Published in final edited form as:

Anim Behav. 2015 September 1; 107: 115–123. doi:10.1016/j.anbehav.2015.05.014.

Establishing the reliability of rhesus macaque social network assessment from video observations

Eric Feczko^{a,b,c,*}, Thomas A. J. Mitchell^{a,b,c}, Hasse Walum^{a,b,c}, Jenna M. Brooks^{a,b,c}, Thomas R. Heitz^{a,b,c}, Larry J. Young^{a,b,c}, and Lisa A. Parr^{a,b,c}

^a Department of Psychiatry and Behavioral Science, Emory University School of Medicine, Atlanta, GA, U.S.A.

^b Yerkes Primate National Research Center, Emory University, Atlanta, GA, U.S.A.

^c Silvio O. Conte Center for Oxytocin and Social Cognition, Center for Translational Social Neuroscience, Emory University, Atlanta, GA, U.S.A.

Abstract

Understanding the properties of a social environment is important for understanding the dynamics of social relationships. Understanding such dynamics is relevant for multiple fields, ranging from animal behaviour to social and cognitive neuroscience. To quantify social environment properties, recent studies have incorporated social network analysis. Social network analysis quantifies both the global and local properties of a social environment, such as social network efficiency and the roles played by specific individuals, respectively. Despite the plethora of studies incorporating social network analysis, methods to determine the amount of data necessary to derive reliable social networks are still being developed. Determining the amount of data necessary for a reliable network is critical for measuring changes in the social environment, for example following an experimental manipulation, and therefore may be critical for using social network analysis to statistically assess social behaviour. In this paper, we extend methods for measuring error in acquired data and for determining the amount of data necessary to generate reliable social networks. We derived social networks from a group of 10 male rhesus macaques, *Macaca mulatta*, for three behaviours: spatial proximity, grooming and mounting. Behaviours were coded using a video observation technique, where video cameras recorded the compound where the 10 macaques resided. We collected, coded and used 10 h of video data to construct these networks. Using the methods described here, we found in our data that 1 h of spatial proximity observations produced reliable social networks. However, this may not be true for other studies due to differences in data acquisition. Our results have broad implications for measuring and predicting the amount of error in any social network, regardless of species.

* Correspondence: E. Feczko, Yerkes Primate National Research Center, Emory University, 954 Gatewood Rd NE, Atlanta, GA 30329, U.S.A. efeczko@emory.edu (E. Feczko)..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

nonhuman primate; reliability; rhesus macaque; social behaviour; social network

The social environment of animals shapes the effects of exogenous factors on animal health and behaviour. For example, conspecific social organization may influence how chimpanzees, *Pan troglodytes* (Hobaiter, Poisot, Zuberbuhler, Hoppitt, & Gruber, 2014) or dolphins (*Tursiops aduncus*; Mann, Stanton, Patterson, Bienenstock, & Singh, 2012) learn from other individuals within the group. Human social environments can modulate the spread of infectious diseases (Lindquist, Ma, van den Driessche, & Willeboordse, 2011) and improve social skills of autistic individuals (Kasari, Rotheram-Fuller, Locke, & Gulsrud, 2012). Social support also influences obesity rates in humans (Sawka, McCormack, Nettel-Aguirre, Hawe, & Doyle-Baker, 2013) and physical fitness in dolphins (Frere et al., 2010). The social rank of individual primates modulates vulnerability to stress and disease (Czoty, Gould, & Nader, 2009; Qin et al., 2013).

In turn, the social environment is shaped by the behaviours of individual animals. Personality traits of individual male rhesus macaques, *Macaca mulatta* (McCowan et al., 2011), as well as their alliances with other males (Beisner, Jackson, Cameron, & McCowan, 2011) can influence the stability of the society in which they live. Environmental factors can alter the roles of individual mice within a group, which in turn alter the social structure (Shemesh et al., 2013). Dolphin social organization depends on environmental and demographic factors (Blasi & Boitani, 2014). Collective behaviour of schooling fish both influences and is influenced by individuals' decisions (Miller, Garnier, Hartnett, & Couzin, 2013; Sumpter, Krause, James, Couzin, & Ward, 2008).

Taken together, these findings suggest that studying the global and local properties of the social environment, such as the stability or efficiency (Watts & Strogatz, 1998) of a social group, and the roles played by individuals within the social environment, respectively (Guimerà, Mossa, Turtschi, & Amaral, 2005; D. Lusseau & Newman, 2004), may lead to insights across multiple domains, from animal health and behaviour to cognitive and social neuroscience. Measurements of the global and local properties of social environments are often referred to as social network analysis. At its core, social network analysis measures the properties of graphs composed of nodes and edges (Newman, 2010). In the context of social networks, the nodes represent individuals while the edges represent dyadic interactions between them (Wasserman & Faust, 1994). Different types of dyadic behaviours may reflect different properties of the social environment, or even separate, but connected, social networks. For example, a network composed of displacements, which is an agonistic behaviour occurring between two primates (i.e. a dyad), would be an agonistic social network (Beisner et al., 2011). On the other hand, edges within a social network could represent spatial proximity durations, which, for the purposes of this paper, are defined as the amount of time that two primates spend in proximity to one another. Such a network may reflect social tolerance (Brent, Maclarnon, Platt, & Semple, 2013) or community associations (Haddadi et al., 2011), the properties of which may differ from an agonistic network. Studies that incorporate social network analysis may provide greater insight into

social organization, which influences overall health and rates of disease. In and of itself, these studies are also critical for improving the management of colonies, because their findings may help to develop strategies that reduce rates of disease and wounding (Makagon, McCowan, & Mench, 2012). Quantifying social networks may be useful in exploring manipulations that modulate social cognition and behaviour in preclinical studies of human and nonhuman primates, for example, to examine pharmacological approaches that may improve social function in disorders such as autism.

