Animal Behaviour 104 (2015) 1–11

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

The consequences of unidentifiable individuals for the analysis of an animal social network

Matthew J. Silk [a, *], Andrew L. Jackson [b], Darren P. Croft [c], Kendrew Colhoun [d], Stuart Bearhop [a, *]

[a] Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter Penryn Campus, Penryn, U.K.
[b] Department of Zoology, School of Natural Sciences, Trinity College Dublin, Ireland
[c] Centre for Research in Animal Behaviour, University of Exeter Streatham Campus, Exeter, U.K.
[d] Wildfowl and Wetlands Trust, Comber, County Down, U.K.

ARTICLE INFO

Article history:
Received 7 October 2014
Initial acceptance 11 November 2014
Final acceptance 23 February 2015
Published online
MS. number: 14-00795R

Keywords:
accuracy
bias
fission–fusion dynamics
network metric
partial network
precision
social strategy
social structure

Social network analysis is pervasive in understanding animal social systems, and provides information about how individuals vary in their social strategies. Many long-term studies comprising uniquely marked individuals use social network analysis as an analytical tool. However, the assumption that it is possible to make inferences using network metrics calculated using a subset of the population has yet to be investigated in an animal social network. We use a simulation study of networks derived from social interactions in a typical fluid fission–fusion social system to determine the precision and accuracy of measures of individual social position based on incomplete knowledge. We show that individual social positions measured in partial social networks correlate strongly with positions in the full social network. This correlation typically becomes stronger as the size of the simulated population is increased and is largely not affected by network density. The choice of network metric has an important effect on the precision of partial networks only when they include a small subset of the population and therefore caution is advised using some of these measures in small partial networks. This work demonstrates that valid inferences about individual social position and strategy can be made using partial networks in a wide range of animal social networks, highlighting the value of applying these methods in large long-term study populations.

Social network analysis is now widely used to study animal social behaviour (Farine, Garroway, & Sheldon, 2012; Pinter-Wollman et al., 2013; Sih, Hanser, & McHugh, 2009; Wey, Blumstein, Shen, & Jordán, 2008), and is fast developing beyond being a descriptive tool to become fundamental in quantifying behavioural interactions and their subsequent consequences in a wider social context (Dey, Reddon, O'Connor, & Balshine, 2013; Formica et al., 2012; Kohn, King, Dohme, Meredith, & West, 2013; Sueur, Jacobs, Amblard, Petit, & King, 2011; Wey, Burger, Ebensperger, & Hayes, 2013). There is a growing focus on understanding how an individual's personality, phenotype and condition interact to influence its social decision making and social strategy (Aplin et al., 2013; Croft et al., 2005, 2009; Silk, Croft, Tregenza, & Bearhop, 2014; Wilson, Krause S., Dingemanse & Krause 2013). By studying variation in individual social position, it is possible to make inferences about both the mechanisms that drive population social structure in a study system (Connor, Heithaus, & Barre, 2001; Stanley & Dunbar, 2013; Wittemyer, Douglas-Hamilton, & Getz, 2005) and the consequences of following particular social strategies for individual fitness (Formica et al., 2012; McDonald, 2007; Wey et al., 2013). Extracting individual level metrics from networks is therefore a major application of social network analysis in behavioural ecology, and being able to apply network analysis to infer individual social behaviour in a wide range of systems becomes highly significant.

Recent studies have linked social position to home range (see Pinter-Wollman et al., 2013 for a review), social status (Sueur & Petit, 2008), age (Patriquin, Leonard, Broders, & Garroway, 2010), sex (Carter Seddon, Frère, Carter, & Goldizen 2013; Gilby & Wrangham, 2008), genetic relatedness (Archie, Moss, & Alberts, 2006; Carter, Seddon, Frère, Carter, & Goldizen, 2013; Schülke, Wenzel, & Ostner, 2013), the acquisition of social information (Aplin, Farine, Morand-Ferron, & Sheldon, 2012; Aplin et al., 2014;

* Correspondence: M. J. Silk and S. Bearhop, Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter Penryn Campus, Penryn, Cornwall, TR10 9FE, U.K.
E-mail addresses: matthewsilk@outlook.com (M. J. Silk), S.Bearhop@exeter.ac.uk (S. Bearhop).

http://dx.doi.org/10.1016/j.anbehav.2015.03.005
0003-3472/

Claidiere, Messer, Hoppitt, & Whiten, 2013), disease status (Weber et al., 2013) and reproductive success (Wey et al., 2013). Additionally, there has been increasing recent interest in social network position as a personality trait or part of a wider behavioural syndrome (Krause, James, & Croft, 2010; Wilson, Krause, Dingemanse, & Krause, 2013). The role of these potential factors in influencing social position and the consequences to an individual of that social position can be investigated using a simple correlative approach at an individual level (Formica et al., 2012; Wey et al., 2013). This makes data from many marking or tagging projects highly valuable in generalizing our understanding of what processes drive individual social strategies and social evolution in many types of fission–fusion social system, and how this then contributes to variation in overall population social structure.

Historically, social network analysis was largely limited to systems in which all (or the vast majority of) individuals were individually identifiable (Connor et al., 2001; Croft et al., 2005; Drewe, Madden, & Pearce, 2009; Ramos-Fernández, Boyer, Aureli, & Vick, 2009). However, it is now increasingly used in a wide range of long-term ecological studies in which this is not the case. In particular, there are now many examples of social networks being used in study populations in which individuals are only identifiable once captured and tagged, a restriction that can substantially reduce the identifiable component of the population (Aplin et al., 2012; Farine & Milburn, 2013; Hamede, Bashford, McCallum, & Jones, 2009; Oh & Badyaev, 2010; Weber et al., 2013) as a result of difficulties in catching all individuals or limitations imposed by the number of different marks available. Additionally, some of these long-term study populations will be larger than other populations in which network analysis has been used previously or have a higher turnover of individuals, increasing the likelihood of a subset of the population being used to build networks. The sampling of social networks in this way is of concern because social networks are relational data, and the relations among the members of a sample will only be a subset of their full set of relationships (Alba, 1982; Croft, James, & Krause, 2008). This means that relational data could be expected to respond more unreliably to sampling from a population than other data types, as missing individuals may exert a strong influence on the social measures of individuals sampled.

This issue has been investigated elsewhere in the social networks literature (Borgatti, Carley, & Krackhardt, 2006; Lee, Kim, & Jeong, 2006; Stumpf & Thorne, 2006; Wang, Shi, McFarland, & Leskovec, 2012). Many of these studies focus on the accuracy of network level measures such as degree correlations or degree distributions (e.g. Borgatti et al., 2006; Frantz, Cataldo, & Carley, 2009; Lee et al., 2006) that are rarely used in animal social networks. The presence of unidentifiable individuals, and subsequent subsampling of the true number of nodes present in the network, is directly analogous to the measurement errors imposed by random node sampling (Lee et al., 2006). Wang et al. (2012) explored the consequences of various types of measurement error on the correlation of node level metrics in an observed network and the true network on which these observations were based and found that for false negative nodes (equivalent to unidentifiable individuals in animal social networks) the metrics tested showed resilience to error. In support of this, Stumpf and Thorne (2006) found that inferences could be made using subsamples of molecular networks generated through random node sampling, suggesting that partial networks could provide useful information if relative values rather than accurate values of node level metrics are required. Animal behaviour research is typically interested in the consequences of variation in network position among individuals (e.g. Aplin et al., 2012; McDonald, 2007; Wey et al., 2013), and therefore being able to make precise inferences about relative differences in node level metrics among individuals would be sufficient.

