

Inferring social structure from temporal data

Ioannis Psorakis^{1,2} · Bernhard Voelkl³ · Colin J. Garroway³ · Reinder Radersma³ ·
Lucy M. Aplin³ · Ross A. Crates³ · Antica Culina³ · Damien R. Farine^{3,4,5} ·
Josh A. Firth³ · Camilla A. Hinde^{3,6} · Lindall R. Kidd³ · Nicole D. Milligan³ ·
Stephen J. Roberts¹ · Brecht Verhelst³ · Ben C. Sheldon³

Received: 6 January 2015 / Revised: 11 March 2015 / Accepted: 13 March 2015 / Published online: 1 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Social network analysis has become a popular tool for characterising the social structure of populations. Animal social networks can be built either by observing individuals and defining links based on the occurrence of specific types of social interactions, or by linking individuals based on observations of physical proximity or group membership, given a certain behavioural activity. The latter approaches of discovering network structure require splitting the temporal observation stream into discrete events given an appropriate time resolution parameter. This process poses several non-trivial problems which have not received adequate attention so far. Here, using data from a study of passive integrated transponder (PIT)-tagged great tits *Parus major*, we discuss these problems, demonstrate how the choice of the extraction method and the temporal resolution parameter influence the appearance and properties of the retrieved network

and suggest a modus operandi that minimises observer bias due to arbitrary parameter choice. Our results have important implications for all studies of social networks where associations are based on spatio-temporal proximity, and more generally for all studies where we seek to uncover the relationships amongst a population of individuals that are observed through a temporal data stream of appearance records.

Keywords Social networks · Group detection · Flocks · Gathering events · Great tits

Introduction

Social network analysis has become a major analytic technique in behavioural ecology. The key motivation for employing network analysis is that the web of interconnections between individuals can provide us with invaluable insights into the underlying mechanisms that govern the system under study, as well as understanding processes that depend on emergent social structure (Newman 2010). Rooted in graph theory, it provides a wealth of measures for studying the overall properties of a web of interconnected individuals and characterising the positions of single individuals within the network. Yet, despite the advantages of the network paradigm and the wealth of analytical and computational tools available for network analysis, the problem of capturing any given system as a graph is not always trivial. Not all systems possess an obvious web-like structure (such as the Internet), where the interconnections between participating entities are apparent from direct observation (computers that are connected through physical cables). In behavioural ecology, we usually make the assumption that consistent physical proximity of individuals is a proxy for social affiliation (Wilson 1975). This approach is indiscriminative regarding the underlying causes for the spatio-temporal associations. Based

Communicated by D. P. Croft

Electronic supplementary material The online version of this article (doi:10.1007/s00265-015-1906-0) contains supplementary material, which is available to authorized users.

✉ Bernhard Voelkl
bernhard.voelkl@zoo.ox.ac.uk

- ¹ Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ Oxford, UK
- ² Thought Machine Ltd, EC2 3HU London, UK
- ³ Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS Oxford, UK
- ⁴ Department of Anthropology, University of California, Davis, CA 95616, USA
- ⁵ Smithsonian Tropical Research Institute, Ancon, Panama
- ⁶ Department of Animal Sciences, University of Wageningen, 6700 AH Wageningen, NL, The Netherlands

on co-occurrences, a link is drawn between individuals using various indices of associations (Ginsberg and Young 1992; Beijder et al. 1998; Whitehead 2008).

Traditional approaches of discovering network structure from occurrence data are based on a discretisation of the observation stream given an appropriate time resolution parameter. There are two major problems intrinsic to these approaches. The first one is referred to as the “Gambit of the Group” (Whitehead and Dufault 1999), which is the assumption that social relationships are transitive and that all individuals occurring at a locus within a certain time period are associated with each other. Problems that arise from such an assumption have been discussed extensively (Croft et al. 2008; Whitehead 2008; James et al. 2009; Franks et al. 2010). The second problem concerns assumptions about what constitutes an appropriate time resolution within which biologically meaningful social associations can be identified. This can be crucial as the choice of the time window size has marked effects on the retrieved network metrics—even when the window size falls within a range that is arguably of biological relevance. Given the weaknesses of traditional link creation approaches, Psorakis et al. (2012) proposed a methodology that exploits key statistical properties of the data stream in order to reveal a temporal modular structure of gathering events that allows retrieving groups of connected individuals. The purposes of this study are to (i) highlight the frequently overlooked problems associated with the choice of the graph construction method, (ii) compare the different available methods and (iii) discuss which approach is most appropriate for time stamped detection data that are typical of those collected by many studies (Krause et al. 2013). There is also the general problem of selecting both a temporal and a spatial resolution for our data. However, in this work, we only consider cases where there is a predefined number of sites where individuals are observed, and the concept of “co-appearance” at a given site between a set of individuals is defined only within a certain temporal distance.

Methods

Study site and data collection

Data used in this study have been collected as part of an ongoing long-term field study of great tits near Oxford, UK. All data used in this study were collected at Bagley Woods (51°42' N, 1°15' W) from a grid of 35 automated feeding stations from December 10th 2011 to February 26th 2012. Since spring 2007, all breeding adults and fledged nestlings as well as a significant proportion of immigrant birds have been fitted with small passive integrated transponder (PIT) tags in addition to metal identification rings from the British Trust for Ornithology. PIT tags are a form of radio frequency

identification (RFID) that allow automatic identification through the contactless transfer of data. RFID technology enables standardised data collection for large samples of individuals, which has broad applications for ornithological research (Garroway et al. 2015), but see Gibbons and Andrews (2004), Krause et al. (2013) and Smyth and Nebel (2013) for potential caveats of this technology. An antenna at a feeding station is used to interact with the electric circuit of the PIT tag via electromagnetic induction. In turn, the PIT tag emits a signal which encodes its unique ID; this is received by the external antenna and decoded by an attached PIT tag reader.

