Behav Ecol Sociobiol (2011) 65:1659–1668
DOI 10.1007/s00265-011-1193-3

METHODS

Determining association networks in social animals: choosing spatial–temporal criteria and sampling rates

Hamed Haddadi · Andrew J. King · Alison P. Wills · Damien Fay · John Lowe · A. Jennifer Morton · Stephen Hailes · Alan M. Wilson

Received: 5 November 2010 / Revised: 28 February 2011 / Accepted: 11 April 2011 / Published online: 3 May 2011

Abstract Social Network Analysis has become an important methodological tool for advancing our understanding of human and animal group behaviour. However, researchers tend to rely on arbitrary distance and time measures when defining 'contacts' or 'associations' between individuals based on preliminary observation. Otherwise, criteria are chosen on the basis of the communication range of sensor devices (e.g. bluetooth communication ranges) or the sampling frequencies of collection devices (e.g. Global Positioning System devices). Thus, researchers lack an established protocol for determining both relevant association distances and minimum sampling rates required to accurately represent the network structure under investigation. In this paper, we demonstrate how researchers can use experimental and statistical methods to establish spatial and temporal association patterns and thus correctly characterise social networks in both time and space. To do this, we first perform a mixing experiment with Merino sheep (*Ovis aries*) and use a community detection algorithm that allows us to identify the spatial and temporal distance at which we can best identify clusters of previously familiar sheep. This turns out to be within 2–3 m of each other for at least 3 min. We then calculate the network graph entropy rate—a measure of ease of spreading of information (e.g. a disease) in a network—to determine the minimum sampling rate required to capture the variability observed in our sheep networks during distinct activity phases. Our results indicate the need for sampling intervals of less than a minute apart. The tools that we employ are versatile and could be applied to a wide range of species and social network datasets, thus allowing an increase in both the accuracy and efficiency of data collection when exploring spatial association patterns in gregarious species.

Keywords Spatial–temporal associations · Social networks · Sampling rate · Social behaviour · Sheep · Flocking

Communicated by J. Krause

H. Haddadi · A. J. King (✉) · A. P. Wills · J. Lowe · A. M. Wilson
Structure and Motion Laboratory,
The Royal Veterinary College, University of London,
Hatfield, Hertfordshire AL9 7TA, UK
e-mail: ajking@rvc.ac.uk

D. Fay
Computer Laboratory, University of Cambridge,
William Gates Building,
Cambridge CB3 0FD, UK

A. J. Morton
Department of Pharmacology, University of Cambridge,
Tennis Court Road,
Cambridge CB2 1PD, UK

S. Hailes
Department of Computer Science, University College of London,
Gower Street,
London WC1E 6BT, UK

Introduction

Social and life scientists are often interested in first tracking, and then quantifying, describing and comparing the structure of spatial–temporal associations among their subjects. To track subjects over time, researchers have traditionally used scan or focal animal sampling of marked or individually identifiable individuals by direct observation (Altmann 1974). More recently, several technological innovations have revolutionised the way in which social behaviour is studied. Researchers can now monitor indi-

vidual positions using Global Positioning Systems (Nagy et al. 2010; Riding et al. 2009; Yasuda and Arai 2005), video tracking software (Delcourt et al. 2009; Miller and Gerlai 2008), radio-frequency identification tags and bluetooth devices (see Krause et al. 2011 for a review). Once this positional data has been collected, powerful new techniques—collectively known as Social Network Analysis (SNA)—can be used to analyse these data (Vital and Martins 2009; Whitehead 2009) and explore the fine-scale social structure of their study systems (Krause et al. 2009; Wey et al. 2008). Briefly, SNA allows researchers to represent individuals (or groups, or populations) as 'nodes' and relationships among nodes as 'edges'. This basic structure provides a framework from which scientists can and make comparisons within or across groups and populations of individuals (Kasper and Voelkl 2009; Faust and Skvoretz 2002). SNA, thus, has the potential to answer a variety of fundamental behavioural, ecological and conservation led questions.

To generate a spatial network, a biologist will typically take spatial data and extract the number of occasions an individual is observed in the same 'space' as another individual over a given sampling regime. A filter may be applied to a network to categorise associations as 'above average' and 'below average', hence reducing the spatial relations to different clusters or (sub)groups (Franks et al. 2010). Presenting spatial information in this way can be extremely informative where researchers are interested in relational data for animal systems that display high fission–fusion dynamics (Aureli et al. 2008). For example, colonies of forest-dwelling big brown bats (*Eptesicus fuscus*) and Bechstein's bats (*Myotis bechsteinii*) consist of collections of individuals that roost in different trees. Network analyses of these roosting associations has revealed preferential associations between individual bats, despite the fact that party compositions at roosting sites frequently change (Kerth and Konig 1999; Willis and Brigham 2004).

For groups that display lower levels of fission–fusion dynamics, but high variability in the spatial relations within groups, a dichotomous 'present' versus 'absent' description of associations between individuals is unlikely to be informative. For example, troops of primates can show extremely stable group membership through time, but high variability in the spatial relations within these groups (see Aureli et al. 2008 for a review). Understanding how this non-random mixing of individuals impacts on the fine structure of animal groups, communities and populations is important, since it can have important consequences for understanding information or disease transmission or the evolution of individual differences in behaviour (Krause et al. 2010; Sueur et al. 2011). Two crucial questions then arise: (1) What spatial–temporal criterion is appropriate for differentiating among spatial associations within a stable group structure, and (2) how often should an observer record spatial association according to this criterion once defined?

Where researchers want to quantify spatial associations because they are interested in exploring the spread of information or disease in their study system, a spatial–temporal criterion and sampling frequency that defines a spatial link between two nodes can be defined explicitly. In the case of airborne transmission of an infectious disease for example, the criteria and sampling frequency are linked to the biological features under investigation (Star 1999). However, it is not always possible to collect these data unless direct observations are performed for types of associations (Drewe 2010). Furthermore, in some cases, association distances are determined by rather inaccurate wireless range of measurement tools (e.g. range of Radio-Frequency Identification (RFID) tag contact distances: Pásztor et al. 2010).

