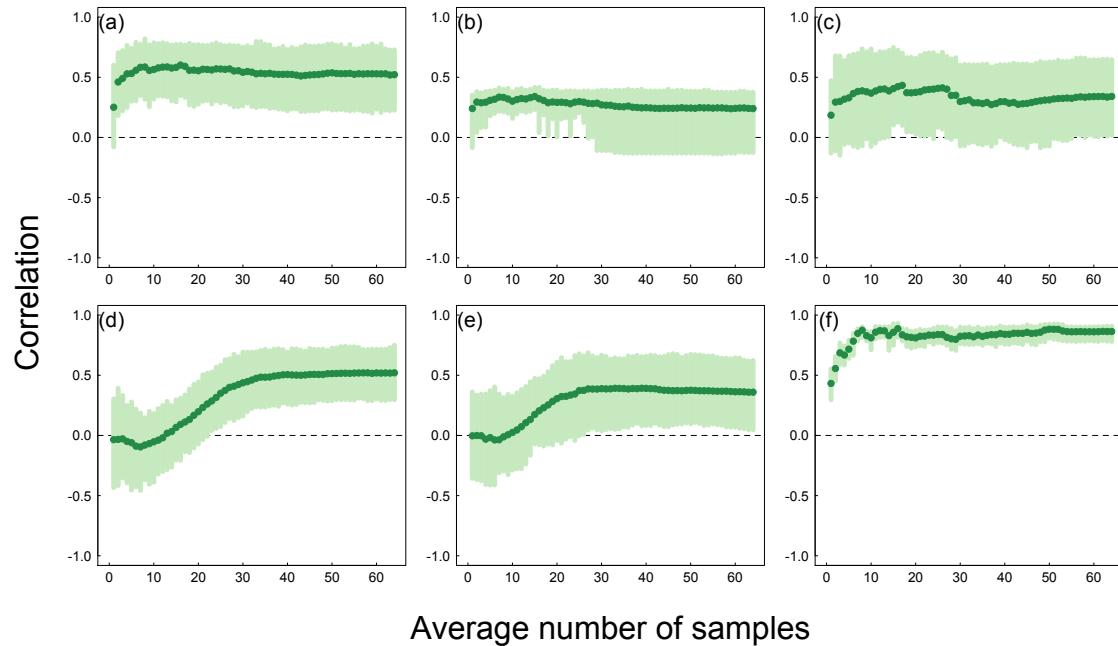


Figure 4. Mean and 95% confidence intervals of the rank correlation of betweenness centrality in the estimated vs. complete networks. Plots show how the rank correlation varies as a function of sampling intensity (average of 1-64 focal samples per individual) for: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) group-level scan sampling methods (see Table 1 for details).

The effect of sampling errors on estimated network robustness

We then repeat our analyses introducing sampling errors--namely mistaken identities (wrongly identifying individual pairs) and missing observations (individuals that should have been recorded but were not). In general, we see an effect of these errors in data collection on the correlations between the estimated and the complete networks (Figure 5). Misidentifications reduce the correlation between the estimated and the complete networks overall, especially when using betweenness metrics (Figure 5g-l). Overall, the estimated networks generated using scan

sampling were comparatively less affected by misidentifications than networks constructed based on focal sampling methods. However, as more observations were added (1 to 64 samples per individual baboon), the misidentification errors accumulate, and the accuracy of the estimated network structure deteriorates, regardless of sampling method (see also Figures S4-S15). We discuss how error accumulation results in this declining correlation further in the Appendix. Data collection errors involving missed observations have less impact on the robustness of the estimated networks than misidentification errors (Figure 5a-f, see also Figures S16-S21). This is especially true for the scan sampling method, where the estimated networks remain accurate even when 50% of the group is not recorded (Figures S16, S18, and S20). In all cases, the effect of observation errors is most pronounced for betweenness compared to the other network metrics. The match between estimated and complete networks further decreases with increasing levels of observational error (i.e.- more misidentified individuals and more missed individuals, see supplemental figures S4-S21).

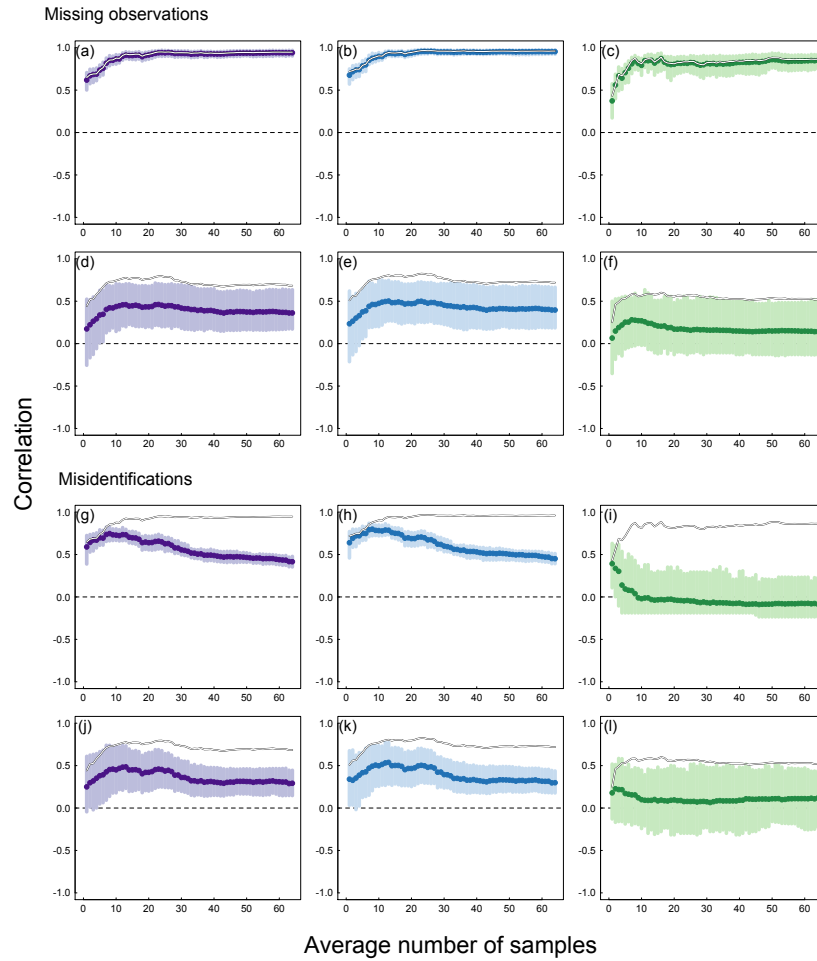


Figure 5. The mean and 95% confidence intervals of the rank correlation in the estimated vs. complete networks with sampling errors for missing observations (panels a-f, where 10% of all individuals are randomly missed and not recorded) and misidentifications (panels g-i, where 2 pairs of individuals are randomly misidentified). The original mean correlation values without any levels of error are marked in each panel with the black and white line. The rank correlation is evaluated across three network metrics: degree (purple panels a, d, g, j), eigenvector centrality (blue panels b, e, h, k), and betweenness centrality (green panels c, f, i, l). Estimated networks are generated using group scan sampling (panels a-c and g- i) and focal samples using the threshold1 method (panels d-f and j-l). Plots show how the rank correlation varies as a function of sampling intensity, from 1-64 average samples per individual.

