University of Bath

PHD

On the Multivariate Analysis of Animal Networks

Mlynski, David

Award date:
2016

Awarding institution:
University of Bath

Link to publication

On the Multivariate Analysis of Animal Networks

David Thomas Mlynski

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2015

COPYRIGHT

Table of Contents

List of Abbreviations

Abbreviation	**Description**
DO	Direct observation
DSP	Double-semi partialling (a form of residual-based permutation based on explanatory variables)
ERGM	Exponential random graph model (p* model)
FLSP	Freedman-Lane semi-partialling (a form of residual-based permutation based on the response variable)
GoG	Gambit of the Group (assumption that repeat co-occurrence in a social group represents social affiliation)
GPS	Global positioning system
LED	Light emitting Diode
MEV	Median edge value
MLR	Multiple linear regression
NMBHT	Null model based hypothesis testing
OLS	Ordinary Least Squares
RA	Reciprocal agreement (when two proximity loggers concurrently register one another)
RFID	Radio frequency identification
RS	Remote sensing
PIT	Passive integrative transponder
QAP	Quadratic assignment procedure (node-label permutation)
TLCA	Time since last common ancestor (in a phylogeny)

Declarations

This dissertation is the result of my own work and includes work done in collaboration, which is specifically indicated here and in the text. All figures included from other sources are described as “Inserts”.

Chapter 4:

- Sections 4.2 to 4.4: The idea behind the study, that interspecific hybridisation can be treated as a networks problem in certain taxa, came from James, Priest and Eugene McCarthy. The plumage and song data in this chapter was collected by Charlotte Pindar. The data required to produce the *Galliformes* hybrid network was collected by Samantha Berry. Phylogenetic data for the same species was measured by James. The data required to produce the *Anseriformes* hybrid network was collected by Miranda Bane and Kat Holmes. The data required to produce the *Paradisaeidae* hybrid network was collected by Charlotte Pindar. The panel figure (figure 6) presented in section 4.5 was created by Charles Cunningham, with panel *D* contributed by Miranda Bane.

- Section 4.5: The suggestion of using sightings taken of individual species as a proxy of sampling effort for hybrids came from discussions with Jason Wolf and Nick Royle.

Chapter 5:

- Section 5.3: The figures in this section have been reproduced, with permission, from Thibault Lengronne’s (2013) doctoral thesis titled: “Social Dynamics in Paper Wasps: The case of nest-drifting in *Polistes canadensis*” published by the University of Lausanne.

- Section 5.2 to 5.7: The materials in these sections are adapted from a manuscript submitted to *eLife* (under review) titled: “Adaptive drivers of inter-nest interactions in a eusocial insect”, co-authored by Thibault Lengronne and David Mlynski, Solenn Patalano, Richard James, Laurent Keller and Seirian Sumner. The fieldwork and data-processing for this study was conducted by Lengronne, Patalano and Sumner.

Chapter 6:

- Section 6.3: The “static correction” for association times recorded by Sirtrack® radio-proximity loggers was developed by Darren Croft. The experimental deployment and testing of these loggers was conducted with Natasha Boyland, who also created the figures presented in this section. The analytical model presented was developed by Richard James. The material in section 6.3 was published in 2013 under the title: “Spatial proximity loggers for recording animal social networks: consequences of inter-logger variation in performance” published in *Behavioural Ecology and Sociobiology*, 67 (11), pp. 1877-1890. Authored by Natasha Boyland, Richard James, David Mlynski, Joah Madden and Darren Croft. The procedure in section 6.3 also features in Natasha Boyland’s doctoral thesis “The influence of social networks on welfare and productivity in dairy cattle” published in 2015 by the University of Exeter.

- Section 6.4: The analysis of video footage was assisted by Ruth Shen and Primrose Manning.

Acknowledgements

I would like to acknowledge the Biotechnology and Biological Sciences Research Council (BBSRC) as the source of the studentship that funded my research.

I would like to thank Dr. Richard James for the monumental role he has had in my professional development and Dr. Nicholas Priest for giving me this opportunity and sourcing a stream of talented project students to assist in my endeavours.

I would like to thank those project students: Charlotte Pindar, Miranda Bane, Samantha Berry, Charles Cunningham, Ruth Shen and Primrose Manning.

I would like to thank Natasha Boyland for food, board and co-authorship and James Sciberras for being a good nurse.

Finally, I would like to thank my family for their support and tolerance. No more Kickboxing.

Abstract

From the individual to species level, it is common for animals to have connections with one another. These connections can exist in a variety of forms; from the social relationships within an animal society, to hybridisation between species. The structure of these connections in animal systems can be depicted using networks, often revealing non-trivial structure which can be biologically informative.

Understanding the factors which drive the structure of animal networks can help us understand the costs and benefits of forming and maintaining relationships. Multivariate modelling provides a means to evaluate the relative contributions of a set of explanatory factors to a response variable. However, conventional modelling approaches use statistical tests which are unsuitable for the dependencies inherent in network and relational data. A solution to this problem is to use specialised models developed in the social sciences, which have a long history in modelling human social networks.

Taking predictive multivariate models from the social sciences and applying them to animal networks is attractive given that current analytical approaches are predominantly descriptive. However, these models were developed for human social networks, where participants can self-identify relationships. In contrast, relationships between animals have to be inferred through observations of associations or interactions, which can introduce sampling bias and uncertainty to the data. Without appropriate care, these issues could lead us to make incorrect or overconfident conclusions about our data.

In this thesis, we use an established network model, the multiple regression quadratic assignment procedure (MRQAP), and propose approaches to facilitate the application of this model in animal network studies. Through demonstrating these approaches on three animal systems, we make new biological findings and highlight the importance of considering data-sampling issues when analysing networks. Additionally, our approaches have wider applications to animal network studies where relationships are inferred through observing dyadic interactions.

I. Introduction

1. Introduction

1.1 Thesis Overview

Networks are often used to model interconnected systems across a range of scientific disciplines. The broad interest in networks has been one of the reasons why the field of network science has progressed so rapidly. However, the types of networks studied and the questions asked of them vary considerably across different scientific disciplines; as such, care is required when new terminology, metrics or analytical techniques cross over disciplines (Croft et al. 2008; James et al. 2009; Krause et al. 2014). In biology, particularly whole-organism biology, we are often interested in explaining a network in terms of a number of explanatory variables; identifying the factors which shape network structure. Describing a network in terms of more than one explanatory variable principally requires statistical modelling. However, the dependencies inherent in network data prevent the use of many conventional statistical tests (Krackhardt 1988; Snijders 2011). These tests assume that data-points are independent from each other; this is not true for networks or often the explanatory variables used to describe them (see: Snijders 2011). Biologists have thus looked to other fields that have had a head-start in the multivariate statistical analyses of network data. This thesis concerns ways in which approaches used to model networks may be taken from the social sciences and adapted for the study of animal networks.

Numerous multivariate network modelling techniques have been developed in the social sciences (Goldenberg et al. 2010). Taking these predictive, multivariate models and applying them to animal networks is attractive (see Pinter-Wollman et al. 2014). In studies of animal networks, statistical approaches have often been descriptive as opposed to predictive and univariate as opposed to multivariate; this has been due, in part, to the difficulties in observing animal interactions and producing reliable networks (Croft et al. 2008). In the social sciences, multivariate network models have been developed primarily for the study of human social networks, where the researcher has had more control over the sampling of the study system and has been able to determine the presence or absence of ties between individuals with a high degree of certainty. In contrast to humans, animals cannot be surveyed to find out with whom they are socially affiliated or otherwise. Taking models developed for human networks and applying them to animal networks could place undue trust in unreliable data, given the difficulty in inferring social ties between animals (see Croft et al. 2008; Krause et al. 2014).

This thesis proposes novel approaches to determine the factors which shape networks associated with animals. Three animal networks are studied to illustrate each approach: the first representing hybridisation between warbler species (family: *Parulidae*) (chapter 4); the second depicting workers that drift between nests in populations of the paper-wasp *Polistes canadensis* (chapter 5) and the third representing associative relationships within a herd of dairy cattle *Bos taurus* (chapter 6). The first of these networks is not strictly a social network, as interactions take place between different species and consist of mating events; however, all of these networks present challenges to multivariate analysis which stem from observing wild animals. In the first network, the presence of edges (reported hybridisation between two species) is relatively reliable, but the absence of edges cannot be trusted. In the second, the network is more reliable, but currently available statistical procedures are too generic for the study system and consequently need modification. In the third system, the method used to generate the network also biases the strength of edges (association time between pairs of cows), which has to be controlled prior to modelling.

The major focus of this thesis is to find means to apply the modelling techniques developed in the social sciences on animal networks, thus allowing multivariate analyses to be conducted on animal networks whilst addressing the issues in data-quality. After considering network fundamentals (chapter 2) and how network approaches are typically conducted (chapter 3), this thesis focuses on three animal network case studies (chapters 4-6). In each of these case studies, a unique approach to facilitate the use of multivariate analytical models is demonstrated. The first of these approaches involves restricting the use of multivariate models to just the explanatory variables (chapter 4) and using these modified explanatory variables to perform more complex and powerful hypothesis testing (see Croft et al. 2011). Explanatory variables may be easier to quantify reliably in animal network studies than the interactions which form the response variable, the network itself. An example being that if we expected a social network to be structured such that similarly-sized individuals were more likely to be connected, measuring the sizes of individuals and then calculating the difference might be easier than identifying patterns of interactions between all individuals in the population. The second approach is based on modifying facets of the models themselves to control for some of the issues associated with animal data (chapter 5), in particular, adjusting the null models behind these procedures from those common to human social networks towards those often used in animal networks. The third approach is to take steps to increase the reliability of animal network data before modelling takes place so that data are evenly sampled; in performing this step, we bring the quality of animal network data closer to the human social network data for which many multivariate network models were developed (chapter 6).

1.2 The Importance of Studying Networks

There is a ubiquitous familiarity with the term “network” in society (Castells 2011). The increase in use of online social networking and social media has brought the concept of networks to the forefront of everyday life. According to the Pew Research Centre (PewResearch 2014), 74% of online adults use social networking sites. Electrical power grids (Xu et al. 2004), transport networks (Sen et al. 2003) and the world-wide web (Tadić 2001) can all be considered as networks with which we might be familiar. At the same time, networks are popular in science; at present, over 63,000 articles in the multidisciplinary journal Nature contain the term “network” (www.nature.com, accessed: 21-06-15). Networks offer a departure from the idea that the world can be described via the objects it is made up of and their properties, highlighting the importance that the relationships between objects can have. Network studies are part of a wider trend (Rhoten 2004) in science towards interdisciplinary research, uniting numerous scientific fields including maths, physics biology, economics and the social sciences (Porter & Rafols 2009). Through the collaborative study of networks, we have made rapid progress in understanding of physical (Iida 1999; Faloutsos et al. 1999; Carreras et al. 2002), social (see: Borgatti et al. 2009) and biological processes (see: Proulx et al. 2005; Wey et al. 2008).

A network, or graph, can be defined as a system of discrete entities, known as nodes, connected by ties, known as edges. In a diagrammatic representation of a network, nodes are depicted using symbols (circles, squares etc.) and edges are depicted as lines connecting pairs of nodes. Networks are used to model systems where the pattern of the connections is heterogeneous and this structure can affect both the local and global properties of a system. The importance of the structure of interactions is otherwise missed through studying pairwise interactions in isolation of each other or analysing the average properties of the system as a whole. Some broad examples of network include: social networks, where individuals may form the nodes of the network linked through identified social relationships (Otte & Rousseau 2002); physical networks, where nodes may take the form of circuit components or powers stations and edges represent physical connections such as wiring (Belevitch 1962), or data connections in a computer network (Faloutsos et al. 1999); economic networks (see: Allen & Babus 2010), where nodes which could take the form of banks (Iori et al. 2008; Gai & Kapadia 2010), businesses (Coviello & Munro 1997) or countries (Smith & White 1991; Garlaschelli & Loffredo 2005) and edges may represent trade (Garlaschelli & Loffredo 2005) or monetary lending (Masi et al. 2011) between them, and biological networks, where nodes may take the form of individual organisms (Lusseau 2003) or the biological components that make up an organism (Jeong et al. 2001; Abouheif & Wray 2002; Papp et al. 2004) and edges the physiological, behavioural, or evolutionary connections they may have. Networks can also be purely theoretical, not directly representing any real-world system at all. Simulated typical networks are constructed to study emergent properties associated with different network topologies- helping us identify properties of networks observed in the real-world (e.g. Bolland 1988; Watts & Strogatz 1998; Pastor-Satorras & Vespignani 2001; Hałaj & Kok 2013).

Uncovering the biophysical explanations underlying the structure of animal systems is important given both the inherent value in understanding the natural world and the potential applied uses of network theory. Social networking alone has been one of the fastest growing subject matter areas in the US patent office. In 2003, only a handful of patents had been published regarding social networks, by 2011 there were at least 3500 (Nowotarski 2011). Through the study of biological networks, a new field of network medicine has arisen from the holistic understanding of the interactions between proteins, genes and metabolic agents (Barabási et al. 2011) leading to the earlier identification and targeted treatment of new disease modules and pathways. Further, the study of epidemiological networks has led to the theoretical development of different network-based targeted vaccination strategies, creating “herd immunity”, where a disease cannot propagate in a society with the lowest vaccination effort (Pastor-Satorras & Vespignani 2002; Christakis 2004; Fine et al. 2011). In machine learning and cognitive science, artificial neural networks form a method of statistical learning based on animal neural networks, excelling in tasks such as speech and visual recognition which otherwise pose a challenge to conventional rule-based programming (Hagan et al. 1996). National security services can be concerned with how organised crime syndicates and terrorist cells operate, network theory has contributed to their understanding of these systems (Krebs 2002; Qin et al. 2005; Xu & Chen 2005; Natarajan 2006; Ressler 2006). In domestic security, networks can enrich our understanding of phenomena such as riots and social epidemics- such as the spread of extremist ideologies (Kennedy et al. 1997; Patten & Arboleda-Flórez 2004; Amblard & Deffuant 2004; Franks et al. 2008; Martins 2008; Radil et al. 2010). The study and refinement of logistic networks has made the transport of goods more efficient and less damaging to the environment (Bell & Iida 1997; Richardson 2005). In the field of engineering, inspirations from biological networks have been applied to the design of sensor networks (Dressler et al. 2005; Barbarossa & Scutari 2007), communication networks (Dressler 2005; Carreras et al. 2007) and swarm intelligence (Webb 2002). However, that is not to say everything about networks is known or that we are getting the most benefit from what we do know. A lack of predictive multivariate models suitable for use on animal networks has impeded our ability to both effectively analyse many networks occurring in nature and apply what we learn for the purposes of improving conservation and agricultural practises.

1.3 Networks in Evolution, Ecology and Behaviour

The networks studied in this thesis are biological, with nodes consisting of either whole organisms or species; this section serves as an introduction into some of the ways networks have been used to study biological systems. In particular, this section highlights how the structure of interconnected biological systems can be important to their properties and function, and how these structures may have evolved. For a review of the role of networks in studying biological systems, the reader is pointed towards: Proulx et al. (2005) for a biochemistry and ecology-based perspective, and Krause et al. (2009) for perspective from behavioural biology.

The theory of evolution has been highlighted as the main unifying concept in Biology (Smocovitis 1996). The currency of evolutionary biology is fitness, the amount of genes an organism passes down to successive generations; this can be directly, through having offspring, or indirectly through increasing the reproductive success of closely related kin, which share some of the same genes, as outlined in inclusive fitness theory (Hamilton

1964; Bijma & Wade 2008). Typically, heritable traits that confer a fitness advantage spread throughout populations, whereas those that confer a disadvantage are lost. If the structure of a biological system can vary, and is somewhat heritable, then it is a viable target for selection and evolution (Hall 2008). Biologists may seek to understand how evolutionary forces shape network structure and whether certain networks structures observed in nature are adaptive, conferring fitness advantages to all or most of their members (Proulx et al. 2005). Biological networks can range from those governing sub-cellular processes through to those happening at the level of the organism and beyond. Molecular biologists may be concerned with whether the structure of gene-regulatory, metabolic, or protein interaction networks may be robust to random faults that happen across the genome (von Dassow et al. 2000; Edwards & Palsson 2000; Jeong et al. 2001). Behavioural biologists may seek to understand how the underlying social network of group-living animals affects individual and collective behaviour (Krause et al. 2009). Ecologists may examine the interactions between collections of organisms, such as populations of species, to better understand ecosystem function and fragility- informing conservation action and furthering our understanding of biodiversity (Iida 1999; Solé & Montoya 2001).