Where relational data are not being used to explore transmission of disease of information, but being used to investigate whether a particular social system displays assortative mixing (and whether this confers an adaptive advantage: Krause et al. 2010; Sueur et al. 2011), the temporal–spatial criteria and appropriate rate of sampling are even less clear. Once again, association distances are often defined according to the methodological tools being used. For instance, when studying association patterns of academic researchers in a conference environment Hui et al. (2005) were only able to look at variability within the ranges defined by bluetooth contact distances, since data were collected by mobile devices carried by participants. Otherwise, spatial associations are defined by initial observations of group sizes and behaviour. For example, Croft et al. (2004) studied social networks in the guppy (*Poecilia reticulata*), and all fish that were found together in a shoal (defined as two or more fish within four body lengths) were deemed to have a direct network connection. This binary assumption was based on the fact that guppy shoals are sufficiently small to allow all individuals in a shoal to interact directly (Croft et al. 2003). If association distances are defined incorrectly or sampling rates are too low, then this will affect network measures in different ways, resulting in network properties being misrepresented (Perreault 2010; James et al. 2009).

In this paper, we propose novel optimisation methodologies for choosing spatial–temporal criterion and sampling rates. We present positional data for an archetypical gregarious mammal—the merino sheep—collected at 1-s intervals using a novel individual-mounted Global Positioning System (GPS) Inertial Measurement Unit (IMU). Using a simple mixing paradigm and network statistics, we first identify a spatial–temporal criterion for relational data. Specifically, we created three groups of sheep from a large founder flock, and isolated them for 2 weeks. We then mixed these three groups together into a single cohort. We

expected that the familiar individuals would be able to recognise another (Kendrick et al. 1996, 2001; Ligout and Porter 2004) and so, pre- and post-mixing, we tracked the spatial associations of all individuals (at 1-s intervals). Using these data, we attempted to identify the distance and time period over which we can best identify clustering of the familiar individuals (in each of the three groups) when mixed. This was achieved using k-means community detection algorithm for picking out natural divisions or subgroups in networks (Girvan and Newman 2002; Newman 2002). Using this defined spatial–temporal distance, we then tracked the network structure of the flock over the first 4 h when they were mixed. This period incorporated four distinct periods of sheep activity. Initially, the newly mixed sheep flock was held together in a holding pen. Then, the flock was herded 1 km toward a novel field. The flock was then released into an open field. The data when the mixed flock was released into the novel field was divided into two distinct activity periods: entry to the field—high activity—and settling in the field—low activity. Over these four distinct activity periods, we explored how the properties of the sheep spatial network altered and determined a minimum sampling regime required to accurately capture network properties of our dataset across these different time phases.

Methods

Study subjects and site

Field experiments were undertaken at the South Australian Research & Development Institute (SARDI) at Turretfield, South Australia, during 2 months in 2010. Study subjects (n=46) were taken from a flock of n=300 merino sheep (*Ovis aries*) grazed at SARDI, and split into three groups: group A (n=10), group B (n=18) and group C (n=18). Sheep were kept in these three groups for 2 weeks in identical sized 0.9-km^2 rectangular fields, and given ad libitum access to hay and water. The three groups were then mixed together (see Mixing experiment).

GPS positional data

Positional data were collected via a GPS detector/IMU data logger carried by individual sheep. Components were mounted and housed in a sealable plastic box and attached to a standard sheep harness, which was worn by all subjects for the entire duration of the experiment (Fig. 1). Together, these had a total mass 530 g (150 g data logger, 381 g harness), which was 1% of mean sheep body mass: sheep mass±standard deviation mass was 52±6 kg, and has been shown not to significantly alter key locomotion parameters of sheep within this managed population (Hobbs-Chell et al., under review). To attach the loggers to individual sheep, groups were herded with the aid of a sheep dog into a holding pen each morning (08:30), and their individual data loggers were attached to the harness using Velcro (Fig. 1) before the sheep were released back to their original field. At the end of the day (17:00), the sheep were temporarily returned to the holding pen for the data loggers to be removed, and positional data was downloaded. Each logger had a unique serial number matched to sheep identity, inserted into the header of all data files and file names produced by the logger.

Fig. 1 GPS modules fitted to harnesses on sheep

Data loggers are an in-house design and comprise a GPS module capable of recording single frequency L1 raw range data at 10 Hz (uBlox LEA-4T GPS module), an IMU comprising a three axis MEMS accelerometer, three axis of MEMS gyroscope and three axis of magnetometer and a GPS patch antenna, MSP430 microcontroller and a rechargeable 2,200 mAh lithium polymer battery. An earlier version is described and evaluated in Tan et al. (2008). Each unit logged pseudo range, Doppler and carrier phase data (RAW GPS data) for each satellite at 1 Hz and stored the data to a micro-SD card. A Novatel FlexPak G2L/OEM4 GPS base station was also mounted with a clear sky view on top of a grain silo at the location (approximately 6 m above ground level). Pseudo range, Doppler, carrier phase and ephemeris data were simultaneously recorded at the base station. GPS data for loggers and base station were post-processed in differential mode using Waypoint GrafNav v8.10 (www.novatel.com). This approach allows carrier phase ambiguity resolution/a fixed integer kinematic solution (Kaplan and Hegarty 2008) and an absolute positional accuracy of about 10–20 cm. Much of this error will be consistent across loggers and Gaussian in nature so the relative sheep positions will be more accurate especially after smoothing. Data were then processed using Matlab version R2010. Infrequently, losses of GPS resolution occurred, which results in a drop in accuracy. These points were removed and remaining points interpolated, so that there were no missing points or abnormal jumps in the GPS data. (The positional data were of high quality with fixed ambiguities about 80% of the time and the remainder good quality floating ambiguity DGPS fixes). Data synchronisa-

tion was facilitated by recording GPS time with each position record.