During the winter, PIT tag-based data logging systems are attached to automated bird feeders that are evenly distributed throughout the study site. Feeders are fitted with RFID antennae in place of perches for two access holes. When an individual takes a seed from the feeder, its identity is transmitted by the tag and stored with a timestamp by the attached data logger. Therefore, each datum consists of identity, location, date and time of detection.

By aggregating records from all feeding locations, the data generated from this scheme consists of a long stream of time stamped observations. Our spatio-temporal data D can be represented in the form $D = \{ID_z, t_z, l_z\}_{z=1}^Z$, where Z is the total number of records in our database (i.e. the number of detected visits). If we take a single record $\{ID_z, t_z, l_z\}$, we read it as “bird ID_z appeared at time t_z at feeding location l_z . Additionally, a given bird i (out of total N birds) may appear more than once in the data. That is, there can be many records z in our data $\{ID_z, t_z, l_z\}$ for which $ID_z = x$, for a given individual x . The time window approach involves placing a link between two individuals, i and j , if they are observed in the data stream at the same location and within a given time interval Δ_t . The more times the two individuals were seen together within such *time windows*, the stronger the link weight $a_{i,j}$ between them. In the following, we present three different ways in which to define the appropriate time intervals.

Fixed time windows

For the fixed time window approach, the data stream is discretised into a series of intervals of fixed length Δ_t , and for all observations $\{ID_z, t_z, l_z\}$ that fall within each time window, a link is created between the corresponding pairs of individuals (Fig. 1a). The output of the process is an undirected weighted matrix A where $a_{i,j}$ is the number of time intervals Δ_t within which individuals i and j co-occurred.

Variable time windows

For the variable time window approach, we pick an individual i and we place an “influence zone” of size Δ_t around each one of its observations (Fig. 1b). Every other individual that was observed at the same location within this time zone is assumed to be

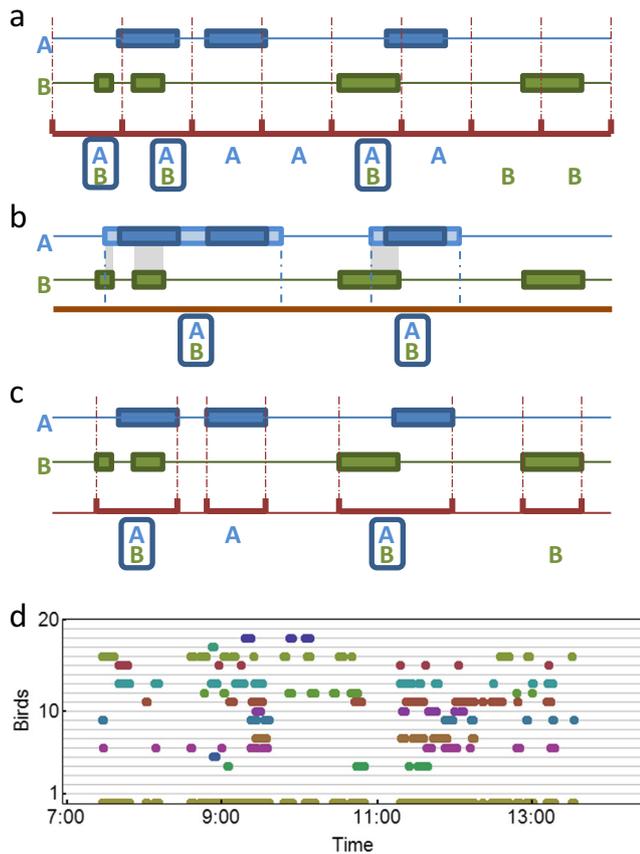


Fig. 1 Hypothetic example of the temporal data stream and the way how discretisation methods identify links of co-occurrence between individuals A and B. *Dark bars* give the real-time presence of a bird at one of the two perches of a feeder. *Letters A and B* below the time axis denote the discrete visitation events of the individuals identified by the corresponding method. *Ovals around the letters* indicate links between individuals. **a** Fixed time windows. **b** Variable time windows. **c** GMM time windows. **d** Example data stream of great tits visiting a feeder. Each line shows the presence of a specific individual; the *bottom line* gives the sum of all recordings

connected to i . In cases where we have successive observations of the same individual, with overlapping influence zones, we simply merge the intervals, as done in Fig. 1b, leading to a *variable time window scheme*. Similar to the fixed time window method, we end up with an undirected weighted graph described by an adjacency matrix A , where each $a_{i,j}$ is the number of times i and j co-occurred within a fixed temporal distance of Δ_r .

Gaussian mixture model

Under a time window model, interactions $a_{i,j}$ are defined within a given temporal distance Δ_r . The selection of this scale parameter is crucial for the extracted topology of the network: Insufficiently small time windows may omit important co-occurrences, while unreasonably large ones lead to an over-estimation of the population's social connectivity. So far, the choice of the parameter was completely arbitrary and either

justified by intuition of what might be biologically plausible, or—more commonly—not at all. Psorakis et al. (2012) suggested a more “data-driven” approach that reduces this arbitrariness by taking the choice out of the hands of the researcher. This is achieved by exploiting the heterogeneous feeder observation profile, where bird visitations are not uniformly spread across time but occur in “bursts”; periods of intense feeding activity where many individuals gather around the feeder, followed by long periods of inactivity. We consider such series of temporally focussed aggregations of birds, with long zero-observation periods between them, as *gathering events* of foraging individuals. More details regarding the identification of gathering event structure is provided in the [Supplementary Material](#).