547

548 *The effect of sampling biases on estimated network robustness*

549

550 Biased errors during data collection (here, spatially peripheral individuals being misidentified
551 more frequently than their spatially central group mates) have a smaller effect on estimates of
552 network structure than errors arising from randomly failing to sample all group members (i.e.,
553 missed individuals), regardless of the sampling regime used (Figures S22-S27). Using scan
554 sampling, even when 50% of individuals are missed in each scan, estimated and complete
555 network rank correlations remain relatively unchanged. As with random errors, the effects of
556 biases in missing observations is greater when using focal sampling than scan sampling, but the
557 overall impact is relatively small.

558

559 *Inferring relationships between social network position and biological traits*

560

561 The ability to infer a relationship between social network position and a biological trait (e.g.,
562 body size, age, parasite load) varies depending on the type of behavioral sampling used to
563 estimate the underlying social network structure, even when there is a perfect relationship
564 between these two variables (degree: Figure 6, eigenvector centrality: Figure 7, and betweenness:
565 Figure 8; dark blue data points). However, in real biological systems the ‘true’ relationship
566 between a trait and network metrics is often relatively noisy. When using degree, we found that
567 for almost all sampling methods, the relationship between the estimated degree and the simulated
568 trait is robust to additional variance (Figure 6). One exception was threshold2, which could be
569 because a large threshold distance connecting individuals (here 20m) creates too dense a network

and therefore removes much of the variation in degree among individuals. The patterns are consistent for eigenvector centrality (Figure 7) and betweenness (Figure 8), although the impact of additional variance is greater for our metric representing the whole network (betweenness) relative to metrics representing local neighborhoods (eigenvector centrality). In general, adding additional samples improves the ability to recover a relationship between network metrics and a biological trait, particularly when additional variance is present (which will be the case for most field studies).

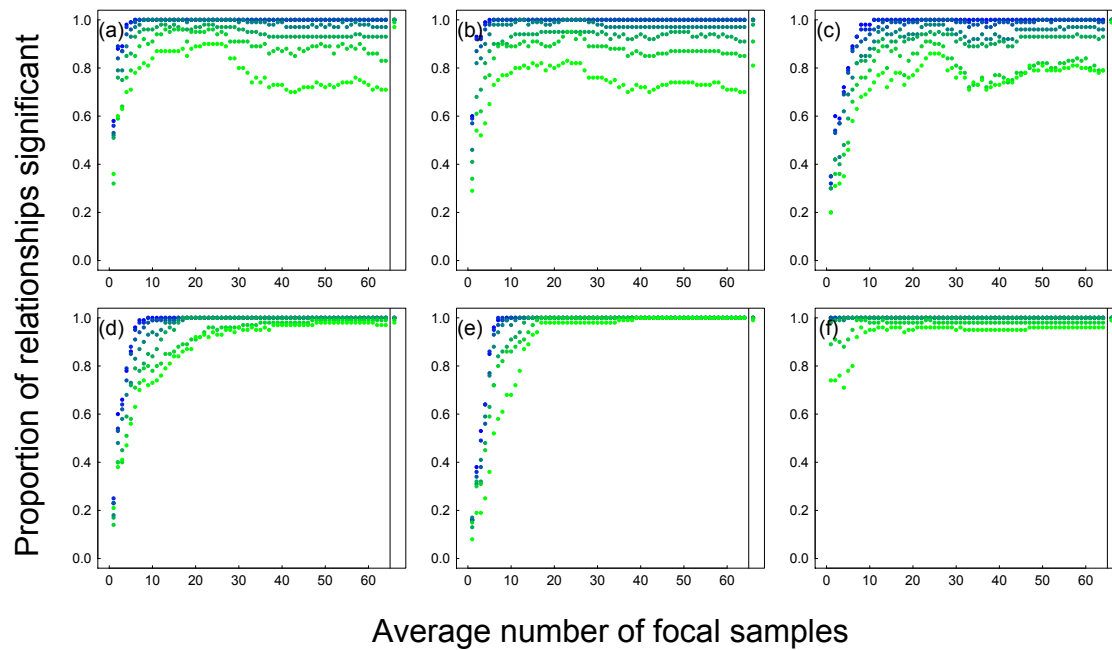


Figure 6. The proportion of models where the correlation between a simulated trait (generated based on each individual's rank in degree centrality in the complete network) and rank in degree centrality in the estimated networks is significant. The proportion of significant models is calculated for each of the six sampling methods: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) scan sampling (see Table 1 for details) across each level of sampling effort. For each complete network, an individual level trait was generated to perfectly correlate

with degree. Dark blue points represent values with little noise added to this relationship ($SD = 0.05$), while lighter green colors represent more noise (SD up to 0.3). Points to the right of the solid vertical black line are the proportion of relationships between the simulated trait and degree centrality in the complete network that are significant, representing the expected rate of significant relationships (as more noise is added, the proportion of models where the simulated relationship is significant reduces). Values on the left of the line should match those on the right of the line if the estimated network can robustly replicate the simulated relationship.