At the molecular-level, genes are the fundamental blueprints of living organisms, regulating and encoding the production of different proteins. However, genes are continually subject to mutations, many of which get enzymatically repaired, whilst some of which persist (Alberts et al. 2002). Mutations are occasionally beneficial, conferring a fitness advantage, but most often deleterious (Peck 1994). Mutations of genes affect the production of proteins, which can guide metabolism through acting as catalysts of different biological reactions (Alberts et al. 2002). Although mutation rates vary across different parts of the genome (Wolfe et al. 1989) - they are broadly assumed to occur at random with respect to individual genes. Therefore, gene, protein and metabolic networks would theoretically benefit from having a structure robust to the random loss of their nodes (the individual genes, proteins, or metabolic agents). Both gene-regulatory and protein-protein interaction networks have been characterised (Uetz et al. 2000; Jeong et al. 2001; Abouheif & Wray 2002; Mittler et al. 2004), showing links between network structure and evolution at the molecular level. In protein interaction networks, Jeong et al. (2001) showed that the most highly connected proteins were three times more likely to be essential for survival than were weakly connected proteins. Subsequent studies found that these proteins were also more pleiotropic (Promislow 2004), evolved more slowly (Fraser et al. 2002), and were less likely to be lost over evolutionary time (Krylov et al. 2003). The gene-regulatory network that governs sex determination may have evolved robustness in Drosophila; maintaining its function in the face of a relatively large variety of different mutations (MacCarthy et al. 2003). In contrast, the gene-regulatory network governing worker-wing development in ants may be more fragile. In species where workers develop wings the gene regulatory network appears to have been highly conserved for millions of years, however, in each of four species which do not develop wings, a unique single gene was disrupted in each case (Abouheif & Wray 2002). The structure of metabolic networks can offer additional resilience to changing environments than genetic redundancy alone (Wagner 2005). The metabolic network of *Escherichia coli* for example, is highly robust to damage. For most enzymes in the network, a change in concentration or complete loss has little effect on overall network function (Edwards & Palsson 2000; Lemke et al. 2004).

The behaviour of an individual can be governed by both intrinsic and extrinsic factors (Vallerand & Bissonnette 1992). Intrinsic factors may be any attribute of that individual, its gender or size for instance. Extrinsic factors take the form of both its abiotic and biotic environment (Croft et al. 2008). For many animal species, an important aspect of an individual’s biotic environment is its social environment, often made up of non-random heterogeneous interactions and relationships (Hinde 1976; Krause & Ruxton 2002). Sociality is widespread in the animal kingdom, considered to be one of the major evolutionary transitions of life (Maynard Smith & Szathmary 1995). Networks allow behavioural biologists to quantify an individual’s social environment to allow a better understanding of how it might influence an individual’s behaviour. An individual’s behaviour also may in turn influence its social environment. The structure of social interactions and relationships within a social group may also have global effects on the fitness of its constituent members (Bergstrom 2002). Further, the position of an individual within its native social network can also influence its fitness (Von Rueden et al. 2010). An animal’s social environment has been shown to influence a wide range of behaviours from mate-seeking and mate-choice (McDonald 2007), predator-inspection (Croft et al. 2006), foraging (Ramos-Fernández et al. 2006), and the development and maintenance of cooperative behaviours (Croft et al. 2006). An element of evolutionary game theory underpins the study of animal social networks, where the payoffs to the strategies adopted by individuals depend on the strategies of other individuals in its social environment (Gibbons 1992; Ohtsuki et al. 2006; Slikker & den Nouweland 2012).

In ecology, networks can be used to represent the feeding, or trophic, interactions between different species, also known as food webs (Kormondy 1969). Mutualistic networks can be used to represent symbiotic associations between organisms, such as plant species and their pollinators (Bascompte et al. 2003). The structures of food webs and mutualistic networks are thought to have a key role in determining the long-term stability and resilience of ecological communities (Solé & Montoya 2001; Neutel et al. 2002; Krause et al. 2003; Emmerson & Yearsley 2004; Montoya et al. 2006; Thébault & Fontaine 2010). Historically, ecological communities have been mathematically modelled under the assumption that all species interact with each other, that there is uniform or saturated network structure (Berlow et al. 2004). Through these approaches, mathematical ecologists have explored how the size and connectivity of food webs affect community stability in the face of fluctuations in population densities (May 2001), the introduction of an invasive species (Case 1990), and the long-term persistence of the community under non-linear population dynamics (Hastings & Powell 1991). However, real-world ecological networks are not uniform and network theory has increased our understanding of them through incorporating heterogeneity (Proulx et al. 2005). Ecologists have identified that networks with more links often show decreased strength in those links, thus decreasing the dependency of one species on any other and increasing robustness to the loss of individual species (Vázquez et al. 2007). Further, cascading extinctions are less likely in compartmentalized ecological networks, where the network is divided into relatively independent sub-networks, on the basis of body size or spatial location for example. As effects of species losses are limited to the original compartment, the overall community is isolated to the loss (Krause et al. 2003).

In epidemiology, both compartmental models have been used to understand the dynamics of the spread of disease across populations (Amitai et al. 2004). The term compartmental comes from the assignment of members of a population into one of a number of compartments, perhaps the most famous compartmental model is the SIR model developed by Kermack and McKendrick (1927). Within a population N at a time t, S_t represents the number of susceptible individuals, I_t the number of infectious infected individuals and R_t the number of removed individuals (i.e. having recovered and now have immunity, or having died). A key assumption underpinning these equations is that there is homogeneous mixing in the population N, meaning that all individuals have an equal probability of contracting the disease with an infection rate of β (Anderson et al.

1992) . Therefore, an infected individual makes contact and is able to transmit the disease with βN others per time unit and the fraction of contacts by an infected with a susceptible is S/N (Brauer et al. 2001). However, contacts within many human and animal populations often form heterogeneous networks- with structure that can affect the spread of disease through a population (Morris 1993; Keeling & Eames 2005). For instance, the sexual contact networks of humans exhibit a power-law like distribution (see section 2.5), with most individuals having one sexual partner each year, but few having in the tens or even hundreds; this allows diseases with very low transmission rates, like the Human Immuno-Deficiency Virus (HIV), to spread rapidly through a population (Andersen & May 1988; Liljeros et al. 2001; Pastor-Satorras & Vespignani 2001; May & Lloyd 2001). Conversely, clustered networks with dense local connections between individuals (see section 2.2) slow the rate of contagion throughout a population, as clustering increases both the size of the epidemic, but also the threshold for an epidemic to occur decreases (Newman 2002b ;Shirley & Rushton 2005).

2. Network Theory

In the substantive chapters (4-6) of this thesis, network-specific metrics and terminology are used which might be unfamiliar to a reader who does not have a background in studying networks. The purpose of this second chapter is to provide a basic introduction to some aspects of network theory. In particular, we highlight three important issues. Firstly, that dependency inherent in network-data prevents us from using modelling approaches common in the biological sciences. Secondly, that the history of network science leads us to look for models developed in the social sciences as a solution. Thirdly, that borrowing models from the social sciences comes with a big warning label as animal and human networks are not as similar as we might initially suspect. The following section provides an introduction to some of the terms, metrics and properties associated with networks. It is not intended as an exhaustive review of networks, just a presentation of the information required to understand the forthcoming chapters. For a more complete description of networks from the perspective of a behavioural biologist (the perspective perhaps most relevant to this thesis) see Croft et al. (2008).

2.1 Fundamentals

Typically, a network is depicted in diagrammatic form as a set of dots or symbols for the nodes, joined by lines or curves for the edges (Trudeau 2013). Networks can be referred to as “undirected” (figure 1 A) or “directed” (figure 1 B) on the basis of the interactions or relationships they represent. An undirected network is one where the nature of a tie or edge is symmetrical, such as two nodes being in a given proximity of each other (Krause et al. 2013) or persistently occurring in the same social groups (Whitehead & Dufault 1999). In directed networks, or “digraphs”, interactions take place from one node to another, in other words, there may be both a “sender” and “recipient” of the interaction. In a directed network, arcs are a term which may be used to denote a directed interaction between nodes, although “edge” or “directed edge” can also be used. Examples of directed edges could be advice seeking (e.g. Creswick & Westbrook 2006) or sent emails (e.g. Newman et al. 2002), or lending between banks (e.g. Iori et al. 2008).

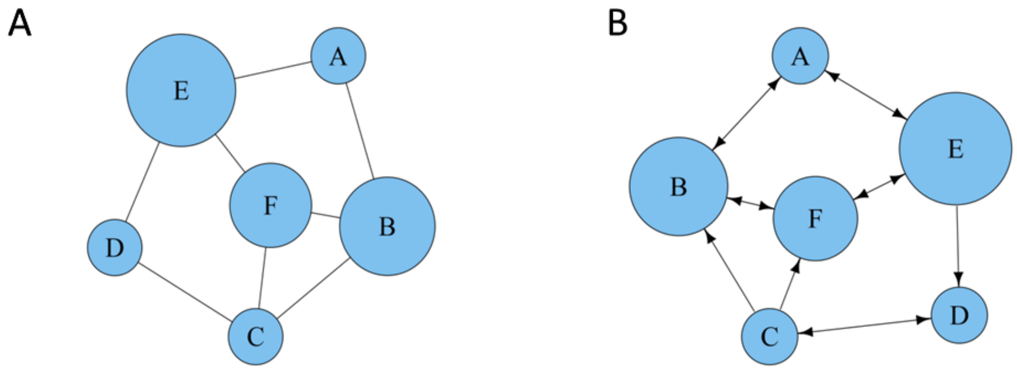


Figure 1 - *A*: An example of an undirected network. The filled circles are the nodes of the network, and the lines connecting nodes are edges. This network is unweighted and can also be described as a “simple graph”. *B*: An example directed network, the arrows between nodes indicated the direction of the given interaction. Interactions can take place in both directions in this example as depicted by the edges between nodes C and D, or in a single direction, as depicted by nodes C and F.

At this point we make the distinction between two types of data: relational and attribute. Attribute data are measures assigned to individual agents, such as body-length, or gender. Relational data considers the relations or interactions between a set of actors. A network is a (structured) relational dataset and so is edge-attribute data (such as measures of the similarity between actors), in the sense that any data-point belongs to two nodes and one node can share data with many others (Croft et al. 2008). Relational data are commonly stored in matrices where each row and corresponding column refers to a specific actor, known as actor-by-actor matrices, so any element will represent a measure between the actors on the corresponding row and column.

An adjacency matrix (or a sociomatrix in the social sciences) refers to a matrix which represents the connections within a set of nodes. Although an adjacency matrix is typically an actor-by-actor matrix, in certain instances, such as when the nodes of networks can take two distinct forms (known as a bipartite network – see Dormann et al. 2009), the number of rows and columns may differ- an example would be a pollinator network where one set of nodes would represent pollinators and the other plants (e.g. Vázquez et al. 2007). A non-zero value in an element of an adjacency matrix tells us that there is an edge between the nodes in the corresponding to the row and column. If the non-zero value is a number other than 1 (i.e. the matrix is non-binary), then this value may imply the strength of an edge in a weighted network (figure 2 A) or the number of edges between nodes (figure 2 B). If the non-zero value lies on the lead diagonal of the matrix, then that may infer that there is a looped or self- edge, as the row and column number will be the same (figure 2 B). In an undirected adjacency matrix, there is symmetry about the lead diagonal, so the values across row *a* will be the same as the values down the column *a*. In other words, if a node in row *a* shares a value of 1 with the node in column *b*, then the node in row *b* also shares a value of 1 with the node in column *a*. In a digraph, there does not have to be symmetry about the lead diagonal, the row usually refers to the sender of a tie or interaction, and the column refers to the receiver (see Wasserman & Faust 1994). A limitation of using an adjacency matrix format to store network-data is that it is inefficient when the number of nodes in the network becomes very large relative to the number of edges. In these instances, a mapping of connections between nodes may be used instead (i.e. node A to node F), which may be called an “edgelist” (e.g. Csardi & Nepusz 2006). On the other hand, an advantage of using adjacency matrices is that there is a large body of mathematical techniques which can be applied to them with direct use in network analysis (see: Mohar 1997). In animal networks, nodes are often in the hundreds (Croft et al. 2008) as opposed to the hundreds of thousands or millions (e.g. the internet (Faloutsos et al. 1999)), for this reason, adjacency matrices are a sufficient means of storing animal network data.

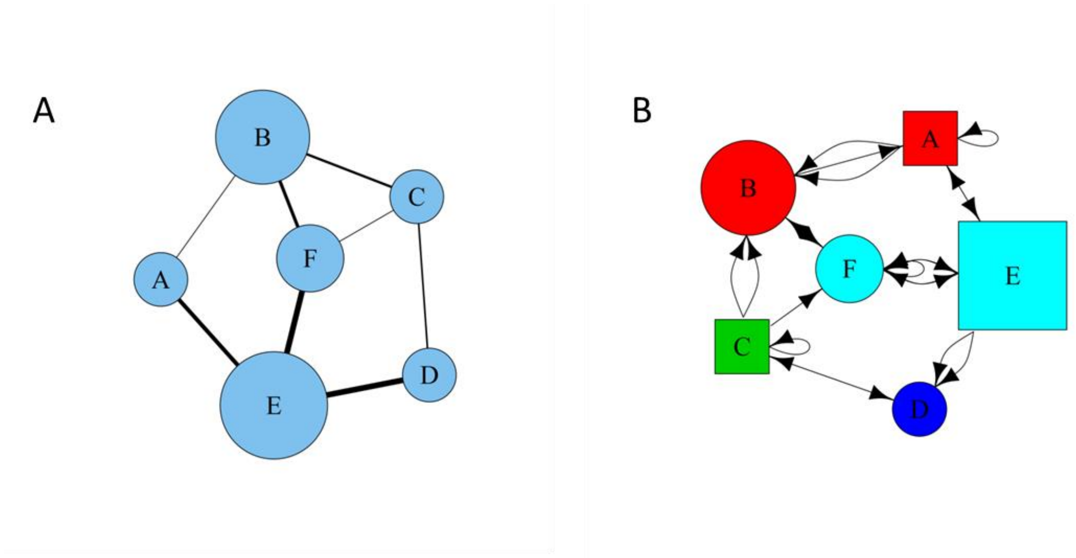


Figure 2- *A*: An example of a simple weighted network, the various thicknesses of edges connecting nodes is used to depict the strength of an edge (i.e. the number of times individuals made contact in a given period). *B*: An example of a "multigraph", where multiple edges can exist between nodes and self-edges or "loops" can be present (node C for example). In this example, the variable size of nodes might be used to represent a given continuous node attribute (such as body-size), whereas the shape of nodes could be used to depict a categorical node attribute (such as gender).

In some networks, edges can be both directed and undirected; these are referred to as mixed graphs (e.g. Richardson 2003). Furthermore, some graphs can have loops, representing an interaction with itself, such as auto-feedback in gene regulatory networks (e.g. Tsang et al. 2007) or multiple edges between any two nodes, often referred to as a "multigraph" (figure 2 B). If a graph is undirected and has no loops it is often referred to as a simple graph (Bollobás 1998a). Thus far, we have only described edges or arcs in the sense that they are present or absent. However, in many cases it is important to consider the strength of the tie or interaction which forms an edge or arc. Networks or graphs which contain this edge attribute data are called weighted networks (figure 2 B). We can thus have weighted versions of directed, undirected and mixed networks (Fletcher et al. 1991).