The Gaussian mixture model (GMM) method works by firstly detecting regions of increased activity using a Gaussian mixture model, and next, the data stream is clustered into non-overlapping gathering events (Fig. 1c). Each observation $\{ID_z, t_z, l_z\}$ is assigned to one gathering event and networks are inferred, equivalently to the time windows approach, by linking all individuals that are members of the same gathering event. The purpose of the GMM is to estimate the number of independent gathering events, their centres of mass and their “borders” to their neighbouring events. This is achieved by considering each time stamp t_z as drawn from a mixture of K Gaussian distributions, and inference on number of mixtures, their centroids, density parameters and cluster membership is achieved via a Variational Bayes (VB) algorithm (more details in the [Supplementary Material](#)). MATLAB code for performing GMM analyses can be found at <https://github.com/ipsorakis/GMMevents>.

In the multiple location setting, we run GMM at each location l_z of the data stream $D = \{ID_z, t_z, l_z\}_{z=1}^Z$ separately, as a gathering event is defined only within the context of a particular location. The whole social network is then reconstructed simply by aggregating the gathering events from all locations, and connecting individuals based on their co-occurrences at those events. The connections are stored in an adjacency matrix A , where if $a_{i,j} \neq 0$, then individuals i, j are connected.

The GMM approach has not only the advantage of not requiring the researcher to guess the appropriate time window size Δ_r , but also drops the constraint that Δ_r is constant over the whole observation period. The two assumptions upon which GMM is based are that individuals visit feeders in flocks (non-uniform feeder visitation profile) and that flock membership signifies social associations to the other flock members.

Association index

Networks extracted directly from the weighted adjacency matrix A have an intuitive interpretation, as every link is weighted by the total number of times two individuals were observed together. However, such networks neither consider different

detection rates of individuals nor different levels of “gregariousness”. Different encounter rates might be due to different activity patterns, different spatial distributions or different levels of habituation to the detection devices (e.g. feeding stations), amongst other factors. Not taking encounter rates into account will lead to an underestimation of social relationships of rarely encountered individuals and to an overestimation of association strengths amongst frequent visitors (Whitehead 2008). To resolve this issue, various indices of association have been suggested (Ginsberg and Young 1992; Beijder et al. 1998; Whitehead 2008). Our approach consists of evaluating the three different scenarios that take place in our visitation data at a single feeder: (i) both individuals i and j were observed within time bin (i.e. time window or gathering event) z , (ii) only one of the two individuals was observed in time bin z and (iii) neither i nor j were present within time bin z . As we used a fully automated detection system relying on RFID tags, we can ignore the potential issue of observer bias and assume that all PIT-tagged individuals are detected with equal accuracy. We therefore used a *simple ratio association index*, which in this case is equivalent to the half weighted index (Cairns and Schwager 1987)—except for a scaling factor of 2. The index is given by $r_{i,j} = \frac{x_{i,j}}{x_i + x_j}$, where $r_{i,j}$ is the association index for individuals i and j , $x_{i,j}$ is the number of times individuals i and j were observed in the same time bin, x_i is the total number of times individual i was observed and x_j is the total number of times individual j was observed.

Gambit of the gathering

Any construction of social networks that makes use of spatio-temporal visitation records is based on the assumption that co-occurrences signify some form of social affiliation. This assumption was dubbed the “Gambit of the Group” (Whitehead and Dufault 1999), as one deliberately ignores the possibility that a co-occurrence amongst two individuals i,j can be the result of a common social tie to a third individual k but with no actual social association between i,j . By grouping several individuals, which arrived at a feeder in a sequential order and by drawing links between all the members of this group, we are assuming social ties of the same quality between consecutively arriving individuals and between individuals with more distant arrival times. Likewise, as with the gambit of the group, we are pooling links with different levels of uncertainty, and thus, we can refer to this dilemma, in analogy, as a “Gambit of the Gathering”. In our case, it is, furthermore, likely that a certain proportion of detected co-occurrences are purely incidental, not only because of the inherent stochasticity of ecological environments but also because feeding stations as employed in our study act as “attractors” to foraging individuals. An important question to be addressed in this respect is how sensible the different network extraction

methods are to certain levels of “gambitness” (i.e. incidental co-occurrences). We, therefore, seek to compare the quality of the extracted social networks given both the GMM and the time window approach and compare them against a ground-truth network. Unfortunately, ground-truth network structure is not available to us in real-world scenarios. Yet, we can compare the effect of the *Gambit of the Group* assumption on networks generated by the GMM and time window approaches using simulated data streams, which can generate bird visits given a fully observed graph. In such artificial logger data, individuals that are connected with each other in the ground-truth graph are placed in close temporal proximity, while individuals with no associations are placed further apart. We also induce a certain level of noise, or *gambitness*, so that individuals with no social affiliation are positioned close to each other with a given probability. We apply both the GMM and the fixed time window method on simulated data streams and compare the extracted networks versus the ground truth one.

Example data

In this paper, we analyse a single sample of data collected from the study population, as described above. The data were collected on Saturday February 4, 2012 and Sunday February 5, 2012 and involve 72,319 detections of 195 individual great tits (other tit species were also tagged and detected, but are not considered further) at 35 different loggers. This weekend was chosen as an example because it was a weekend with the highest numbers of visits per weekend for this season (range, 26,297–72,319) and the highest number of recorded great tits per weekend (range, 97–195).

Results

Impact of window sizes

First, we constructed weighted networks using a fixed time window approach for time window sizes from 1 to 16 s and furthermore for 20, 30, 60, 100, 150, 200, 300, 600, 1000, 1500, 2000, 3600, 5000, 9000, 18,000 and 36,000 s. For each resulting network, we calculated four different network measures: mean degree centrality (Wasserman and Faust 1994), mean edge weight disparity (Barthelemy et al. 2005), mean betweenness centrality (of the unweighted graph, Freeman 1979) and mean clustering coefficient (Watts 1999). Additionally, we constructed networks using variable time windows with influence zones of 1–15 s and furthermore for 20, 25, 30, 45, 60, 120, 250, 500, 1000, 1500, 2500, 3600, 5000, 7500, 10,000, 15,000, 18,000 and 20,000 s on each side of the visitation bout and calculated the same network measures likewise.