The effect of using 'partial networks' constructed using a random subset of the population on the properties of individual metrics in animal social networks has received little attention (Croft et al., 2008; Cross et al., 2012). Franks, Ruxton, and James (2010) investigated how various network properties depended upon the number of censuses and the proportion of individuals sampled in each census. They concluded that the former was more important, and that network level properties were fairly resilient to a low number of individuals being detected on each census. While this focuses more on the concept of the detectability of individuals rather than the effects of unidentifiable individuals, it does perhaps suggest that there is potential for social network studies to make valid inferences using only a sample of the population.

In this study we used a simulated fluid fission–fusion social structure parameterized using a long-term study of light-bellied brent geese, *Branta bernicla hrota*, to investigate how properties of the social network vary as the proportion of identifiable individuals in the population changes. Light-bellied brent geese are a species with typical fluid fission–fusion dynamics with considerable spatiotemporal variation in group size (Fig. 1). The outcome of this is a fairly well connected social network that contains variation among individuals in social network positions, making it possible to test structural hypotheses at an individual level. We examined how changing proportions of identifiable individuals in a population alters the relationship between an individual's apparent network metrics and its real social metrics (i.e. network metrics when the entire population consists of identifiable individuals).

We looked at a selection of key social metrics to determine whether the choice of network metric in a given study could affect the accuracy and precision of the conclusions that can be drawn. Different network metrics vary in the extent to which they are influenced by relational data depending on to what extent they are global (take into account the structure of the entire network) or local measures. We investigated four of the most commonly used measures of centrality and one commonly used measure of transitivity (Table 1), which between them provide a range of important measures of social position. We expected that local measures (e.g. degree, strength) would be less affected by subsampling from networks than more global measures such as betweenness (the reasons for this are explored in more depth in the Discussion). This study therefore highlights the importance of understanding the consequences of random node deletion for the analysis of animal social networks.

METHODS

Generation of Population

The simulated populations of light-bellied brent geese analysed in this study were generated using data from a long-term ecological study on the Irish wintering population of this subspecies.

(1) Three simulated populations of geese containing 37, 104 and 591 individuals were generated, with a population defined here as all individuals that could potentially be included in a network, possible methodological constraints notwithstanding. These populations were formed of family units, where a family unit was defined as an unpaired (unassociated) individual, a nonreproductive pair or family (parents with dependent juveniles). Family units remained together throughout the sampling process. The three population sizes used were selected to represent a range of realistic population sizes likely to be used in long-term studies, in which use of partial networks is much more likely.

(2) The frequency of family units (the percentages of juveniles in our three simulated populations were 13.5%, 8.7% and 12.0% compared to 15.5 ± 13.3% from Madsen, Cracknell, & Fox, 1999) and

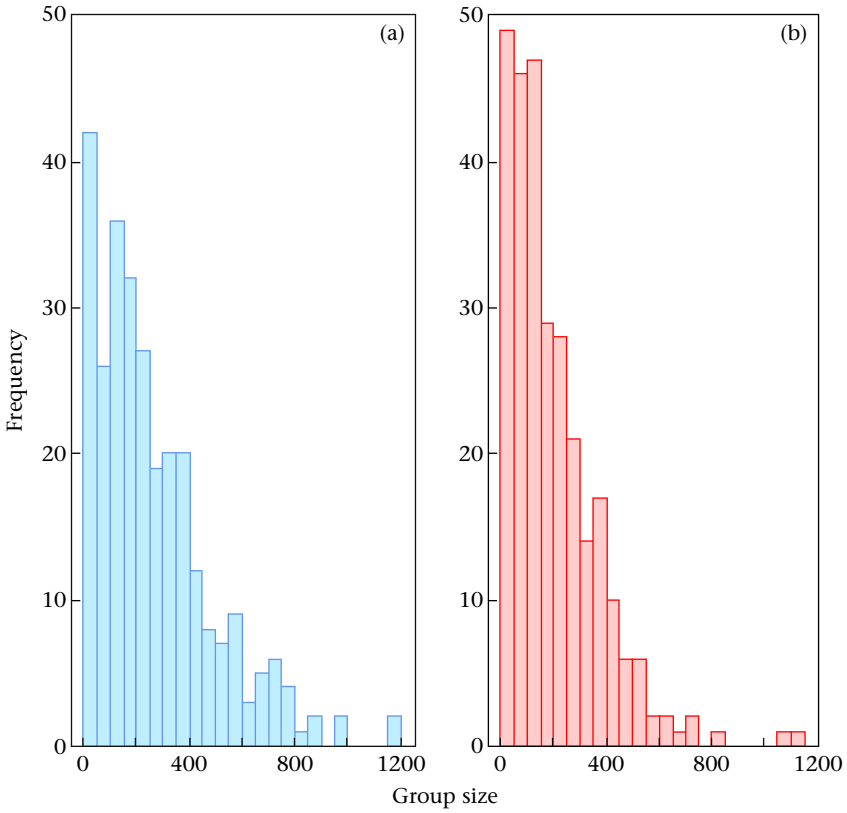

Figure 1. (a) The group size distribution of the Dublin staging population of ECHA light-bellied brent geese in January–March 2012. (b) An example histogram of the corresponding negative binomial group size distribution used in our models. The sample drawn from the distribution is equivalent to the observed number of groups ($N = 283$).

Table 1
Details of the five key network metrics used in the study

Metric	Type	How is it calculated?	What does it measure?
Degree	Centrality	A count of individuals to which a focal individual is connected in the network	The number of social connections an individual has
Strength	Centrality	A sum of the total weight of all a focal individual's connections in the network	Accounts for the value of the connections an individual has as well as the number of individuals to which it is connected
Betweenness	Centrality	A count of the shortest paths between other individuals in the network that pass through a focal individual	The importance of an individual in connecting different parts of the network
Eigenvector centrality	Centrality	The focal individual's value in an eigenvector that corresponds to the maximum eigenvalue from the network in matrix form	The influence of an individual in the network taking into account its second-order connections
Clustering coefficient	Transitivity	The proportion of individuals connected to the focal individual that are interconnected themselves	The embeddedness of the individual within its social clique

the number of juveniles in each family (Fig. 2) were specified from distributions based on real data.

(3) Each individual and family unit were assigned a unique identity for use in social group generation and subsequent social network construction.

Generation of Social Groups

At each iteration of the algorithm the population was split into different social groups. The group size distribution for these groups was based on fitting a negative binomial distribution to real data from a focused socioecological study of the Dublin wintering population in January–March 2012 (Fig. 1) with size = 1.4 and μ = population/10. Social groups were assigned at a family unit level to ensure that family units remained cohesive as observed in the real data.

(1) All family units in the population were randomly reordered.

(2) A group size was determined randomly from the negative binomial distribution outlined above, with the population size used to calculate μ equalling the number of family units in the population.

(3) The corresponding number of family units was taken from the reordered population set.

(4) Steps 2 and 3 were repeated until the total number of family units allocated equalled or just exceeded the number of family units in the population.

(5) If the latter case was true, this last group was not included and any remaining family units were allocated randomly to groups already created.

