Animal Behaviour xxx (2017) 1–12

Contents lists available at ScienceDirect

Animal Behaviour

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

Special Issue: Social Networks

Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups

William J. E. Hoppitt [a, 1], Damien R. Farine [b, c, d, *]

[a] School of Biology, University of Leeds, Leeds, U.K.
[b] Department of Collective Behaviour, Max Planck Institute for Ornithology, Konstanz, Germany
[c] Department of Biology, University of Konstanz, Konstanz, Germany
[d] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, U.K.

ARTICLE INFO

Article history:
Received 17 May 2016
Initial acceptance 30 June 2017
Final acceptance 24 July 2017
Available online xxx
MS. number: SI-16-00437

Keywords:
affiliation
animal social network
interaction
social network analysis
social structure

Social network analysis has provided important insight into many population processes in wild animals. Constructing social networks requires quantifying the relationship between each pair of individuals in the population. Researchers often use association indices to convert observations into a measure of propensity for individuals to be seen together. At its simplest, this measure is just the probability of observing both individuals together given that one has been seen (the simple ratio index). However, this probability becomes more challenging to calculate if the detection rate for individuals is imperfect. We first evaluate the performance of existing association indices at estimating true association rates under scenarios where (1) only a proportion of all groups are observed (group location errors), (2) not all individuals are observed despite being present (individual location errors), and (3) a combination of the two. Commonly used methods aimed at dealing with incomplete observations perform poorly because they are based on arbitrary observation probabilities. We therefore derive complete indices that can be calibrated for the different types of incomplete observations to generate accurate estimates of association rates. These are provided in an R package that readily interfaces with existing routines. We conclude that using calibration data is an important step when constructing animal social networks, and that in their absence, researchers should use a simple estimator and explicitly consider the impact of this on their findings.

© 2017 Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour.

A foundation of animal social network analysis is estimating the frequency that two individuals associate or interact. Social networks are typically a description of interconnections that are formed by relationships (edges) among multiple individuals (nodes). Social network analysis is a set of tools that can be used to describe the patterns formed by these interconnections or evaluate these against hypotheses (Farine & Whitehead, 2015; Whitehead, 2008). One feature of social network analysis that is perhaps unique to studies on animal populations is that researchers rarely have a complete record of all interactions or all associations (but see Boogert, Farine, & Spencer, 2014; Farine, Spencer, & Boogert, 2015; Strandburg-Peshkin, Farine, Couzin, & Crofoot, 2015). Thus, relationships are often imperfectly sampled, which can introduce uncertainty in the social network. To account for variation in sampling effort and observation frequency, Cairns and Schwager (1987) outlined commonly used association indices. These indices convert the number of observations of pairs of individuals seen associating or interacting into an association rate, representing their propensity to associate or their probability of being observed together.

Incomplete sampling of animal interactions or associations can occur due to a range of different reasons. We can classify data sets as having two possible types of missing data (Cairns & Schwager, 1987): (1) single or few observers can only collect data on one or a few groups at a time and miss many simultaneous associations or interactions occurring elsewhere, and (2) individuals are difficult to observe or identify and missed even when they are present. In type (1), while a number of pairs of individuals (also known as dyads) are being observed together in one or more groups, the status of other individuals in the population is unobserved. In type (2), when one or more groups are being observed they are incompletely sampled, resulting in data that suggest that certain dyads were not interacting or associating even when they were and could have been observed

* Correspondence: D.R. Farine, Department of Collective Behaviour, Max Planck Institute for Ornithology, 78457 Konstanz, Germany.
E-mail address: dfarine@orn.mpg.de (D. R. Farine).
[1] *E-mail address:* W.J.E.Hoppitt@leeds.ac.uk (W. Hoppitt).

http://dx.doi.org/10.1016/j.anbehav.2017.08.029
0003-3472/© 2017 Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour.

Please cite this article in press as: Hoppitt, W. J. E., & Farine, D. R., Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups, Animal Behaviour (2017), http://dx.doi.org/10.1016/j.anbehav.2017.08.029

doing so. In both cases, the relationships inferred from the observed data are likely to be influenced by the amount of data that was missed. However, the propensity for each type of missing observations to impact our estimates of association or interaction rates and social network structure remains to be properly explored.

Controlling for missed observations is one of the most important steps in social network analysis. Using simulated data, Franks, Ruxton, and James (2010) identified the impact of missing observations when constructing social networks. They found that missing observations between known individuals was more problematic than missing individuals altogether, and concluded that social network sampling should maximize the amount of data collected about known individuals rather than maximizing the number of individuals sampled. One reason for this is because a key component of social networks, weak edges, are often disproportionately likely to be missed, and leaving these out can have profound implications on the structure of the social network (Granovetter, 1973). These findings are also supported by the work of Silk, Jackson, Croft, Colhoun, and Bearhop (2015) who explored the effect of completely missing individuals in the social network. They found that, with adequate sampling, having as few as 30% of individuals known can be enough to produce informative networks for hypothesis testing.