Directed edges are often depicted using arrows as opposed to lines in the network (figure 1 B), the arrows point from the sender to the receiver of the interaction that constitutes the arc. Edges can also take varying thickness to indicate their weight as previously described (figure 2 B), such as to indicate the amount of trade between countries (e.g. Manna 2008) . Further, nodes can be sized differently to represent continuous node attribute data, such as an individual's body length (e.g. Croft et al. 2006). The use of different symbols or colours can also be ascribed to different nodes to represent categorical attribute data such as gender and disease status (e.g. VanderWaal et al. 2013) (see figure 2 A). Edges may also be dashed, dotted or coloured differently to represent a different category of interaction in a network, such as gene co-expression and protein-protein interactions in a gene network (e.g. Obayashi et al. 2009) (see section 1.3).

Thus far, we have only described what are known as static networks. However, dynamic or temporal networks can be used represent systems where nodes, or the edges between them, may change throughout the observation period. The transient nature of both edges and nodes in these systems may affect the dynamics of process which occur on them (Wey et al. 2008). If we supposed that node F caught a disease in figure 1 A, we might expect it to be able to spread to nodes B, C, and E. However, if F's contacts had happened with B, C and E, before F became infected, then the disease would not be able to spread. Temporal network data are considered in chapter 6, where close-proximity events are measured over time in a herd of Dairy Cattle, *Bos taurus*.

2.2 Network measures

In this thesis, we use certain network measures to explain some of the variation in the structure of animal systems (see chapter 4 and 5). For a comprehensive review of network measures, the reader is directed towards Wasserman & Faust (1994) - a perspective from the social sciences. For a review of network measures from the perspective of behavioural biologists, Croft et al. (2008) provide a review within their book and Wey et al. (2008) have published a review article on this subject. In this section of the introduction, some network measures relevant to the rest of this thesis are introduced. An ongoing theme in this thesis is that given animals are harder to observe than humans, animal networks can be rife with uncertainty and the sampling of the data used to construct the network often cannot be assumed to be free from bias (Lusseau et al. 2008; Rendell & Gero 2013). Different network measures have varying sensitivity to missing and unreliable data - a common issue associated with animal data (see James et al. 2009 for a commentary) which we revisit later in section 2.5.

A number of metrics can be used to measure and describe a network at different levels, ranging from those relevant to the individual node to those relevant to the whole network. The first network measures we might consider are paths and walks, as these also form the basis of other network measures. A path is a sequence of nodes connected by edges that can be followed to from one node to another in the network without any node or edge being used twice. We are often interested in identifying the shortest path between any two nodes - the path which uses the fewest number of edges to get from one node to another. Path length is the number of edges traversed in a path, and the shortest between any two nodes in a network is also referred to as their "geodesic distance" (Bouttier et al. 2003). The distribution of path lengths in a network gives us an impression of how easy the network is to navigate (i.e. how efficiently one node can be reached from another) which in turn affects how quickly things can spread across the network (Watts & Strogatz 1998; Shirley & Rushton 2005).

2.2.1 Node-level network metrics

Most node-based measures used in networks consider phenomena known as centrality. Centrality is a measure of how influential or how involved an individual is in the greater network structure. The simplest node-based measure of centrality is "degree". Degree measures the number of connections a given node has, in a directed network. Degree may be split into in-degree, the number of incoming connections from other nodes, and out-degree, the number of connections being sent out to other nodes. In a weighted network, the term "strength" can be added to all of the above terms to indicate that the sum of edge-weights is being used instead of the raw number of connections. Degree strength can also be referred to as node-strength; in-degree can be called in-strength and out-degree called out-strength. Two other prominent node-centrality measures are "closeness centrality" and "betweenness centrality". Closeness centrality measures the shortest path length between a focal individual and all other members of the social group. Betweenness centrality refers to the number of shortest paths between all pairs of nodes in the network that pass through a given node.

2.2.2 Intermediate level metrics

From node-level measures we can progress to what Wey et al. (2008) describe as intermediate measures of network structure, these include the "clustering coefficient" and "cliquishness". In undirected networks, clustering coefficient is used to measure how close a node's immediate network neighbourhood is to forming a clique, or a complete graph (Watts & Strogatz 1998); in other words, how densely (or sparsely) the network is clustered around the focal individual. Cliquishness describes to what extent the network is divided into cohesive subgroups. A clique is a set of nodes where each node is directly tied to each other (Luce & Perry 1949). Both of these terms link into the concept of network "communities" and the presence of subgraphs or motifs. A community is a group of nodes densely connected to each other, but sparsely connected to the rest of the network, they are identified using algorithms such as random walks - a random walk will spend a long time within a community, but a short time between communities (Pons &

Latapy 2006); this is because the high clustering within communities makes them slow to cross (Watts & Strogatz 1998). Within a community there may be more triangle motifs or subgraphs. Motifs and subgraphs are small structural patterns that may be present in the wider network; a triangle or triad being a set of three nodes with three edges, in this sense a triangle is a clique. Another intermediate network measure is the “component”, a component is could be described as a network fragment, consisting of a set of nodes interlinked with each other, but not interlinked with the rest of the network. A researcher may be interested in the number of components in a network or the size of the largest component, sometimes referred to as a giant connected component (*GCC*) (see Bollobás 2001).

2.2.3 Global network metrics

There are many global measures which can be used to describe the properties of whole networks. Perhaps the most simple of these measures is edge or network “density”; this refers to the proportion of edges that exist out of those possible. In an undirected network, the mean degree of nodes is two times the number of edges over the number of nodes, as an edge is always connected to two nodes. The mean degree and edge density in a network are thus closely linked as the total number of possible edges in a network is the n^2- n where *n* is the number of nodes (Newman 2003). Instead of taking a single statistic (such as mean degree) to examine the connectivity of the network, the distribution of node degrees across the network (the degree distribution) can reveal interesting features of the network, such as a high abundance of highly connected nodes (sometimes called “hubs”) which may otherwise be overlooked (see chapter 4). The assortativity of a network is a correlation coefficient which depicts the tendency for nodes to be linked by edges on the basis of a given node attribute, such as their degree (Newman 2003). Degree assortativity is the tendency for nodes to connect to others on the basis of having either a similar (positive assortativity) or dissimilar (negative assortativity, or dissortativity) number of connections. Both types of degree assortativity can generate meso-scales network features. Positive degree correlation causes a dense core to form in the centre of the network called a core-periphery structure, whereas negative assortativity can cause decentralised network structures with hubs distributed evenly across the network (Newman 2003;Rombach et al. 2014). A final global measurement to mention is the global clustering coefficient, a measure of how many closed triangles there are relative to how many connected triplets of nodes there are (which look like “V”s) in the network (see Wasserman and Faust 1999); this gives a network wide perspective on how often the connections of individual nodes are also connections of each other.

In both chapters 4 and 5 we look for assortativity in networks based not on structural metrics such as degree, but external explanatory factors such as the physical distance between nodes (a particular type of assortativity called propinquity) or the similarity of nodes for a given trait (a particular type of assortativity called homophily) (see McPherson & Smith-Lovin 1987). Given that there are expected resource costs to edges existing, be it time spent maintaining a social relationship, and/or lost reproductive output, knowing which factors promote nodes to be connected can inform us as to the potential fitness payoffs of having or not having connections (see Krause & Ruxton (2002) for review of costs and benefits associated with group living). Given that many network features are known to depend on edge-density (see Croft et al. 2008), it is something that we preserve when testing certain features of our observed networks in chapter 4 and 6 (see chapter 3.5).

2.3 Dependence

Recall that one of the main themes of this thesis is to explain network structure in terms of a number of explanatory variables. One way in which this can be achieved is through statistical modelling, which allows the effect of each explanatory variable to be evaluated in light of other explanatory variables thought to influence the presence and strength of ties in the network. Modelling relational data, such as networks, requires a different set of statistical approaches to those commonly used for attribute data; due to the various sources of dependence in networks (see Snijders 2011 for a review). In this section, we are introduced to one widely-used multivariate modelling approach; the multiple linear regression (MLR). When an MLR is used with attribute data, the significance of each explanatory variable and the model as a whole are often evaluated using tests which have distributional assumptions (such as the t and F-test) (see Sokal & Rohlf 1987). In the first part of this section, the MLR is briefly introduced using a hypothetical example based on attribute data. Some of the assumptions that underpin both the fitting and the testing stage of the model are highlighted- including those based on independence. One set of independence assumptions are nearly always violated in relational data, which make conventional significance tests invalid. The second half of this section covers the causes and consequences of dependence in relational and network data and one way in which significance testing is conducted in light of them.

2.3.1 Modelling attribute data

To demonstrate why and where approaches to model relational data differ from those used to model attribute data, consider a hypothetical example: suppose we wanted to understand the factors which determine the mass of a species of solitary fish in a pond, we have reason to suspect that mass is related to mean daily caloric intake (DCI) and age. Body mass measurements, DCI and age are attribute data, as they are measured from individual fish in the population. The fish have been randomly and independently sampled and can more or less be considered as representative of the pond population.

One generic statistical modelling approach used to evaluate the contribution of a set of explanatory variables to a response variable is multiple linear regression (MLR); the multivariate extension of simple linear regression (SLR) (see Sokal & Rohlf 1987). If we take just part of our fish example, where the response variable body mass Y is described in terms of just one explanatory variable, daily caloric intake (DCI) X_1, a SLR would involve fitting a regression line with the formula $Y=\beta_1 X_1+c$ $(+\mathcal{E})$, to depict the relationship between body mass and DCI. In this equation, β_1 is the gradient of the regression line, representing the strength of the effect of DCI (X_1) on body mass Y, c is the intercept (the mean body mass when all predictor variables are set to 0) and $\mathcal{E}$ is an error term representing the spread of the data around that line, or the final set of residuals left after the model is fitted.

The gradient β_1 of the regression line between body mass and DCI may be fitted using ordinary least squares (OLS) estimation; this gradient is useful for biological interpretation as it provides us with a measure of the strength of the effect of the explanatory variable on the response variable, i.e. how much body mass is expected to increase with increasing DCI. In OLS estimation, the sum of the squared value of the residuals is used to fit the gradient of the line through the data, with the best fit minimising the mean squared error (the average of the squares of the residuals, see Neter et al. 1996). The extension for a multiple linear regression, where there can be two or more explanatory variables. In our example, we now include a second explanatory variable, age X_{2}. Using a common intercept term c, the same estimation procedure is performed for X_2 using the residuals left from the first regression with X_1 as the new response variable; this provides a gradient for the second regression line β_2 and an effect-size for X_2. Note that the intercept term c may not always be biologically interpretable if either of the explanatory variables cannot feasibly take a value of 0.

There are a few assumptions associated with MLR worth highlighting at this point. These assumptions can be split into those which affect the “mechanics” of the OLS estimation procedure which will lead to inaccurate parameter estimation (the βs and the c) and poor model fit, and those which predominantly affect inference using parametric statistical tests. These tests rely on distributional assumptions allow certain properties of a distribution to be estimated using approximations; a key example being the standard error of the mean being approximately the sample standard deviation divided by the square root of the sample size. It should be noted that inaccurate estimates of parameter values from poor OLS estimation will, in turn, lead to poor statistical inference, but that poor inference does dot effect parameter estimation using OLS (although it might lead to the researcher making poor choices regarding model specification).

The major assumptions that affect the mechanics of the OLS estimation procedure are as follows:

Linearity

OLS estimation makes some assumptions about the nature of the relationship between the explanatory variables and the outcome, namely that the relationships are linear with respect to the βs. By this, we mean that the explanatory variables can be transformed (i.e. squared), but the relationship between the final explanatory inputs into the model and response variable has to be linear (i.e. a unit increase of X of leads to the same increase (or decrease) in Y over all values of X). In MLR, it has been shown that if the relationship between an explanatory variable and a response variable is in fact not linear, then the relationship between the two variables will be under-estimated by OLS estimation. At the point of statistical inference, this carries the risk of increased type 2 error (falsely accepting the null hypothesis) with respect to the non-linear variable, and increased type 1 error (falsely rejecting the null hypothesis) for any other variables in the regression which shares some variance with the non-linear variable in question (Osborne & Waters

2002). The linearity of the relationship between two variables can be checked by studying the residuals from a simple linear regression to see if the variation in the size of residuals remains zero-centred over all predicted values of Y- known as a zero conditional mean (Pedhazur 1997).

Negligible error in explanatory variables

The primary assumption of OLS estimation is that there is zero or negligible error in the independent variable, given that the OLS method only attempts to minimise the mean squared error in the dependent variable when fitting the model.

At the point of inference, if there is a high amount of error, or unreliability in any of the measured variables (including the response variable), then the strength of a relationship between the response and explanatory variable(s) will be underestimated by OLS estimation- increasing the risk of type 2 error (accepting the null hypothesis when it is in fact false). High type 2 error for one variable is linked to a higher risk of type 1 error with any variables which share some variance with the unreliable variable; this is because each succeeding variable entered in the MLR has the opportunity to claim part of the error variance left over by the unreliable variable(s) in the OLS estimation. In MLR, if the unreliability in the variables can be estimated, using Cronbach alphas (Cronbach 1951) for example, then the analyses can be corrected to avoid error. Osborne and Waters (2002) provide a correction for unreliable measurements where the correlation between a dependent and independent variable is divided by the square root of the product of reliability estimates (which range between 0 and 1) for each of the two variables; this has the effect of increasing the strength of the correlation to compensate for unreliable data- a β in the model is a multiple of a correlation coefficient.

Independent explanatory variables

The assumption of independence between explanatory variables can be violated in any multivariate model. Collinearity, also known as multicollinearity, refers to situations where any two explanatory variables have a linear relationship with each other (see Wold et al. 1984; Alin 2010)- violating the assumption of independence between explanatory variables. Due to the way in which OLS estimation is calculated (see

Davidson & MacKinnon 1993), small changes in the data or model can lead to erratic changes in parameter estimation when variables are highly collinear. Further, the effects of highly collinear variables on the response variable cannot be separated, leading to model non-identifiability (Dormann et al. 2013). Collinearity may not reduce the power of the model as a whole when predicting within the range of data sampled (Harrell 2013), however, extrapolation is likely to be erroneous if the relationship between collinear factors does not remain constant (Meloun et al. 2002). Further, collinearity can be problematic because it inflates the variance of regression parameters and hence potentially leads to the wrong identification of significant predictors in the model- as some important factors may be falsely rejected from the model (type 2 error) (Wheeler 2007).

Common causes of collinear variables can be: intrinsic, where two measures are linked through a common factor, such as arm and leg-length both being reflective of height; compositional, such as proportion-data, where a high proportion of factor *A* in a sample means that there will be lower proportion of factor *B* and *C*; incidental, where collinearity occurs by chance due to poor experimental design (such as a small sample-size); or the result of model misspecification, where two variables are in fact the measure of the same thing, such as age and date-of-birth (an example of “perfect” collinearity) (Dormann et al. 2013). There are numerous tests for collinearity, including the condition number test (Goldstein 1993) and the Farrar-Glauber test (Farrar & Glauber 1967). Solutions to the presence of collinear variables are numerous, including: mean centring data, omitting one of the variables when they are highly collinear, to using a single latent (unobserved) variable to represent multiple collinear variables (see Dormann et al. 2013 for a review).

Normality

Normality is the first of a set of distributional assumptions which we will now consider that primarily affect statistical inference based on distributional assumptions (see Winer et al. 1971), with lesser or irrelevant effects on model fitting using OLS estimation.

An MLR typically assume that there is a multi-normal distribution of the data (normal for each variable). OLS estimation has been shown to be relatively robust to moderate departures from normality (Berry & Feldman 1985), but given that each data-point is treated equally in OLS estimation, large outliers can bias estimation if not removed prior to fitting. Analyses by Osborne (2001) showed that removing these outliers reduces the probability of type 1 and type 2 errors in significance testing and improves accuracy of estimates.

One of the assumptions of significance tests, such as the F-test and t-test, is that the distributions of errors in the model to be normal, and as such may be sensitive to departures from normality. Given that the effects on significance testing depend on both the test being considered and the way the errors in the model depart from normality, the reader is directed to Box & Watson (1962) for review on how robust regression tests are to this assumption.