Each of the three populations was randomly split into social groups using this method five or 10 times to provide the association information used to generate social networks of two different densities (edge densities for all networks are provided in Appendix Table A1). This was sufficient to generate considerable variation in

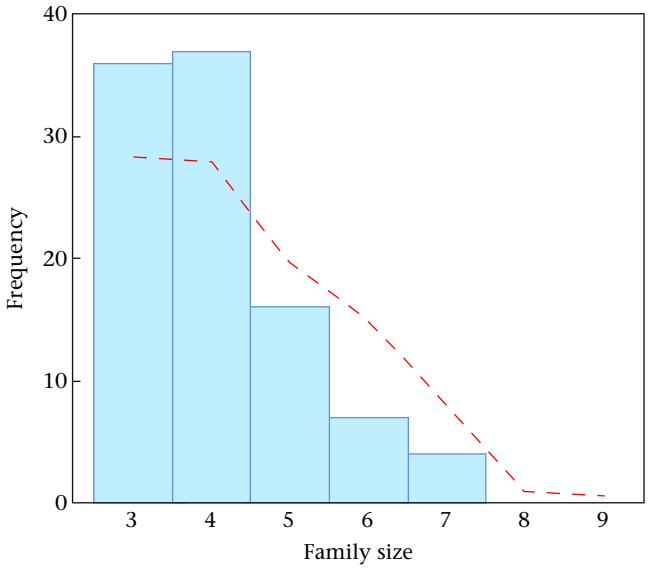

Figure 2. The distribution of the number of juveniles in a family group in the Irish wintering population of ECHA light-bellied brent geese as a percentage of the total number of families (dashed line) plotted against an example histogram of the simulated binomial distribution used in our models ($N = 100$).

edge weight while preventing networks from being completely connected, as this would impact on the ability to use certain key network metrics to differentiate between the social network position of individuals in binary networks. By doing this we could determine any effect of network density on the results obtained.

Construction of Social Networks

Social networks were constructed using the gambit of the group assumption (Croft et al., 2008; Franks et al., 2010; Whitehead & Dufault, 1999), with edges connecting individuals that had co-occurred in the same social group. Weighted association matrices were calculated, with weights of edges in the weighted networks being the number of times individuals had co-occurred in a social group.

Definition of Identifiable Individuals

Simulations were run for each combination of the three population sizes and two network densities (six in total), then analysed with five different levels of assigned marking effort (90%, 70%, 50%, 30%, 10% of individuals made identifiable) in the two larger populations. In the smallest population the case where only 10% of individuals were identifiable was excluded as networks constructed using only three or four individuals would not be informative. Individuals were made identifiable at random, and individuals that had not been made identifiable were removed from the association matrix to create a partial network containing only identifiable individuals (making this process equivalent to random node deletion from the full network). This process is outlined in more detail below. Ten repeats of each level of marking effort were run, generating either 40 or 50 partial networks for each full network, depending on the population size.

To assign individuals as identifiable we (1) assigned each individual a number from a random uniform distribution between 0 and 1 and (2) defined individuals with a number equal to or lower than the assigned marking effort as identifiable. To generate variation between replicates in the identity and precise number of individuals made identifiable, steps 1 and 2 were repeated 10 times to produce 10 different subsets of the original network.

Calculation of Network Metrics

For each combination of identifiability, population size and network density, measures of strength, degree, betweenness, eigenvector centrality and clustering coefficient (Table 1) were calculated for each node in the weighted full network and each of the weighted partial networks (280 partial networks in total).

Statistical Analysis

Precision

Correlation coefficients between network metric values in each partial network and the corresponding full association matrix were calculated. This approach (rather than matrix correlation methods) was used so that the magnitude of the effect size could be calculated to provide information on accuracy (see below). For each level of marking effort in each run of the simulation, the mean and variance of the 10 correlation coefficients generated was calculated for each network metric. These values were then used to illustrate how marking effort, population size and network density affect the precision of estimates of each network metric in partial networks. This enabled us to investigate how the proportion of individually identifiable individuals in a population may affect our ability to use that partial network to infer their real social position.

Accuracy

The mean slope of the regression was calculated for each of the five network metrics for each level of marking level in each run of the simulation. The slope of the regression describes how the value (rather than its relative value) of network metrics calculated in each partial network is related to the value in the real network. This was used to illustrate the effect of marking effort on the accuracy with which different network metric values in a partial network estimated an individual's real network metric values.

Bias

The relationship between the real (simulated) social positions and y residual (both its actual value and the magnitude of this value) of an individual was then found for each regression. This enabled us to detect any potential bias towards individuals in particular social positions having their network metrics under- or overestimated when using partial networks to infer social position.

Generation of the simulated populations and all social network analysis was carried out using R 2.15.1 (R Development Core Team, 2013). Network analysis used the packages tnet (Opsahl, 2009) and igraph (Csardi & Nepusz, 2006).

RESULTS

Precision

The results for the medium and large populations are presented in Fig. 3 and for the small population in Fig. 4. The mean correlation between partial and full networks declined nonlinearly and the variance in the correlation coefficients calculated increased nonlinearly as the proportion of identifiable individuals decreased (Figs 3, 4). The precise nature of these changes depended on the size of the simulated population, the density of the social network and network metric being investigated (Figs 3, 4).