Mixing experiment

Data loggers recorded the individual positions of sheep in their respective groups (A, B and C) from 9 AM to 5 PM on pre-mixing day. On the day of mixing, the three groups were brought together in a holding pen for a period of 20 min, before being herded toward a large (unfamiliar) 2.5 km^2 field which took a further 60 min. Once the sheep flock arrived at the novel field, they were released and monitored for 190 min, which could be divided into two distinct activity periods: entry to the field—high activity—lasting 20 min, and settling in the field—low activity—140 min. These two activity periods were determined on the basis of initial analyses of network properties over this period (see Sampling regime and network entropy rate).

Inter-individual sheep distances

Using our GPS data, we created spatial matrices detailing the straight line distance between all dyads in the flock (when housed individually as three flocks, and when mixed as a single flock). These data provided us with the total number of seconds sheep *i* spent at a certain distance from sheep *j*. We created spatial adjacency matrices at a number of pre-defined spatial–temporal scales. Preliminary observations of sheep flocking at the site and previous research on Merino sheep suggested that individuals tend to be spaced 1–3 m during normal activity (Lynch and Hinch 1992). We, therefore, calculated adjacency matrices for 30 different spatial–temporal scales that ranged from two individuals spending 1 min at 1 m from one another, to five consecutive minutes at 3.5 m. Dyads were therefore defined as 'associating' every time that criterion was met.

K-means clustering algorithm

We expected that upon mixing in a large open field, all else being equal, sheep would initially associate with those individuals whom they are most familiar. Thus, we expected to be able to identify clustering of sheep into three distinct sub-groups within the network when mixed. We ran a k-means clustering algorithm using our un-weighted (binary) adjacency matrices at 30 different spatial–temporal scales. The clustering output from each simulation was then compared to the composition of the original groups, and the accuracy of the clusters identified expressed as the proportion of individuals correctly assigned. Using these values, we were able to identify the spatial–temporal distance at which the original three groups of sheep could be identified within the larger, mixed flock over our whole dataset. Since the sheep movements were partially restrained in the first 50 min of our experiment, which could degrade the accuracy of the k-means algorithm, we repeated our analyses with these first 50 min removed. Data for the two periods were qualitatively similar (i.e. the best performing spatial–temporal scale was the same in each case), so we present our analyses over the whole experimental period. Sociograms representing our data (network graphs) were drawn using Pajek software (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

Sampling regime and network entropy rate

Having identified the spatial–temporal scale which appeared to represent meaningful spatial associations among individuals, we explored how different sampling regimes a researcher may employ in the field would affect the 'information' captured about our networks. We use the term information loosely, to represent how well an agent (a cue, a signal or a disease, for example) can diffuse through a network (Gomez-Gardenes and Latora 2008). Estimating an appropriate sampling rate to capture the information contained in a dynamic graph is not a straightforward issue. If the data are sampled at too low a rate then, information that may be of use (community structure for example), may be lost. However, interesting information is typically only known aposteriori (i.e. after the samples have been examined). Here, we took the approach of examining how the information in the network is preserved after sub-sampling. This approach thus makes no assumptions about what might be important but rather looks at how much information is preserved at a particular, lower, sampling rate. We measured preserved information by calculating the 'entropy rate', *h*, of our networks. The entropy rate is calculated by first considering the 'walk Laplacian' of the network. The walk Laplacian represents the network as a Markov chain in which a move may be made from one node, *i* (in this case, individual), to another, *j*, with a probability equal to the weight on the edge from *i* to *j* divided by the total weight of edges leaving *i*. Thus, for our binary (un-weighted) networks, a walk Laplacian, $\pi_{i,j}$, is constructed as:

$$\pi_{ij} = \frac{1}{\sum_i A_{i,j}} \tag{1}$$

where $A_{\{i,j\}}$ is the adjacency matrix of the graph. From (1) the entropy rate of a Markov chain may then be calculated as:

$$h = -\sum_{i,j} \pi_{j,i} \times W_i^{\bullet} ln(\pi_{j,i}) \tag{2}$$

where W_i is the stationary distribution of the Markov chain which is the first eigenvector of π.

We calculated the entropy rate for every second of our dataset, thus giving us a time series 'signal' of individual entropy rates. The aim was to examine how this signal degrades as the sampling rate decreases. To do this, we used Fourier transforms. Fourier transforms are a method for converting a signal (in this case, associations amongst sheep) from time domain to the frequency domain, enabling a complex set of associations to be decomposed to a simple sequence of components (Burrus 2010).

Results

Spatial–temporal criterion

Using a k-means detection algorithm, we identified familiar sheep in a large mixed flock with accuracy ranging from 0.37 to 0.85 (possible range 0–1, i.e. none to full accuracy) at a variety of different spatial–temporal scales. Small distances (<2 m) at relatively short time periods (<4 mins) performed poorly, whilst larger distances (3 to 3.5 m) performed well regardless of time periods (0.65–0.85 accuracy). However, the best performing spatial–temporal scales were in the mid-range distance and time periods of the criterion we chose to explore, with the best performing spatial–temporal scale overall being 2.5 m for three consecutive minutes (Table 1). This distance was also found to be the distance at which sheep most commonly associate both pre- and post-mixing (Fig. 2), and we generated a network of associations for each individual flock and the mixed flock using this criterion (Fig. 3).

Entropy rate

We calculated the entropy rate of our sheep association network when our three groups were mixed to examine how different sampling regimes impact on the 'information' represented in the networks. We found that the entropy rate fluctuates frequently, and more broadly, varied in accordance with the four activity phases of the sheep ('holding pen', 'herding', 'entry into field', 'in field': Fig. 4). Specifically, when the mixed sheep were confined to a holding area (10× 10 m), their opportunity for varied spatial interaction was limited, and hence, the network entropy is highest. When the sheep were moved toward their new field, the entropy rate decreases, and as they enter the field, the entropy rate fluctuates and is at its lowest level. As the mixed flock familiarises with the new environment, entropy rate then stabilises. We were able to demonstrate what these different entropy levels look visually by producing representative sociograms for each of these phases (Fig. 5). Surprisingly, the sociograms also revealed differences in the amount of mixing through these time phases. During the first activity period (holding pen), sheep showed no attraction to their familiar flockmates (Fig. 5a). When the sheep moved as a flock towards their new field (moving), the sheep showed high degree of attraction and moved with their familiar flockmates (Fig. 5b). However, once the flock entered the field, and the spatial network quickly became sparse (Fig. 4), the sheep spread out and very few dyads met our 2.5-m distance criteria of spatial association (Fig. 5c). As the sheep became familiar with their new environment, the network gradually converged and formed a connected component again, reaching similar entropy rates as observed in the holding pen (Fig. 4), and with some attraction to familiar individuals persisting (Fig. 5c).