Homoscedasticity

One assumption of an MLR which mainly pertains to significance testing is that the error distribution from the model is assumed to be homoscedastic. Homoscedasticity refers to the variance in the error from an MLR being constant across all values of the independent variable. Berry and Feldman (1985) state that slight departures from homoscedasticity (known as heteroscedasticity) have little effect on significance testing, but marked heteroscedasticity can lead to an increase in the possibility of type 1 error. One way in which heteroscedasticity is identified is through looking at the distribution of residuals (error) on the Y-axis for all predicted values of the response variable (shown on the X-axis), if there is an even scatter around the line Y=0 then there is no heteroscedasticity. Tests for certain patterns of heteroscedasticity also exist, such as the Goldfeld-Quandt test (Goldfeld & Quandt 1965) (when variance is proportional to the response variable) and the Glejser test (Glejser 1969) (when variance is greatest at the extremes of the observations). One solution to heteroscedastic error distributions in an MLR is through the use of mixed effects models, which contain terms which can control for unobserved causes of heterogeneity in the error distribution, i.e. by allowing the variance of error to be different at different sampling sites, or for the variability in error to be the response of an independent variable (see Zuur et al. 2009). Other solutions involve using tests which do not require distributional assumptions (introduced in section 2.3.2).

Heteroscedasticity is not a violation of the assumption of OLS *per se* (estimators will still be unbiased), but it does violate the assumptions of significance tests in potentially three ways, firstly modelling error is not homoscedastic (by definition), secondly, there will most likely be some form of dependence in modelling error and finally, the distribution of modelling error may depart from normality (Berry & Feldman 1985; Osborne & Waters 2002).

Independence of errors

One of the assumptions used when for significance testing in an MLR is that each observation of the response variable is independent from all others; more specifically, that the residuals or errors from the model are independent from each other. The assumption of independent errors is a particularly important in terms of this thesis, as it is one of the principal reasons why some network-based modelling approaches use a permutation-based approach to infer significance (which does not require distributional assumptions) (see Manly 1997 for a review of this type of significance testing). The assumption of independent errors may be violated in a number of scenarios, examples include: time-series data, when an observation at a given point in time may be dependent on those before it, known as temporal autocorrelation; model-misspecification, where the response variable is dependent on an explanatory variable not included in the model and in relational data, where data consists of comparisons or interactions between individuals in a group. In all cases, significance tests based on assumptions of independence are invalid. In time-series data, the mean square error may be seriously underestimated. The impact of this is that the standard errors are underestimated and the partial t-tests used to determine the significance of variables inflated - leading to type 1 error (Pindyck & Rubinfeld 1998). The Durbin-Watson test (Durbin & Watson 1950; Durbin & Watson 1951) is one means with which to identify non-independence of this sort and the methods proposed by (Judge et al. 1985) a means to correct this issue.

The non-independence of network observations often resulting in underestimated variability of a sample, which in turn increases the chance of type 1 error (Hanneman & Riddle 2005). In relational data in general, there can be autocorrelation amongst the rows and columns of an actor-by-actor matrix, as all measures in a row or column depend on a specific actor. Krackhardt (1988) showed that row-column autocorrelation can seriously undermine tests based on assumptions of independence. The incidence of type 1 error was shown to increase drastically as the degree of correlation along rows and columns of an actor-by-actor matrix increased, from 10% at $r<0.2$ to 60% at $r\geq0.7$ when an F-test was used to determine significance of a MLR.

2.3.2 Modelling relational data

To illustrate how significance testing is conducted in an MLR using relational data, consider a new species of social fish in a pond; this species of fish has the strange habit (for fish) of feeding each other, trophylaxis. Trophylaxis is by its nature a relational behaviour, as it involves both a donor and recipient. For the sake of using OLS as a means to fit the model, we are particularly interested in the weighted trophylaxis network, not just whether trophylaxis has taken place (if the network was binary maximum likelihood estimation might be used- see Huber 1967). We believe that trophylaxis can be explained by the relatedness of individuals (their genetic similarity) and preferentially takes place between of similar sizes- perhaps for reasons relating to prey handling. For reasons highlighted in this section, each observation of the response variable cannot be assumed independent of the each other. We have highlighted in section 2.3.1 that this may have little effect on the fitting of a model using OLS estimation, but will invalidate the distributional assumptions of parametric tests, such as t and F-tests.

Statistical independence broadly means that each data-point in a dataset, however we define them, conveys no information about any other. A number of dependencies can be present in networks. In this thesis, these are categorised into the “structural” factors where existing network structure affects new network structure i.e. the presence of some edges affect where others might form; “external” factors, those which are not driven by the structure of the network itself, but may drive network structure, and “nuisance” factors, those which serve no useful information about why a network looks the way it does, but affects our choice of statistical analyses. For a description of the effect of nuisance factors see Krackhardt (1988), for an excellent review on structural dependencies in networks and how they can be incorporated into network models, see Snijders (2011).

At a fundamental level, a network is a structured form of a relational dataset. If there is a reason to believe a network is not random structure, there may be merit in trying to explain it in terms of another relational variable, as in our example, where fish trophylaxis might be explained by kin structure and size similarity. Recall that an actor-by-actor matrix is a common way of storing relational data. In cases where an actor-by-actor matrix contains information regarding the differences (genetic difference in for example) between a set of actors, it might be referred to as a distance matrix. Two data points on the same row of a distance matrix, belonging to dyad (a,b) and dyad (a,c) are not independent of each other as both points involve the same actor in the measurement; the same is true between two data points on the same column. To return to our new example, if fish *a* is the largest in the network then all of the values along row *a* of the distance matrix will be positive, and all values down column *a* negative - except for the diagonal value, which will be 0 (as the fish is being compared with itself). Thus there is often correlation amongst rows and columns of a distance matrix, with the values in a given row or column being more similar to each other than with other values in the matrix; this is known as row-column or structural autocorrelation (Krackhardt 1988). Row-column autocorrelation (a "nuisance" factor by our definition) both violates the assumption of independent data-points which underlies a t-test or F-test and also creates heteroscedastic error distributions- where the variability of a variable is unequal across the range of values of a second variable that predicts it (see figure 3). The effects of heteroscedasticity on model inference are described in section 2.3.1.

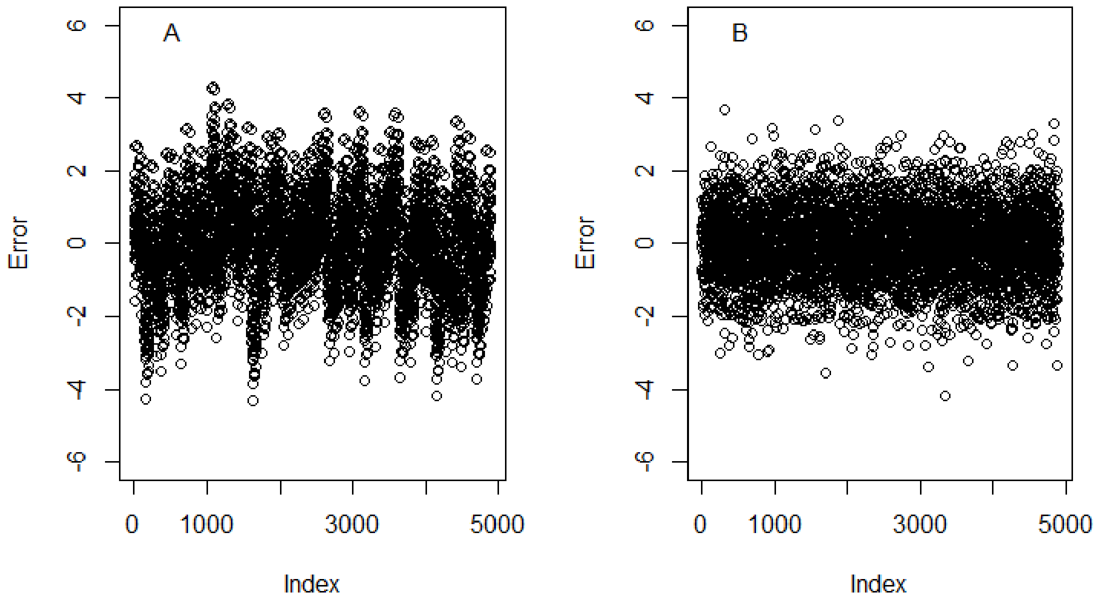


Figure 3- *A*: A heteroscedastic error distribution from a 70-actor distance matrix, generated by comparing the differences between individuals randomly assigned a number from a normal distribution with a mean of 0 and standard deviation of 1 (n=4900). The striped pattern of errors is one characteristic signature of row-column autocorrelation. *B*: A homoscedastic error distribution generated by randomly assigning 4900 data-points from a normal distribution with a mean of 0 and standard deviation of 1.

There are a number of other potentially informative dependencies that have been found through the empirical study of networks (see Snijders 2011 for a review). If we return to the idea of an external explanatory factor influencing the presence of edges between nodes, there have been numerous studies which have found that ties are more likely between nodes that are similar on the basis of a given trait - known as homophily (McPherson et al. 2001). Homophily is a specific example of assortativity (which also includes nodes being more likely to attach if they are different on the basis of a given trait). Assortativity can be both an "external" or "structural" factor in our classification system. In the fish trophylaxis example, nodes of similar size would be more likely to be connected in the network- an external factor. However, if nodes were to assort on the basis of degree or a structural measure (how many fish a given fish feeds and is fed by), then assortativity is, in this instance, a structural factor. Homophily can also take place relative on the categorical status of nodes - such as nodes of the same gender being more likely to feed each other.

Another form of structural dependency that can be observed in many directed networks is reciprocation (Garlaschelli & Loffredo 2004). Reciprocation refers to the presence of one tie increasing the likelihood of another being present in the immediate network neighbourhood. In the fish trophylaxis example, one fish feeding another would result in a greater tendency for the fish to be fed back by the recipient. A simple measure of the amount of reciprocity in a network is the number of bi-directional links divided by the total number of links in the network (see Wasserman & Faust 1994). This phenomenon doesn't have to be contained to pairs of nodes, but can consist of longer cycles. Individual a could feed b, which increases the likelihood that b feeds c, who may then subsequently feed a - creating a cyclic triad (see: Boyd & Richerson 1989; Molm et al. 2007). Reciprocity violates the notion that the occurrence of one event does not affect the probability of the other, as once a link between a and b is formed there is a higher probability of a link leaving b and a link returning to a.

Networks can also exhibit transitivity and clustering. Transitivity refers to the tendency for there to be closed triangles in a network. If two nodes a and b have an edge between them, as do nodes a and c, then there is a higher likelihood that nodes b and c will also have an edge between them, closing the triangle (Girvan & Newman 2002). A network with a high number of closed triangles present will also be highly clustered. The term transitive triangle refers to a motif in a directed network where the exact pattern of ties is *a* to *b*, *b* to *c*, then *a* to *c*. A transitive triangle is different from a three-cycle where *a* is linked to *b*, *b* to *c*, then *c* to *a* (like the above reciprocal feeding example). In a transitive triangle, there is a clear dominant node, in our example *a*, the same cannot be said in a three-cycle. Hierarchies are another form of dependency in some directed networks exhibited by a high number of transitive triangles and few three-cycles. In our fish feeding example, feeding can be directed from low degree individuals to high degree individuals. In animal networks hierarchies can present themselves when an alpha male or female tends to dominate over a beta male or females and both tend to dominate over omega male and females (see Drews 1993); this kind of social structure is common, for example, in primates (e.g. Sade et al. 1988; Sapolsky 2005).

The presence of hierarchies in networks leads onto another structural form of network dependence known as degree differentials (Snijders 2011). If we return to our trophylaxis example, every time a fish gets fed it may firstly gain a little more energy, perhaps increasing its ability to solicit more food. The fish may now attract even more food from low-degree individuals in what is sometimes known as the rich-gets-richer phenomenon (de Solla Price 1976). Conversely, a fish may get known as a feeder and attract more and more hungry fish. Assortativity of the basis of degree is a form of homophily when positive, i.e. high degree nodes being linked to high degree nodes, but it can also be negative (Newman 2002a). Social networks generally have positive assortativity (see Rombach et al. 2014), while biological and technological networks are generally dissortative (i.e., negative assortativity) (Stanton & Pinar 2011). Both forms of assortativity of the basis of degree mean that the connections in a network are not independent of one another.

The presence of all types of structural and nuisance dependencies in real-world networks and their corresponding relational data mean that a common assumption of statistical tests, that data are independent and identically distributed *iid*, does not hold (see section 1.4 of Clauset 2011 for a description of *iid*). Indeed, using models which rely on distributional assumptions for inference on relational data with a moderate amount of correlation amongst column and rows (a structural autocorrelation) biases significance testing to such an extent that it is not uncommon for type I errors of *t* -statistics to exceed 50% (Krackhardt, 1988). Network models may take one of two main approaches to deal with this issue. The first is to specify a model with parameters which represent aspects of network dependency, so that the remaining error in the model is said to be conditionally independent given these parameters and suitable for conventional significance tests (see Snijders' (2011) description of the "p2 model" as an example). Another approach is to use reference distributions, generated through a resampling procedure, instead of distributional assumptions (see Good 2012 for a general review on reference distributions and resampling). Reference distributions are created from creating numerous "null" datasets (4999 in this thesis), where in each dataset, some aspects of the original dataset has been randomised, and other aspects constrained to provide a reflection of the scenarios expected under a null hypothesis. The null datasets can provide a distribution of the test statistic (measured from each null dataset) which can be used to determine if the observed statistic is especially high or low without making any assumptions as to what that distribution should look like (see Manly 1997).

2.4 A Brief History of Network Science

To understand why a large number of the network modelling techniques come from the social sciences and why many network measures and related algorithms come from the physical sciences, it is important to understand the timeline of some of the historical developments in network science.

Network theory has its earliest roots in graph theory- a branch of mathematical combinatorics. Graph theory began with prominent mathematician and physicist Leonhard Euler in the early eighteenth century. Euler created the abstract mapping of nodes and edges which we would describe as a network or graph. Early applications of graph theory involved physical (see Oldham (2008) for an example), chemical (e.g. Cayley 1875) and mathematical systems (e.g. Lhuilier 1813). The book "*Graph Theory, 1736-1936*" covers the "pre-social science" history of network studies (Biggs et al. 1976). Biology was certainly established in the nineteenth century, resembling the modern day discipline (see Mayr 1982; Hull & Ruse 1998). However, the inherent variability and randomness in biological data combined with the difficulty in observing some biological systems high likelihood of data inaccuracy may have posed an insurmountable challenge for mathematical techniques available at this time (Von Mering et al. 2002). The first attempts to incorporate randomness into mathematical modelling did not arise until the late nineteenth century with the work of Thiele (1880), whom described the mathematics behind Brownian motion- a stochastic process.

The first applications of network science to a social system was conducted by psychologist Moreno in the early 1930s (Moreno & Jennings 1934), over half a century before the first network studies of animals (such as Sade et al. 1988). Moreno hoped to explain why there had been an unusually high incidence of girls whom had ran-away from Hudson school in upstate New York – 14 within a period of two weeks. Moreno suspected that the spate of runaways was more to do with the position of the girls in an underlying social network rather than individual personality traits or circumstance. Moreno developed a quantitative method for measuring social relationships known as “sociometry”. Using this technique, he created a mapping which depicted the patterns of social relationship between the girls at the school; this mapping was a network, which Moreno coined as a “sociogram”. Moreno proposed that social influence and information would flow through this network and that it was the girl’s network position that determined whether they ran away. Moreno had come up with “the social network”, which was to become a fruitful area of research in the social sciences (see Borgatti et al. 2002) and later be applied to animal societies (see Croft et al. 2008) - providing a means to model the social environment of the individual.

