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
Available online 9 April 2015
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