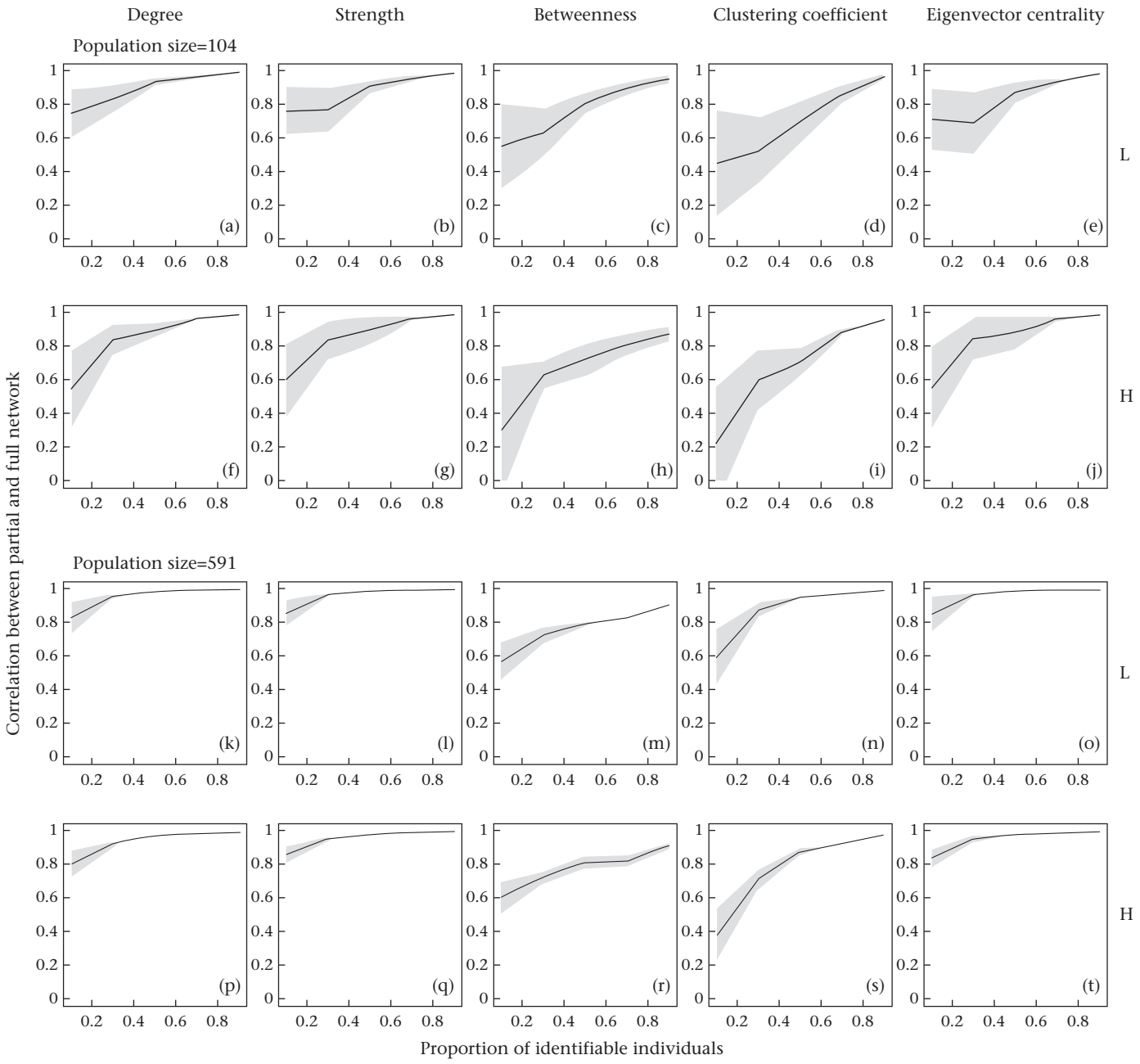

Figure 3. The relationship between the proportion of identifiable individuals and the correlation between five different social metrics in the partial social network (containing only these individuals) and full network in two simulated populations of different sizes (our medium and large populations): 104 individuals (a-j) and 591 individuals (k-t). Rows L (a-e and k-o) and H (f-j and p-t) for each population refer to the low- and high-density networks for each population size. Solid black lines join the mean correlation coefficients at each level of marking effort, with the light grey area bounded by the mean ± SD at each level of marking effort when these values are between 0 and 1.

Population Size

An interaction between population size and the proportion of individuals in the partial network was important in explaining the correlation between metric values in a partial network and their corresponding values in the full network (Figs 3, 4). As the size of the simulated population increased (compare Fig. 3a–j with Fig. 3k–t and Fig. 3 with Fig. 4), the magnitude of the changes to the mean and variance decreased, especially when the proportion of identifiable individuals in the population was low. Even in the smallest population (37 individuals) precision only declined substantially when a low proportion of individuals were identifiable and networks were constructed from a very small number of individuals (Fig. 4). This pattern was similar for all metrics investigated, although the scale of the effect did vary. This suggests that the precision of partial networks to infer individual social network positions increases in larger study populations.

Network Density

Network density generally had a limited effect on the precision of inferences from partial networks, especially as population size increased. However, estimates of several metrics in the smallest population size (Fig. 4) and of clustering coefficient from all high-density partial networks were consistently less precise (Figs 3, 4). The result for clustering coefficient is consistent with decreased

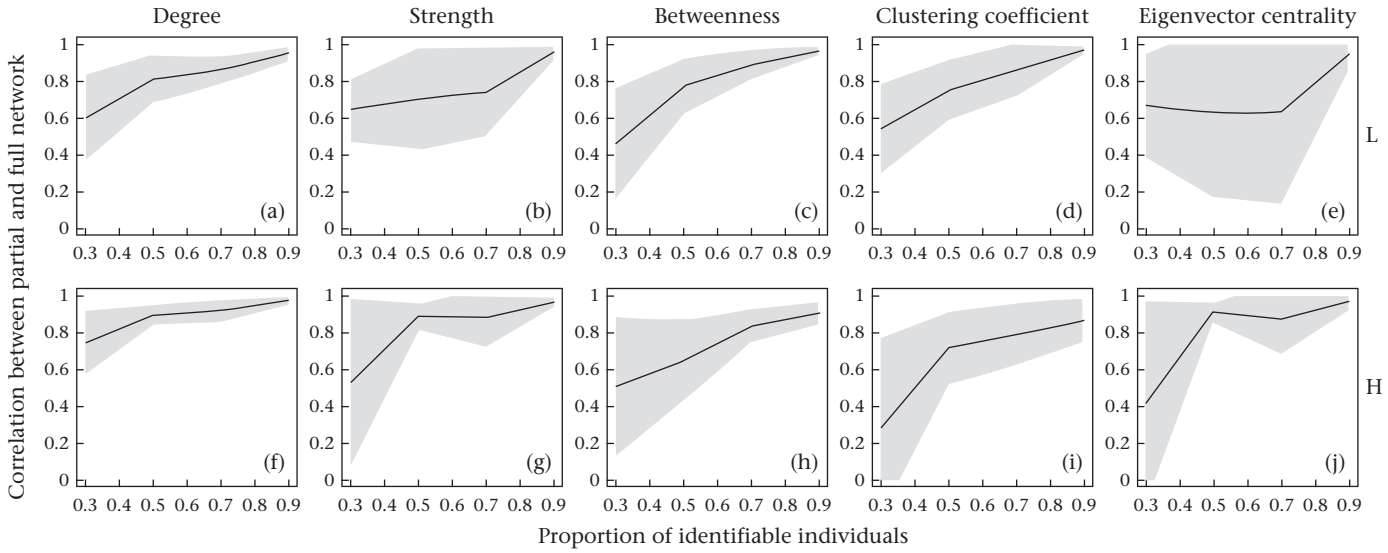

Figure 4. The relationship between the proportion of identifiable individuals and the correlation between five different social metrics in the partial social network (containing only these individuals) and full network in a population of 37 individuals. Rows L (a-e) and H (f-j) refer to the low- and high-density networks. Solid black lines join the mean correlation coefficients at each level of marking effort, with the light grey area bounded by the mean ± SD at each level of marking effort where these are between 0 and 1.

precision of more global measures. Additionally, there was increased variance in the precision of partial networks at low marking efforts in the smallest population size investigated (Fig. 4).

Network Metrics

The correlation of all network metrics between the partial and full network declined as the proportion of individuals included in the partial network decreased. For all metrics there was also a clear change in the relationship when the partial network contained fewer than approximately one-third of individuals, below which correlation coefficients declined more rapidly. However, there was substantial variation between different network metrics (compare across columns in Figs 3 and 4) in the extent of these changes. When 50% or more individuals in the population were identifiable, the values for all network metrics in partial networks were correlated very closely with the simulated real situation in the two larger populations investigated (Fig. 3) and for all metrics apart from eigenvector centrality in the smallest population (Fig. 4). Measures of centrality (degree, strength and eigenvector centrality) in partial networks remained highly correlated with those in the full network even when only 10% of individuals in the population were identifiable for the two larger population sizes investigated. This pattern remained in the small population for degree and strength, but eigenvector centrality became unreliable in most partial networks, suggesting that is a poor measure to use in very small networks in general (Fig. 4). The strength of correlations for betweenness and clustering coefficient declined much more substantially, especially in the two smaller population sizes used (Figs 3, 4). For clustering coefficient in particular, there was also a substantial increase in variance, indicating a lack of reliability in values in a partial network. This was especially true in networks of higher density.