A significant development in network science came with the introduction of random graph theory in 1959 – which united probability theory and graph theory. Random graph theory arose from the independent works of Gilbert (1959) and Erdӧs and Rényi (1959). A random graph is one which is generated via a random process. In Gilbert’s *G(n,p)* random graph, edges occur independently of each other between n nodes with a probability *p* between 0 and 1. In *G(n,p)* random graphs, the total number of edges between nodes is not fixed. In contrast, Erdӧs - Rényi *G(n.m)* graphs do have an explicitly fixed number of edges. In *G(n,m)* random graphs, any version of the graph which satisfies the condition of having n nodes and m edges has an equal probability of occurring. When n is large, *G(n,m)* and *G(n,p)* graphs are nearly equivalent, where $m \approx pN$ where N is the maximum possible number of edges for the graph (see Bollobás 1998b). The addition of random graph theory suited the analysis of systems which were not deterministic and those which exhibit complex features difficult to formalise mathematically. Random graph theory would aid the study of social networks, where nodes may represent individual humans are often small in size with irregular structure.

In the early 1970s, a number of conceptual advances were made in the study of social networks by social scientists. One of the advances that is touched upon in this thesis (chapter 5) regards the importance of weak ties in social networks (Granovetter 1973). The idea was that strong ties between individuals are often clumped together, in the sense that an individual's close contacts tend to know each other. As a result, some of the information they pass along is redundant. However, weak ties representing acquaintances can easily be unconnected to the clique of the individual, and therefore more likely to be sources of novel information. Weak ties were linked to better jobs and faster promotions-leading to the notion of individuals having social capital (Inkpen & Tsang 2005).

The 1970s also saw social scientists begin to develop model networks and network models. Krause et al. (2014) later distinguish these two terms on the basis that a model network is one generated at random with rules or constraints used to preserve some typical properties, whereas a network model is a statistical modelling procedure focussed towards network data. Model networks would provide the basis for some upcoming network models, being used to produce reference distributions for significance testing (see section 2.3.2). Producing these reference distributions required substantial computational power relative to what was available at the time, whereas using more conventional statistical tests based on distributional assumptions (many of which were developed in the 1940s) could be conducted without computation. The rapid advances in computational power made through the 1970s and 1980s facilitated the greater use of computationally derived reference distributions by academics, which was no doubt a catalysing factor in the development of both model networks and network models.

One type of model network developed at this time was conditionally uniform graphs (Holland & Leinhardt 1976), which feature in both chapters 5 and 6 of this thesis. Conditionally uniform graphs (CUGs) are "random" networks which all exhibit a given set of statistical property constrained by the researcher. Every version of a network that satisfies the constraints has an equal probability of being included in the set of CUGs, whereas networks which don't satisfy the constraints have no probability of being included. In a conditionally uniform model, the significance of a test statistic is inferred via the "conditionally uniform distribution" (a reference distribution, see section 2.3.2), in other words, through seeing how frequently the test statistic could occur in the set of CUGs. For example, we might want to preserve the number of reciprocated ties when testing if reciprocation is more likely between nodes with a similar trait. The conditionally uniform approach is somewhat limited by only being descriptive in nature (not capable of providing effect sizes); the researcher would know whether the observed feature of the network is significant, but not how it would fit into a predictive model in light of other explanatory factors. Another limitation of the approach is the difficulty in producing sets of CUGs with a high number or complex set of conditioning statistics (the things being constrained). Due to this, CUGs were not commonly used in the social sciences as more powerful predictive models superseded the approach in immediate years to follow. However, CUGs were to find use in the later study of animal social networks, where they provided a means to control for uneven sampling protocols common in animal network studies and produce null models which were more representative of the expectation of the null hypotheses being tested (see Croft et al. 2011).

The 1980s saw the first multivariate attempts to statistically model networks by social scientists and statisticians. Many of these approaches are covered in detail in Snijders (2011). Some prominent examples of network models developed at this time included variations on latent-space models. Latent space models developed in the 1960s provided a foundation to model unknown sources of dependency as unobserved latent variables (those that are not directly observed but are rather inferred). The p1 model was the first latent-space type approach to be suitable for use in directed networks. It used two parameters to describe the potential for each node to send and receive edges, as well as parameters for the total number of ties and the tendency toward reciprocation. A weakness of the model was that it required a large number of explanatory factors, at least 2 for each node which led to issues regarding over-parameterisation of the model, it also assumed that dyads were independent of each other (see section 2.3.2 as to why this is often not the case) (Holland & Leinhardt 1981). Various modifications were proposed to reduce dimensionality and later the p2 model would broadly replace the p1 model altogether (Robins et al. 2007). In the p2 model, the presence or absence of ties would be regressed on the actor and dyad-based explanatory variables with the addition of terms to partition the error in the model on the basis of both the “sender” and “receiver” of the tie (making the error distribution conditionally independent, see section 2.3.2); thus the p2 model is a random effects model (Snijders 2011).

An important class of multivariate statistical method developed in the social sciences at this time was the MRQAP (Multiple Regression Quadratic Assignment Procedure) (see Krackhardt 1988). The MRQAP used OLS estimation to fit a multiple linear regression (MLR) as previously described (see section 2.3.1). The significance testing in the model was achieved through the use of a permutation procedure to produce reference distributions of the model statistics; specifically a node-label randomisation called the quadratic assignment procedure (QAP) (Hubert & Schultz 1976). Unlike many other approaches, the MRQAP was not developed with the goal of assigning effect-sizes to the structural causes of dependency in networks (see section 2.3.2), only the external factors which explain the presence and absence and/or the weight of edges. An advantage of the approach over the "p" models was that it allowed the analysis of almost any type of relational dataset: weighted, binary, directed, non-directed, and networks with loops. The MRQAP (and its variations) is the central network model used for many purposes in this thesis.

Another type of network model developed around this time could also provide meaningful effect sizes for the structural causes of network dependency; this family of model was originally named the p* model (Wasserman & Pattison 1996) and later the "exponential random graph model" (ERGM) (see Snijders 2011). ERGMs enabled the examination of the underlying mechanisms of network factors and processes that generate non-random network structures (Anderson et al. 1999, Robins et al. 2007). ERGMs, closely related to logistic regression, use stochastic modelling to determine the probability that a social link exists among individuals based on a set of predictor variables, which initially could only take the form of structural measures, such as number of edges or counts of subgraphs such as triangles and stars (see Robins et al. 2007). ERGMs represented the dependency between ties directly, instead of conditioning on latent variables. When using structural measures as sufficient statistics (i.e. that these measures can capture all the dependence in the network), ERGMs were based on conditional independence assumptions between the observed tie variables. A key conceptual advance underpinning ERGMs was the assumption that the observed network was generated by a process where existing ties could affect the formation of new ties - which replicates how we might expect real networks to form. Given this assumption, ERGMs had the ability to meaningfully test for the effect of structural sources of dependency on the observed network.

Moving away from multivariate statistical models, the return of physicists to the study of networks in the latter portion of the twentieth century transformed the types of question asked of networks. Physicists were suited to describe the overall shape and statistical properties of empirical networks- i.e. those based on measures of real-world phenomena (Newman 2008). Drawing upon a background in statistical mechanics, they could provide analytical results for those networks which were large enough that both the number of nodes and edges could be assumed infinite and the ratio between them as constant (Hill 1963; Albert & Barabási 2002; Castellano et al. 2009). As pointed out in Newman (2008), the addition of physicists to the study of networks provided new theories, analytical techniques, models and algorithms; all but the most-recent of these are summarised in Newman (2003).

Network theory has seen appreciable uptake by behavioural biologists to study the behaviour of group living animals over the last two decades (see Croft et al. 2008; Krause et al. 2014). Networks provided a means to represent an animal's social environment and further determine how that environment influences individual behaviour. One main difference between studies of human and animal social systems is in the sampling and construction of the social network. Humans can self-identify their social bonds, whereas animal networks have to be inferred from observing social interactions (behaviours such as allogrooming) or patterns of association. Constructing animal social networks was, and still is challenging, requiring rich relational datasets often hard to obtain from wild animals (Whitehead & Dufault 1999; Croft et al. 2008). One particular framework developed to infer social relationships from repeated associations was the "Gambit of the Group" (GoG) (see Franks et al. 2010). The idea behind GoG is to infer social ties between individuals through seeing how often they occurred in the same social group over repeated observations. In early studies, observations had to be carried out directly by human observers, which introduced a number of biases; animals more accustomed to the presence of human observers may be sighted more often and unsighted animals may still be in a group together (Croft et al. 2008). Indices primarily developed for use in ecology were employed to determine the strength of association (edges) between two individuals, such as the half-weight index (Cairns & Schwager 1987), which provided an estimate to the likelihood that two individuals would be seen together compared with the likelihood of seeing any of the two individuals upon encountering a group.

The most recent development in network science involves the study of the temporal dynamics of networks (see Hanneke et al. 2010; Blonder et al. 2012; Holme & Saramäki 2012). Temporal networks consider the order in which edges are formed, maintained and dissolved over time within a potentially changing set of actors. Temporal networks have particular importance when the ties between actors are transient or when the network's structure evolves and changes substantially with time. Static networks do not consider the order of ties between actors, just a representation of those that have occurred, and perhaps the strength of their occurrence over a given time period. If these ties are not persistent, then the order in which they occur could affect spreading and navigation processes on the network (e.g. Gauvin et al. 2013).

Recent advances in radio-frequency identification technology have allowed interactions in animal systems to be monitored in real-time; this has both been referred to as the remote-sensing and the reality mining of social interactions (see Krause et al. 2013). For example, insects can be tracked travelling between nest-colonies using passive radio frequency identification tags (RFID) (e.g. Sumner et al. 2007), active RFID tags such as radio-proximity loggers can be used to identify when larger animals are within a given proximity of each other (e.g. Hamede et al. 2009; Drewe et al. 2013; Weber et al. 2013), and the reduction in size of GPS tracking systems, although not a radio-based, can be used to provide the positions of wide-ranging animals in space (e.g. Wolf et al. 2007; Zbinden et al. 2011). These technologies have provided the high resolution data required to study the temporal aspects of animal networks in fine detail. They have also allowed the contacts between animals to be monitored over longer continuous timeframes than previously possible (Ryder et al. 2012). A number of modelling approaches have been developed for temporal networks (e.g. Hanneke et al. 2010). Relational events models are one recently developed means to model the sequence of events that unfold in a network, with the particular advantage of taking into account different ways in which data can be collected (e.g. Patison et al. 2015). Temporal networks are still a relatively young and developing branch of network science; for now, many metrics are generalisations and models extensions of existing techniques for static networks (see Holme & Saramäki (2012) for a comprehensive review).

2.5 Lost in Translation

The multidisciplinary background of network science has fostered a range of different approaches to describe and analyse networks (see reviews by Wasserman & Faust 1994; Newman 2003; Snijders 2011; Croft et al. 2011). Given the history of the network science, many of the techniques and metrics used in the study of animal networks have been borrowed from other disciplines with a longer pedigree in network analysis. As highlighted in Croft et al. (2008) and James et al (2009), care has to be taken when techniques, measures and terminology are used outside of the context for which they were developed. This is especially true in the study of animal networks, which sit apart from human and physical networks on account that they are often unevenly sampled and rife with uncertainty (Lusseau et al. 2008). In a thesis where statistical approaches developed in the social sciences are taken and applied them to animal networks, it is worth recapping some of the issues which have been highlighted by others in the field.

2.5.1 Borrowed Metrics

There are a wide variety of metrics which exist to measure certain network features, some of which might be very sensitive to the way in which animal network data are sampled. Animal network data can often be biased in certain ways which should be considered when choosing metrics. For a review of these issues and their effects on the choice of network metrics, the reader is directed towards James et al. (2009) and more recently Farine and Whitehead (2015); in this section, a couple of relevant issues from that article are highlighted.

The first issue concerns how sampling protocol can influence network metrics. Consider a typical animal network, constructed via gambit of the group (GoG). It has been recognised that this method of generating networks results in a higher incidence of transitive closure (James et al. 2009); this means the networks produced are more clustered. Due to this, when determining whether there is significant clustering in this network, the manner in which it was constructed needs to be recognised as contributing some clustering regardless of the “true” social structure; this can be achieved through comparing the observed network to null networks constructed under the same protocol as the observed from a permuted null dataset (Bejder et al. 1998; Krause et al. 2011). Note that null models such as the QAP from the social sciences do not control for sampling biases in such a manner, and treat the structure of edges in the network as given (see Croft et al. 2008).

The second issue concerns how certain network metrics have varying sensitivity to missing or unreliable data. For illustrative purposes, we will consider the sensitivity of two node-based measures: degree and betweenness (see section 2.2), to misplaced or misallocated edges. If a hypothetical network is partitioned into two communities, where one individual acts as a bridge between the two; this individual will have a very high betweenness, featuring in any short path between two individuals of different communities. Other individuals in this particular individual's immediate network neighbourhood (which may also be referred to as the "first-order zone" - see Wasserman & Faust 1994) may also have high betweenness as many paths will pass through them on their way to traversing this bottleneck in the network. However, now consider that an interaction has been missed which also linked these two communities in the network. In terms of degree, this is not disastrous, as two individuals in the network have one fewer edges attached to them then they should. However, a new route between the two network communities could have near global consequences to the betweenness values attributed to individuals (James et al. 2009; Farine & Whitehead 2015). Perhaps this is a worst case scenario, but it does highlight that we should consider the protocol used to determine the presence of edges when borrowing metrics from the social sciences, where "betweenness" originated (Freeman 1977). Degree is by no means safe either if there are issues pertaining to how nodes are defined. Falsely lumping actors in a network could artificially inflate their degree to an extent where they perhaps become a hub (see chapter 4); hub individuals are very important in network processes such as the spreading of information or disease (Pastor-Satorras & Vespignani 2001; Pastor-Satorras & Vespignani 2002).

2.5.2 Borrowed Models

Models developed in the social sciences present themselves as powerful predictive tools for the study of networks. However, just like the above-stated network metrics, many of these models were developed in the social sciences for evenly-sampled, reliable datasets. Rendell and Gero (2014) and Krause et al. (2014) both highlight the potential dangers in taking models developed for human social networks and applying them to animal networks which may be unevenly sampled and rife with uncertainty (Lusseau et al. 2008).

If we take a p* or ERGM as an example (Wasserman & Pattison 1996) (note that these models are not used in the substantive chapters of this thesis), we must first appreciate that these models are essentially just an extension of a logistic regression (where the presence of a network tie is the "event") conditioned on some structural measures, such as subgraph counts (triangles, stars etc.). Sometimes, with animal data, the presence of edges are known with confidence, but questions may be raised as to whether the absences of edges result from the sampling protocol or the true biology of the system. In such a situation the researcher should consider the appropriateness of using a model based on a response variable of 1s and 0s when the 0s are not trustworthy. Logistic regression implicitly assumes that both sensitivity and specificity of the response variable are 100% (Copeland et al. 1977; Quade et al. 1980; Magder & Hughes 1997). By sensitivity, we mean that we can detect the presence of edges with perfect resolution- i.e. that there are no false negatives. By specificity, we mean that there are no false positives i.e. misallocated or falsely allocated edges. When either sensitivity or specificity is less than 100%, estimated logistic regression coefficients tend to be too close to zero; this makes it harder to detect relationships between explanatory and response variables. At the same time, the precision of the estimated coefficients is overstated, resulting in confidence intervals that tend to be too small (Magder & Hughes 1997).

For another example, if we take a MRQAP (Krackhardt 1988) (these models feature heavily substantive chapters of this thesis), the node-label permutation procedure (the QAP) assumes that the structure of edges in the networks is fixed. Croft et al. (2008) highlight that the QAP randomisation procedure, which is focussed on permuting the aspects of the network, can have limitations when the network is derived structure generated from sampling associations between animals. They highlight that null models centred on permuting the raw-data can better account for sampling biases and point to the works of Whitehead & Dufault (1999) and Whitehead et al. (2005). However, these raw-data or "data-stream" (specifically used to refer to GoG raw-data) (see Bejder et al. 1998) based permutation procedures are not used in an MRQAP model and, as such, we should be wary about applying these models to animal networks based on measures of associations as they currently stand. In this thesis, the edges in the network are not derived from group associations, but instead are measured directly (in this sense they are raw-data); however, this does not mean that the data are free from sampling issues. In chapter 4, the use of a MRQAP would require what is essentially missing data to be assumed as indicative of an absence of an edge, whereas in chapter 5, the null model behind a QAP would allow physically impossible null networks to occur, reducing the relevance of the reference distribution for significance testing.