Prior studies of nonhuman primate social networks have measured static, but not dynamic, social networks derived from dyadic behaviours, frequently combining hundreds of hours of observation collected over a period of several months into a single network (Beisner et al., 2011; Brent, Heilbronner, et al., 2013; Brent, Lehmann, & Ramos-Fernandez, 2011; Brent, Maclarnon, et al., 2013; Brent, Semple, Dubuc, Heistermann, & Maclarnon, 2011; Hobaiter et al., 2014; McCowan et al., 2011; Qin et al., 2013). Such approaches limit the ability to measure changes in the social environment, which may reflect changes in exogenous factors (Perreault, 2010), throughout that period. For example, a recent study by Brent, Maclarnon, et al. (2013) suggests that rhesus macaque social environments change seasonally; earlier work investigating how rhesus macaques form associations demonstrates changes between birth, interim and mating seasons (Altmann, 1967). Measuring such dynamics may be important in determining the effect of a treatment (e.g. pharmacological intervention or removal of an individual from the group) on the social environment, or whether certain individuals may be more receptive to a treatment than others, depending upon each individual's role. For example, a pharmacological intervention to improve social communication in autism may cause an individual to become more of a social hub within the network. While social network dynamics has been an issue for studying primate social networks, the same issues apply for studying any social network.

In the present paper, we provide methods for estimating the reliability of dyadic behaviours in order to derive the dynamic changes in social networks. Before one can accurately measure the dynamics of a social network at the highest temporal resolution, one must first determine the fewest number of behaviours, and therefore the amount of observation time, required to measure a reliable social network. After acquiring data to produce a social network, one must then determine the amount of error associated with the duration of observation periods. Prior studies have validated multiple strategies for measuring the amount of error in the data acquired or for predicting the amount of error one may expect. Some studies have examined randomization of individual dyadic associations (Croft, Madden, Franks, & James, 2011; James, Croft, & Krause, 2009; Sundaresan, Fischhoff, & Dushoff, 2009; Voelkl, Kasper, & Schwab, 2011; Whitehead, Bejder, & Andrea Ottensmeyer, 2005) or compared the observed network to modelled networks (Handcock & Gile, 2010) to measure error at the level of individual associations. Others have measured the error of network properties, by using bootstrapping to establish confidence intervals for observed network metrics (Lusseau, Whitehead, & Gero, 2008), to determine the sampling rate required to capture variation in communities (Haddadi et al., 2011) or to examine the effect of sampling rate on network metrics from artificial networks (Perreault, 2010).

Here we extend these methods by measuring the amount of error observed across the entire social structure, as opposed to individual associations or network metrics. This resembles a study that used simulated data to examine how much data is required to achieve a specific correlation between estimated and true social structure (Whitehead, 2008). However, our methods use a similarity metric that captures variation in the magnitudes of associations, and we simulate the observation period from resampled data instead of randomizing individual dyadic associations. Taken together, these methods further one's ability to measure the amount of error in acquired data and to predict the amount of error one may expect in future observations. We chose to examine the amount of error as a function of the observation time for several behaviours, because behaviour varies in its frequency of expression. Some behaviours, like spatial proximity, may occur frequently, whereas others, such as bared-teeth displays, may occur more infrequently. We used a similarity statistic: η^2 to determine the proportion of variance that one network explains in another (Keppel, 1991). We measured network similarities using three different analyses of data from 10 monkeys that were simultaneously recorded from video. Importantly, the described techniques extend prior studies of social network error and reliability and can be applied to any number of data sets for measuring the amount of error in acquired data and for predicting the amount of data necessary to construct a reliable social network.

METHODS

Subjects

The subjects of this study were 10 male rhesus macaques (average age 3.5 years) born into large social groups (over 100 individuals) at the Yerkes Field Station in Lawrenceville, Georgia, U.S.A. These subjects were mother-reared for the first three years of life, after which they were removed from their respective social groups and placed into peer groups to reduce aggression and inbreeding and to mimic male migration typical of this species. Some of these subjects were familiar with each other prior to peer housing (e.g. two separate groups of three individuals, where each group comprised individuals born into the same social group) whereas the remaining individuals were previously unfamiliar with one another. Individuals were given unique dye markings prior to the start of the study to aid in identification.

The peer enclosure comprised an indoor–outside run to which subjects had full access during data collection (Fig. 1). The outside run included several perches along the walls as well as swings and enrichment devices. The inside area was approximately one-third the size of the outside area and consisted of two interior rooms divided by a concrete wall with an animal access door in between. Subjects spent most of their time in the outside run during the video-recording sessions.