Table 1 Performance of k-means in detecting familiar individuals once mixed together into one larger flock at 30 different spatial–temporal scales

		Distance (meters)					
		1	1.5	2	2.5	3	3.5
Time (minutes)	1	0.56	0.58	0.60	0.59	0.85	0.79
	2	0.49	0.57	0.46	0.58	0.64	0.72
	3	0.37	0.55	0.60	0.85	0.69	0.70
	4	0.52	0.58	0.81	0.80	0.70	0.70
	5	0.51	0.56	0.57	0.81	0.80	0.75

Warmer colours in the plot represent higher accuracy

Effect of sampling regimes

Taking our entropy calculations, we then used a fast Fourier transform to look at the frequency content of our data in a time series. We explored the frequency content for the whole 4-h period, and also for the different phases of mixing (holding pen, herding, entry into field and in field) individually. The frequency content was found to be a typical inverse function, meaning that the frequency content of the signal captured (in this case the adjacency matrix entropy) declined rapidly with reduced sampling regime. To illustrate this, Fig. 6 shows the accumulated frequency content as a function of sampling frequency over the four different phases of mixing. It shows that if we were to take our positional data at a rate of 0.2 Hz (i.e. every 5 s), 90% of the frequency content is preserved, no matter what the activity of the sheep flock. However, for frequency content of 70% or 80% to be achieved, the required sampling rate is more variable according to flock activity.

Discussion

In this paper, we used novel optimisation methodologies for choosing spatial–temporal criterion and sampling rates for social animals, using spatial data from sheep flocks as a

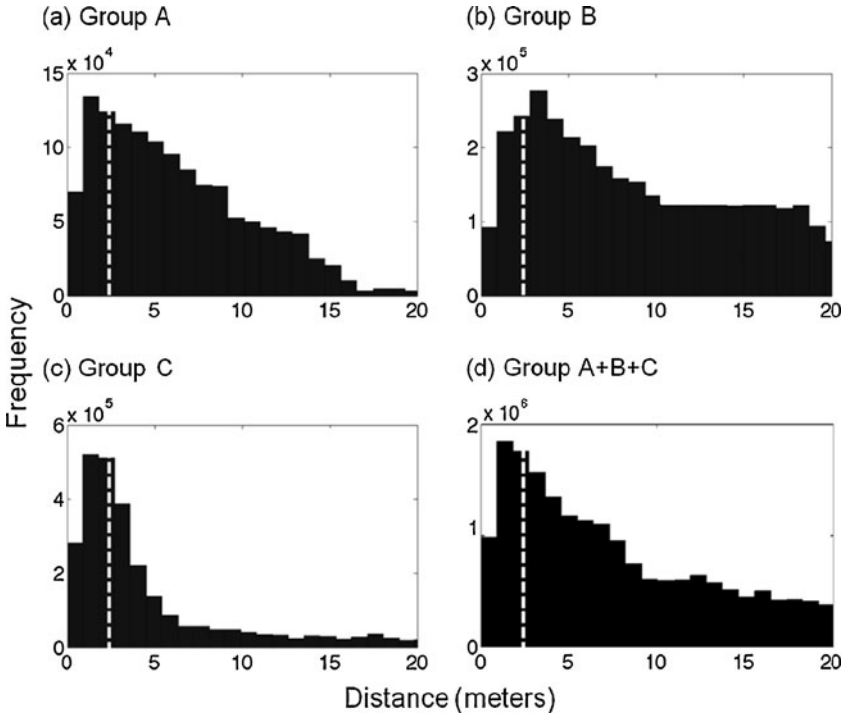

Fig. 2 Frequency histograms of spatial associations among sheep at a range of 0–20 m. Inter-sheep distances are recorded at 1-s intervals over a single day (9 AM to 5 PM period) providing data from 28,800 s. $N=10$ sheep are represented in **a**, $N=18$ in **b** and **c**, and **d** represents data from where sheep groups A, B and C were mixed together. Note that whilst frequency histograms for inter-sheep distances are highly variable across groups, each shows a high frequency of inter-sheep distance associations of 2.5 m. This is indicated by the *white dashed line*

case study. We used a k-means detection algorithm to identify the distance and time period, over which we can best identify clustering of the familiar individuals (three groups) when mixed. The algorithm performed poorly at identifying clusters at low time distance thresholds. This suggests that associations at this spatial–temporal scale are somewhat random. The most appropriate spatial and temporal indicators of an association between familiar sheep were 2.5 m or less, for 3 min or longer. Since community structure is a coarse-grained property, it is