Figure 2a–c shows example networks based on time window sizes of 1, 5 and 3600 s (1 h), respectively. As can be seen, the three graphs differ markedly from each other despite the fact that all three are derived from exactly the same set of raw data. The differences between the three graphs a–c results only from cutting the data stream into time windows of different lengths. The impact of this on the resulting network was clearly reflected in the computed network measures, which varied widely depending on window size (Fig. 2d–g). As these data show, betweenness centrality shows strong fluctuations up to time windows of approximately 30 s, while mean clustering coefficient and degree centrality show the largest increase over that time window range.

A similar picture emerges when we compare networks created using variable time windows with influence zones of different sizes (Fig. 2h–j) and calculate the corresponding network measures (Fig. 2k–n). Comparing the results for fixed and variable time windows, one can note that the estimates based on variable time windows seem to fluctuate a little bit less, though the overall patterns are the same. Thus, for this specific case, we would argue that any time window size between 30 s and 1 h would produce rather similar graphs while networks based on smaller or larger time windows may deviate substantially.

Applying the GMM method, we get 1064 gathering events at 22 different locations on the first day and 1194 gathering events for 23 locations on the second day, which makes on average 100 gathering events per location for the whole weekend with an average duration of one gathering event of 232 (SD±157)s. The retrieved network based on the gathering event method (Fig. 3) has a density of 0.08, edge weight disparity of 0.16, betweenness centrality of 209, degree centrality of 16.9 and clustering coefficient of 0.81.

One general problem of all data stemming from a continuous time stream is that successive data points from a single individual are not statistically independent. How much the presence of a bird at a feeder at a time point $t + \Delta_t$ is determined by its presence at time point t depends on the size of Δ_t and can be estimated by the temporal autocorrelation coefficient r . We calculated for each bird and each day of logging the autocorrelation coefficient for time lags from one to 240 s (Fig. 4a). As expected, the correlation is highest for very small time lags: If a bird is recorded at a feeder, it is likely that it will be still at the feeder a second later. The temporal autocorrelation reaches a minimum around 20 s, reflecting the feeding behaviour of the birds, which do not stay at the feeder once they recovered a seed, but fly off to a nearby tree in order to open and eat it there. Thereafter, the correlation coefficient rises again before it slowly decreases, which is due to the birds staying close to the feeders and repeatedly retrieving seeds for an extended period. This temporal autocorrelation has consequences for the time window approach, as the presence of birds in two consecutive time bins will always be correlated,

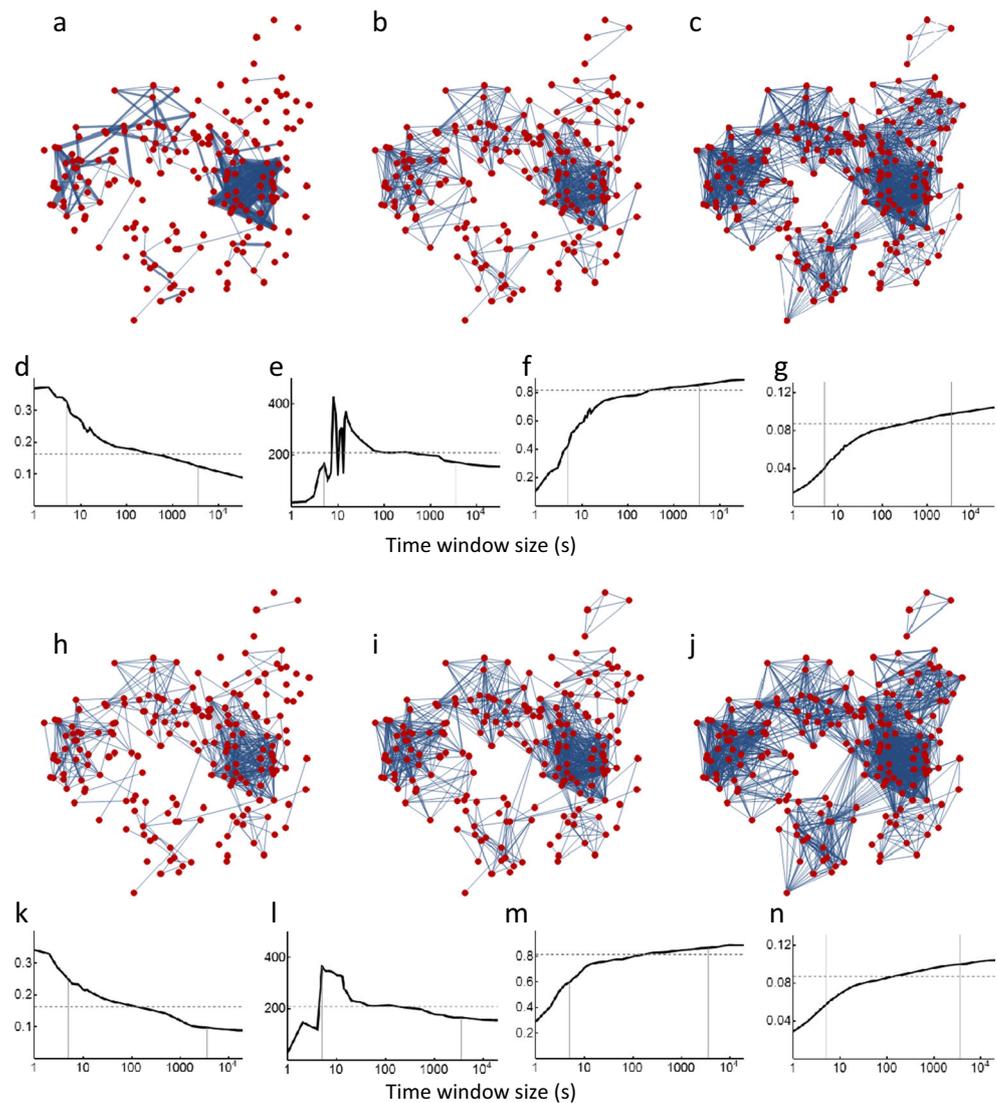
irrespective of the length of the time window. Figure 4b shows the autocorrelation for a time window of 10 s, respectively, suggesting that by taking such a time window and sampling only every second interval, one could substantially reduce the temporal autocorrelation. Yet, this would mean discarding half of the data. Note that, in addition to temporal dependencies, there are also statistical dependencies between data of different birds due to the nature of social interactions. These dependencies are not accounted for here.