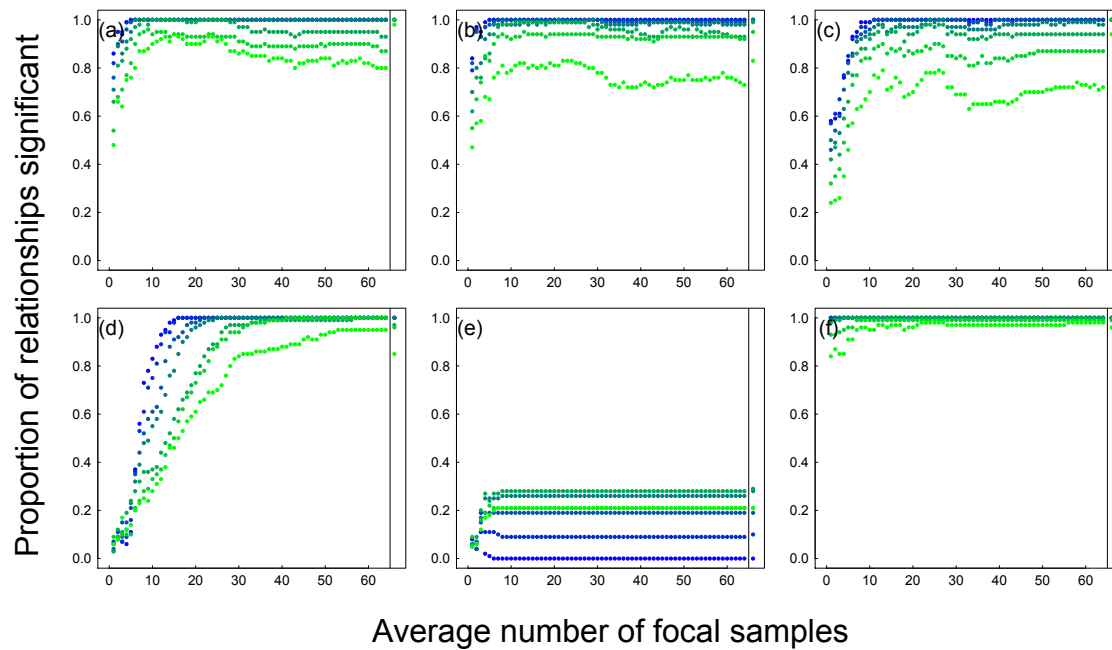


Figure 7. The proportion of models where the correlation between a simulated trait (generated based on each individual's rank in eigenvector centrality in the complete network) and rank in eigenvector centrality in the estimated networks is significant. Proportion of models significant is calculated for each of the six sampling methods: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) scan sampling (see Table 1 for details) across each level of sampling effort. For each complete network, an individual level trait was generated to perfectly correlate

with eigenvector centrality. Dark blue points represent values with little noise added to this relationship ($SD = 0.05$), while lighter green colors represent more noise (SD up to 0.3). Points to the right of the solid vertical black line are the proportion of relationships between the simulated trait and eigenvector centrality in the complete network that are significant, representing the expected rate significant relationships (as more noise is added, the proportion of models where the simulated relationship is significant reduces). Values on the left of the line should match those on the right of the line if the estimated network can robustly replicate the simulated relationship.

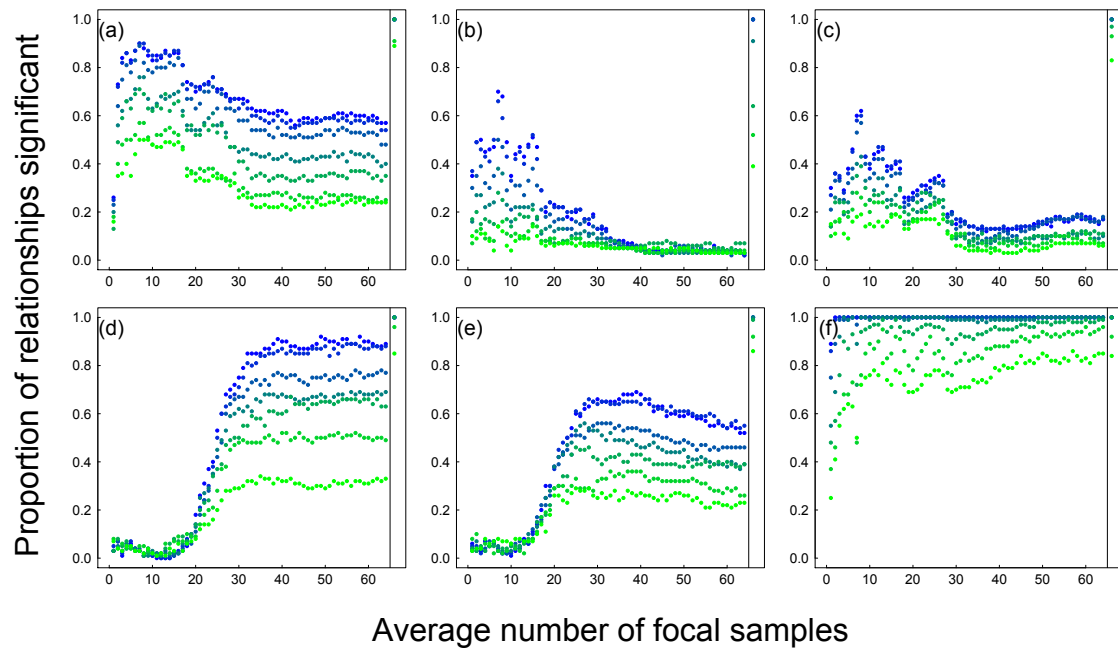


Figure 8. The proportion of models where the correlation between a simulated trait (generated based on each individual's rank in betweenness in the complete network) and rank in betweenness in the estimated networks is significant. The proportion of significant models is calculated for each of the six sampling methods: (a) threshold1, (b) threshold2, (c) chain rule, (d) NN1, (e) NN2, and (f) scan sampling (see Table 1 for details) across each level of sampling

effort. For each complete network, an individual level trait was generated to perfectly correlate with eigenvector centrality. Dark blue points represent values with little noise added to this relationship ($SD = 0.05$), while lighter green colors represent more noise (SD up to 0.3). Points to the right of the solid vertical black line are the proportion of relationships between the simulated trait and betweenness centrality in the complete network that are significant, representing the expected rate of significant relationships (as more noise is added, the proportion of models where the simulated relationship is significant reduces). Values on the left of the line should match those on the right of the line if the estimated network can robustly replicate the simulated relationship. The decrease in proportion of relationships significant using some observer methods is because in small groups such as these baboons, increasing the number of samples saturates the network (it becomes dense and fully-connected), meaning that everyone is equally connected to everyone else (hence the relationships are undifferentiated).

Discussion

Social network analyses are a powerful tool for quantifying the complex inter-connected structure of animal communities. However, constructing social networks requires large quantities of data, as the number of connections in a network is generally high even when there are few individuals (nodes) present. In this study, we test which methods of data collection are most effective, how much sampling effort is required to produce robust estimates of animal social network structure, and the effects of errors and biases in data collection on network reliability. Importantly, we note that our data comes from GPS-generated proximity social networks in a

population of olive baboons living in semi-open savanna habitats. While the general principles uncovered by our study should provide useful guidelines for future studies, we caution researchers against using the exact values produced in our study. We emphasize that researchers should carefully consider the biological system and social network of interest to make an informed decision about the best sampling method and amount of sampling effort needed to create accurate social networks.

Overall, we found that when using a local measure of network affiliation (i.e. weighted degree), the networks inferred from ‘observational’ data were generally robust, regardless of whether focal or scan sampling methods were used or how proximity was defined. Many analyses of social connectedness focus on the number of social partners an individual has, and this finding is important because it demonstrates that two common approaches to defining proximity and sampling such relationships—nearest neighbor focal follows and threshold-based group scan sampling—can accurately capture local connections within social networks. However, when global metrics of the network are of interest, our results suggest that decisions about sampling methodologies are particularly important. Certain sampling strategies used in observational data collection can generate spurious structure such that the networks produced have little or no correlation to the underlying or “real” social structure. Global metrics—in particular, betweenness—are also more sensitive to insufficient sampling.