Missing observations that could have been recorded can have large impacts on the social network that is generated, and these impacts are made worse when particular individuals are missed more often than others. Farine and Whitehead (2015) recently demonstrated how small differences in the likelihood of observing individuals of different classes can introduce systematic biases in their social network. They first simulated observations of individuals associating with preferred and avoided associates. They then introduced a small observation bias, in this case reducing the probability of observing one of two classes of individuals to 80% by removing 20% of the observations of those individuals. This resulted in a significant effect of class on degree (the sum of the association strengths in the nodes with intact data was higher than in the nodes where data had been removed). This means that the social network estimated for the individuals in this population is incorrect.

In this paper, we theoretically re-evaluate existing association indices and derive new measures to deal with missing observations of groups, missing individuals in groups, and the combination of these. We show that the extent that existing association indices adjust estimates of association strength is entirely arbitrary, and are as likely to overcorrect any bias that might occur as they are to reduce it. Existing association indices can also perform poorly at estimating relative association strengths, which has implications for many social network studies. We instead derive improved association indices that enable researchers to correct properly for the biases arising from group location error and individual identification error, and discuss how to collect appropriate calibration data. Finally, we provide an R package 'assocInd' that allows researchers to calculate accurate association indices for pairs of individuals from their observation data, and to simulate the effects of different types of errors on estimates of associations.

THE SIMPLE RATIO AND THE HALF-WEIGHT INDEX

The purpose of an association index is to estimate the proportion of time any two individuals, a and b, spent associated. Association indices typically range from 0 (the two individuals were never observed together) to 1 (the individuals are always seen together). The resulting association rates are often used as a proxy to quantify the propensity for pairs of individuals to interact (Farine, 2015; Whitehead & Dufault, 1999), although the assumption that individuals interact in proportion to their association rate should be considered on a case-by-case basis (Castles et al., 2014). Association data are collected by repeatedly sampling the population, and recording who is associating with whom in each sampling period. For any two individuals, we can then calculate:

x : the number of sampling periods with a and b observed associated.
y_a : the number of sampling periods with just a identified.
y_b : the number of sampling periods with just b identified.
y_{ab}: the number of sampling periods with a and b identified but not associated.
y_{Null}: the number of sampling periods with neither a nor b identified.

In an ideal scenario, every individual is seen and correctly identified in every sampling period, such as in many captive populations, or at least we have the situation where $y_{Null} = 0$. Intuitively, in the ideal scenario, researchers can validly use the simple ratio index (SRI), $x/(y_a + y_b + y_{ab} + x)$, as an estimate of the proportion of time a and b spend together. However, when errors arise from missing observations, it is less clear that the simple ratio is appropriate. The most commonly used approach for correcting association indices to account for missing observations is to reduce the weighting given to observations of just one individual (because we have a lower confidence in these). Because missing observations are widespread in behavioural research, many researchers use the half-weight index (HWI): $x/\left(\frac{1}{2}(y_a + y_b) + y_{ab} + x\right)$. This index is believed to correct for the biases arising from such error, in particular when individuals are relatively more likely to be detected when they are apart than when they are together. When investigating the performance of association indices, Cairns and Schwager (1987) found that the HWI resulted in lower bias and lower error for a given estimate than the simple ratio when observations were missed. However, while this has served as useful justification for many researchers, it is also important to note that Cairns and Schwager (1987) reported up to four times greater error in the HWI than what they achieved using a maximum likelihood function (see also below). Furthermore, they noted a number limitations of association indices arising from hidden assumptions.

Here we revisit some of the assumptions of the half-weight index. Notably, we show that the extent to which the half-weight index adjusts estimates of association is entirely arbitrary, and is as likely to 'overcorrect' any bias that might occur as it is to reduce that bias. Note that an alternative variant to the HWI, the twice-weight index (TWI), $x/(2(y_a + y_b) + y_{ab} + x)$, is a monotonic function of the HWI and thus we do not investigate it in this paper. Ginsberg and Young (1992) previously raised the issue that the HWI and TWI use arbitrary weightings, and they predicted that association indices will continue to be widely used. Indeed, the HWI is still the most commonly used index in animal social network studies.

A first consideration is to determine what constitutes an association. Researchers commonly make the simplifying assumption that all individuals observed together (and apart from other individuals, i.e. in a group) are associated with each other (the gambit of the group). However, the appropriate definition of 'group' will depend on the species being studied and the question being addressed (Whitehead, 2008). For some questions, a 'group' need not be a social group but may simply be an aggregation of individuals in close proximity (e.g. when determining who is likely to have had an opportunity to interact with, or observe and learn socially from, another individual). When using this gambit to collect association data, researchers list the individuals in each group found for each sampling period (henceforth 'group-level data'). For simplicity, in our discussion and terminology we assume

Please cite this article in press as: Hoppitt, W. J. E., & Farine, D. R., Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups, Animal Behaviour (2017), http://dx.doi.org/10.1016/j.anbehav.2017.08.029

Download English Version:

https://daneshyari.com/en/article/8488705

Download Persian Version:

https://daneshyari.com/article/8488705

Daneshyari.com