A method for selecting the appropriate window size

The exploratory analysis done so far suggests that very small time windows may omit important co-occurrences and introduce temporal dependencies, while unreasonably large ones can lead to an over-estimation of the population's social connectivity (compare Fig. 2a and c). Yet, so far, we have no means of telling which, out of all different network topologies that result from varying Δ_t , is the most appropriate. Therefore, in cases where we have no expert knowledge of the temporal scale of our data, we have to examine multiple time windows and select the one that satisfies some performance metric. Although there are many graph quantities to consider, such as the ones we examined in Fig. 2, these are more descriptive variables of a particular topological structure, rather than a fitness score of how well a given time window produces an appropriate network given the data stream at hand. We, therefore, define our performance metric based on some form of *deviation from randomness*, instead.

Let us consider a randomised version of the data stream, where for each location we have performed a shuffling of bird labels while maintaining the order of timestamps. Such a scheme maintains key characteristics of the data set such as number of observations per individual, location popularity, temporal distribution of records, but breaks all dependences in the observation sequence induced by social structure. Given a certain time window Δ_t , we produce one network from the original and one network from the randomised data stream and compare them based on a link-by-link mean square error (MSE) metric. [Other metrics have also been considered, although they do not significantly alter the results. Such metrics are based on the mean square error of (a) the link-by-link weight difference, (b) node-by-node difference in clustering coefficient and (c) node-by-node difference in degree centrality. All metrics have been appropriately normalised to keep consistent scaling. See [Supplementary Material](#) for more information.] We repeat the randomisation process for a given number of times and produce an empirical distribution of the MSE scores between the observed and the randomised data. We then perform such a scheme for a range of candidate windows Δ_t , keeping track of the dissimilarity between observed and randomised data.

Fig. 2 Networks of co-foraging based on different fixed time window sizes **a** 1 s, **b** 5 s and **c** 3600 s and on variable time windows with an influence interval of **h** 1 s, **i** 5 s and **j** 1800 s. Vertices represent single individuals. An edge was drawn between two individuals if they have been observed in the same time window. Edge thickness represents absolute frequency of co-occurrences. Vertex position indicates the approximate spatial position of the birds given by the centre of gravity of their activity area. A small random error was added to the vertex positions in order to avoid complete overlap of birds that visited only a single feeder. Change of four network metrics: mean edge weight disparity (**d**), mean betweenness centrality (**e**), mean clustering coefficient (**f**), mean degree centrality (**g**) in dependence of fixed time window size and of mean edge weight disparity (**k**), mean betweenness centrality (**l**), mean clustering coefficient (**m**), and mean degree centrality (**n**) in dependence of the influence interval for variable time windows. Horizontal grey dotted lines indicate the estimates for network measures based on the GMM approach, vertical grey lines indicate the values for the example networks



In Fig. 5, we plotted the dissimilarity between the null and observed networks across a range of different time window values. As we would expect, for Δ_t values close to zero there is minimal difference between the networks generated from the null and observed data streams, as the time window is so strict that no social ties can be defined. Similarly, for inappropriately large Δ_t (close to the total observational time span), every individual is connected to all others, and the actual ordering of visits does not matter; thus, the observed and null networks converge. Between the two extreme cases, a maximally non-random structure emerges for some intermediate Δ_t value, as shown in Fig. 5. In our case, we get a maximal dissimilarity between networks based on observed and randomised data for a fixed time window size of the order of 100 s (102 s for weight difference, 89 s for clustering coefficient and 101 for degree centrality). Selecting this

value as the optimal time window size is a reasonable first step in analysing such data streams, as it does not require any prior or expert knowledge about the scaling parameter Δ_t , though it makes the strong assumption that the “interaction radius” between individuals (and, hence, the appropriate time window size) is fixed throughout the period of data collection. Additionally, the process of performing multiple runs and network extractions can be computationally demanding, especially in cases of large data streams. Finally, a note of caution has to be made at this point: While choosing the time window based on maximal dissimilarity is a feasible way to objectify the choice, it does not guarantee that this window size is biologically relevant. As such, we have no empirical indication that evolution selects for networks that are maximally different from random, nor can we think of any theoretical reasons why this should be so.