To simultaneously record all behaviours from these 10 subjects, observation sessions were videotaped using a Lorex (Edge 2) Digital Video Surveillance Recorder (L-DVR) and four wireless indoor–outdoor recording cameras (Lorex LW2281) (<http://www.lorextechnology.com/>). Each camera was mounted in a corner of the outdoor enclosure (e.g. Fig. 1c) inside a metal box with a clear Lexan front to protect the cameras from monkey tampering while providing a clear view of enclosure. The four cameras were

positioned towards the centre of the enclosure and provided clear views along the length of the outdoor enclosure. The combined camera views ensured that there were no blind spots, and each angle overlapped with the others to offer multiple viewing angles and optimize visualization of subjects regardless of their position in the enclosure. The L-DVR recorder captured footage from these cameras in separate video files and saved them to a one-terabyte hard drive that could provide over 1 month of continuous recording. This hard drive was placed in a compartment outside the enclosure, allowing researchers easy access to establish recording sequences and copy data files with minimal disturbance to the animals.

Observation Schedule and Coding

Video recordings were made of the group over a 6-week span during the months of November and December of 2013. For these 6 weeks, recordings were programmed to occur during a 'Do Not Disturb' period that was enacted twice per week between 1100 and 1300 hours to minimize human disturbance during the observational periods. We then selected 10 one-hour observations between 1130 and 1230 hours from these files to be coded for the present reliability analyses. The 10 observations were spaced as evenly as possible across the 6-week period, accounting for poor weather, unforeseen disturbances (e.g. compound work or human interference) where the resulting observations occurred at an average span of 4.6 days apart. All individuals were present in the social group during the recording period.

The video footage was saved from the hard drive and converted to Windows Media format for behavioural coding in Noldus Observer XT10.1 (<http://www.noldus.com/>). All footage was coded by a single coder (A.J.M.) who could reliably identify all animals and their behaviour from the video footage. We created an ethogram of specific behaviours, including spatial proximity, grooming, play (solo, chase and wrestle) and mounting, using both duration (in seconds) and point events (frequencies). Spatial proximity was coded as an interaction when two subjects sat within arm's reach of one another. An unknown category was included if an animal went inside during a session, which occurred rarely. Only spatial proximity, grooming and mounting behaviours were analysed for this study.

Data Analysis

The development of adjacency matrices—We extracted each set of coded data for each hour of observation from Observer. We then converted spatial proximity, grooming and mounting behaviours into discrete adjacency matrices using in-house scripts via MATLAB 2012a (<http://www.mathworks.com>; requests for the individual scripts can be sent to the corresponding author). Each adjacency matrix comprised a 10 × 10 grid, where the rows and columns reflect the identity of the subjects and each cell within the matrix represents the duration (grooming or contact) or frequency (mounting) of the behaviour that occurred between every subject dyad. For both grooming and mounting behaviours, the matrices reflected mean duration of behaviours in seconds. These matrices were asymmetrical across cells or dyads, where the rows represented the groomers or the mounters and the columns represented the recipient of these behaviours. Thus, these matrices contained 90 cells representing the behaviour between every dyadic combination of subjects. For spatial proximity, the matrix was symmetrical since two monkeys were in contact with each other by definition. Thus, this matrix contained only 45 cells or dyads of

behaviour. We derived the overall average networks for each behaviour by taking the average of each of the 10 one-hour adjacency matrices. Therefore, the average networks were derived from the entire 10 h of observation.

We then divided these average matrices in half (to be used for a split-half reliability analysis; see below) by splitting each hour of observation into 'even' and 'odd' minutes, one reflecting the behaviours observed during the odd minutes and the other reflecting the behaviours observed during the even minutes. Because differences in the duration of behaviours may influence the similarity estimates from a split-half analysis, we also measured the mean duration of grooming and spatial proximity bouts to ensure that high similarity scores were not driven by differences in the duration of behaviours.

Similarity measures—We assessed similarity in the behaviours observed across each of the 10 one-hour observations by comparing measures of η^2 between these matrices. These comparisons were done separately for each behaviour. η^2 provides a measure of the similarity between two data sets (Keppel, 1991). Specifically, $\eta^2 = \text{SSB}/\text{SST}$, where SSB reflects the between-sample sum of squares and SST reflects the total sum of squares. A value of one indicates that 100% of the variance in one data set is explained by a separate data set, while a value of 0% indicates that the information from one data set cannot be explained by the other. We chose to use η^2 as the similarity metric because, unlike a Pearson's correlation coefficient, η^2 takes into account differences in the magnitude of associations between matrices. This helps reduce overestimation of similarity for the split-half analysis, because η^2 will be reduced as the duration of a behaviour becomes less evenly divided between even and odd minutes. However, we caution that using η^2 does not completely eliminate potential confounds from the split-half analysis. We used two approaches to measure the amount of error in the social networks derived from behaviours via η^2 . We used a third approach to provide some a priori predictions regarding the amount of error one can expect in a social network collected from this group.

(1) Are social networks derived from behaviour that is reliable within 1 h observation sessions?:

To determine the amount of error from a social network derived from a single hour of video observation, we conducted a split-half reliability analysis. We measured η^2 values between the odd and even adjacency matrices for spatial proximity, mounting and grooming separately. This produced a set of 10 within-session η^2 values per behaviour. We used a one-sample two-tailed t test against an η^2 value of 0.7 to determine whether within-session η^2 values were significantly reliable. The η^2 value of 0.7 was chosen arbitrarily because it indicates that at least 70% of the variance in one network is explained by the other network; such a value would be roughly equivalent to a Pearson correlation coefficient of 0.84. P values were Bonferroni-corrected for the three assessed behaviours (corrected P value = 0.0167) to maintain an alpha of 0.05.