Accuracy: Network Metrics and Network Density

The accuracy of node level metrics from partially sampled networks was highly dependent on the metric being used (Fig. 5, Appendix Fig. A1). The accuracy of local measures, degree (not illustrated as the results obtained were nearly identical to strength) and strength decreased linearly in direct proportion to the proportion of identifiable individuals in the population (Fig. 5a, Appendix Fig. A1a). In contrast, the accuracy of eigenvector centrality did not depend on marking effort in any networks, although the estimates from partial networks tended to be more accurate in high-density networks (Fig. 5d) and became inaccurate when low numbers of individuals were identifiable in the smallest population investigated. For both betweenness (Fig. 5b, Appendix Fig. A1b) and clustering coefficient (Fig. 5c, Appendix Fig. A1c) the effect of the proportion of identifiable individuals on the accuracy of estimates from partial networks was strongly influenced by the density of the network being studied. Estimates of betweenness remained highly accurate at all levels of marking effort in low-density networks, but accuracy declined in direct proportion to marking effort in high-density networks for the two larger networks investigated (Fig. 5b). In the smaller population the same result was apparent for high-density networks (Appendix Fig. A1c), but accuracy was unpredictable in low-density networks (Appendix Fig. A1c) This is likely to be caused by betweenness in more sparse networks being a global measure closely related to path length with raw values not affected predictably by random node deletion, while in higher density networks it is a more local measure closely linked to degree. The accuracy of clustering coefficient declined in proportion to marking effort in low-density networks and was independent of marking effort in high-density networks (Fig. 5c, Appendix Fig. A1c), most likely as a result of the network being close to fully connected in higher density networks and there being restricted variation among individuals in clustering coefficient. This also seemed to make the accuracy highly dependent on the size of the simulated population in these high-density networks (Fig. 5c). These problems serve to highlight a key issue with the use of clustering coefficient for association data, in which networks are often highly connected.

Bias: the Effect of Social Network Position

The regression slopes for the correlation between an individual's real (simulated) social network position and the raw values of its y residuals for the relationship between that and its position in a partial network were close to zero (all slopes less than 10^{-16}) for all network metrics in all runs of the simulation. In

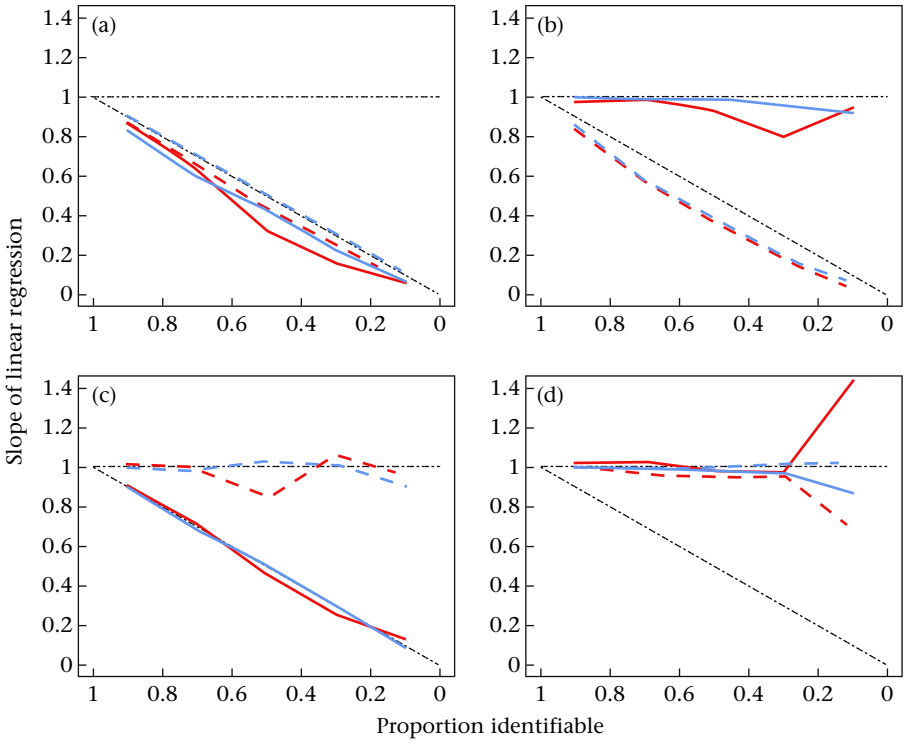

Figure 5. The relationship between the proportion of identifiable individuals in a partial network and the slope of the linear regression for four different social metrics: (a) strength, (b) betweenness, (c) clustering coefficient and (d) eigenvector centrality. Solid lines represent low-density networks and dashed lines high-density networks. Blue lines represent the large population (591 individuals) and red lines the medium population (104 individuals). In (a) and (b) lines are jittered to enable them to be distinguished. The dot-dash black lines represent the case when accuracy is independent of the proportion of identifiable individuals ($y = 1$) and the case when accuracy decreases linearly with the proportion of identifiable individuals in a partial network ($y = 1 - x$).

addition, correlations between an individual's real (simulated) social network position and the magnitudes of its y residuals for the relationship between that and its position in a partial network were weak for all network metrics in all runs of the simulation. The effect sizes of this relationship for cases in which 10% and 50% of individuals are identifiable are presented in Table 2. These results imply that the effect of an individual's social network position on the reliability of using a partial network to infer its real social position is minimal regardless of population size, network density or the proportion of marked individuals in the population.

DISCUSSION

Concerns have been expressed on the reliability of using social networks constructed using a subset of the population (Croft et al., 2008; Cross et al., 2012). However, the impact of using these 'partial networks' in lieu of full information to infer individual social network metrics has yet to be investigated in animal social networks. Although there are studies looking at this process of random node sampling (Lee et al., 2006) elsewhere in the social networks literature, they have not really focused on the ability to make predictions about individual level network metrics. This study investigated how the precision and accuracy of social metrics calculated using social network analysis are affected by varying the proportion of individually identifiable animals in a simulated network. The results indicate that, contrary to previous expectations, in social networks based on fluid social interactions it is possible to make precise inferences about individual social position when not all individuals in a population are individually identifiable (i.e. partial networks). This is despite the lack of imposed

Table 2
The mean effect sizes of 'real' network metric value on the magnitude of the y residual for the relationship between individual network metric values in partial and full networks

% Identifiable	Network metric	Small		Medium		Large	
		Low density	High density	Low density	High density	Low density	High density
10	Degree	NA	NA	0.003	−0.005	−0.001	−0.005
10	Strength	NA	NA	0.012	−0.0001	0.003	0.001
10	Betweenness	NA	NA	0.020	0.018	0.019	0.017
10	Clustering coefficient	NA	NA	0.070	−0.018	−0.104	0.049
10	Eigenvector centrality	NA	NA	0.092	−0.009	0.051	0.023
50	Degree	−0.057	−0.031	−0.00001	−0.030	−0.003	−0.012
50	Strength	−0.003	0.019	0.017	0.009	−0.0004	0.002
50	Betweenness	0.066	0.085	0.095	0.112	0.096	0.078
50	Clustering coefficient	−0.055	−0.091	−0.002	−0.017	−0.012	−0.006
50	Eigenvector centrality	0.048	0.120	0.073	0.054	0.008	0.009

The table shows the effect sizes in high- and low-density networks when 10% and 50% of individuals are identifiable in all three populations (37, 104 and 591 individuals). All metrics are calculated in weighted networks.

variation in individual social network positions, which is likely to reduce variation in network metrics relative to many real-world animal dynamic social systems, and means these results may be somewhat conservative. The relative precision of metric values in partial networks is especially robust for more local measures of social network position, with the precision of metrics such as betweenness that account for global properties of the network being more affected in small partial networks. Significantly, there was an increase in reliability as the simulated population grew in size. Thus, in long-term study populations, which typically have larger study populations than those traditionally used for network analysis, the results obtained using partial networks can be useful in providing key insights into variation in social network position among individuals. Previous work elsewhere in the social network literature has demonstrated the potential for changes in network structure to influence the consequences of sampling from a network (Frantz et al., 2009), so a crucial next step will be to extend these results to the more highly structured animal social networks such as those found in primates.