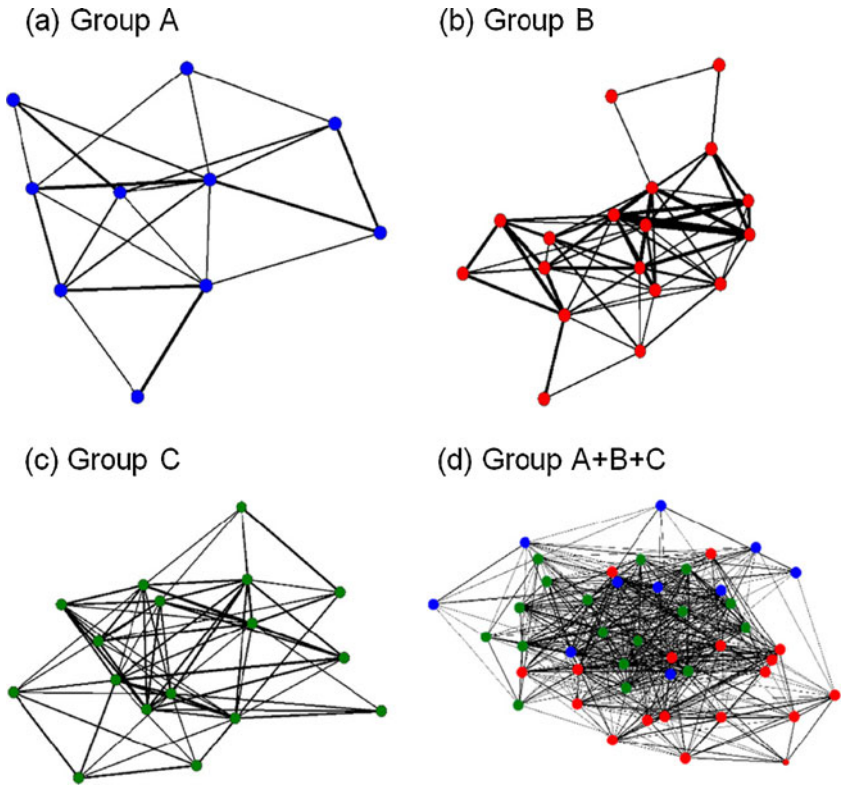

Fig. 3 Sociograms depicting spatial associations of the three individual groups before mixing **a**, **b** and **c**, and the mixed group on day of mixing **d**. In **d**, *blue nodes* represent sheep in group A, *green nodes* represent sheep in group C and *red nodes* represent sheep in group B. In all cases, the thickness of the *lines* (edges) indicates the frequency of associations between each dyad, and the network is filtered so that only links above the group mean average are shown for ease of illustration

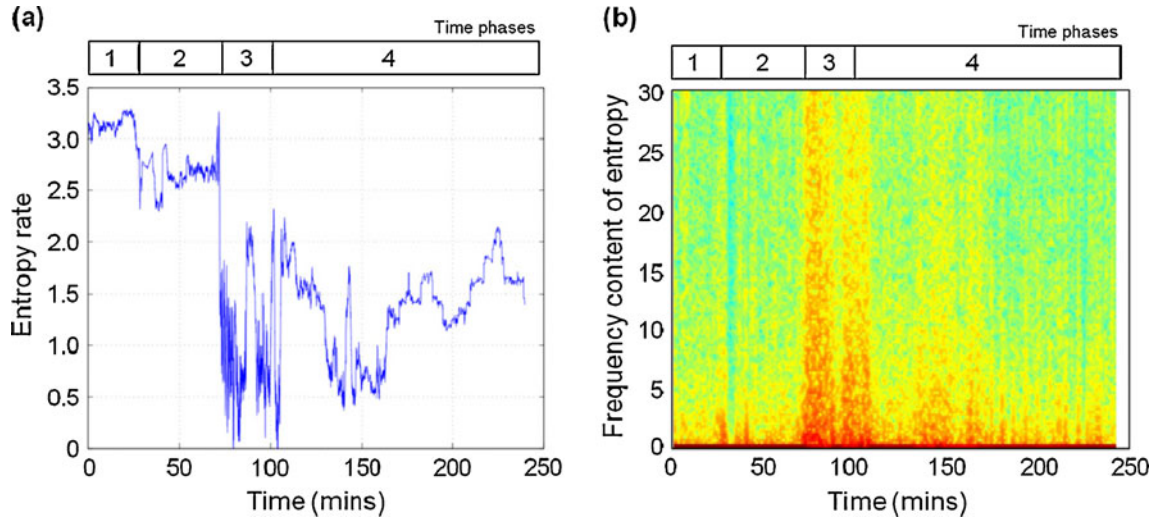

Fig. 4 **a** Entropy rate over time and **b** frequency content of entropy (a spectrogram) depicting the changing structure of the flock over time. In each figure, four distinct activity periods are labelled as time phases: 1='holding pen'; 2='herding'; 3='entry into field'; 4='in field' (see Methods for more details). At low entropy, rates indicate the flock is very dispersed, while a high entropy rate indicates a highly associated flock (also see Fig. 6). The spectogram shows that the frequency content illustrates the highly variable structure of the flock (*warm colours*) during phase 3 when the sheep flock enters the novel new field

expected that larger inter-individual distances would perform well. However, the largest distances we tested did not perform as well as intermediate distance thresholds, suggesting that associations at distances greater than 3 m may represent simple 'same space' associations. Overall, the high performance of k-means at intermediate distances (85%) suggests that these criterions represent a genuine association, underlined by the fact that this distance corresponded to the most frequent association distance for dyads across our datasets (Fig. 2).

Our exploration of the network throughout the period, which the sheep were mixed, demonstrated variability in the attraction of individuals to familiar sheep. Measuring the degree of attraction between familiar and unfamiliar individuals offers an exciting area for future research, and will be of use to researchers studying individual recognition (e.g. Taubert 2010; Thom and Hurst 2004; Wilkinson et al. 2010). Future work could use mixing experiments like we describe here to investigate how group size, familiarity (duration of associations) and context might influence association patterns among familiar and unfamiliar individuals, and provide quantitative tests of discrimination in naturalistic settings. However, it will not always be practical to perform a similar mixing paradigm as we performed with