The amount of data required to generate robust estimates of network structure and to detect relationships between position in the network and individual-level biological traits was surprisingly low (but see the caveats discussed below). Importantly, because it captures the

presence/absence of many edges per scan, scan sampling generally outperformed other methods across all of the tests we implemented, while focal animal sampling that defined proximity relationships based on local thresholds (with an appropriately-chosen distance) and nearest neighbor approaches (with correct edge definitions) performed reasonably well in most cases (although the latter require more data).

Our study suggests that data collection methods should attempt to maximize the number of potential interactions (edges) recorder per observations, rather than the number of observations. In our proximity study, scan sampling produced more accurate and robust networks, especially when using global metrics. This is because scan sampling provides data on proximity network associations of most individuals (even if some are out of sight) across the group at each observation. Focal sampling conversely, measures the ties of single individuals in turn (in essence, a “star-shaped” network), and reconstructs the whole network from these partial snapshots. In doing so, it can take long periods of time before a changed relationship is sampled, requiring one of the individuals to become a focal (which in large groups could take several days). A potential consequence of this is a mismatch between the timing of the predictor of the effect and the time when the outcome is detected. We recommend researchers review the possible number of edges sampled per observation as indicated in Table 1, and thus how often each edge is sampled on average. When sampling under the focal method, there is only a single individual that can have $N-1$ possible edges. All other individuals can have at most a single edge. However, with scan sampling, every individual in the network can have $N-1$ possible edges at each sample point. Scan sampling thus produces more information (i.e. edge connections) for the entire network at one time.

How do our findings inform decisions about the trade-offs between different sampling methods that could be of use when designing a field study? Sampling decisions have a significant impact on the outcome of data collection, and thus on the appropriateness of a dataset to address questions about group and individual behavior (Fragaszy, Boinski, & Whipple, 1992). Focal samples are typically used to examine individual-level characteristics and patterns of behavior. During focal follows, the observer records behaviors that are expressed by or directed at the focal animal (Altmann, 1974), which provides excellent data on individual characteristics and dyadic relationships, but limits the number of interactions across the group that are being sampled at once (see Table 1). By contrast, scan sampling captures a greater extent of the overall group structure by sampling the entire group at one time (rather than re-observing the same dyads many times). Focal sampling is advantageous for understanding individual-level traits (such as social connectedness or bond strength), while scan sampling provides a bigger picture of a system's network structure by sampling many potential interactions at once. We also emphasize that focal sampling is likely less prone to errors in data collection than scan sampling methods, an important trade-off we discuss below. When using focal nearest neighbor data to construct networks, we found that the definition used to calculate edge weights can also have a significant impact on the accuracy of the resultant network produced. If studying very local relationships (i.e. using only degree), then both approaches to defining nearest neighbors (NN1 and NN2) when calculating edge weights perform well. However, as soon as a more global perspective is taken ('friends of friends' Brent, 2015), then calculating edge weights as rates (NN1) rather than proportions of total (NN2) results in major differences in the resulting quality of the estimated networks.

While our study emphasizes the importance of capturing a large amount of data across all possible interacting individuals to generate accurate social networks, we note that the best methods to collect this data may vary based on whether researchers are interested in interaction (i.e.—aggression, grooming, etc.) versus association (i.e.—proximity) networks. We examine this effect in spatial networks where scan sampling collected more data across all group members at each sample point than focal sampling and resulted in more robust proximity social networks. However, we acknowledge that some specific interactions related to key biological processes will be important to collect, such as the role of grooming in forming and maintaining social bonds and alliances (Crofoot, Rubenstein, Maiya, & Berger-Wolf, 2011b; J. B. Silk et al., 2010). Often in studies of these interaction networks (including grooming, aggressive, and vocal networks), scan sampling methods rarely capture the instantaneous behaviors crucial to evaluating these networks. Focal sampling, or some form of all-occurrence sampling across the group, may then provide a better avenue to obtain sufficient data. We encourage researchers to carefully choose a sampling scheme that both maximizes the number of potential interactions recorded in one observation and also collects enough data on the behaviors of interest.

Our study also explores how sampling effort affects the accuracy of social network analysis. For most edge definitions, smaller sample sizes still provide similar information on individuals' positions within the social network as larger sample sizes (Figures 2, 3, and 4). For example, correlations between estimated networks and complete networks reached 0.8 after approximately 10 samples across all network metrics when using scan samples. However, focal sampling methods that record only nearest neighbor associations (NN1 and NN2) require more sampling effort to produce higher correlation values between estimated and observed social networks

(around 30 focal samples per individual) than sampling regimes that record all close associated within a certain threshold range. This is likely because a single neighbor/associate is recorded each scan, resulting in fewer total sampled associations. Ultimately, our overall result that small sample size can still capture the real network structure in most sampling regimes is likely because few observations are needed to capture the main (strongest) edges in a network. Feczko et al. (2015) similarly found that a minimum of 1 hour of spatial proximity observations produced reliable social networks using video observations of rhesus macaques. Their observations were conducted over a single day, and it is likely that only 1 hour of observation for that day was sufficient to reduce errors in the accuracy of edge weights. We note an important distinction between their study and ours: Feczko et al. (2015) evaluated the accuracy of edge weights, whereas our study examines the reliability and robustness of the overall network. Both studies agree that increased sampling intensity may not always be a priority.

Furthermore, we recommend that researchers carefully consider two important aspects of their system when conducting network studies: the dynamicity of the system (changes in network structure over time) and the rate of interactions between individual animals. In a situation where a network is invariant in time and researchers have sufficient time to collect data across all interacting animals, then both focal or scan sampling methods will produce similar results. However, in a situation where a network has some dynamics over time (such as individuals entering or leaving the network, or where important outcomes arise from rare interactions, we strongly recommend that researchers maximize the density of the data collected to produce more accurate representations of network dynamics (for further discussion on network dynamic versus non-dynamic networks see Farine, 2018). For instance, in our study scan sampling consistently

produced more robust networks than focal sampling as when GPS collars dropped out from the study system because the sampling of individuals was unbiased relative to their occurrence in the complete data (see Figure S28). We suggest that authors carefully consider questions such as: what are the number of edges observed per focal on average? What are the number of edges observed per scan on average? How many focal samples are needed to fill up a network at the same density as scan sampling? At what rate is the network structure expected to change during the study period (e.g. via demographic changes)? These questions will help with the design of data collection. For example, in forest-living species, visibility may limit the feasibility of scan sampling and thus all-occurrence sampling of behaviors (if unbiased) or focal sampling with multiple observers may be a better methodological choice.

Although a small number of observations were often sufficient for producing estimates of network structure that were highly correlated with the underlying, ‘real’ network, no sampling method yielded a perfect reconstruction, no matter the number of observational samples collected per individual baboon. This is because all subsampling methods miss some network connections (i.e. those happening in between scan samples, or among individuals not associating with the focal animal), and this persists even as the number of samples accumulates. To increase the correlation between the estimated and the true networks, observers would need to conduct more samples per individual per unit time rather than increasing the number of focal observations by conducting them over longer periods of time. We suggest that there is a trade-off between the resolution at which data are sampled and the number of samples obtained that warrants greater consideration in field studies.