We then conducted a permutation analysis separately for each behaviour to ensure that the split-half reliability findings were not confounded or biased by how the data were split (e.g. by dividing the average matrices into odd and even minutes), because the duration of specific behaviours may influence the η^2 values observed. We conducted 10 000 permutations on the matrices derived per 1 h observation session. We then randomly

assigned each minute of data per permutation to one of two groups without replacement. Adjacency matrices were derived from each group for spatial proximity, grooming and mounting behaviours. For each behaviour, we then calculated η^2 values between the two randomly created matrices. We computed the average η^2 value across the permutations for each 1 h session and used a one-sample two-tailed t test against an η^2 value of 0.7 to determine whether within-session η^2 values were significantly reliable. P values were Bonferroni-corrected for the three comparisons (corrected P value = 0.0167).

(2) Is each hour of behaviour similar to the average matrix?: To determine whether 1 h of recorded behaviour reflected the social network derived from 10 h of observation, we calculated η^2 values between each 1 h observation session and for the overall network averaged over 9 h of data, excluding the 1 h used for comparison. We used a one-sample two-tailed t test against an η^2 value of 0.7 to determine whether networks derived from 1 h of observation were significantly reliable. These tests were conducted separately for the spatial proximity, grooming and mounting networks. P values were Bonferroni-corrected for the three comparisons (corrected P value = 0.0167).

(3) How much data is required to produce reliable social networks?: To determine the amount of error one may predict in a social network derived from this group, we used a bootstrapping technique to compare simulated networks to the average network. Such a method provides some a priori predictions for future studies of social networks on the same group. Briefly, we divided each 1 h session into 30 min blocks and sampled sets of blocks with replacement, then averaged them to produce a simulated network. The simulated networks ranged from 30 min to 10 h and did not necessarily reflect the observed network. For example, a single bootstrap sample for 2 h of data comprised four 30 min sessions, which were randomly selected and averaged together; the same 30 min session may be selected multiple times within a single bootstrap. The η^2 value was computed for the bootstrapped adjacency matrix relative to the average matrix derived from 10 h of observation. We calculated 10 000 bootstraps for each time block to construct 95% confidence intervals for the η^2 values per block. The intervals allowed us to evaluate the amount of error we might expect from a given amount of data collected from a similar group using video observations. Such information provides an a priori justification for how much data should be acquired in future studies, although one should be careful not to assume that such a justification may hold true for all social groups. As with prior analyses, these intervals were computed separately for networks derived from spatial proximity, grooming and mounting.

Ethical Note—All procedures adhered to ASAB/ABS Guidelines for the use of animals in research and were approved by the Institutional Animal Care and Use Committee of Emory University (protocol no. YER-2002259-071016GA).

RESULTS

Within-session Similarity for Spatial Proximity, Grooming and Mounting Behaviours

Across the 10 h, different behaviours occurred at different frequencies (mounting: 160 observations; grooming: 240 observations; spatial proximity: 2306 observations). Per observation, the average spatial proximity duration was 27.7 s, whereas the average grooming duration was 30.9 s.

Figure 2 depicts the η^2 values for the split-half reliability analysis. Sessions are ordered by date and time to visualize potential progression of the social networks (i.e. whether dates closer to each other showed a more similar social structure than dates further apart). To visualize within-subject η^2 values for the three behaviours, we constructed bar graphs (Fig. 2d) from the within-subject values (Fig. 2a, b, c, along the diagonal, bordered by black lines). The last row and column of each matrix depicts the similarity between each individual session and the average matrix. η^2 values for within-session comparisons revealed that mounting social networks were not significantly similar (3 out of 10 sessions had an η^2 value greater than 0.7; $\eta^2 = 0.67 \pm 0.068$; $t_9 = -0.47$, $P = 0.65$; Fig. 2a, d). However, both grooming and spatial proximity networks demonstrated significant similarity (grooming: 7 out of 10 sessions had an η^2 value greater than 0.7; $\eta^2 = 0.87 \pm 0.048$; $t_9 = 3.5$, $P = 0.0068$; Fig. 2b, d; spatial proximity: 10 out of 10 sessions had an η^2 value greater than 0.7; $\eta^2 = 0.98 \pm 0.004$; $t_9 = 66$, $P = 2 \times 10^{-13}$; Fig. 2c, d).

Although the results of this first analysis suggest that 1 h of grooming or spatial proximity data has minimal error, it is clear from Fig. 2 that the between-session similarity was extremely high for spatial proximity (35 out of 90 comparisons; mean $\eta^2 = 0.72$), but not for grooming (6 out of 90 comparisons; mean $\eta^2 = 0.56$) or mounting (2 out of 90 comparisons; mean $\eta^2 = 0.57$) networks. Although grooming may be extremely similar within sessions, the low similarity between observation sessions suggests that social networks derived from 1 h of grooming data may have as much error as social networks derived from mounting networks.