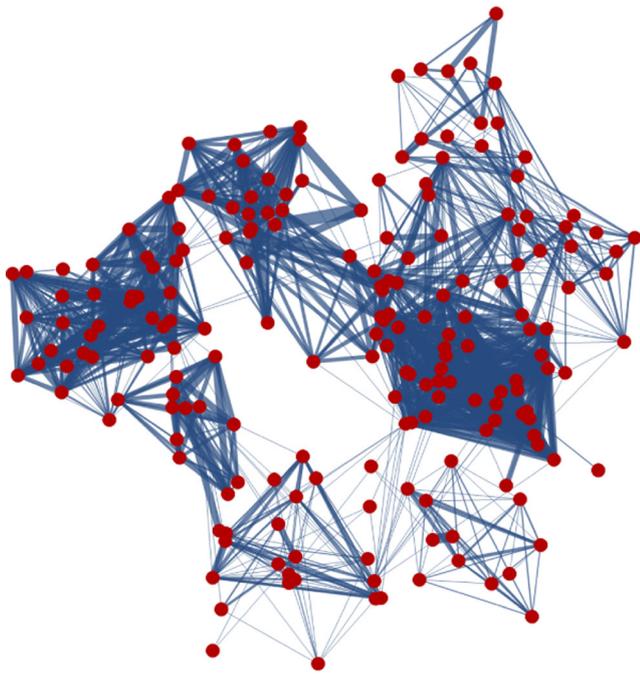


Fig. 3 Networks of co-foraging based on the GMM approach. Vertices represent single individuals. An edge was drawn between two individuals if they have been observed in the same gathering event. Graphs are drawn as in Fig. 2

Analysis of gambitness

In order to study how strongly the “gambit of the gathering” affects the constructed networks, we generated artificial data streams and extracted each time two networks, one using the GMM method and another one using the fixed time window. We parameterised the time window method using the maximum-dissimilarity scheme presented previously. For producing the data stream, we first generated a reference (ground-truth) network of a hypothetical social network of 128 birds and 4 flocks, using the Girvan–Newman random graph model (Girvan and Newman 2002). We then converted the graph to data stream form, so that connected individuals (members of the same flock) appear in close temporal proximity. Additionally, we allow a certain level of “noise”, by placing a number of randomly selected individuals across the artificially generated data

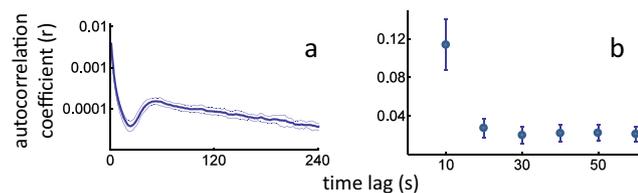


Fig. 4 Temporal autocorrelation for presence at a feeder. **a** Mean autocorrelation coefficient for all individuals over 24 days of data collection (**bold line**) and standard deviation (**thin line**) for time lags of 1–240 s. **b** Mean autocorrelation coefficient for fixed time windows of 10 s. Error bars give the standard deviation

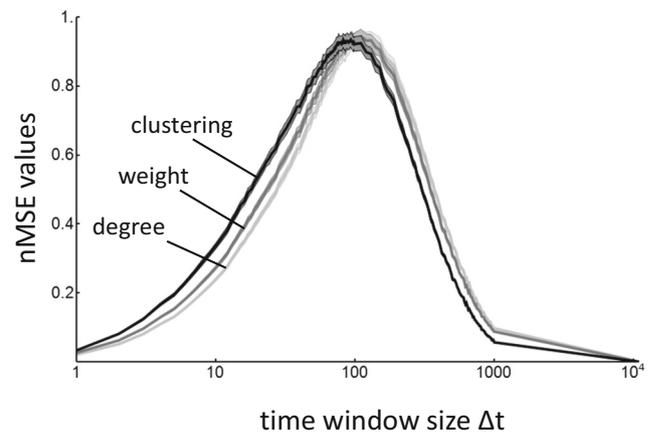


Fig. 5 Three dissimilarity metrics between the observed network and the null network across various time window sizes. Link-by-link weight difference (*dark grey*), node-by-node difference in clustering coefficient (*black*) and node-by-node degree centrality difference (*light grey*). Each point gives the mean difference between the observed and $R=100$ randomisations of the data stream of visits of great tits on one example day. Error bars indicate standard deviation

stream. Algorithmic and computational details for the artificial data stream generator are presented in the online [Supplementary material](#). For each network, we extracted its community structure (i.e. the groups of individuals that are closely connected to each other) and compared it against the corresponding community structure of the ground-truth graph via the normalised-mutual-information (NMI) score (Danon et al. 2005). The quantity NMI takes values from 0 to 1, with two networks yielding 1 when they have identical community structure and zero if there is no similarity in the grouping pattern of their members. In Fig. 6, we illustrate the effect of *gambitness*, i.e. the probability that unaffiliated individuals will appear in close temporal proximity in the data stream, based on the quality of the extracted networks from the GMM and time window methods. We can see that, across a range of noise levels (0 for no probability of unaffiliated individuals co-occurring and 1 for a completely random sequence of occurrences), the GMM method produces a considerably more accurate extraction of the underlying ground-truth graph, as reported by 100 cases of artificially generated data streams per noise level. Notably, the relative performance of the GMM approach, versus a fixed time window approach, improves as the noise level increases. Finally, we were interested if the algorithm produces stable solutions across multiple runs on the same data. For this purpose, we considered 1000 runs of GMM for each of 100 randomly generated data streams. We compared the similarity of those 100 different solutions via their NMI values. We can see that GMM possesses excellent solution stability, with most NMI values ($x \pm SD$) falling within 0.9916 ± 0.0022 for weekend 1 and 0.992 ± 0.002 for weekend 2 ([Supplementary material](#)).