As more samples are collected, we sometimes detected a slight decrease in correlation values or in the ability to extract the simulated relationships (e.g. Figure 8). There are two effects at play here. First, when collecting a few samples, we are more likely to observe strong links. Thus, the essentially binary edges (edges present vs. not present) are quite likely to match the strong edges in the real network, generating a correlation in the networks. This does not necessarily mean the network is *accurate*. Second, as more observations are added, the network can become saturated and dense—where individuals are, for the most part, connected to everyone else. This means that there is no differentiation among individuals, and small amounts of noise can have a large impact on their ranks. Some studies have suggested that thresholding edges (removing or setting edges below a certain weight to zero) may overcome limitations of fully-connected networks (Franks et al., 2010), however this is unlikely to be a suitable solution given the issues with picking a weight below which to remove edges; such an approach risks a propensity for ‘fishing’ for a value that generates the most significant results (see Farine, 2014; Farine & Whitehead, 2015).

Mistakes happen—this is a reality of all data collection, but one that can be particularly prevalent in field conditions. It is therefore important to understand how different types of mistakes will impact the results of a study. By simulating errors and biases that may be common in field-based observational sampling methods, we show that such errors affect networks constructed using focal sampling methods more than those using scan sampling. Previous research has shown that estimating social positions is relatively robust to missing individuals completely from the study (M. J. Silk et al., 2015). Our work expands this finding by demonstrating that networks are robust even when individuals are missed during observations, and is a reassuring result for field studies where groups are widely spread and difficult to always observe. By contrast,

misidentifying study animals had a strong negative impact on the robustness of network inferences, regardless of which sampling methods were used. This is because misidentifications change the network structure by potentially adding edges where none should exist. In real terms, this suggests a benefit to focal follows if this approach facilitates more accurate identification of individuals than scan sampling methods. For example, one long-term study of white-faced capuchins requires researchers to maintain 97% accuracy in sampling observations (Perry & Manson, 2008), resulting in approximately 1 in 100 lines of data containing missed observations and 1 in 1000 lines of data containing misidentified individuals (Perry, personal communication). Our results suggest that if researchers are uncertain about the identity of an individual in the field during data collection, it is better to skip that individual than misidentify it, as misidentifications have a greater effect than missed observations on the robustness of social network analysis.

We also investigated the effects of biased, or non-random, observation errors on estimated social network structure, by introducing a representative bias such that the probability of an individual being missed during behavioral sampling was proportional to its distance from the center of the group. The results from this simulation suggested that biased sampling did not impact network robustness as strongly as randomly generated errors. Although this finding seems non-intuitive, it arises because such biases disproportionately affect one part of the network. Most observational biases have a strong impact on a small number of individuals, whereas random errors affect the entire network equally. While we evaluated only one type of observational bias, we believe this result is applicable across many potential sources of bias in behavioral data collection that cause certain individuals to be more likely to be missed than others.

822 We found that the choice of distance threshold for designating edge connections between
823 individuals significantly impacts the robustness of the estimated networks. For example, in our
824 troop of baboons, 10m (most studies of baboon networks use a 5 or 10 m threshold, see Table
825 S1) represents about 20% of the average group spread. Thus, if following a similar sized group
826 of primates that only spread over 20m, our threshold1 results could be equivalent to a 5m
827 threshold in that species. Another consideration when choosing distance thresholds is the density
828 of the group. For example, we found that a 20m threshold (threshold2) performed relatively
829 poorly because it captured too many edges and made the resulting network too dense. Thus, in
830 species that form dense groups, a smaller distance threshold might be more appropriate.
831 Knowledge of the study system will play an important role in making these types of sampling
832 decisions. We suggest that one useful approach to make such decisions might be to identify the
833 typical size (i.e. diameter) of visually-obvious subgroups. Then, use the average diameter size of
834 subgroups to designate appropriate distance association thresholds. Alternatively, previous
835 studies (Farine et al., 2016; Henzi, Lycett, & Weingrill, 1997; King et al., 2011; J. B. Silk,
836 Seyfarth, & Cheney, 1999) have shown that baboons have a distinct number preferred associates.
837 Thus, measuring the average distance to that number of associates could provide a biologically
838 meaningful threshold. In our study, threshold1 resulted in a mean of 4.1 associates per focal
839 observation, which is consistent with previous studies that have shown baboons have, on
840 average, 4-5 preferred social partners (J. B. Silk et al., 2006; J. B. Silk et al., 1999). By contrast,
841 the networks constructed using threshold2 had an average of 9.1 associates per focal observation,
842 which exceeds the suggested number of important associates in the literature, and generates
843 much denser networks with lower differentiation among social relationships. We recommend

844 that researchers carefully consider what is a biologically meaningful edge definition for their
845 study animals.

846

847 We hope the general principles uncovered by our study, which we summarize below, will
848 provide guidance to those designing effective data collection protocols for studies that involve
849 social network analysis. Significantly, our findings suggest that maximizing the number of
850 interactions that could potentially be observed at each sample point should be a priority for
851 constructing group social networks, a result that should be robust to different species and
852 systems. For proximity social networks, scan sampling produced the best estimates of the true
853 proximity network of the baboon group. Scan sampling provided proximity association data on
854 multiple interacting individuals simultaneously across the group, while focal sampling measured
855 snapshots of each individual's spatial ties in turn. As above, we note that while scans worked
856 well in our study of proximity, studies of certain interaction networks (like vocal or aggression)
857 likely will obtain more data using focal samples or all-occurrence sampling across the group.
858 Additionally, we found fewer samples were needed to produce robust social networks when
859 using observational methods that record more data across the majority of nodes in the network
860 (i.e. scan sampling for proximity networks). When using a local measure of network affiliation
861 (i.e. weighted degree), the estimated networks were overall robust regardless of observational
862 sampling method. Global measures of association (i.e. eigenvector centrality and betweenness)
863 were more sensitive to insufficient sampling across the group and errors in data collection.
864 Increased sampling was necessary for the nearest neighbor methods and they produced less
865 accurate social networks when using global measures of association. For proximity network
866 studies, we found that the choice of distance between two individuals to consider them

connected, in our case 10m versus 20m, can significantly impact the effectiveness of the resulting network. We further found that networks produced using scan sampling methods were less affected by sampling errors and biases in data collection. Relatively dense networks, like that of a close social group like baboons, are more robust to errors than one might expect. Biased errors, like more often missing individuals on the periphery of the group, impacted network estimates less than random errors. Our results also suggest that misidentifying an individual can decrease the robustness of estimated social networks more than missing that individual in an observation, a result that likely holds true across different types of network studies (aggression, grooming, vocal, etc.).