2.5.3 Summary

The concept of known-knowns (what we know that we know about a network), known-unknowns (what we know that we don't know about a network) and unknown-unknowns (things that exist about the network which we haven't thought about or yet been able to characterise) should be at the forefront of anyone analysing animal networks. These concepts have been popularised in statistical thinking by Silver (2013). When considering metrics and analyses of animal networks, choices should be geared to be as reliant upon known-knowns as possible, whilst being as robust to known-unknowns as possible. There is perhaps nothing that can be done with regards to unknown-unknowns, only to accept that overtime some of these will become known unknowns and as such we should be adaptable in the techniques we choose and prepared to re-evaluate what we know and the techniques we use as this information comes available.

3. Typical approaches to analyse a network

3.1 Preface

This chapter introduces some of the typical questions asked of networks across different disciplines and some of the broad methodologies used to answer them. The unique challenges faced in the study of animal networks are highlighted, along with some of the current approaches used to overcome them. The beginning of this chapter is roughly partitioned on the basis of discipline, although it should be noted that many network studies are interdisciplinary in nature. The weighting assigned to each discipline in this section increases as we go on; this is not a reflection of the relevant contributions of each discipline. The weighting instead reflects the goals of this thesis, to incorporate network modelling approaches from the social sciences into the study of animal networks. At the end of this chapter, we introduce the three approaches we will use in this thesis to facilitate the multivariate statistical analysis of animal network data and some of the relevant computational tools used in these approaches.

3.2 Network studies in Physics

The network studies conducted by physicists are often based on empirical studies of large real-world networks, such as the internet. The large size of these networks allows the use of approximations developed in the field of statistical mechanics (Hill 1963; Croft et al. 2008; Newman 2008). In contrast to animal networks, the networks studied by physicists often contain little uncertainty; their edges being inferred from electronic (e.g. Faloutsos et al. 1999), physical (e.g. Crucitti et al. 2004) or chemical signatures (e.g. Uetz et al. 2000). Physicists are often interested in the broader scale properties of a network, its shape and its statistical properties (Newman 2008). The focus on network properties and behaviour differs from the interests of social scientists, whom may be concerned with how the position of individuals in a network might affect their behaviour and the factors which determine the structure of social networks. A review into the interests and developments in studying networks by physicists is provided in Newman (2008).

Physicists, computer scientists and mathematicians all have their own ways of constructing models which can help us to understand the important structural properties of an empirical network. These models can help us understand the interplay between a network's structure and the processes that take place on it, such as the flow of traffic in a transport system (e.g. Holden & Risebro 1995) or the spreading of a computer virus (e.g. Pastor-Satorras & Vespignani 2001; Newman et al. 2002). The majority of network models constructed and studied by physicists are often directly practical. If a network model is to be used over the observed network, the model will be based on some important features or dynamics observed in the real-world. Newman (2008) summarises the principal approach as being to list possible mechanisms that might be responsible for determining the shape or statistical property of a network and then make a model incorporating some or all of those mechanisms. The networks produced by the model can be examined and used for further modelling, such as how different immunisation strategies might affect the spreading of disease (Pastor-Satorras & Vespignani 2002) or how robust the flow around a transport network is to road closures (Jenelius et al. 2006).

3.3 Network studies in Social Sciences

Many social networks exhibit neither perfectly random nor perfectly regular structure, which can make them difficult to model algebraically, and they are often too small for some of the approximations that physicists use (Newman 2008). In the physical sciences, a key research goal has been describing the global characteristics of large empirical networks. In contrast, social scientists have tended to focus on the variation in structure across a network, using these variations to explain differences in individual characteristics and the outcome of individual actions. Another common goal of network studies in the social sciences is to explain the formation of ties between individuals in terms of a number of explanatory factors (Borgatti et al. 2009). These factors may be structural in nature (e.g. degree assortativity), or take the form of non-network derived relational data, such as the differences in individual's salaries. A number of statistical models have been developed (see Snijders 2011) to facilitate the multivariate analysis of social networks. In this section, we will particularly focus on how one model, the multiple regression quadratic assignment procedure (MRQAP) (Krackhardt 1988) and its extensions (Dekker et al. 2007; Butts 2014) are used to investigate the structure of social networks.

Let's now consider an example network, Y, to provide an example of a network with features typical to those studied in the social sciences. *Y* is a static network consisting of fifty nodes representing individual academics; the edges between nodes have been inferred through summing the email messages sent to and from colleagues in an academic institution. For the sake of argument, let’s pretend we know the explanatory factors which determine the formation of ties in this network: ties are more likely between: colleagues in the same department X_1; colleagues with a similar Erdös number X_2 (the distance in a citation network to Paul Erdös, a network scientist and mathematician with the most publications of any mathematician in history); colleagues with similar salaries X_3, and ties are likely to form closed triangles X_4. We can summarise this by saying that $Y{\sim}f(\beta_1 X_1, \beta_2 X_2, \beta_3 X_3, \beta_4 X_4) + \mathcal{E}$, in other words that Y is a function f of the four explanatory variables- f is a logistic function in this case. The betas β denotes the contribution of each explanatory variable to the outcome. X_4 is the only factor in this hypothetical example that is purely generated through network structure. We will add that individuals can belong to more than one department with regards to X_1 and that there is some substantial collinearity between salary X_3 and Erdös number (Grossman 2002) X_2. Recall that collinearity is where the two explanatory variables have strong covariance. Anderson & Robinson (2001) and Dekker, Krackhardt, and Snijders (2007) showed that collinearity can also cause problems when permutation procedures are used to test for the significance of explanatory variables.

There are a number of models a social scientist could use to determine the relationship between the response variable Y and the four explanatory variables. One choice which may be considered is the MRQAP (Krackhardt 1988). The MRQAP consists of two parts; the first is a multiple linear regression, fitted using typical procedures such as ordinary least squares (OLS) estimation and partial regression coefficients. The second part is a Quadratic Assignment Procedure (QAP) (node-label) permutation, used to evaluate the significance of the coefficients (denoted as β_1 through to β_4) and the fit of the model as a whole. The QAP leaves the structure of edges in-tact; this is attractive as it preserves the row-column interdependence in each permuted dataset used to evaluate the observed data. The original MRQAP procedure was developed by Krackhardt (1988) under the restrictive assumption of independence between all variables (including between the independent variables and the response variable), but it was still valuable in the sense that it provided a way of testing significance which didn’t require distributional assumptions such as *iid*.

An advantage of this approach is it is based on a linear regression (but has also been extended to a logistic regression in Butts (2014)) so the researcher could use the number of emails to create a weighted network Y, retaining as much information as possible about the system. In the MRQAP, the response and explanatory variables are stored as square matrices. Categorical variables such as “belonging to the same department” X_1 can be scored as a 1 or 0 to denote membership of the same or different departments respectively. If the researcher believed that the tendency to send and receive emails was different between departments, then this could only be modelled using a set of dummy variables. Dummy variables are categorical variables which can shift the intercept of the model with the Y-axis, in this case, depending on the department.

A minor limitation of the MRQAP might be that there are no “random-effect” capabilities which could otherwise be used to attach terms to partition and represent the error in the model. In our email example, we might want to use a random effect to describe the different error around a model fit in different departments. This would be informative if we expected that one department had a large variation in the emails people sent and received compared to another, but that each department’s pattern of emails was driven by the same combination of factors. The lack of random-effects in a MRQAP is not limiting for significance testing as no assumptions are required about the error distribution, as the QAP generates a reference distribution to test the significance of the fit of the betas and the model as a whole. One of the uses for random effects in “conventional” mixed effects models (which typically use distributional assumptions to infer significance) is that the error in the model can be made conditionally independent and normally distributed through the use of these variables (see Zuur et al. 2009).

A more pronounced limitation of the MRQAP is that the only way to incorporate network-based dependencies such as X_4, reciprocity, would be to have them represented as covariates. A dichotomous version of matrix Y could be reflected along the lead diagonal to represent X_4. However, the dynamics of social networks are complicated because structural network dependencies such as reciprocity are endogenous feedback effects (i.e. as the network forms, the presence of edges early in that process affects where the latter edges are placed) - so a model such as an MRQAP, which only considers the end-product of this process could only ever be descriptive of these traits (see Snijders 2011). In contrast, the membership of an individual to an academic department will remain largely constant throughout the formation of the network.

When a QAP is used to evaluate the contribution of each explanatory factor in the model, the null hypothesis varies depending on whether the *X* variables or the *Y* variable is permuted. When a given *X* is permuted, the null hypothesis is that this factor has no relationship with *Y*, but the other variables (we will refer to these as the "controlled variables" in the regression) may still have a relationship with *Y*. X-permutations destroy the relationship between the *X* variable being permuted and the controlled variables in the regression; this is justified when there is no relationship between explanatory variables, but not when there is collinearity (see section 2.3.1). Collinearity means that an X-permutation would violate what is known as the "ancillarity principle" (Godambe 1982), which states that any dependence between independent variables should be kept intact. The violation of this principle can lead to incorrect type I error rates (Welch, 1990; Ter Braak, 1992; Anderson & Legendre, 1999). The null model when permuting *Y* is subtly different, that *Y* is not related to the *X* variable at hand and not related to the controlled variables in the regression. It hence is a test for the hypothesis that the β being tested is 0 as is the coefficient of the controlled factors in the regression (Dekker et al. 2007).

Different researchers have taken different stances on which variable should be permuted by the QAP. Manly (1997) recommended that the response variable should be permutated, but others have argued against this (Kennedy and Cade 1996), especially when non-pivotal statistics are used (a pivotal statistic is a function of observations and random effects and has a distribution not influenced by any unknown "nuisance" parameters). However, partial regression coefficients (βs) and t-statistics often used to access significance are often pivotal under multivariate normal distributions. I.e. the probability distribution of the statistic under the null hypothesis does not depend on unknown or nuisance parameters. The t-statistic may become non-pivotal when *Y* shows strong row-column autocorrelation (see section 2.3.2) or if it depends on the "latent" coefficient describing the collinearity between two X variables i.e. X_2 and X_3, for instance.

Due to the fact that there definitely is collinearity in our theoretical example, we would opt for a residual-based QAP permutation known as a "double semi partialling" QAP (QAP-DSP) (see Dekker et al. 2007 for a more detailed explanation). The QAP-DSP and another similar approach, Freedman-Lane Semi-Partialling (FLSP) (Freedman & Lane 1983) are both residual permutation procedures. In this thesis, we use the QAP-DSP (chapter 5) so we will use this approach as our main example. In essence, the QAP-DSP first converts the *X* variable being tested into its residuals through modelling *X* against the controlled variables (the other *X* variables combined); this partials out the effect of the control variables on *X* and thus conditions *X* to be independent of the control variables. The X-residuals are then permuted and then regressed against the *Y* and the controlled variables. As such the controlled variables enter the regression twice, once against the X-variable being tested, and again when the residuals of *X* are regressed against *Y*, hence the term "double." The QAP-DSP approach thus minimizes the correlation between the focal variable and the control variables under permutation. The QAP-DSP therefore does not violate the ancillarity principle through conditioning on the nuisance statistics- any correlation between any given *X* and the controlled variables. The DSP and FLSP approaches are the most resilient of the QAP type approaches, identified to be robust to row- column autocorrelation, and all but extreme levels of spuriousness (the effect of a confounding factor) and skewness in the data (Dekker et al. 2007).

To truly gauge the effect that reciprocity X_4 has on the system (given one important assumption), a different approach to the MRQAP is required. Exponential random graph models (ERGMs) or p* models (Holland & Leinhardt, 1981; Frank & Strauss, 1986; Wasserman & Pattison, 1996; Snijders et al. 2006) are often used to model networks in terms of their structural dependencies (see section 2.3.2), and on this basis warrant a mention in this chapter even if not explicitly used in this thesis. In theory, ERGMs can provide predictive analysis of network effects such as reciprocity X_4 on a single static network *Y*. ERGMs are a family of models for dichotomous or other discrete network data with a greater focus on modelling the specific network-related dependence structure. The logic behind an ERGM is that the observed network is the outcome of a stochastic process, where a network is grown edge-by-edge through a combination of chance and perhaps other network effects and node attributes. In other words, the network is treated as a self-organising system of relational ties. The observed network is thus regarded as one realisation from a larger set of possible networks with similar important characteristics; this is a fundamentally different standpoint to an MRQAP which regards the observed network as a fixed entity. To begin with, we don't know what the exact process which generated *Y* was, so we seek to form a model based on a plausible and principled hypothesis as to what this could be (see Robins et al. 2007 for an introduction).

We will follow a simple example where we are just interested in the structural effect of reciprocation X_4 on the observed network Y. To help determine whether it is worth including X_4 in the ERGM model, we could take a turn back to the MRQAP. Given our example $Y \sim f(\beta_1 X_1,\ \beta_2 X_2,\ \beta_3 X_3,\ \beta_4 X_4) + \varepsilon$, the MRQAP may have identified that reciprocated ties in the network were associated with edges, alternatively, we could compare the observed amount of reciprocated ties with a suite of random networks to show that the amount observed rarely happens by chance. If so, we can conclude that the structural characteristic of reciprocated ties is the outcome of a social process and not chance. Then, as a starting point, we may posit a stochastic network model with two parameters, one that reflects the propensity for ties to occur at random and one that reflects our suspicions that there is a propensity for reciprocation to occur. Further constraints can be imposed on the model such as keeping the same set of nodes and the same number of edges. Adding reciprocation into the model reflects an expectation about what sort of networks are more likely to occur. The range of possible networks, and their probability of occurrence under the model, is represented by a probability distribution on the set of all possible graphs that satisfies the constraints of the model i.e. the number of nodes and edges. In this distribution of graphs, those with a high amount of reciprocation will have higher probabilities than those with little reciprocation. The precise probabilities associated with these graphs, of which the observed graph is a part, depend on the value of the reciprocity parameter. At this point, we don't know what that value of this parameter is. Akin to logistic regression, a form of likelihood estimation is used to assign values to parameters so that the most probable amount of reciprocation in the ensemble of networks generated matches that which is seen in the observed network. Monte Carlo maximum likelihood techniques are often used for estimation (Snijders 2002). The Monte Carlo approach simulates a distribution of random graphs from a starting set of parameter values, and subsequently refines the parameter values through repeated comparison of the distribution of graphs produced against the observed graph, with this process repeated until the parameter estimates become stable (known as convergence). We can then explore the range of network outcomes predicted by the model and make inferences about model parameters. We can infer whether any model parameter is significantly different from zero through examining whether the corresponding configuration is present in the observed graph to a greater or lesser extent than expected by chance, given other parameter values. In the example, a model that is a good fit to the data in terms of reciprocation would be expected to have a positive reciprocity parameter. A well-fitted

ERGM will predict networks structurally similar to the observed network using few parameters.

The MRQAP and ERGM each has its strengths and weakness, both models were developed with high quality data in mind. Although both of these models are multivariate and have been described as predictive, it is worth noting that cause and correlation can be very hard to determine without experimental manipulation of the system (as highlighted in: Croft et al. 2008; Pinter-Wollman et al. 2014; Farine and Whitehead 2015). In our example, Y could very well be the factor that causes X_2 and X_3. Individuals that send more emails may very well end up in higher paid jobs and engage in more collaborations than those that shun their inbox.