Accuracy of Individual Level Metrics

Some network metrics remain accurate in partial networks, or are inaccurate in a predictable manner, even when a low proportion of a population is identifiable. The strong relationship between the proportion of identifiable individuals and the accuracy with which local measures (degree and strength) predict the actual value of an individual's centrality in reality is notable, as it means it is likely to be possible to correct for this effect if the level of marking effort in a population is known. For example, if 40% of a long-term study population of 200 individuals was known to be uniquely identifiable and an individual in the social network constructed had a degree of 12, it could hypothetically be inferred to be likely to have a degree of 30 in a full network. It is also clear that measurements of eigenvector centrality remain similar regardless of the proportion of a population in a partial network, so that these values could be universally considered fairly accurate. This knowledge could be useful, for example, in cases where social networks of two study populations varying in the proportion of marked individuals were to be compared. However, for both betweenness and clustering coefficient there is a complex interaction between marking effort and network density that would make using corrections to estimate an individual's real values for these metrics highly unreliable in natural study populations, and making correcting for the effect of subsampling on values of these metrics unfeasible using these methods. Elsewhere in the social networks literature there has been a greater focus on correcting for missing data in networks (e.g. Huisman, 2009; Koskinen, Robins, Wang, & Pattison, 2013; Smith, 2012), and it seems likely that incorporating these approaches that use imputation (Huisman, 2009) or exponential random graph models (Smith, 2012) could be of great benefit to improving the robustness of social network analysis in animal social systems.

Precision of Individual Level Metrics

In the majority of animal behaviour research understanding the relative differences between individuals in social network position will be key, making our conclusions about the precision of network metrics in partial networks especially important. Both the precision of the predictions made using partial networks and the lack of bias caused by variation in metric values are resilient to the number of identifiable individuals in the population, making our results comparable to the findings of Franks, James and Ruxton (2010) that reduced detectability (rather than identifiability) also had a limited effect on various network measures. If half or more of the individuals in a population can be recognized then all five of the social network metrics investigated can be used informatively with high levels of confidence in all population sizes investigated (with the exception perhaps of eigenvector centrality in very small networks). Below this level of marking effort, simple measures of centrality remain reliably precise at all population sizes, while betweenness and clustering coefficient become consistently less precise in the predictions they produce. This effect is more pronounced in smaller populations for both metrics, and in high-density networks for clustering coefficient. While the mean correlations between partial networks and the real situation remain relatively strong even for these metrics, the high levels of variation that occur around these values would mean their use as predictive measurements of individual social position for single networks would have a high risk of being unreliable. As a result, care is advised in selecting which social metrics should be used in these circumstances, especially in small study populations.

Individual Level Metrics in Small Partial Networks

This work reveals differences between social metrics in their precision when a low percentage of individuals in a population are identifiable (30% or less). Although basic measures of centrality remain highly informative when few individuals in the population are individually identifiable, the precision of betweenness and clustering coefficient declines considerably, especially if the study population is relatively small. It is likely that these measures perform far less well in small populations and marginally less well in large populations as they are more global measures and connections between other individuals in the network are of more significance, thereby increasing the issues associated with subsampling relational data (Alba, 1982; Croft et al., 2008). Measures of betweenness, for example, can be highly dependent on the presence or absence of a single edge.

This is important as these two metrics are integral in understanding some key aspects of an individual's social position. Betweenness measures the importance of an individual in connecting different parts of the network by measuring the number of shortest paths between other individuals that pass through it (Freeman, 1977; Wey et al., 2008). This makes it a key metric for investigating the spread of disease or information through a network (Aplin et al., 2012; Hamede et al., 2009). For example, in many epidemiological studies, individuals with high betweenness are termed super-spreaders (Craft, Volz, Packer, & Meyers, 2011; Weber et al., 2013). Clustering coefficients are based on how well interconnected first-order connections are and measure the embeddedness of an individual within its local network structure, so could be important in considering repeated social interactions and familiarity (Croft et al., 2008; Wilson et al., 2013).

Given the unreliability of these metrics as indicators of real social position in small partial networks, finding alternative metrics or combinations of metrics that do not display these properties to corroborate any inferences made would be beneficial. For clustering coefficient, one option is to use the mean and coefficient of variation of the weight of an individual's network connections. Individuals that tend to interact with the same subset of other individuals have a higher mean and more variable weight of social associations, as well as higher clustering coefficients. Given the high precision of both degree and strength at all proportions of identifiable individuals, this method offers particular promise in small partial networks. Using these metrics may also avoid the lack of variation generated by using clustering coefficient in association-based networks, which typically have a relatively high density of weak connections as highlighted in the high-density networks in this study. A possible alternative to using betweenness may be to

use flow betweenness instead (Freeman, Borgatti, & White, 1991; Weber et al., 2013). This measures the proportion of total paths between all pairs of nodes passing through a focal node, and thus seems likely to be less susceptible to reductions in the proportion of identifiable individuals in a population and potentially more relevant to key ecological questions as well. Additionally, measuring the efficiency of an individual's egonetwork (Hanneman & Riddle, 2005) may provide some idea of the importance of an individual in connecting different parts of a network, and a simpler measure is also less likely than betweenness to be affected by missing individuals.

Applications and Further Work

Although this research is limited to association-based social networks (constructed using the gambit of the group assumption rather than focal observations of behavioural interactions) parameterized using a single species, there is no reason why these conclusions should not hold more generally for many other dynamic social systems, especially for larger study populations in which unidentifiable individuals are likely to be more prevalent. The social networks generated in this study varied in their density (the number of completed edges), and individual social network metrics were also variable. Furthermore, the light-bellied brent geese used as a model system form a fluid fission–fusion social system (Silk et al., 'n.d.'), typical of many similar social foragers (Aplin et al., 2012; Conklin & Colwell, 2008; Sundaresan, Fischhoff, Dushoff, & Rubenstein, 2007). As such, the study should be seen as an important starting point in researching the potential of using social network analysis in animal populations in which a substantial proportion of individuals are not individually identifiable. Further work generalizing these conclusions across a range of biologically meaningful network structures is likely to prove highly rewarding. Varying network topologies have certainly been found to influence the accuracy of network level metrics (Frantz et al., 2009), so may influence the precision of individual level metrics as well. Thus it could be speculated that the different properties of these types of network may make inferences made about individual social positions more or less susceptible to subsampling effects. One network level property that has been shown to have an adverse effect on the precision of node level metrics in partial networks is a positively skewed degree distribution (Wang et al., 2012). Although many association-based networks would not be expected to display strongly positively skewed degree distributions (e.g. the citation network used in the example above), a few animal social systems are likely to provide exceptions, especially when a small number of individuals move between cohesive social groups. Therefore, investigating similar questions about 'partial networks' in more highly structured social systems, such as those found in primates (Lehmann & Dunbar, 2009) and dolphins (Cantor et al., 2012; Connor et al., 2001; Lusseau, 2003) would be especially beneficial.