Because grooming bout durations were long, the high within-session similarity for both the grooming and spatial proximity networks may have been driven by how the data were initially split (e.g. odd versus even minutes). Therefore, we extended the split-half reliability analysis by conducting a permutation analysis without replacement. Permuted within-session similarity remained significant for spatial proximity ($\eta^2 = 0.91 \pm 0.034$; $t_9 = 19.8$, $P = 9.9 \times 10^{-9}$), but not for mounting behaviour ($\eta^2 = 0.65 \pm 0.15$; $t_9 = -0.95$, $P = 0.36$; Fig. 3). There was a trend for the similarity of grooming networks within observation sessions ($\eta^2 = 0.78 \pm 0.11$; $t_9 = 2.37$, $P = 0.0421$; Fig. 3), but the trend was not significant after correcting for multiple comparisons. Furthermore, within-session similarity was significantly higher for spatial proximity networks than for grooming ($t_{18} = 3.4$, $P = 0.0032$) or mounting ($t_{18} = 5.19$, $P = 6.2 \times 10^{-5}$) networks. Taken together, the permutation analysis confirmed that networks derived from spatial proximity had little error after 1 h of observation time, while mounting networks did not. The permutation analysis also throws into question whether a single hour of data is sufficient to derive a social network from grooming behaviours.

Similarity of Individual Adjacency Matrices to the Average Matrix

We compared the similarity of individual 1 h observation sessions to the average network for spatial proximity, grooming and mounting to determine the amount of error observed from each network derived from 1 h of data. The mean η^2 values for each individual session relative to the average are shown in Fig. 4. Spatial proximity matrices using 1 h of data (averaged across the η^2 values calculated between each individual session and the average matrix) were significantly similar to the average matrix ($\eta^2 = 0.78 \pm 0.013$; $t_9 = 6.2$, $P = 5 \times 10^{-6}$; Fig. 4). However, grooming ($\eta^2 = 0.56 \pm 0.014$; $t_9 = -10.1$, $P = 4.7 \times 10^{-9}$) and mounting ($\eta^2 = 0.58 \pm 0.02$; $t_9 = -7.0$, $P = 1.3 \times 10^{-6}$) matrices were significantly dissimilar relative to the average matrix (Fig. 4).

Taken together with the prior analysis, the data suggest that a single hour of grooming or mounting data may have too much error to derive a reasonable social network. While grooming may be internally similar, within a 1 h observation session, it may occur too infrequently or too inconsistently across observation periods to reflect the average network produced over 9 h of observations. However, it is possible that with more data, sampled over a longer time frame than the hour-long bins analysed here, grooming may better reflect the average (9 h) network. Next, we performed bootstrap analysis (see below) to provide insight into how much data we might need to significantly reduce the error for mounting and grooming data.

Bootstrap Analyses

As more data (hours of observation) were incorporated into an individual adjacency matrix, the mounting and grooming networks appeared to become reliable, exceeding the 0.7 η^2 threshold after 3–4 h of data (Fig. 5a). It is unclear, however, whether any network derived from 3–4 h of observations for this group would be more reliable than random chance with 3–4 h of data. Therefore, the bootstrapping analysis, sampling with replacement, allowed us to measure the similarity between the average and potential networks with 95% confidence. For spatial proximity networks, the network produced by 1 h of data was significantly similar to the average network (Fig. 5b). This result is consistent with our prior analyses, suggesting that spatial proximity behaviours produce reliable social networks within only 1 h of observation. Mounting networks were significantly reliable when compared to their average network after combining at least 6 h of observation (Fig. 5c). Grooming networks were significantly reliable when compared to the average network after combining at least 4 h of observation (Fig. 5d).

DISCUSSION

Differences in network structure may arise due to insufficient data or because networks may change over time (Perreault, 2010). In this paper we provide several complementary methods for determining the amount of error observed in social networks depending on how much data is acquired, which allows one to examine, post hoc, whether a change in a social network may result from changes in social structure. Using the data provided from video observations, we show that different behaviours require different amounts of data to produce meaningful social networks for behaviour recorded simultaneously from a group of 10

monkeys. These differences may reflect variations in the quantity and frequency of behaviours and their respective distributions across dyads. Our findings complement prior literature investigating the effects of error, thresholding and small sample size on the reproducibility of dyadic associations (Croft et al., 2011; Handcock & Gile, 2010; James et al., 2009; Sundaresan et al., 2009; Whitehead, 2008; Whitehead et al., 2005), metrics representing global and local network properties (Haddadi et al., 2011; David Lusseau et al., 2008; Perreault, 2010; Voelkl et al., 2011) and social matrices. Our approach extends these methods by measuring error across an entire social structure and by predicting the amount of data necessary to derive reliable social networks. The permutation testing approach, performed without replacement, measures how much error may be present in the data acquired. The bootstrapping analysis, with replacement, estimates how much error one might expect from subsequently acquired data. It is important to emphasize that such estimates must be interpreted with caution. One should keep in mind that the results for this social group may not necessarily apply to others, because of differences in group composition, the types of behaviours measured or how the observations were recorded. For example, the data used for the present analysis were derived from a 10-member group. Larger group sizes have a greater range of potential interactions, and more observational data may be required to achieve a similar level of error. Fortunately, these methods allow others to examine the minimum amount of data necessary to measure a reliable social network in any social group.

Spatial Proximity Networks

We found that 1 h of video data was sufficient to derive a reliable social network involving spatial proximity. Spatial proximity networks showed very strong split-half reliability, suggesting that the same social network was captured within an hour-long span. Hour-long sessions were similar to the average network produced from 10 h of behaviour coded over a 6-week period, and the bootstrapping analysis showed that just 1 h of data can explain 90% of the variance in the same behaviour averaged over 10 h. We do not think that the long duration of spatial proximity explains the high similarity observed in these analyses, because the average grooming bout was even longer, but grooming networks were more dissimilar across sessions than were spatial proximity networks. In any case, the permutation test further supports our finding that 1 h of observation data is sufficient to derive meaningful spatial proximity networks in our group.