3.4 Network studies in in Behavioural Biology

Behavioural biologists may be motivated to study many of the same phenomena in animal social networks as social scientists study in human social networks. However, the study of animal social networks requires fundamentally different methodologies than the study of human social networks. Animal networks, like human social networks, are often small and neither completely random nor completely regular. An initial interest of a behavioural biologist might be to investigate how the societies of group- living animals are organised. However, there are challenges unique to studying animals as opposed to humans or physical systems- many of which are covered in Croft et al. (2008). The first hurdle that may present itself is how individuals are identified, animals do not come with an Internet Protocol (IP) address like most computers, nor do they have faces which we have evolved to recognise. As such a mechanism is needed to identify individuals, or even groups of individuals. In many studies individuals have been identified through their markings. To use an example from Croft et al. (2008), it will be easier for species such as giraffes than it will be for ants, however, ants can be painted with nail varnish to make unique markings and contained in a laboratory setting- giraffes cannot. There is always an underlying risk that individuals may be misidentified in the wild when marking-based identification is used. An alternative is using tags to identify individuals; however, even this type of approach can lead to missing or misplaced data. Tags are designed to be unobtrusive to the animal, but practical to the researcher, this trade off means that sometimes an individual may be orientated so that its tag cannot be seen, the tag may get dirty, or fall off entirely (as experienced in chapter 6).

Once individuals have been identified, the researcher may then choose the manner in which he or she wants to construct the social network of the group. Defining the group is often more complex for animals than humans, as humans social interactions may be partitioned into easily identifiable groupings, such as school classes, companies or other institutions (Croft et al. 2008). In cases where these groupings may be questionable, techniques such as snowball sampling can be used (see Wasserman & Faust 1994). In human systems this type of sampling might be easier to achieve than with animal networks, where data-collection can be laborious and sampling limited by resources.

Unlike humans, animals cannot be surveyed for their social ties directly, so the social network has to be inferred through studying their behaviour as a proxy. This can range from directly observing direct interactions such as trophylaxis or allogrooming (e.g. Carter & Wilkinson 2013), to inferring social ties through associations (such as proximity (e.g. Zhang & Horvath 2005; Handcock et al. 2009; Haddadi et al. 2011), shared use of space (e.g. Bailey et al. 2001; Dyo et al. 2010), repeated co-occurrence in social groups (Franks et al. 2010). Associations carry the risk that there is indeed no social relationship and that animals are linked only by routine or resource aggregation. However, as Rendell and Gero (2013) state "It is difficult to envisage a situation where social structure would ever exist independently of spatial ecology—the former evolves within the constraints of the latter". Many primate studies have employed the first approach (O'Brien 1993), constructing social networks from ethograms outlining the behaviour of individuals with a particular interest in the social interactions (such as allogrooming); this requires a high amount of human observation and a single observer is limited by how many individuals can be observed concurrently. Social networks constructed through observations of social interactions are often small in size, in the tens of individuals. Recall that inferring ties through association via repeated co-occurrence in the same social groups is known as using "the Gambit of the group" (GoG) (Franks et al. 2010); this approach has been used on lots of aquatic organisms (Krause et al. 2014) such as guppies *Poecilia reticulata* (Croft et al. 2004), bottlenose dolphins *Tursiops truncates* (Lusseau 2003), and sperm whales *Physeter macrocephalus* (Lusseau et al. 2008). The addition of radio-proximity technology such as RFID tags and radio proximity collars and tags to studies of animal behaviour has allowed the continual monitoring of individuals for instances of close proximity. A social network can be constructed under the assumption that individuals who spend a lot of time within a pre-defined distance of each other are social partners (e.g. Hirsch et al. 2013). Further, these technologies provide interaction data in continuous time, and have been used in the study of temporal animal contact networks and disease transmission dynamics in general (e.g. Böhm et al. 2008; Böhm et al. 2009;Hamede et al. 2009; Marsh et al. 2011; Drewe et al. 2013; Hirsch et al. 2013b).

In animal network studies, we are often left with data which has some undesirable properties. Early studies had the choice either had very small sample sizes as the result of having to observe animals directly for interactions, or used GoG to construct a network from samples taken at intervals (e.g. Croft et al. 2004; Gero et al. 2008; Godfrey et al. 2009; Best et al. 2013; Aplin et al. 2013; Farine & Milburn 2013). It is understandable therefore, that there has been a reluctance to use many of the approaches developed in social sciences for evenly-sampled, rich datasets where the social network has been constructed through the amalgamation of self-identified social ties (Rendell & Gero 2013; Krause et al. 2014). To demonstrate some methods used to analyse which of a set of variables influence the structure of an animal network, consider a new example network Y, *the* social network of fifty animals created using the GoG. For simplicity, we assume that the animal society has fission-fusion dynamics, and as such each set of observations (separated by fission events) of group co-occurrence can be counted as independent of others (see Croft et al. 2008 for more details). We have reason to believe the presence of ties in Y is determined by kinship X_1, gender X_2, and dominance X_3 and that there are five predominant kinship groups or families in the network. The explanatory variables X_1 and X_2 have been inferred through taking saliva samples upon fixing standard identification tags to the individuals. The saliva samples have provided genetic information used to identify gender and also determine relatedness between individuals and infer kinship grouping. A proxy for dominance has been obtained through observing feeding order on a set of independent occasions X_3.

A common methodology to statistically analyse an animal social network is “Null Model Based Hypothesis Testing” (NMBHT) (outlined in Croft et al. 2011), which could be described as a design-based approach. A design-based approach is purely descriptive and inference is restricted to the finite population being studied (see Sterba 2009). In the context of animal networks, NMBHT begins with choosing a statistic that best represents the network-feature of interest we want to test. If we return to our example, we may want to test whether there are stronger ties between closely related animals (X_1). The test statistic chosen may therefore represent the relationship between edge-strength and relatedness in the network. The statistic may also be chosen to depend on the most reliable aspects of the observed network whilst being robust to suspected sources of uncertainty and bias (Lusseau et al. 2008). For example, if a network was constructed from video footage of a laboratory population, the absence of ties between individuals may be as trustworthy as the presence of ties. In this instance, we may want to take a test statistic that was informed by both the presence and absence of edges. In contrast, if the network was constructed from aerial photographs using GoG, unreported edges may not necessarily mean that social ties are absent. Associations could have occurred in-between sampling periods or outside of the reach of the sampling protocol (such as under canopy or cloud cover). In which case, a test statistic might be chosen which bears little weighting on absent reports. Incidentally, missed associations are a reason why Franks et al. (2010) advise keeping GoG-derived networks weighted where possible.

The statistic chosen in our example is the sum of the edge-weights in Y multiplied by their corresponding kinship values X_1 (where higher values indicate closer kin), which we will call "*wk*". We hypothesise that the value of *wk* is higher than expected by chance, reflecting a tendency for strong ties to occur between closely related individuals. To assess whether *wk* is statistically significant, the observed value of *wk* can be compared to an ensemble of *wk*s generated by a randomisation procedure known as a null model - a general outline to randomisation-based significance testing is provided in Manly (1997). The null model is a procedure to generate networks which could occur as the outcome of a given null hypothesis; typically that the observed statistic is the result of chance. In our example, each null network is then measured against the kinship data X_1 to produce a null version of *wk*. The process is repeated a large number of times (typically between one thousand and ten thousand times) to produce a probability distribution of *wk* under the null hypothesis; this distribution is also known as a reference distribution. The p-value of the observed *wk* can be deduced by measuring the proportion of null *wk* values with a value equal to or greater than the observed value of *wk*. The observed value of *wk* is included in the null distribution, as it must be a viable product of the null hypothesis; this also ensures that the p-value from the permutation test can never be 0 in line with the recommendations of Smyth and Phipson (2011). In our example, if the observed *wk* was in the 99th percentile of the reference distribution, we would determine that the observed value of *wk* was significantly high with a p-value of 0.01.

In NMBHT, there will be choices made by the researcher as to which aspects of the observed network to constrain (e.g. the number of nodes) and which to randomise (e.g. the allocation of edges) in the null model (Croft et al. 2008). The design of the null model employed is very important (see Croft et al. 2011); a null model which is too generic, such as an unconstrained edge randomisation (see section 2.4), may prove to be a "straw-man" null model i.e. a weak representation of the null hypothesis (Croft et al. 2011; Pinter-Wollman et al. 2014). A "straw-man" null model might, for example, allow edges to exist in null networks which could never occur in nature due to physical constraints, or fail to capture some fundamental biological aspects of the system. A reference distribution which contains measures from impossible network combinations means that the p-value loses its interpretation (see Croft et al. 2008; Croft et al. 2011). Controlling for the sampling procedure used to produce the observed network is perhaps one of the most important constraints when testing animal networks. The sampling protocol can drastically effect the properties of animal networks (James et al. 2009; Krause et al. 2011; Farine & Whitehead 2015). Controlling for the sampling protocol is usually achieved through constrained randomisations of the raw-data instead of the observed network (Croft et al. 2011). Bejder et al. (1998) showed permuting a GoG-derived animal social network directly can easily lead to biases and overestimates of statistical significance.

To demonstrate how sampling protocol can be controlled for via the use of a raw-data permutation procedure, we return to our GoG dataset Y. The raw-data used to construct Y might typically look like a "stream" of point-samples over time. Each point-sample would contain a list of individuals partitioned into the social groups they were observed in at that time. The raw dataset for a particular point-sample would thus form a bipartite network, with individuals as one set of nodes and the groups they belonged to as another. There might be temporal autocorrelation in the "data-stream", where two neighbouring point-samples are more likely to contain similar data. Furthermore, each point-sample may contain a different set of individuals, along with a different number of social groups of different sizes. The size and number of groups observed and the frequency with which different individuals were observed in these point-samples are important factors which can influence the structure of Y (Croft et al. 2008). A "data-stream permutation" (Farine & Whitehead 2015) may preserve the number and sizes of the groups as well as the set of individuals observed in each point-sample; randomising the allocation of individuals between these groups (Bejder et al. 1998; Croft et al. 2011). From each permutated data-stream, an actor-by-actor matrix (the social network) would be constructed using the same procedure as the observed dataset. Employing the GoG assumption and using an index which takes into account how many times a pair of animals were observed in the same group relative to how much they were observed in any group (see Cairns & Schwager 1987).

A data-stream permutation null model ensures that the influence of the sampling protocol on the observed network is preserved in every null network and corresponding null value of *wk* (Bejder et al. 1998; Croft et al. 2011). We can expand this null model further to test more complex hypotheses. If we suspected that the observed *wk* is actually just the product of the tendency for ties to form between individuals in the same kinship group with no extra preference beyond, we may add a constraint to the null model which requires the kinship group structure to be represented in every null network, i.e. that individuals can only be swapped between groups into the positions occupied by another member of the same kinship group. This constraint may shift the reference distribution of *wk* towards the observed value, increasing the chance that the observed *wk* could occur under this new null hypothesis. Using NMBHT to control for one factor on top of more fundamental constraints (like the sampling protocol) is possible. However, NMBHT becomes increasingly computationally complex to perform as the number of factors we wish to control for gets large, resulting in its limited multivariate capacity (similar to conditionally uniform graphs- see section 2.4).

Thus far, we have considered an example where the network has been constructed from association data. Animal networks may also be constructed from interactions. Interactions are a strong form of evidence for a relationship given that they require a choice to be made by the animal (Croft et al. 2008). In the case where Y is an undirected network created through observations of affiliative interactions between individuals, the raw-data and observed network may be the same i.e. an edge with a weight of 5 would represent 5 observed social interactions. In which case, a simple and commonly used null model in human and animal network studies is a node-label permutation (a QAP). If a QAP is used to examine the significance of *wk*; this type of test is a single variable Mantel test (1967). The test statistic *wk* is a z-score and the two matrices Y and X_1 are distance matrices - each is symmetric about the lead diagonal. Mantel tests can also be stratified so that nodes may only be permuted between individuals in the same block (e.g. family group); this could be used to investigate whether the observed relationship between strength of tie and kinship does not extend beyond kinship group. A Mantel test can also use a correlation coefficient as a test statistic, when this statistic is a Pearson's product correlation coefficient the Mantel test becomes almost identical to a MRQAP with one explanatory variable. However, a MRQAP multiplies the Pearson's correlation coefficient according the unit values of the dependent variable to make it a regression coefficient β (so for every unit change of X, you see a β change in Y).

One method of testing for a correlation between two variables while accounting for a third variable is to use a partial Mantel test (see Legendre & Legendre 2012). Partial Mantel tests provide a means to examine the relationship dominance X_3 and Y whilst partialling out the effect of kinship X_1 on Y. Partial Mantel tests also permute the response variable and can also have a block structure- which could be used such that nodes labels were only reassigned to individuals of the same gender, to provide a new example. A partial Mantel test is a step towards performing a multivariate analysis on animal data as one factor can be evaluated in light of another through partialling.

3.5 Multivariate analyses in Behavioural Biology

There are many problems associated with the analysis of animal networks, many of which stem from the methods used to collect animal network data. Over the last decade, the availability and quality of animal network data has gone through somewhat of a revolution through the miniaturisation of radio-tracking and global position system (GPS) devices to the point where they can be mounted directly upon animals. At the same time, advances in video analysis software has allowed us to track and monitor multiple individuals simultaneously (see Krause et al. 2013; Krause et al. 2014). The data generated through these approaches are still perhaps a way off the quality of data which can be obtained in the social sciences (Prange et al. 2006; Watson-Haigh et al. 2012; Drewe et al. 2012; Boyland et al. 2013; Rutz et al. 2015, see chapter 6). Despite this, some believe that we are at a point where models from the social sciences can be used on animal networks, and many do not. Pinter-Wollman et al. (2014) rallied for the greater adoption of modelling approaches from the social sciences to be used in animal networks; this call was criticised as being overly-enthusiastic by Krause et al. (2014) and Rendell and Gero (2014). The main argument rightly being that animal networks are not sampled evenly, are rife with uncertainty (Lusseau et al. 2008) and can have other forms of dependence (see Whitehead 1995 for a good example) not accounted for in models developed for human social networks.

Data censoring is a central problem in animal network analysis. In the social sciences, datasets can be complete in the sense that the relevant individuals can be accurately determined, the presence or absence of all possible associations sampled-evenly and network boundaries identified (see Hanneman and Riddle 2005). Animal networks are often incomplete in the sense that individuals, or relationships between individuals may be missed, especially in non-laboratory populations; certain network metrics and structural features can be very sensitive to the inclusion of missed data (James et al. 2009; Farine & Whitehead 2015). In contrast to human networks, animal network data can also contain biases relating to the sampling regime used and the attributes of nodes, such as variation in identifiability (one sex of a species may be more conspicuous) or personality of the individuals studied (whether they are neophyllic or neophobic for instance) (Croft et al. 2008). Therefore, applying these models as they stand without addressing sampling issues common in animal data is ill-advised, yet the power of these approaches is a dangling carrot which has become the motivation for this thesis (and perhaps the recent review by Farine and Whitehead (2015)). The remainder of this chapter considers some rudimentary steps which can be taken to better marry these models to animal network data.

3.5.1 Improving animal network data

Whitehead and James (2015) took a MRQAP type approach to generate network indices from remotely sensed data. The approach was to model the duration of association between animals in terms of a number of confounding factors, such as having similar patterns of use of habitat in time and space, gregariousness and differential association rates among age/sex classes. The confounding factors would form a model to predict the association time, the residuals between the actual association times and those predicted by the confounding variables was taken as a more direct measure of social preference.

This example shows that the MRQAP can be employed to remove many biases in network data. For many types of network enquiry the researcher can be more certain he or she is examining a social network instead of one depicting concurrent non-social activity- such as feeding at the same location. However, using the residuals from one MRQAP type approach in another MRQAP instead of including the confounding and “useful” explanatory factors in one model will mask potential relationships between confounding and “useful” variables, which may lead to violations of the ancillarity principle (see section 3.4).

Remote-sensing technologies present the most potential for the collection of rich animal association data used which can be used to construct a social network. Active radio-frequency identification (RFID) technology allows instance where individuals are within a predefined distance to be captured, or can be used to provide continuous distance approximations. In theory, if a social group can be identified and fitted with loggers, then we have complete coverage for associative relationships. What we may not have is even-sampling (see chapter 6 and Boyland et al. 2013). However, given that the readings on one collar or tag should be reciprocated by the other, we show that it is possible to characterise and correct for the biases of these individual tags and collars. Through *post-hoc* corrections of logger data, we might be able to drastically improve the quality of data gathered through these approaches and remove most of the sampling bias. Given that most models employed in the social sciences implicitly assume even-sampling, removing the bias from animal network data would improve their suitability for these approaches.