Conclusions

By influencing behavioural traits, foraging success and disease status, social network position will often be fundamental in determining body condition, stress levels (Brent, Semple, Dubuc, Heistermann, & MacLarnon, 2011) and ultimately reproductive success (Formica et al., 2012; McDonald, 2007; Wey et al., 2013). In this regard, demonstrating the potential of using social network approaches in systems in which many individuals are not identifiable is highly important. Focusing at this individual level also means that social networks do not have to provide accurate information about individuals so long as relative differences between individuals remain precise. This work used social interactions in a simulated population with fluid fission–fusion dynamics, typical of many social animals, and therefore our conclusions that partial networks provide precise information on individual social network position will be widely applicable for many species with dynamic social systems. Additionally, we highlight that the influence of network topology on the ability to use node level measures to predict social position, rather than measuring accuracy of network level properties, has received little attention in the broader networks literature. Therefore, further work extending this research into animal social networks with different topologies remains of great importance, and clearly has important applications for the analysis of animal social networks.

Acknowledgments

M.J.S. is funded by the University of Exeter and a grant from FERA Seedcorn awarded to S.B., S.B. is funded by an ERC grant 310820, D.P.C. is funded by a grant from the Leverhulme Trust (RPG-175). The Irish Brent Goose Research Group acknowledges the support of the WWT, NIEA and NPWS. We thank Jenni Sanderson, Damien Farine and one anonymous referee for useful comments on the manuscript.

References

Alba, R. D. (1982). Taking stock of network analysis: a decade's results. *Research in the Sociology of Organizations, 1*, 39–74.

Aplin, L., Farine, D., Morand-Ferron, J., Cole, E., Cockburn, A., & Sheldon, B. (2013). Individual personalities predict social behaviour in wild networks of great tits (*Parus major*). *Ecology Letters, 16*, 1365–1372.

Aplin, L., Farine, D., Morand-Ferron, J., & Sheldon, B. (2012). Social networks predict patch discovery in a wild population of songbirds. *Proceedings of the Royal Society B: Biological Sciences, 279*, 4199–4205.

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2014). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*. http://dx.doi.org/10.1038/nature13998.

Archie, E. A., Moss, C. J., & Alberts, S. C. (2006). The ties that bind: genetic relatedness predicts the fission and fusion of social groups in wild African elephants. *Proceedings of the Royal Society B: Biological Sciences, 273*, 513–522.

Borgatti, S. P., Carley, K. M., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks, 28*, 124–136.

Brent, L., Semple, S., Dubuc, C., Heistermann, M., & MacLarnon, A. (2011). Social capital and physiological stress levels in free-ranging adult female rhesus macaques. *Physiology & Behavior, 102*, 76–83.

Cantor, M., Wedekin, L. L., Guimarães, P. R., Daura-Jorge, F. G., Rossi-Santos, M. R., & Simões-Lopes, P. C. (2012). Disentangling social networks from spatiotemporal dynamics: the temporal structure of a dolphin society. *Animal Behaviour, 84*, 641–651.

Carter, K. D., Seddon, J. M., Frère, C. H., Carter, J. K., & Goldizen, A. W. (2013). Fission–fusion dynamics in wild giraffes may be driven by kinship, spatial overlap and individual social preferences. *Animal Behaviour, 85*, 385–394.

Claidiere, N., Messer, E. J., Hoppitt, W., & Whiten, A. (2013). Diffusion dynamics of socially learned foraging techniques in squirrel monkeys. *Current Biology, 23*, 1251–1255.

Conklin, J. R., & Colwell, M. A. (2008). Individual associations in a wintering shorebird population: do Dunlin have friends? *Journal of Field Ornithology, 79*, 32–40.

Connor, R. C., Heithaus, M. R., & Barre, L. M. (2001). Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'. *Proceedings of the Royal Society B: Biological Sciences, 268*, 263–267.

Craft, M. E., Volz, E., Packer, C., & Meyers, L. A. (2011). Disease transmission in territorial populations: the small-world network of Serengeti lions. *Journal of the Royal Society Interface, 8*, 776–786.

Croft, D., James, R., Ward, A., Botham, M., Mawdsley, D., & Krause, J. (2005). Assortative interactions and social networks in fish. *Oecologia, 143*, 211–219.

Croft, D. P., James, R., & Krause, J. (2008). *Exploring animal social networks*. Princeton, NJ: Princeton University Press.

Croft, D. P., Krause, J., Darden, S. K., Ramnarine, I. W., Faria, J. J., & James, R. (2009). Behavioural trait assortment in a social network: patterns and implications. *Behavioral Ecology and Sociobiology, 63*, 1495–1503.

Cross, P., Creech, T., Ebinger, M., Heisey, D., Irvine, K., & Creel, S. (2012). Wildlife contact analysis: emerging methods, questions, and challenges. *Behavioral Ecology and Sociobiology, 66*, 1437–1447.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*.