Despite the reliability of spatial proximity, care should be taken when interpreting results of a social network derived from this behaviour. Unlike active behaviours, such as grooming or mounting, spatial proximity may not necessarily indicate a clear affiliative or agonistic relationship, or the specific roles of each individual. Rather, spatial proximity may reflect a basic level of social tolerance between pairs of animals. However, the relationship between the degree to which two animals tolerate one another and the duration of spatial proximity may not be linear. For example, two animals that are in proximity for 20 s during a 30 min observation period does not imply that they are twice as socially tolerant as two animals that are in proximity for 10 s.

A second issue is that the networks produced by spatial proximity are dense. On average, spatial proximity behaviours were coded 230.6 times/h. In our 10 h of observation data, the density of observed interactions (i.e. the number of pair–pair interactions, or edges, observed in the network divided by the total possible number of edges) ranged from 38% to 73%; the spatial proximity network derived from all 10 h had a density of 91%. However, many of these edges were somewhat weak and could simply be a product of random chance. For example, in the average network, only 28% of spatial proximity durations were greater than 1 min within a 30 min span. As a result, some of these weak edges may not be important for measuring the global and local properties of the spatial proximity network, and therefore may not provide a functionally relevant measure of social tolerance. Therefore, a more meaningful examination of the global and local properties of spatial proximity networks may require excluding such edges by thresholding out short periods of contact between individuals. In doing so, one should keep in mind that changing the density of the network through these means may alter its reliability with respect to time.

Grooming Networks

We found that grooming networks may become reliable when they are derived from at least 4 h of observation. Interestingly, the initial split-half reliability analysis suggested that grooming interactions had little error when derived from only 1 h of data, but after the permutation testing it was clear that this similarity was far less than that observed for spatial proximity. Furthermore, social networks derived from 1 h of grooming data were not similar to the average network derived from 10 h of grooming data. The bootstrap analysis suggests that at least 4 h of data are required to measure a reliable grooming network; however, it is possible that more than 4 h may be required, depending on the scarcity of observations. One explanation for the disparate findings is that the grooming networks derived from 1 h of data are sparse and therefore the similarity may fluctuate from session to session. Per hour of data, only 24 grooming observations were noted, and these grooming bouts were concentrated in a few dyads. The density of the grooming networks for 1 h of data ranged from 1% to 16.67%; the network derived from all 10 h had a density of 42%.

Mounting Networks

We found that mounting behaviour networks may become reliable when derived from at least 6 h of observation. The split-half reliability analysis showed that 1 h of data was insufficient to derive a meaningful social network. Mounting networks derived from 1 h of data were not significantly similar to the mounting network averaged over 9 h of data. The bootstrap analysis suggested that at least 6 h of data are required to measure a reliable mounting network. However, the confidence intervals were extremely large; at least 9 h of data may be required to explain 80% of the variance found in 10 h of data. The limited reliability of mounting may be explained by the scarcity of observations. Only 160 observations of mounting were coded across the 10 h of data, and mounting behaviour may have been even more concentrated within a few dyads than grooming behaviour. The density of the mounting networks for 1 h of data ranged from 1% to 11%; the mounting network derived from all 10 h had a density of 49%. While these values may be similar to the ranges observed for grooming networks, the distribution of density across the networks

was very different. Half of the grooming networks had a density over 11%; only one of the mounting networks had a density over 9%.

Measuring and Predicting the Amount of Error in Social Networks

The methods used in this paper are applicable for measuring the amount of error in social networks derived from other observational approaches and for predicting the amount of error one might find in future observations. Although we found that a single hour of spatial proximity behaviour produced a reliable social network when derived from video observations, this may vary from case to case when using human observers via focal sampling, via event sampling or via videorecorded observations. It is also possible that differences in the size, age or sex distribution of the species being studied can affect the frequency of behaviours. For example, the frequency of mounting observed within a group of 10 adolescent macaque males may differ from the frequency of mounting observed in a group of 100 macaques comprising four or five different matriline.

Furthermore, it is also possible that these results may change with more data (Perreault, 2010). Unfortunately, we were limited to only 10 h of observation for this analysis, and the specific story might change with 15 or 20 h of data. Because additional data may reflect a change in the true social structure, methods that examine the similarity of social networks using randomized time points, such as those used in this paper, can test whether the differences observed across time reflect differences observed independent of time (e.g. sampling bias or other sources of error). Future studies can use these methods to better estimate the minimum amount of data necessary to measure the social environment for their specific groups.

Evaluating Changes in Social Networks as a Function of Time/Treatment

Here we extend methods that can be used to evaluate the minimum amount of data necessary to derive social networks. We demonstrated these methods on data derived from video observations from a group of 10 adolescent male macaques, but these methods are also easily transferrable to other studies of social networks in any species. By determining this minimum, one can treat social networks as individual data points, and therefore assess whether different independent variables (e.g. pharmacological manipulations) alter the properties of social environments more than what would occur by chance (e.g. removal of key individuals, seasonal effects, rank changes, drug interventions, etc.). Therefore, the use of such methods is powerful for determining the effects of endogenous and exogenous factors on the social structure of groups, which may have translational application for improving social function in psychiatric conditions such as autism or social anxiety disorders.