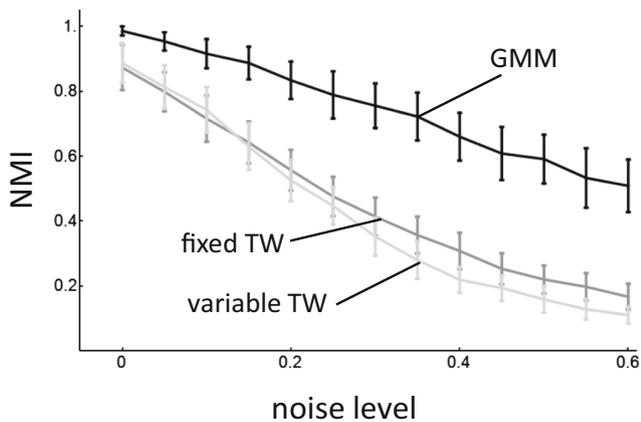


Fig. 6 Quality of community structure recovered from simulated noisy data. Mean normalised-mutual-information (NMI) score is plotted for different noise levels for networks constructed using the GMM approach (black), fixed time window approach (dark grey) and variable time window approach (light grey). Error bars indicate standard deviation based on 100 simulations per noise level

Discussion

Capturing the behaviour of animals (or humans) by continuously observing individuals brings along with it a fundamental problem: The behaviour expressed at a specific time point is not independent from the behaviour expressed only an instance earlier. The same is true for spatial relationships. To deal with these temporal dependencies, one can try to discretise the temporal data stream into discrete and independent behavioural events. How to do this properly for identifying social associations is the main focus of this paper. As an example, we used data from our own work where the presence of PIT-tagged individual great tits at feeding stations was continuously recorded. Yet, the solution that we present is not restricted to a system with PIT-tagged individuals but can also be applied to any other system where proximity data of individuals are recorded continuously (i.e. with a high sampling frequency), as it is the case with Encounternet (e.g. Rutz et al. 2012), sirtrack (Sirtrack Ltd., Hawkes Bay, NZ) or e-obs (e-obs GmbH, Gruenwald, DE).

Constructing a social network from a time-stamped data stream is not a trivial task. Whatever clustering scheme is used in order to discretise the time stream, by choosing very small time bins one risks excluding biologically meaningful associations, while by choosing very large time bins one risks including too much noise. At the same time, small time bins will increase the temporal dependencies of successive bins, and consequently threaten statistical independence, while large time bins will lead to an overestimation of meaningful links up to a point where a ceiling effect obstructs the detection of contrasts. In order to deal with the temporal dependencies, one could arguably set a threshold value of what would be an acceptable auto-correlation value and chose a time bin size that produces temporal auto-correlations just below this

threshold. In order to increase the likelihood of capturing only meaningful associations, we have suggested that comparing how different produced measures are from expectations for random networks and choosing the bin size in a way that maximizes this difference might be one way forward. Yet, while we can argue that biologically meaningful signals should deviate from random patterns, we know of no convincing indication that biological processes necessarily produce maximal deviation from randomness.

Given that, in our example, birds will usually stay at a feeder for a certain time, any fixed time interval will occasionally cut such an individual foraging bout into two and the bird will be recorded as being present in two consecutive time bins. This produces temporally dependent data. For the flexible time window, we defined an influence zone around each recording of a bird, and if influence zones of consecutive visits of a single bird overlapped, these were merged to a single influence zone and counted only once. By doing this, we effectively reduce the temporal dependencies. The GMM deals with the problem in a different way. Instead of defining influence zones for single individuals, it defines the duration of gathering events by considering all birds present simultaneously. Here, it can—in principle—happen that a foraging bout of an individual is cut into two and counted twice in successive gathering events (producing temporal dependent data), though the mixture model effectively minimises the instances of such bout splits. Reducing temporal autocorrelation in social group structure is also particularly important as many permutation procedures aimed at defining non-random social structure assume temporal independence of observed groups (Beijder et al. 1998).

A basic problem intrinsic to all methods for constructing networks based on spatio-temporal association data is the “gambit of the group” (Whitehead and Dufault 1999). The problem might be most apparent in the Gaussian mixture model or for the fixed time window approach with large bin sizes, where often enough a large proportion of the population (i.e. the birds visiting the specific location on that day) is observed in the same time bin or gathering event and links are drawn between all these birds. The effect of the gambit of the group is an overestimation of meaningful associations. Clearly, the problem increases with increasing bin sizes, though it is impossible to quantify it in real social networks, as the underlying structure of “true” social associations is not known. We have, however investigated how the different data extraction methods perform in this respect, by creating artificial networks and simulated sampling. As, in this case, we have the full information about the underlying data, we can estimate the level of overestimation of links and, hence, give a measure for *gambitness* for the different methods. For the hypothetical example network, the GMM method was more reliable in recovering the original graph structure than the time window approach for all levels of added noise. Thus, while we

can never exclude the risk of falsely interpreting random associations as group memberships completely (hence the term “gambit”), the GMM method will usually produce networks that are only slightly affected by this, as long as random associations do not occur too frequently.

Finally, we want to note that the time window approach goes against our biological intuition in at least one important aspect. It assumes that the optimal time window stays constant over the whole observation period. However, we also know that the activity of birds changes over the course of the day (e.g. Lahti et al. 1997), and it is sensible to assume that the optimal time window for defining associations changes with the prevailing activity. Additionally, plenty of observations collected over several decades suggest that, during winter, great tits tend to forage and move around in flocks (Hinde 1952; Saitou 1978; Perrins 1979; Gossler 1993). Ignoring these well-established facts would mean ignoring some well-established facts about the biology of our subjects.

Closing remarks

The data used for this study stem from a long-term monitoring project where we equipped a large proportion of a population of wild birds with PIT tags and recorded their appearance at 35 feeding stations during one winter. This resulted in a data stream of over half a million recordings of birds at the feeders. Given this—for a behavioural study—substantial data set, it might be expected that it would be easy to construct reliable and meaningful social networks out of the temporal data stream of feeder visits. Yet, in contrast, our analyses here demonstrate a remarkable degree of variability in how different the networks looked and in how much estimates of network measures differ depending on the chosen time window size. Although some effects of time window size are expected, the realised magnitude of this effect and the non-linearity and unpredictability of its direction were striking. If a summarising network statistic which scales between zero and one can take on any value between 0.3 and 0.9 for networks based on exactly the same data stream—only depending how we chose the size of the time windows—then we cannot have much confidence in any network based on an arbitrarily chosen fragmentation of a temporal data stream. At its best, this means that the results researchers get might be a random effect of chosen time window size; at its worst, researchers can just choose the time window size which produces measures supporting their preferred hypothesis.