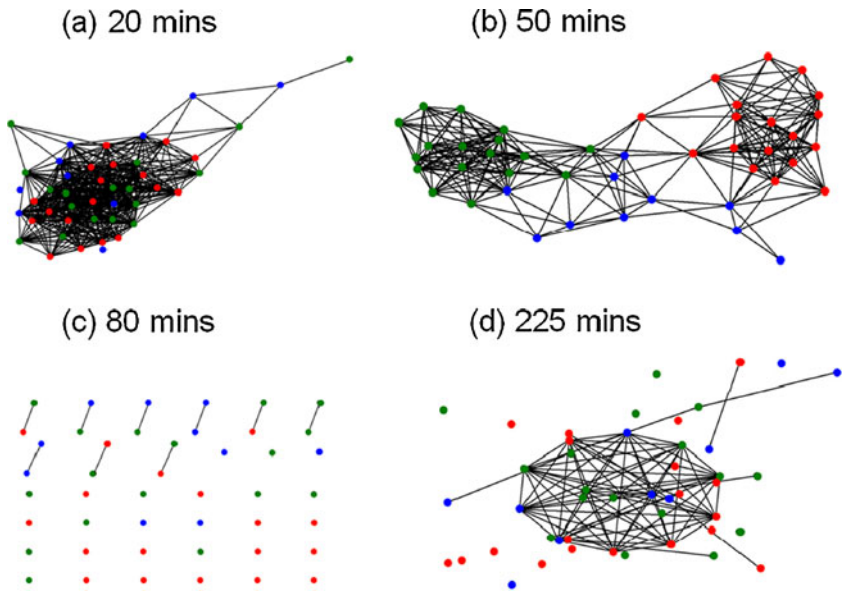

Fig. 5 Sociograms depicting spatial associations of the mixed group taken at four different single second 'snapshots' for a newly formed sheep flock. Nodes represent individual sheep and *lines* (edges) indicate an association between dyads at 2.5 m. Each network's corresponding entropy rate can be seen in Fig. 5a

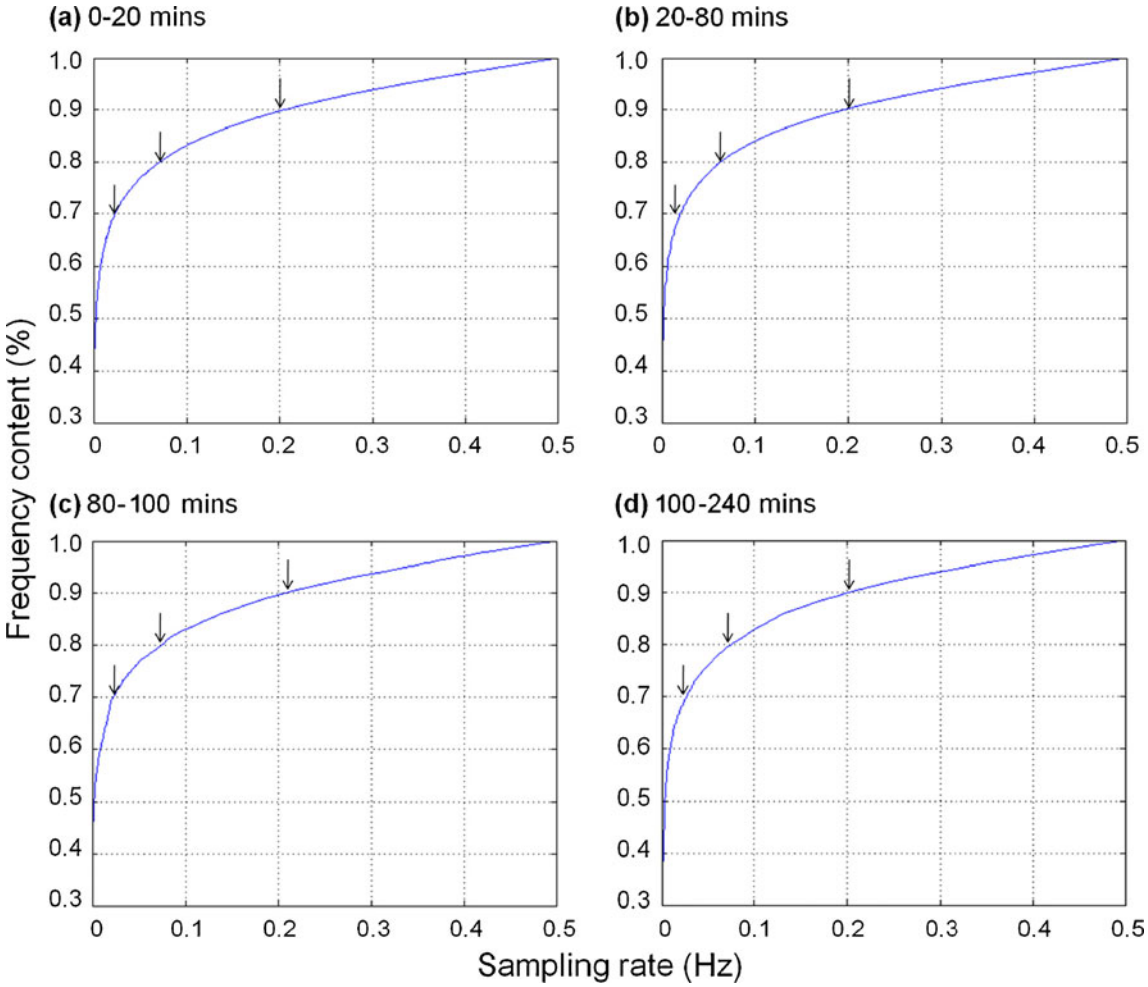

Fig. 6 The accumulated frequency content of a sheep spatial association network as a function of sampling frequency. Frequency content is shown for four distinct activity periods **a** holding pen, **b** herding, **c** entry into field and **d** in field. Note that at a sampling rate of 0.2 (i.e. once every 5 s) approximately 90% of the signal is retained regardless of sheep activity (indicated by *arrows*), but at lower sampling rates, the frequency content preserved during different activity states is more variable (information of 70% and 80% are indicated with *arrows* for comparison). Entropy rates and association networks for these data are provided in Figs. 5 and 6

the sheep here. Given that our best performing spatial criteria corresponded to the most frequently associating distance across dyads (Table 1), irrespective of group size and field size, our findings suggest that a rule-of-thumb, 'use the distance at which dyads most frequently interact to define an association', might well produce meaningful results. This may be particularly useful in the absence of any other prior information.