Acknowledgments

Support for this research was provided by the National Institute of Mental Health grant R01MH096983 (L.J.Y. and L.A.P.). Additional support was provided by the National Institutes of Health grant OD P51OD011132 to Yerkes Primate National Research Center, Emory University. Financial support was provided by the Wenner-Gren Foundations (H.W.).

References

- Altmann, SA., editor. Social communication among primates. University of Chicago Press; Chicago, IL: 1967.
- Beisner BA, Jackson ME, Cameron A, McCowan B. Effects of natal male alliances on aggression and power dynamics in rhesus macaques. *Am J Primatol.* 2011; 73(8):790–801. doi: 10.1002/ajp.20907. [PubMed: 21698659]
- Blasi MF, Boitani L. Complex Social structure of an endangered population of bottlenose dolphins (*Tursiops truncatus*) in the Aeolian Archipelago (Italy). *PLoS One.* 2014; 9(12):e114849. doi: 10.1371/journal.pone.0114849. [PubMed: 25494331]
- Brent LJ, Heilbronner SR, Horvath JE, Gonzalez-Martinez J, Ruiz-Lambides A, Robinson AG, et al. Genetic origins of social networks in rhesus macaques. *Sci Rep.* 2013; 3:1042. doi: 10.1038/srep01042. [PubMed: 23304433]
- Brent LJ, Lehmann J, Ramos-Fernandez G. Social network analysis in the study of nonhuman primates: a historical perspective. *Am J Primatol.* 2011; 73(8):720–730. doi: 10.1002/ajp.20949. [PubMed: 21433047]
- Brent LJ, Maclarnon A, Platt ML, Semple S. Seasonal changes in the structure of rhesus macaque social networks. *Behav Ecol Sociobiol.* 2013; 67(3):349–359. doi: 10.1007/s00265-012-1455-8. [PubMed: 23565026]
- Brent LJ, Semple S, Dubuc C, Heistermann M, Maclarnon A. Social capital and physiological stress levels in free-ranging adult female rhesus macaques. *Physiol Behav.* 2011; 102(1):76–83. doi: 10.1016/j.physbeh.2010.09.022. [PubMed: 20933532]
- Croft D, James R, Ward A, Botham M, Mawdsley D, Krause J. Assortative interactions and social networks in fish. *Oecologia.* 2005; 143(2):211–219. [PubMed: 15682346]
- Croft DP, Madden JR, Franks DW, James R. Hypothesis testing in animal social networks. *Trends in Ecology & Evolution.* 2011; 26(10):502–507. [PubMed: 21715042]
- Czoty PW, Gould RW, Nader MA. Relationship between social rank and cortisol and testosterone concentrations in male cynomolgus monkeys (*Macaca fascicularis*). *J Neuroendocrinol.* 2009; 21(1):68–76. doi: 10.1111/j.1365-2826.2008.01800.x. [PubMed: 19094095]
- Frere CH, Krutzen M, Mann J, Connor RC, Bejder L, Sherwin WB. Social and genetic interactions drive fitness variation in a free-living dolphin population. *Proc Natl Acad Sci U S A.* 2010; 107(46):19949–19954. doi: 10.1073/pnas.1007997107. [PubMed: 21041638]
- Guimerà R, Mossa S, Turtshi A, Amaral LAN. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America.* 2005; 102(22):7794–7799. http://www.ncbi.nlm.nih.gov/beanalyzer/beanalyzer.fcgi?db=pubmed&cmd=Retrieve&dbfrom=pubmed&list_uids=15911778. [PubMed: 15911778]
- Haddadi H, King AJ, Wills AP, Fay D, Lowe J, Morton AJ, et al. Determining association networks in social animals: choosing spatial–temporal criteria and sampling rates. *Behav Ecol Sociobiol.* 2011; 65(8):1659–1668.
- Handcock MS, Gile KJ. Modeling social networks from sampled data. *Annals of Applied Statistics.* 2010:5–25.
- Hobaiter C, Poisot T, Zuberbühler K, Hoppitt W, Gruber T. Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biol.* 2014; 12(9):e1001960. doi: 10.1371/journal.pbio.1001960. [PubMed: 25268798]
- James R, Croft DP, Krause J. Potential banana skins in animal social network analysis. *Behav Ecol Sociobiol.* 2009; 63(7):989–997.
- Kasari C, Rotheram-Fuller E, Locke J, Gulsrud A. Making the connection: randomized controlled trial of social skills at school for children with autism spectrum disorders. *J Child Psychol Psychiatry.* 2012; 53(4):431–439. doi: 10.1111/j.1469-7610.2011.02493.x. [PubMed: 22118062]
- Keppel, G. Design and analysis: A researcher's handbook. 3rd ed.. Prentice-Hall; Englewood Cliffs, NJ: 1991.
- Lindquist J, Ma J, van den Driessche P, Willeboordse FH. Effective degree network disease models. *J Math Biol.* 2011; 62(2):143–164. doi: 10.1007/s00285-010-0331-2. [PubMed: 20179932]