Although our findings provide useful guidelines for estimating how much data is required to generate robust social networks using different observation methods, we also highlight some additional considerations that we could not address using our data. The primary caveat to our findings is that relationships, and thus networks, can change over time. We captured data only during 14 days, during which no major changes in social relationships were likely to have occurred (e.g. the dominant male remained the same throughout this period). Our simulated data collection is also very intensive relative to most field studies (we mimicked 3 observers who never had lunch and required no bathroom breaks). Acquiring this many focal observations may take several months in the field, and if any changes in the social structure of the study group occurred during this time, then the robustness of the resulting network could be lower than our estimates. For example, although in our simulations we controlled for the drop-out in some GPS collars over the 14 days (by calculating dyadic association rates as a function of the time the collars collected data on both individuals simultaneously), we still found a weak declining

correlation between estimated and complete networks as more data were collected, matching the time when the first wave of collars failed. This result allowed us to explore how realistic changes in the network (i.e. the disappearance of a few individuals) influence our ability to create robust network estimates. Importantly, a number of biological and social processes induce changes in social networks over time. At long time scales, individuals are born, age into social networks, emigrate, and die. Over shorter time scales, individuals' social bonds may evolve or change (like that of grooming networks). At even shorter time scales, the state of an individual may influence their association and interactions patterns (e.g. a hungry individual may groom less than a satiated individual who can devote more time to socializing). Researchers should thus consider the biological process(es) that they are interested in to inform the study duration and sampling intensity. For instance, our 2-week study would probably be insufficient for a study on how the development of social relationships impacts proximity associations. We note that it is not the changes in the system that cause inaccurate networks, but rather our limited ability to detect and account for those changes. Researchers may further consider segmenting longitudinal data into time windows to check for significant changes in the network structure.

Social network analysis provides a shared and relatively well-agreed-upon methodology for studying social behavior and its consequences. In this paper, we explore how the sampling decisions and potential sources of observational error impact the ability to construct meaningful social networks based on observational data. We conclude that robust quantification of networks, particularly at the local neighborhood scale (using metrics such as weighted degree) but also across entire networks (e.g. using eigenvector centrality or betweenness), is an achievable goal. Our simulations provide useful insights into how we can improve our decisions about what kind

of definitions to use when collecting data and generating networks, notably on how to decide on a distance threshold for defining proximity. However, our simulations also highlight a potential trade-off between collecting more samples and extending the observation period into times when social structure changes. We therefore emphasize that the resolution of data collection and the number of samples collected are two distinct considerations. We encourage further investigations, and even more modeling, to continue to test the methods we use and improve best-practices in the field.

Acknowledgements

We thank Kenya National Science and Technology Council, Kenyan Wildlife Service and Mpala Research Centre for permission to conduct research. We thank M. Wikelski, E. Bermingham, D. Rubenstein and M. Kinnaird for logistical support; R. Kays, S. Murray, M. Mutinda, R. Lessnau, S. Alavi, J. Nairobi, F. Kuemmeth, and W. Heidrich for assistance; and Gisela Kopp and the Farine Lab for comments on various drafts of the paper. This study was funded by the NSF (EAGER-IOS-1250895; IIS- 1514174). Data collection was supported by funding from the Max Planck Institute for Ornithology, the Smithsonian Tropical Research Institute, and Princeton University. DRF received additional support from BBSRC (BB/L006081/1 to B.C. Sheldon). GHD received additional support from a UC Davis Provost's fellowship. Data are deposited at [www.datarepository.movebank.org/](http://www.datarepository.movebank.org/(doi:10.5441/001/1.kn0816jn))(doi:10.5441/001/1.kn0816jn).

References

- Adelman, J. S., Moyers, S. C., Farine, D. R., & Hawley, D. M. (2015). Feeder use predicts both acquisition and transmission of a contagious pathogen in a North American songbird. *Proceedings of the Royal Society B-Biological Sciences*, 282(1815). doi: 10.1098/rspb.2015.1429
- Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013). Network-based diffusion analysis reveals cultural transmission of lobtail feeding in humpback whales. *Science*, 340(6131), 485-488. doi: 10.1126/science.1231976
- Altmann, J. (1974). Observational study of behavior- sampling methods. [Article]. *Behaviour*, 49(3-4), 227-267. doi: 10.1163/156853974x00534
- Ancillotto, L., Serangeli, M. T., & Russo, D. (2012). Spatial Proximity between Newborns Influences the Development of Social Relationships in Bats. *Ethology*, 118(4), 331-340. doi: 10.1111/j.1439-0310.2011.02016.x
- Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, 518(7540), 538-541. doi: 10.1038/nature13998
- Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cole, E. F., Cockburn, A., & Sheldon, B. C. (2013). Individual personalities predict social behaviour in wild networks of great tits (*Parus major*). *Ecology Letters*, 16(11), 1365-1372. doi: 10.1111/ele.12181
- Aplin, L. M., Farine, D. R., Morand-Ferron, J., & Sheldon, B. C. (2012). Social networks predict patch discovery in a wild population of songbirds. *Proceedings of the Royal Society B-Biological Sciences*, 279(1745), 4199-4205. doi: 10.1098/rspb.2012.1591
- Boogert, N. J., Farine, D. R., & Spencer, K. A. (2014). Developmental stress predicts social network position. *Biology Letters*, 10(10). doi: 10.1098/rsbl.2014.0561
- Brent, L. J. N. (2015). Friends of friends: are indirect connections in social networks important to animal behaviour? *Animal Behaviour*, 103, 211-222. doi: 10.1016/j.anbehav.2015.01.020
- Brent, L. J. N., Heilbronner, S. R., Horvath, J. E., Gonzalez-Martinez, J., Ruiz-Lambides, A., Robinson, A. G., . . . Platt, M. L. (2013). Genetic origins of social networks in rhesus macaques. *Scientific Reports*, 3. doi: 10.1038/srep01042
- Brown, C. R. (1986). Cliff swallow colonies as information centers. *Science*, 234(4772), 83-85. doi: 10.1126/science.234.4772.83
- Carter, A. J., Lee, A. E. G., & Marshall, H. H. (2015). Research questions should drive edge definitions in social network studies. *Animal Behaviour*, 104, e7-e11. doi: <http://dx.doi.org/10.1016/j.anbehav.2015.03.020>
- Castles, M., Heinsohn, R., Marshall, H. H., Lee, A. E. G., Cowlshaw, G., & Carter, A. J. (2014). Social networks created with different techniques are not comparable. *Animal Behaviour*, 96, 59-67. doi: 10.1016/j.anbehav.2014.07.023
- Cheney, D. L., Silk, J. B., & Seyfarth, R. M. (2016). Network connections, dyadic bonds and fitness in wild female baboons. *Royal Society Open Science*, 3(7). doi: 10.1098/rsos.160255
- Clark, F. E. (2011). Space to choose: network analysis of social preferences in a captive chimpanzee community, and implications for management. *Am J Primatol*, 73(8), 748-757. doi: 10.1002/ajp.20903