In order to eliminate this arbitrariness in the parameter choice for the time window, we introduced a simulation method that finds the time window size which maximises dissimilarity of generated networks with the same underlying data structure. In addition to being computationally burdensome, this method makes the assumptions that the optimal time window size is constant over the whole observation period and

that maximal potential dissimilarity is biologically meaningful. Both assumptions are not necessarily warranted. We therefore opted for an alternative approach (Psorakis et al. 2012), which allows variation of the time window size over time and does not rely on maximising a statistical dissimilarity measure but instead incorporates assumptions about the bird’s biology—namely that birds move around in flocks (producing a non-uniform observation profile). This method has already been successfully applied in several studies of this population (e.g. Farine et al. 2012; Aplin et al. 2013).

In summary, we can say the following. (1) Analyses that require a discretisation of a temporal data stream must be distrusted if the choice of the time window size is arbitrary and not further justified. (2) The GMM approach can better reconstruct the underlying social affiliation patterns, by incorporating information about the animals’ social biology and exploit the non-uniform or flocking nature of bird visits at the feeders. As such, it does not commit to a rigid temporal resolution. (3) For applications where the visitation profile is uniform, we have proposed an appropriate methodology for setting time window size, by selecting the one that gives rise to the most “non-random” network structure. These observations hold for all social networks reconstructed from temporal data (which is the case for most networks in behavioural ecology), but potentially also for any other analysis that makes use of time windows for discretisation of a data stream.

Acknowledgements We thank J. Howe, T. Wilkin, S. Evans, A. Hinks and A. Grabowska for assistance in the field and three anonymous reviewers for valuable comments on the manuscript. This research was funded by an ERC grant to BCS (AdG 250164) and a Microsoft Research Grant to IP. LMA was also funded by an Australian Postgraduate Award and by an International Alliance of Research Universities travel grant.

Ethical standards All authors declare that the present study complies with the current laws in the UK.

References

- Aplin LM, Farine DR, Morand-Feron J, Cole EF, Cockburn A, Sheldon BC (2013) Individual personalities predict social behaviour in wild networks of great tits (*Parus major*). *Ecol Lett* 16:1365–1372
- Barthelemy M, Barrat A, Pastor-Satorras R, Vespignani A (2005) Characterization and modeling of weighted networks. *Physica A* 346:34–43
- Beijder L, Fletcher D, Brager S (1998) A method for testing association patterns of social animals. *Anim Behav* 56:719–725
- Cairns SJ, Schwager SJ (1987) A comparison of association indices. *Anim Behav* 35:1454–1469
- Croft DP, James R, Krause J (2008) Exploring animal social networks. Princeton University Press, Princeton
- Danon L, Diaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. *J Stat Mech-Theory Exp* 09, P09008
- Farine DR, Garroway CJ, Sheldon BC (2012) Social network analysis of mixed-species flocks: exploring the structure and evolution of inter-specific social behaviour. *Anim Behav* 84:1271–1277

- Franks DW, Ruxton GD, James R (2010) Sampling animal association networks with the gambit of the group. *Behav Ecol Sociobiol* 64:493–503
- Freeman LC (1979) Centrality in social networks. I: conceptual clarification. *Soc Networks* 1:215–239
- Garroway CJ, Radersma R, Hinde CA (2015) Perspectives on social network analyses of bird populations. In: Krause J, Croft DP, James R (eds) *Animal social networks: perspectives and challenges*. Oxford University Press, Oxford, pp 171–183
- Gibbons JW, Andrews KM (2004) PIT tagging: simple technology at its best. *Bioscience* 54:447–454
- Ginsberg JR, Young TP (1992) Measuring associations between individuals or groups in behavioural studies. *Anim Behav* 44:377–379
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99:7821–7826
- Gossler A (1993) *The great tit*. Hamlin, London
- Hinde R (1952) The behaviour of the great tit (*Parus major*) and some other related species. *Behav Suppl* 2:1–201
- James R, Croft DP, Krause J (2009) Potential banana skins in animal social network analysis. *Behav Ecol Sociobiol* 63:989–997
- Krause J, Krause S, Arlinghaus R, Psorakis I, Roberts S, Rutz C (2013) Reality mining of animal social systems. *Trends Ecol Evol* 28:541–551
- Lahti K, Koivula K, Orell M (1997) Dominance, daily activity and winter survival in willow tits: detrimental cost of long working hours? *Behaviour* 134:921–939
- Newman MEJ (2010) *Networks: an introduction*. Oxford University Press, Oxford
- Perrins CM (1979) *British tits*. Collins, London
- Psorakis I, Roberts SJ, Rezek I, Sheldon BC (2012) Inferring social network structure in ecological systems from spatio-temporal data streams. *J R Soc Interface* 9:3055–3066
- Rutz C, Burns ZT, James R, Ismar SMH, Burt J, Otis B, Bowen J, St Clair JH (2012) Automated mapping of social networks in wild birds. *Curr Biol* 22:R669–R671
- Saitou T (1978) Ecological study of social organization in the great tit *Parus major* L. I. Basic structure of winter flock. *Jpn J Ecol* 28:199–214
- Smyth B, Nebel S (2013) Passive integrated transponder (PIT) tags in the study of animal movement. *Nat Educ Knowl* 4:3
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge
- Watts DJ (1999) *Small worlds: the dynamics of networks between order and randomness*. Princeton University Press, Princeton
- Whitehead H (2008) *Analyzing animal societies*. The University of Chicago Press, Chicago
- Whitehead H, Dufault S (1999) Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Adv Stud Behav* 28:33–74
- Wilson EO (1975) *Sociobiology: the new synthesis*. Harvard University Press, Cambridge