We have also shown that entropy rate varies considerably according to the behaviour and activity of the sheep (Fig. 4). Since entropy rate characterises the heterogeneity within a network as a single number, it can provide a robust and transferable representation of how well information or disease can spread through a network. In the case of information, it may enable researchers to model the transmission of information through a social system (Franz and Nunn 2009), not only based on the number and type of interactions (Voelkl and Noe 2010), but how variable/predictable these are over time. For disease transmission, measures of entropy can work in the same way since infectious processes may be driven by a network of contacts that is generally structured by the organization arising from behavioural and spatial heterogeneities within the group (Naug 2008). Entropy measures may also be particularly important for management of livestock. Specifically, it may offer a tool for understanding how different housing arrangements impact upon inter-individual association patterns (Buijs et al. 2010; Febrer et al. 2006; Leone and Estevez 2008).

Examining the frequency component of our entropy measure has shown that for an accurate representation of the sheep social network (and its variability), sampling more frequently than 1-min intervals is required. This rate of sampling is higher than that used in field studies of highly social animals (Altmann 1974; Rose 2000; Tyler

1979). But it is not an unreasonable rate of recording, especially where researchers employ modern automated data collection and processing tools like those used here (Tan et al. 2008), and are interested in changes in network structure over short periods. In our sheep flock example, we have shown that to achieve 70% frequency content (i.e. to capture 70% of the information about the variability in spatial associations in the flock) over short periods, a sample of sheep spatial positions would be required every 50 s (0.02 Hz; Fig. 6a, b, c). However, where researchers are interested in examining variability in spatial associations over longer periods and wish to maintain a high frequency content, sampling rates need to increase. For our sheep data, sampling at a rate of once every 15 s would be required to achieve 70% frequency content when the sheep were in their new field (phase 4, 80–240 mins; Figs. 4; 6c). Further investigations using frequency content of entropy in a network also has potential for the design of data logging devices that record spatial position. Where the aim is to produce accurate data over long sampling periods, small differences in frequency content that are observable according to activity states may enable researchers to vastly reduce power consumption in low activity periods. Indeed, sampling rates can be set to be dynamic according to the level of activity in the network, which can be correlated (and thus inferred) from acceleration data. (The data logging system employed here has a three-axis accelerometer, and movement-dependent logging rate is an area of active research in our group.)

In summary, we have presented the first attempt at empirically analysing the appropriate spatial and temporal association distances that should be used to construct and analyse social networks of highly gregarious individuals (human or non-human). The methods presented are not limited to analysis of social network in sheep and could be applied in a variety of different domains (e.g. information dissemination, disease transmission or spatial coordination for example). However, our results emphasise the need to choose sampling criteria appropriate for the species to be studied. Otherwise, as demonstrated by our k-means analysis, associations that simply represent ‘random’ or ‘same space’ associations will be included in the data. Conversely, if an incorrect sampling criterion is used then genuine associations can also be excluded. We are aware that it is not always possible to perform such controlled mixing experiment that we present here, especially in wild animals where capturing and releasing the subjects is not a viable option. In such scenarios, our data suggest that the frequency of associations may provide a useful ‘rule-of-thumb’ for defining associations in the absence of any other information and our entropy analyses demonstrate the utility of adopting as higher sampling rate as is possible. We hope that the tools that we present here can be incorporated into future studies that use social network techniques to explore spatial associations so that more meaningful data can be collected from which researchers can ask more detailed questions of their study systems.

Acknowledgements Thanks to the Structure and Motion Laboratory at the Royal Veterinary College, Kyle Roskilly for help with the GPS measurement equipment and post-processing tools and Skye Rudiger and all of the staff at South Australian Research Development Institute (SARDI) for the support of our fieldwork. Thanks also to Declan McKeever for the discussions on transmission distances of infectious diseases and to Julian Drewe and two anonymous reviewers for their constructive feedback on an earlier version of this manuscript. This work was funded by CHDI Inc.

References

Altmann J (1974) Observational study of behavior—sampling methods. Behaviour 49(3–4):227–267

Aureli F, Schaffner CM, Boesch C, Bearder SK, Call J, Chapman CA, Connor R, Di Fiore A, Dunbar RIM, Henzi SP, Holekamp K, Korstjens AH, Layton R, Lee P, Lehmann J, Manson JH, Ramos-Fernandez G, Strier KB, Van Schaik CP (2008) Fission-fusion dynamics new research frameworks. Curr Anthropol 49(4):627–654. doi:10.1086/586708

Buijs S, Keeling LJ, Vangestel C, Baert J, Vangeyte J, Tuyttens FAM (2010) Resting or hiding? Why broiler chickens stay near walls and how density affects this. Appl Anim Behav Sci 124(3–4):97–103. doi:10.1016/j.applanim.2010.02.007

Burrus C (2010) Fast fourier transforms. Connexions. http://cnx.org/content/col10550/1.21/

Croft DP, Arrowsmith BJ, Bielby J, Skinner K, White E, Couzin ID, Magurran AE, Ramnarine I, Krause J (2003) Mechanisms underlying shoal composition in the Trinidadian guppy, *Poecilia reticulata*. Oikos 100(3):429–438

Croft DP, Krause J, James R (2004) Social networks in the guppy (*Poecilia reticulata*). Proc Biol Sci 271:S516–S519

Delcourt J, Becco C, Vandewalle N, Poncin P (2009) A video multitracking system for quantification of individual behavior in a large fish shoal: advantages and limits. Behav Res Meth 41 (1):228–235. doi:10.3758/brm.41.1.228

Drewe JA (2010) Who infects whom? Social networks and tuberculosis transmission in wild meerkats. Proc Biol Sci 277 (1681):633–642. doi:10.1098/rspb.2009.1775

Faust K, Skvoretz J (2002) Comparing networks across space and time, size and species. Sociol Methodol 32:267–299

Febrer K, Jones TA, Donnelly CA, Dawkins MS (2006) Forced to crowd or choosing to cluster? Spatial distribution indicates social attraction in broiler chickens. Anim Behav 72:1291–1300. doi:10.1016/j.anbehav.2006.03.019

Franks DW, Ruxton GD, James R (2010) Sampling animal association networks with the gambit of the group. Behav Ecol Sociobiol 64 (3):493–503. doi:10.1007/s00265-009-0865-8