- Couzin, I. D., James, R., Mawdsley, D., Croft, D. P., & Krause, J. (2007). Social Organization and Information Transfer in Schooling Fishes. In C. Brown, K. Laland & J. Krause (Eds.), *Fish Cognition and Behavior* (pp. 166-185): Blackwell Publishing Ltd.
- Crofoot, M. C., Kays, R. W., & Wikelski, M. (2015). Data from: Shared decision-making drives collective movement in wild baboons: Movebank data repository.
- Crofoot, M. C., Rubenstein, D. I., Maiya, A. S., & Berger-Wolf, T. Y. (2011a). Aggression, grooming and group-level cooperation in white-faced capuchins (*Cebus capucinus*): insights from social networks. *Am J Primatol*, 73(8), 821-833. doi: 10.1002/ajp.20959
- Crofoot, M. C., Rubenstein, D. I., Maiya, A. S., & Berger-Wolf, T. Y. (2011b). Aggression, Grooming and Group-Level Cooperation in White-Faced Capuchins (*Cebus capucinus*): Insights From Social Networks. *American Journal of Primatology*, 73(8), 821-833. doi: 10.1002/ajp.20959
- Croft, D. P., James, R., Thomas, P. O. R., Hathaway, C., Mawdsley, D., Laland, K. N., & Krause, J. (2006). Social structure and co-operative interactions in a wild population of guppies (*Poecilia reticulata*). *Behavioral Ecology and Sociobiology*, 59(5), 644-650. doi: 10.1007/s00265-005-0091-y
- Cross, P. C., Lloyd-Smith, J. O., Bowers, J. A., Hay, C. T., Hofmeyr, M., & Getz, W. M. (2004). Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. *Annales Zoologici Fennici*, 41(6), 879-892. doi: 10.2307/23736148
- Dubosq, J., Romano, V., Sueur, C., & MacIntosh, A. J. J. (2016). Network centrality and seasonality interact to predict lice load in a social primate. *Scientific Reports*, 6. doi: 10.1038/srep22095
- Dufour, V., Sueur, C., Whiten, A., & Buchanan-Smith, H. M. (2011). The Impact of Moving to a Novel Environment on Social Networks, Activity and Wellbeing in Two New World Primates. *American Journal of Primatology*, 73(8), 802-811. doi: 10.1002/ajp.20943
- Farine, D. R. (2014). Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Animal Behaviour*, 89, 141-153. doi: 10.1016/j.anbehav.2014.01.001
- Farine, D. R. (2015). Proximity as a proxy for interactions: issues of scale in social network analysis. *Animal Behaviour*, 104(0), e1-e5. doi: <http://dx.doi.org/10.1016/j.anbehav.2014.11.019>
- Farine, D. R. (2018). When to choose dynamic vs. static social network analysis. *Journal of Animal Ecology*, 87(1), 128-138. doi: 10.1111/1365-2656.12764
- Farine, D. R., Aplin, L. M., Sheldon, B. C., & Hoppitt, W. (2015). Interspecific social networks promote information transmission in wild songbirds. [Article]. *Proceedings of the Royal Society B-Biological Sciences*, 282(1803), 9. doi: 10.1098/rspb.2014.2804
- Farine, D. R., & Milburn, P. J. (2013). Social organisation of thornbill-dominated mixed-species flocks using social network analysis. *Behavioral Ecology and Sociobiology*, 67(2), 321-330. doi: 10.1007/s00265-012-1452-y
- Farine, D. R., & Sheldon, B. C. (2015). Selection for territory acquisition is modulated by social network structure in a wild songbird. *Journal of Evolutionary Biology*, 28(3), 547-556. doi: 10.1111/jeb.12587
- Farine, D. R., & Strandburg-Peshkin, A. (2015). Estimating uncertainty and reliability of social network data using Bayesian inference. *Royal Society Open Science*, 2(9). doi: 10.1098/rsos.150367

- Farine, D. R., Strandburg-Peshkin, A., Berger-Wolf, T., Ziebart, B., Brugere, I., Li, J., & Crofoot, M. C. (2016). Both Nearest Neighbours and Long-term Affiliates Predict Individual Locations During Collective Movement in Wild Baboons. *Scientific Reports*, 6. doi: 10.1038/srep27704
- Farine, D. R., Strandburg-Peshkin, A., Couzin, I. D., Berger-Wolf, T. Y., & Crofoot, M. C. (2017). Individual variation in local interaction rules can explain emergent patterns of spatial organization in wild baboons. *Proceedings of the Royal Society B-Biological Sciences*, 284(1853). doi: 10.1098/rspb.2016.2243
- Farine, D. R., & Whitehead, H. (2015). Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84(5), 1144-1163. doi: 10.1111/1365-2656.12418
- Feczko, E., Mitchell, T. A. J., Walum, H., Brooks, J. M., Heitz, T. R., Young, L. J., & Parr, L. A. (2015). Establishing the reliability of rhesus macaque social network assessment from video observations. *Animal Behaviour*, 107, 115-123. doi: 10.1016/j.anbehav.2015.05.014
- Formica, V. A., McGlothlin, J. W., Wood, C. W., Augat, M. E., Butterfield, R. E., Barnard, M. E., & Brodie, E. D. (2011). Phenotypic Assortment Mediates the Effect of Social Selection in a Wild Beetle Population. *Evolution*, 65(10), 2771-2781. doi: 10.1111/j.1558-5646.2011.01340.x
- Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337. doi: 10.1136/bmj.a2338
- Fragaszy, D. M., Boinski, S., & Whipple, J. (1992). Behavioral sampling in the field: Comparison of individual and group sampling methods. *American Journal of Primatology*, 26(4), 259-275. doi: 10.1002/ajp.1350260404
- Franks, D. W., Ruxton, G. D., & James, R. (2010). Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology*, 64(3), 493-503. doi: 10.1007/s00265-009-0865-8
- Fraser, O. N., Schino, G., & Aureli, F. (2008). Components of Relationship Quality in Chimpanzees. *Ethology*, 114(9), 834-843. doi: 10.1111/j.1439-0310.2008.01527.x
- Furuichi, T. (1983). Interindividual distance and influence of dominance on feeding in a natural Japanese macaque troop. [journal article]. *Primates*, 24(4), 445-455. doi: 10.1007/bf02381678
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360-1380.
- Henzi, S. P., Lusseau, D., Weingrill, T., van Schaik, C. P., & Barrett, L. (2009). Cyclicality in the structure of female baboon social networks. *Behavioral Ecology and Sociobiology*, 63(7), 1015-1021. doi: 10.1007/s00265-009-0720-y
- Henzi, S. P., Lycett, J. E., & Weingrill, T. (1997). Cohort size and the allocation of social effort by female mountain baboons. *Animal Behaviour*, 54, 1235-1243. doi: DOI 10.1006/anbe.1997.0520
- James, R., Croft, D. P., & Krause, J. (2009). Potential banana skins in animal social network analysis. *Behavioral Ecology and Sociobiology*, 63(7), 989-997. doi: 10.1007/s00265-009-0742-5
- King, A. J., Sueur, C., Huchard, E., & Cowlshaw, G. (2011). A rule-of-thumb based on social affiliation explains collective movements in desert baboons. *Animal Behaviour*, 82(6), 1337-1345. doi: 10.1016/j.anbehav.2011.09.017