- Lusseau D, Newman ME. Identifying the role that animals play in their social networks. *Proc Biol Sci.* 2004; 271(Suppl. 6):S477–S481. doi: 10.1098/rsbl.2004.0225. [PubMed: 15801609]
- Lusseau D, Whitehead H, Gero S. Incorporating uncertainty into the study of animal social networks. *Animal Behaviour.* 2008; 75(5):1809–1815.
- Makagon MM, McCowan B, Mench JA. How can social network analysis contribute to social behavior research in applied ethology? *Appl Anim Behav Sci.* 2012; 138(3-4):152–161. doi: 10.1016/j.applanim.2012.02.003.
- Mann J, Stanton MA, Patterson EM, Bienenstock EJ, Singh LO. Social networks reveal cultural behaviour in tool-using [corrected] dolphins. *Nat Commun.* 2012; 3:980. doi: 10.1038/ncomms1983. [PubMed: 22864573]
- McCowan B, Beisner BA, Capitanio JP, Jackson ME, Cameron AN, Seil S, et al. Network stability is a balancing act of personality, power, and conflict dynamics in rhesus macaque societies. *PLoS One.* 2011; 6(8):e22350. doi: 10.1371/journal.pone.0022350. [PubMed: 21857922]
- Miller N, Garnier S, Hartnett AT, Couzin ID. Both information and social cohesion determine collective decisions in animal groups. *Proc Natl Acad Sci U S A.* 2013; 110(13):5263–5268. doi: 10.1073/pnas.1217513110. [PubMed: 23440218]
- Newman, MEJ. *Networks: An introduction.* Oxford University Press; Oxford, U.K.: 2010.
- Perreault C. A note on reconstructing animal social networks from independent small-group observations. *Animal Behaviour.* 2010; 80(3):551–562.
- Qin DD, Dominic Rizak J, Feng XL, Chu XX, Yang SC, Li CL, et al. Social rank and cortisol among female rhesus macaques (*Macaca mulatta*). *Dongwuxue Yanjiu.* 2013; 34(E2):E42–E49. doi: 10.3724/sp.j.1141.2013.e02e42. [PubMed: 23572366]
- Sawka KJ, McCormack GR, Nettel-Aguirre A, Hawe P, Doyle-Baker PK. Friendship networks and physical activity and sedentary behavior among youth: a systematized review. *Int J Behav Nutr Phys Act.* 2013; 10:130. doi: 10.1186/1479-5868-10-130. [PubMed: 24289113]
- Shemesh Y, Sztainberg Y, Forkosh O, Shlapobersky T, Chen A, Schneidman E. High-order social interactions in groups of mice. *Elife.* 2013; 2:e00759. doi: 10.7554/eLife.00759. [PubMed: 24015357]
- Sumpter DJ, Krause J, James R, Couzin ID, Ward AJ. Consensus decision making by fish. *Curr Biol.* 2008; 18(22):1773–1777. doi: 10.1016/j.cub.2008.09.064. [PubMed: 19013067]
- Sundaresan SR, Fischhoff IR, Dushoff J. Avoiding spurious findings of nonrandom social structure in association data. *Animal Behaviour.* 2009; 77(6):1381–1385.
- Voelkl B, Kasper C, Schwab C. Network measures for dyadic interactions: stability and reliability. *Am J Primatol.* 2011; 73(8):731–740. doi: 10.1002/ajp.20945. [PubMed: 21394745]
- Wasserman, S.; Faust, K. *Social network analysis: Methods and applications.* Cambridge University Press; Cambridge, U.K.: 1994.
- Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature.* 1998; 393(6684):440–442. doi: 10.1038/30918. [PubMed: 9623998]
- Whitehead H. Precision and power in the analysis of social structure using associations. *Animal Behaviour.* 2008; 75(3):1093–1099.
- Whitehead H, Bejder L, Andrea Ottensmeyer C. Testing association patterns: issues arising and extensions. *Animal Behaviour.* 2005; 69(5):e1. doi:10.1016/j.anbehav.2004.11.004.

Highlights

- We developed methods to measure the reliability of social networks.
- Social network reliability varies depending on the behaviour and group.
- Future social network studies should use such methods on their data.

Figure 1.
 Layout of the enclosure in which the 10 male macaque subjects were socially housed: (a) aerial diagram of the enclosure; (b) snapshot of enclosure from camera 3.

Figure 2.

Similarity of behavioural matrices between 30 min sessions for (a) mounting, (b) grooming and (c) spatial proximity in rhesus macaques. Sessions are ordered by date. Thick black lines delineate different dates. Thin white lines delineate even- and odd-minute sessions. The value in the last row and column of each matrix reflects similarity between the average matrix and all sessions. (d) Mean $\pm$ SE within-session similarity (η^2) values for spatial proximity, grooming and mounting networks. The dashed line represents an η^2 value of 0.7, which was used to conduct the t tests.

Figure 3. Mean $\pm$ SE within-session similarity (η^2) values derived from permutation analysis of spatial proximity, grooming and mounting behaviours of 10 rhesus macaques during each 1 h observation session. The dashed line represents an η^2 value of 0.7, which was used to conduct the t tests.

Figure 4.

Mean $\pm$ SE similarity (η^2) values between the average matrix and individual session matrices for spatial proximity, grooming and mounting. The dashed line represents an η^2 value of 0.7, which was used to conduct the t tests.

Figure 5.

Similarity (η^2) in spatial proximity, grooming and mounting behaviour of rhesus macaques based on the number of hours of data used to generate a cumulative session matrix: (a) raw data for spatial proximity, grooming and mounting; (b) bootstrapped data for spatial proximity; (c) bootstrapped data for mounting; (d) bootstrapped data for grooming. For comparative purposes, the dashed line in (b, c, d) represents an η^2 value of 0.7.