Dey, C. J., Reddon, A. R., O'Connor, C. M., & Balshine, S. (2013). Network structure is related to social conflict in a cooperatively breeding fish. *Animal Behaviour, 85*, 395–402.
Drewe, J., Madden, J., & Pearce, G. (2009). The social network structure of a wild meerkat population: 1. Inter-group interactions. *Behavioral Ecology and Sociobiology, 63*, 1295–1306.
Farine, D. R., Garroway, C. J., & Sheldon, B. C. (2012). Social network analysis of mixed-species flocks: exploring the structure and evolution of interspecific social behaviour. *Animal Behaviour, 84*, 1271–1277.
Farine, D. R., & Milburn, P. J. (2013). Social organisation of thornbill-dominated mixed-species flocks using social network analysis. *Behavioral Ecology and Sociobiology, 67*, 321–330.
Formica, V., Wood, C., Larsen, W., Butterfield, R., Augat, M., Hougen, H., et al. (2012). Fitness consequences of social network position in a wild population of forked fungus beetles (*Bolitotherus cornutus*). *Journal of Evolutionary Biology, 25*, 130–137.
Franks, D. W., Ruxton, G. D., & James, R. (2010). Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology, 64*, 493–503.
Frantz, T. L., Cataldo, M., & Carley, K. M. (2009). Robustness of centrality measures under uncertainty: examining the role of network topology. *Computational and Mathematical Organization Theory, 15*, 303–328.
Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*, 35–41.
Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: a measure of betweenness based on network flow. *Social Networks, 13*, 141–154.
Gilby, I. C., & Wrangham, R. W. (2008). Association patterns among wild chimpanzees (*Pan troglodytes schweinfurthii*) reflect sex differences in cooperation. *Behavioral Ecology and Sociobiology, 62*, 1831–1842.
Hamede, R. K., Bashford, J., McCallum, H., & Jones, M. (2009). Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters, 12*, 1147–1157.
Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California Press.
Huisman, M. (2009). Imputation of missing network data: some simple procedures. *Journal of Social Structure, 10*, 1–29.
Kohn, G. M., King, A. P., Dohme, R., Meredith, G. R., & West, M. J. (2013). Robust autumn social attributes predict spring courtship skills in juvenile female brown-headed cowbirds, *Molothrus ater*. *Animal Behaviour, 85*, 727–732.
Koskinen, J. H., Robins, G. L., Wang, P., & Pattison, P. E. (2013). Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks, 35*, 514–527.
Krause, J., James, R., & Croft, D. (2010). Personality in the context of social networks. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 4099–4106.
Lee, S. H., Kim, P.-J., & Jeong, H. (2006). Statistical properties of sampled networks. *Physical Review E, 73*, 016102.
Lehmann, J., & Dunbar, R. I. M. (2009). Network cohesion, group size and neocortex size in female-bonded Old World primates. *Proceedings of the Royal Society B: Biological Sciences, 276*, 4417–4422.
Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society B: Biological Sciences, 270*, S186–S188.
Madsen, J., Cracknell, G., & Fox, A. D. (1999). *Goose populations of the Western Palearctic: A review of status and distribution*. Wageningen, The Netherlands: National Environmental Research Institute, Denmark and Wetlands International.
McDonald, D. B. (2007). Predicting fate from early connectivity in a social network. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 10910–10914.
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*, E80–E89.
Opsahl, T. (2009). *Structure and evolution of weighted networks* (Doctoral dissertation). London, U.K.: Queen Mary, University of London.
Patriquin, K. J., Leonard, M. L., Broders, H. G., & Garroway, C. J. (2010). Do social networks of female northern long-eared bats vary with reproductive period and age? *Behavioral Ecology and Sociobiology, 64*, 899–913.
Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., de Silva, S., et al. (2013). The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology, 25*, 242–255.
R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Ramos-Fernández, G., Boyer, D., Aureli, F., & Vick, L. G. (2009). Association networks in spider monkeys (*Ateles geoffroyi*). *Behavioral Ecology and Sociobiology, 63*, 999–1013.
Schülke, O., Wenzel, S., & Ostner, J. (2013). Paternal relatedness predicts the strength of social bonds among female rhesus macaques. *PLoS One, 8*, e59789.
Sih, A., Hanser, S. F., & McHugh, K. A. (2009). Social network theory: new insights and issues for behavioral ecologists. *Behavioral Ecology and Sociobiology, 63*, 975–988.
Silk, M. J., Croft, D. P., Robertson, P. A., Tregenza, T., Fisher, D. N., Sharma, M. D., et al. (n.d.). *The causes and consequences of individual differences in social network position in a fission-fusion social system*. Manuscript in preparation.
Silk, M. J., Croft, D. P., Tregenza, T., & Bearhop, S. (2014). The importance of fission-fusion social group dynamics in birds. *Ibis, 156*, 701–715.
Smith, J. A. (2012). Macrostructure from microstructure generating whole systems from ego networks. *Sociological Methodology, 42*, 155–205.
Stanley, C. R., & Dunbar, R. (2013). Consistent social structure and optimal clique size revealed by social network analysis of feral goats, *Capra hircus*. *Animal Behaviour, 85*, 771–779.
Stumpf, M., & Thorne, T. (2006). Multi-model inference of network properties from incomplete data. *Journal of Integrative Bioinformatics, 3*, 32.
Sueur, C., Jacobs, A., Amblard, F., Petit, O., & King, A. J. (2011). How can social network analysis improve the study of primate behavior? *American Journal of Primatology, 73*, 703–719.
Sueur, C., & Petit, O. (2008). Organization of group members at departure is driven by social structure in *Macaca*. *International Journal of Primatology, 29*, 1085–1098.
Sundaresan, S. R., Fischhoff, I. R., Dushoff, J., & Rubenstein, D. I. (2007). Network metrics reveal differences in social organization between two fission–fusion species, Grevy's zebra and onager. *Oecologia, 151*, 140–149.
Wang, D. J., Shi, X., McFarland, D. A., & Leskovec, J. (2012). Measurement error in network data: a re-classification. *Social Networks, 34*, 396–409.
Weber, N., Carter, S. P., Dall, S. R., Delahay, R. J., McDonald, J. L., Bearhop, S., et al. (2013). Badger social networks correlate with tuberculosis infection. *Current Biology, 23*, R915–R916.
Wey, T., Blumstein, D. T., Shen, W., & Jordán, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour, 75*, 333–344.
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*, 1407–1414.
Whitehead, H., & Dufault, S. (1999). Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Advances in the Study of Behavior, 28*, 33–74.
Wilson, A. D., Krause, S., Dingemanse, N. J., & Krause, J. (2013). Network position: a key component in the characterization of social personality types. *Behavioral Ecology and Sociobiology, 67*, 163–173.
Wittemyer, G., Douglas-Hamilton, I., & Getz, W. M. (2005). The socioecology of elephants: analysis of the processes creating multitiered social structures. *Animal Behaviour, 69*, 1357–1371.

APPENDIX

Table A1
The densities of the six networks used in this study (three population sizes at two network densities) at three levels of filtering

Population size	Edge weight filter	Low density	High density
37	Unfiltered	0.598	0.812
37	Weights >1	0.184	0.394
37	Weights >3	0.014	0.044
104	Unfiltered	0.482	0.791
104	Weights >1	0.130	0.453
104	Weights >3	0.004	0.063
591	Unfiltered	0.548	0.773
591	Weights >1	0.163	0.408
591	Weights >3	0.003	0.039

Unfiltered networks contain all edges, 'weights >1' retains only edges between individuals that associated on two or more occasions and 'weights >3' retains only edges between individuals that associated on four or more occasions.

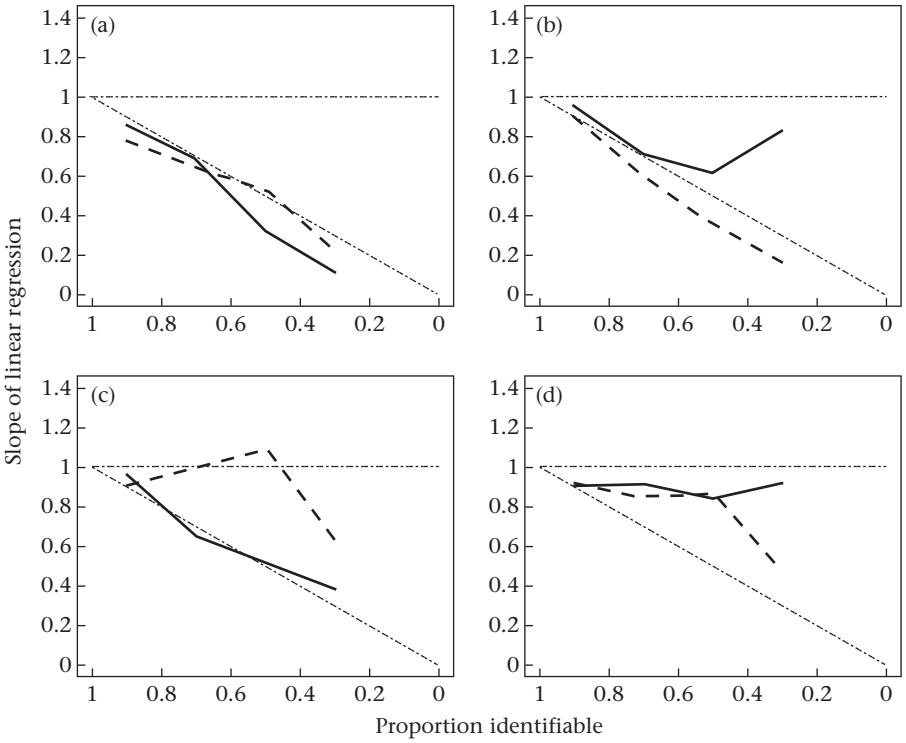

Figure A1. The relationship between the proportion of identifiable individuals in a partial network and the slope of the linear regression for four different social metrics: (a) strength, (b) betweenness, (c) clustering coefficient and (d) eigenvector centrality in a population of 37 individuals. The dot-dash lines represent the case when accuracy is independent of the proportion of identifiable individuals ($y = 1$) and the case when accuracy decreases linearly with the proportion of identifiable individuals in a partial network ($y = 1 - x$).