Franz M, Nunn CL (2009) Network-based diffusion analysis: a new method for detecting social learning. Proc Biol Sci 276 (1663):1829–1836. doi:10.1098/rspb.2008.1824

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821–7826. doi:10.1073/pnas.122653799

Gomez-Gardenes J, Latora V (2008) Entropy rate of diffusion processes on complex networks. Phys Rev E 78(6):065102

Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, Diot C (2005) Pocket switched networks and the consequences of human mobility in conference environments. In: SIGCOMM 2005, Philadelphia, PA

James R, Croft DP, Krause J (2009) Potential banana skins in animal social network analysis. Behav Ecol Sociobiol 63(7):989–997. doi:10.1007/s00265-009-0742-5

Kaplan E, Hegarty C (2008) Understanding GPS principles and applications. Artech House, Norwood

Kasper C, Voelkl B (2009) A social network analysis of primate groups. Primates 50(4):343–356. doi:10.1007/s10329-009-0153-2

Kendrick KM, Atkins K, Hinton MR, Heavens P, Keverne B (1996) Are faces special for sheep? Evidence from facial and object discrimination learning tests showing effects of inversion and social familiarity. Behav Process 38(1):19–35

Kendrick KM, da Costa AP, Leigh AE, Hinton MR, Peirce JW (2001) Sheep don't forget a face. Nature 414(6860):165–166

Kerth G, Konig B (1999) Fission, fusion and nonrandom associations in female Bechstein's bats (*Myotis bechsteinii*). Behaviour 136:1187–1202

Krause J, Lusseau D, James R (2009) Animal social networks: an introduction. Behav Ecol Sociobiol 63(7):967–973. doi:10.1007/s00265-009-0747-0

Krause J, James R, Croft DP (2010) Personality in the context of social networks. Phil Trans Biol Sci 365:4099–4106. doi:10.1098/rstb.2010.0216

Krause J, Wilson A, Croft DP (2011) New technology facilitates study of social networks. Trends Ecol Evol 26:5–6

Leone EH, Estevez I (2008) Use of space in the domestic fowl: separating the effects of enclosure size, group size and density. Anim Behav 76:1673–1682. doi:10.1016/j.anbehav.2008.08.004

Ligout S, Porter RH (2004) The role of visual cues in lambs discrimination between individual agemates. Behaviour 141:617–632

Lynch JJ, Hinch GN (1992) The behaviour of sheep. CAB International

Miller NY, Gerlai R (2008) Oscillations in shoal cohesion in zebrafish (*Danio rerio*). Behav Brain Res 193(1):148–151. doi:10.1016/j.bbr.2008.05.004

Nagy M, Akos Z, Biro D, Vicsek T (2010) Hierarchical group dynamics in pigeon flocks. Nature 464(7290):U890–U899. doi:10.1038/nature08891

Naug D (2008) Structure of the social network and its influence on transmission dynamics in a honeybee colony. Behav Ecol Sociobiol 62(11):1719–1725. doi:10.1007/s00265-008-0600-x

Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208701. doi:10.1103/PhysRevLett.89.208701

Pásztor B, Mottola L, Mascolo C, Picco GP, Ellwood S, Macdonald DW (2010) Selective reprogramming of mobile sensor networks through social community detection. Paper presented at the 7th European Conference on Wireless Sensor Networks (EWSN2010). Coimbra, Portugal

Perreault C (2010) A note on reconstructing animal social networks from independent small-group observations. Anim Behav 80 (3):551–562. doi:10.1016/j.anbehav.2010.06.020

Riding TAC, Dennis TE, Stewart CL, Walker MM, Montgomery JC (2009) Tracking fish using 'buoy-based' GPS telemetry. Mar Ecol Prog Ser 377:255–262. doi:10.3354/meps07809

Rose LM (2000) Behavioral sampling in the field: continuous focal versus focal interval sampling. Behaviour 137:153–180

Star KDC (1999) The role of infectious aerosols in disease transmission in pigs. Vet J 158(3):164–181

Sueur C, Jacobs A, Amblard F, Petit O, King AJ (2011) How can social network analysis improve the study of primate behavior? doi:10.1002/ajp.20915

Tan HL, Wilson AM, Lowe J (2008) Measurement of stride parameters using a wearable GPS and inertial measurement unit. J Biomech 41(7):1398–1406. doi:10.1016/j.jbiomech.2008.02.021

Taubert J (2010) Evidence of human-like, holistic face processing in spider monkeys (*Ateles geoffroyi*). J Comp Psychol 124(1):57–65. doi:10.1037/a0017704

Thom MD, Hurst JL (2004) Individual recognition by scent. Ann Zool Fenn 41(6):765–787

Tyler S (1979) Time-sampling—matter of convention. Anim Behav 27:801–810

Vital C, Martins EP (2009) UCINET for windows: software for social network analysis. Ethology 115(4):347–355. doi:10.1111/j.1439-0310.2009.01613.x

Voelkl B, Noe R (2010) Simulation of information propagation in real-life primate networks: longevity, fecundity, fidelity. Behav Ecol Sociobiol 64(9):1449–1459. doi:10.1007/s00265-010-0960-x

Wey T, Blumstein DT, Shen W, Jordan F (2008) Social network analysis of animal behaviour: a promising tool for the study of sociality. Anim Behav 75:333–344. doi:10.1016/j.anbehav.2007.06.020

Whitehead H (2009) SOCPROG programs: analysing animal social structures. Behav Ecol Sociobiol 63(5):765–778. doi:10.1007/s00265-008-0697-y

Wilkinson A, Specht HL, Huber L (2010) Pigeons can discriminate group mates from strangers using the concept of familiarity. Anim Behav 80(1):109–115. doi:10.1016/j.anbehav.2010.04.006

Willis CKR, Brigham RM (2004) Roost switching, roost sharing and social cohesion: forest-dwelling big brown bats, *Eptesicus fuscus*, conform to the fission-fusion model. Anim Behav 68:495–505

Yasuda T, Arai N (2005) Fine-scale tracking of marine turtles using GPS-argos PTTs. Zool Sci 22(5):547–553