- Lehmann, J., & Boesch, C. (2009). Sociality of the dispersing sex: the nature of social bonds in West African female chimpanzees, Pan troglodytes. *Animal Behaviour*, 77(2), 377-387. doi: <http://dx.doi.org/10.1016/j.anbehav.2008.09.038>
- Matthews, L. (2009). Intragroup behavioral variation in white-fronted capuchin monkeys (Cebus albifrons): mixed evidence for social learning inferred from new and established analytical methods. *Behaviour*, 146(3), 295-324. doi: doi:10.1163/156853909X410937
- Nakamichi, M., & Koyama, N. (1997). Social Relationships Among Ring-Tailed Lemurs (Lemur catta) in Two Free-Ranging Troops at Berenty Reserve, Madagascar. *International Journal of Primatology*, 18(1), 73-93. doi: 10.1023/A:1026393223883
- Oh, K. P., & Badyaev, A. V. (2010). Structure of Social Networks in a Passerine Bird: Consequences for Sexual Selection and the Evolution of Mating Strategies. *American Naturalist*, 176(3), E80-E89. doi: 10.1086/655216
- Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092), 502-505. doi: 10.1038/nature04605
- Perry, S., & Manson, J. (2008). *Manipulative Monkeys: The Capuchins of Lomas Barbudal*. Cambridge: Harvard University Press.
- R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Ramos-Fernández, G., Boyer, D., Aureli, F., & Vick, L. (2009). Association networks in spider monkeys (Ateles geoffroyi). *Behavioral Ecology and Sociobiology*, 63(7), 999-1013. doi: 10.1007/s00265-009-0719-4
- Rimbach, R., Bisanzio, D., Galvis, N., Link, A., Di Fiore, A., & Gillespie, T. R. (2015). Brown spider monkeys (Ateles hybridus): a model for differentiating the role of social networks and physical contact on parasite transmission dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1669). doi: 10.1098/rstb.2014.0110
- Ryder, T. B., McDonald, D. B., Blake, J. G., Parker, P. G., & Loiselle, B. A. (2008). Social networks in the lek-mating wire-tailed manakin (Pipra filicauda). *Proceedings of the Royal Society B-Biological Sciences*, 275(1641), 1367-1374. doi: 10.1098/rspb.2008.0205
- Schlicht, L., Valcu, M., & Kempenaers, B. (2015). Spatial patterns of extra-pair paternity: beyond paternity gains and losses. *Journal of Animal Ecology*, 84(2), 518-531. doi: 10.1111/1365-2656.12293
- Scott, J. (1991). *Social Network Analysis: A Handbook*. London, Newbury Park, CA: SAGE Publications.
- Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science*, 302(5648), 1231-1234. doi: DOI 10.1126/science.1088580
- Silk, J. B., Altmann, J., & Alberts, S. C. (2006). Social relationships among adult female baboons (papio cynocephalus) I. Variation in the strength of social bonds. *Behavioral Ecology and Sociobiology*, 61(2), 183-195. doi: 10.1007/s00265-006-0249-2
- Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., . . . Cheney, D. L. (2010). Strong and Consistent Social Bonds Enhance the Longevity of Female Baboons. *Current Biology*, 20(15), 1359-1361. doi: 10.1016/j.cub.2010.05.067

- Silk, J. B., Seyfarth, R. M., & Cheney, D. L. (1999). The structure of social relationships among female savanna baboons in Moremi Reserve, Botswana. *Behaviour*, 136, 679-703. doi: 10.1163/156853999501522
- Silk, M. J., Jackson, A. L., Croft, D. P., Colhoun, K., & Bearhop, S. (2015). The consequences of unidentifiable individuals for the analysis of an animal social network. *Animal Behaviour*, 104, 1-11. doi: 10.1016/j.anbehav.2015.03.005
- Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, 348(6241), 1358-1361. doi: 10.1126/science.aaa5099
- Suzuki, M., & Sugiura, H. (2011). Effects of Proximity and Activity on Visual and Auditory Monitoring in Wild Japanese Macaques. *American Journal of Primatology*, 73(7), 623-631. doi: 10.1002/ajp.20937
- Szykman, M., Engh, A. L., Van Horn, R. C., Funk, S. M., Scribner, K. T., & Holekamp, K. E. (2001). Association patterns among male and female spotted hyenas (*Crocuta crocuta*) reflect male mate choice. *Behavioral Ecology and Sociobiology*, 50(3), 231-238. doi: DOI 10.1007/s002650100356
- Valente, T. W. (1995). *Network models of the diffusion of innovations* (Vol. 2): Hampton Press Cresskill, NJ.
- VanderWaal, K. L., Atwill, E. R., Isbell, L. A., & McCowan, B. (2014). Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*). *Journal of Animal Ecology*, 83(2), 406-414. doi: 10.1111/1365-2656.12137
- Viscido, S. V., & Shrestha, S. (2015). Using quantitative methods of determining group membership to draw biological conclusions. *Animal Behaviour*, 104, 145-154. doi: <http://dx.doi.org/10.1016/j.anbehav.2015.03.007>
- Voelkl, B., Kasper, C., & Schwab, C. (2011). Network Measures for Dyadic Interactions: Stability and Reliability. *American Journal of Primatology*, 73(8), 731-740. doi: 10.1002/ajp.20945
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge university press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. [10.1038/30918]. *Nature*, 393(6684), 440-442.
- Watts, D. P. (1992). Social relationships of immigrant and resident female mountain gorillas. I. Male-female relationships. *American Journal of Primatology*, 28(3), 159-181. doi: 10.1002/ajp.1350280302
- Wey, T. W., Blumstein, D. T., Shen, W., & Jordan, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour*, 75, 333-344. doi: 10.1016/j.anbehav.2007.06.020
- Wey, T. W., Burger, J. R., Ebensperger, L. A., & Hayes, L. D. (2013). Reproductive correlates of social network variation in plurally breeding degus (*Octodon degus*). *Animal Behaviour*, 85(6), 1407-1414. doi: 10.1016/j.anbehav.2013.03.035
- Whitehead, H. (2008). *Analyzing animal societies: quantitative methods for vertebrate social analysis*. Chicago: University of Chicago Press.
- Wolf, J. B. W., Mawdsley, D., Trillmich, F., & James, R. (2007). Social structure in a colonial mammal: unravelling hidden structural layers and their foundations by network analysis.

1162 *Animal Behaviour*, 74(5), 1293-1302. doi:
1163 <http://dx.doi.org/10.1016/j.anbehav.2007.02.024>
1164