3.5.2 Modifying explanatory variables

In some animal network studies, the explanatory variables may be trustworthy, but the response variable, i.e. the animal network, questionable. If a single or explanatory variable is untrustworthy, then the corresponding β estimated by the model for that variable may also be untrustworthy, but the model as a whole will be increasingly robust to the variable with more explanatory variables included, as long as they do not share variance with the questionable one. However, if the response variable is untrustworthy, then every beta assigned to every explanatory variable will also be untrustworthy. These systems may therefore be the most unsuitable for conventional modelling approaches which regress multiple explanatory variables on a given response variable, but with the right approach, may still provide useful biological insight. In light of a questionable response variable, the flexibility afforded by NMBHT with respect to the choice of test statistic and null model may be the favoured option to test explanatory variables against an untrustworthy response variable (see Croft et al. 2011).

.

However, one limitation of NMBHT is that only one variable can be analysed at a time. If the researcher wants to know how the network is linked to one factor, X_1, after controlling for two or more other explanatory variables, then it gets very complex to develop a null model which can control for this many factors when testing X_1 (see Snijders 2011). If the X-variables are trustworthy, then a MRQAP can be used to model the X_1 in light of the other explanatory variables. A model $X_1 \sim f(\beta_2 X_2 + \beta_3 X_3) + \varepsilon$ can be constructed. The residuals between the observed X_1 and those predicted by the model can become the new variable, rX_1. The values in rX_1 can then be used to make a given test statistic S_1 to represent the unique influence of this factor on the network. A similar type of approach has been employed Whitehead & James (2015), who took the residuals from an association matrix as a form of association index from a model where confounding factors were included as explanatory variables.

The use of MRQAPs to modify explanatory variables would provide the researcher with a multivariate analysis where the contribution of each variable to a response variable has been evaluated whilst controlling for all of the other variables tested. This allows more flexibility for the researcher to tailor the null model used to test the significance of a given statistic, such as S_1, to control for data sampling issues. The final output will multivariate, but still descriptive for each variable given the NMBHT approach. However, through scaling all of the explanatory variables to range between the same values, comparative effect sizes could be obtained.

3.5.3 CRAN R

The primary tool used to conduct all of the analyses in this thesis is the programming language R. R is an open-source, object-orientated programming language geared towards statistical computing and graphics which first appeared in 1993 (R Core Team 2013). R is based on a command line interface; with several graphical front-ends available. R is most commonly used by statisticians and data-scientists, but has seen a broad and substantial increase in popularity in recent years compared to other statistical programming languages (Muenchen 2012). The popularity of R has extended to the biosciences, where several books have been published on the topic of using R for the analysis of biological data (e.g. Gentleman et al. 2006; Gentleman 2008; Logan 2011).

R is such a powerful statistical programming language thanks, in part, to the large community of users who support it. There are over six thousand specialised statistical packages currently available, each package containing a number of functions geared towards a certain goals and published with supporting documentation. Further, several free graphical user interfaces (GUIs) are available, including integrated text-editors such as R-studio.

There are two R packages which have predominantly been used to visualise and analyse networks in this thesis, they are:

“igraph” ; a package (Csardi & Nepusz 2013) which contains all the functions needed to plot networks and measure their features- such as assortativity (used in chapter 5). “igraph” stores network data as a dyadic “edgelist”, which takes the format: “node A-> node B”, “node C -> node E”; this is especially efficient for very large networks, and igraph is especially suited to visualising networks with millions of nodes- far in excess of the size required for animal networks. In this thesis, all of the network visualisation has been conducted using this package.

“sna”; a package (Butts 2014) which contains a range of tools for network analysis, including node and graph-level indices, structural distance and covariance methods, structural equivalence detection, network regression, random graph generation, and 2D/3D network visualization. Importantly, this package contains the MRQAP-based functions “netlm” and “netlogit” which provide the basis for many of the analyses in this thesis.

Pre-made functions offer significant shortcuts to the user- especially when the type of analysis performed is complex. A simple example of a function in R is “sd”, which is used to find the standard deviation of a set of numbers. In R, whatever precedes a set of round brackets is typically the function and whatever is first listed within the round brackets is the data that the function will be performed on. Simply, “sd(c(1,3,3,1))” could be entered into the R console where “c” dictates that the numbers are being combined into a vector and R would return 1.15 – the standard deviation of the sample. Many functions allow the specification of additional arguments which may tailor how they work after the inclusion of the data. An advantage of R is that if the source code behind the function “sd” was required, the function could be entered into the console followed by no brackets and R would return something like “sqrt(var(x))”; this is an example where the function has been made using “lower-level” R-functions. *Sqrt* is the function for square root and *var* is the function for variance. Many, if not most packaged R functions are made from R source code, although some call upon other languages such as C. R therefore offers some of the accessibility of other non-command line based statistical programmes through GUIs and huge amounts of packaged functions, but with much greater flexibility - a basic level of tailoring is offered through the arguments supplied with functions, but often the source code behind a function is visible and can be modified (see chapter 5 for an example with the “netlogit” function). R packages are thus easily extendable, allowing for the timely incorporation of new methods and/or refinements; often based on code that can be inspected (and hence modified) by the community and portable, allowing use by researchers on a variety of computing platforms; and freely available (Butts 2008b).

3.5.4 Adapting the MRQAP

The MRQAP procedures in the “sna” package can currently be specified with any one of eight different null model arguments, including: QAPs on the explanatory variables, a QAP on the response variable, the QAP-DSP, or the choice of some basic CUGs (Butts 2014) (explained in section 2.4). These nulls may all be too generic for many animal systems, having been developed for use in the social sciences. However, through accessing the source codes of the “netlogit” and “netlm” functions, we can include null models which better suit use the animal system being studied. Through this we marry together the call to produce more realistic null models from many behavioural biologists (Krause et al. 2011; Pinter-Wollman et al. 2014; Farine & Whitehead 2015) and another call to use more predictive modelling approaches in animal network analysis. Given the findings of Dekker et al. (2007) and Krackhardt (1988) we would still be careful as to using an animal based CUG with regards to issues such as row-column autocorrelation, skewedness and collinearity. If the response variable is to be permutated, it would also be under the assumption of pivotal test statistics.

Each of the three methods covered in this section (3.5) are now used in the following three data-containing chapters.

II. Multivariate analysis of a report-based network

4. Avian hybridisation networks

4.1 Abstract

There has been growing interest in understanding the behavioural and ecological factors which drive interspecific hybridisation (Randler 2006, Willis 2013). Many previous comparative studies have been limited through having to maintain the assumption that hybrid pairs are independent from one another. In taxa where hybridisation is frequent, we show that cases are often interlinked. To take advantage of rich hybridisation datasets, we need to account for dependency in the data this creates. Here we show that methodologies often used to study social networks can identify historical, ecological, and behavioural factors which underlie the structure of hybrid complexes. Using published reports of hybridisation, we find that the wood warblers, a group of recently radiated passerines, form a single hybrid network encompassing all hybridising species. Through using a network-based analytical approach which addresses both the quality of hybrid report data and the shared phylogenetic history of species, our results suggest that close relation, breeding proximity, song and plumage similarity are associated with hybridization in this system. Although this study is currently limited by the quality of hybrid data available, with the increasing use of genomics to identify introgression between species and identify hybrid offspring, we see promising avenues for the networks-based approach.

4.2 Introduction

Many definitions have been offered for hybridisation (summarised in Harrison 1993). Broadly speaking, hybridisation is where two genetically divergent individuals reproduce. The amount of genetic divergence required to separate hybridisation from typical mating can range from a single genetically heritable character to the amount of genetic variation which separates species, depending on the definition (See Arnold 1997). For the purpose of this chapter, we are interested in hybridisation between incipient species; the term ‘incipient’ being added for reasons later described. Further, we are interested in what Arnold (1997) describes as “natural hybridisation”: that which occurs in a natural setting as opposed to experimentally or in captivity. We only consider natural hybridisation which results in the production of offspring in order to use the presence of hybrid offspring to identify meaningful hybridisation which has taken place in nature. Hybridisation which produces offspring may also be the most important in an evolutionarily sense, as hybrid offspring nearly always have some level of fertility either with each other, or one of the parent species (Grant 1963).

Interspecific hybridisation is not a rare event confined to few species with questionable species boundaries. Mallet (2005) reported that 25% of plant species and 10% of animal species hybridise. Numerous crosses have been reported between species of different genera (see McCarthy 2006) and, rarely, different families in both the wild (Liptack & Druehl 2000) and in captivity (Billingham et al. 1961; Ohno et al. 1964; Kisaka et al. 1997). Of all major animal taxa, hybridisation is perhaps best recorded in birds (Randler 2002; Randler 2004; Aliabadian & Nijman 2007), where between 9.3% (Grant & Grant 1992) and 19% (Randler 2006; McCarthy 2006) have been reported to hybridise. Mallet (2005) noted that most hybridisation occurs between evolutionarily young species; those which have recently diverged from a common ancestor. In this study, we focus on collections of avian species, some of which are evolutionarily young through recent radiation events (see Jetz et al. 2012).

The tendency for evolutionarily young species to hybridise challenges our notions of what constitutes a species. The biological species concept (Dobzhansky 1937; Mayr 1942) defines species as “groups of actually or potentially inbreeding natural populations, which are reproductively isolated from other such groups”. Arnold (1997) noted that a strict usage of this definition means that species do not hybridise. Mayr (1942) argued that if species hybridised than they were in fact sub or semi-species, adding that any offspring produced by two separate species are either sterile or exhibit low fitness and would quickly be removed by natural selection (Mayr 1963). Other prominent species definitions such as the phylogenetic (Cracraft 1989), cohesion (Templeton 1989) and recognition species (Paterson 1985) concepts take a similar standpoint: that hybridisation is either non-existent between species, or produces offspring which are more or less evolutionarily inconsequential. To err on the side of caution, we use the species definitions supplied by most recent molecular phylogenies published for a given taxa and add the term “incipient” to describe those species reported to hybridise; “incipient” referring to the fact that these species are not completely reproductively isolated from others.

The role of hybridisation in producing novel evolutionary lineages in animal taxa may be questionable (see Arnold 1997 for a review), but its role in shaping biodiversity and the evolutionary trajectory of species is well-supported. Firstly, hybridisation can generate genetic variance in a population through the introgression of new genetic material- a regular occurrence in rapidly radiating groups of species (Seehausen 2004; Mallet 2005; Grant et al. 2005). Natural selection acts on genetic variance in populations and introgressed genetic material may traverse species boundaries if it confers a selective advantage (Baack & Rieseberg 2007). Viable hybrid offspring may have some level of fertility (Grant 1963), but often the heterozygous sex are rare or sterile, known as Haldane’s rule (Haldane 1922). The absence or sterility of one sex does however present a formidable obstacle to the formation of a hybrid lineage. Secondly, hybridisation can prompt the evolution of diversity by selecting for character displacement and other isolation mechanisms (Brown & Wilson 1956; Schluter 2001). The “Wallace effect” describes the process of selection for reproductive isolation mechanisms (Grant 1966), otherwise known as reinforcement. In cases where two spatially proximate populations are genetically divergent to the extent that any offspring produced through hybridisation exhibit lower fitness, individuals who do not hybridise have a comparatively higher fitness than those that do. The mechanisms which promote isolation are therefore under selection and, over generations, may spread through the population prompting diversifying speciation (Butlin 1987; Howard 1993). In warblers, Martin et al. (2015) suggests that reinforcement may drive the plumage coloration in migratory warblers.

Hybridisation can also serve to reduce diversity, posing an extinction threat and conservation challenge. Climate change, habitat loss or modification, and the direct introduction of species to new areas can bring recently diverged, previously isolated species into contact with each other (Rhymer & Simberloff 1996). In these situations, hybridisation can lead to genetic mixing, where genetic introgression across a species barrier can cause species to converge genetically- reducing diversity. An example of this is seen in the New Zealand Grey Duck *Anas superciliosa* and the Mallard duck *Anas platyrhynchos*. The introduction of the Mallard duck has been one of the reasons for the decline of the native Grey Duck. Mitochondrial DNA analysis suggests that the loss of the Grey Duck to the Mallard duck is a real possibility (Rhymer et al. 1994), with existing Grey Duck populations becoming increasingly similar to mallard ducks morphologically (Wiegand 1935). Genetic mixing can also lead to outbreeding depression, where co-adapted gene complexes are broken apart and local-adaptions of one or both species are lost (Rhymer & Simberloff 1996). An example of this has been observed in two species of clam, *Mercenaria mercenaria* and *M. campechiens*, which have been introduced into each other's native range and concurrently into new uncolonised areas. Hybridisation between the two species is common, with hybrid offspring more susceptible to the disease gonadal neoplasia, resulting in reduced hybrid fitness through increased mortality and reduced reproduction (Bert et al. 1993).

Despite its importance, most of our knowledge about hybridisation come from studies of focussing on only two species (examples include: Szymura & Barton 1986; Bert et al. 1993; Gante et al. 2004; Tracey et al. 2008; Hayden et al. 2010; Tyler et al. 2013; Huyse et al. 2013). The implicit assumption in these studies is that the focal pair of species is independent from hybridisation with other species. In some situations, this assumption is justified, to return to a previous example, as far as we know there are no reports of hybridisation of either species of *Mercenaria* clam with any other species. However, if we consider a different example, hybridisation between the Hermit warbler *Setophaga occidentalis* and Townsend's warbler *Setophaga townsendi* the same assumption is not justified. Studies (such as Harrison 1990; Rohwer & Wood 1998; Pearson & Rohwer 2000; Rohwer et al. 2001) into genetic variation, asymmetries in behaviour, gene flow and aggression across the Hermit-Townsend's hybrid-zone link their findings to hybridisation between the two species. However, both the Hermit and Townsend's warbler hybridise with another sympatric west-coast migratory warbler species, the Black-throated gray warbler (*Setophaga nigrescens*) (see McCarthy 2006). There is a chance that any effects of hybridisation with this third species are confounding the studies across the hybrid zone. Mutual hybridisation with a third species offers another route for gene flow and may also affect the evolutionary dynamics of processes such as reinforcement, potentially amplifying variations in character displacement and other traits associated with reproductive isolation.

Only a few studies have analysed multiple hybrid pairs simultaneously. An example can be seen in Randler (2006), where a general linear model was used to identify that pairs of Western Palearctic bird species produced more hybrids when they were parapatric, have low levels of paternal investment and at least one species was classified as endangered. To maintain an assumption of the modelling approach used, that each case of hybridisation was independent of any other, multiple phylogenetically independent hybrid types were selected and those crosses linked through mutual hybridisation were discarded. Willis et al. (2014) was perhaps the first published multispecies study to include interconnected hybrid pairs using a methodology which can control for the dependency in the data this can create. The advantage of the approach that Willis *et al.* used was that recently radiated collections of species could be studied in detail at the familial level, whereas Randler’s approach would require the removal of much if not all of the data and is thus restricted to broad-scoped taking few samples from numerous families of organisms.

Networks are often used to model systems where multiple agents are interconnected via multiple ties. A common theme of network studies is a focus on how edges, or interactions, are structured with respect to each other and attributes associated with the nodes, or agents, in the system. Hybrid networks have previously been created to represent intra-species hybridisation in plants (Lenz 1959 depicted in Arnold 1997), and depict avian genera which cross in certain taxa (McCarthy 2006). However, networks have not yet been used to visualise hybridisation between what we might regard as species and network analyses has not yet been applied to hybrid-networks. Network analyses provide a means to consider cases of hybridisation within the context of others; allowing multiple interlinked cases to be compared simultaneously. Further, the visualisation of interspecific hybrid-networks could potentially alter how we think about the speciation process- clades with rampant hybridisation potentially representing a snapshot of a radiation. In a network depicting hybridisation between incipient species, the species would form the nodes of the network and identified cases of hybridisation between them would form the edges.

Avian species which have undergone recent radiations are a good place to begin looking for candidate taxa with underlying hybrid-networks. Firstly, recall that hybridisation is perhaps the best recorded in birds out of all major animal taxa (Randler 2002; Randler 2004; Aliabadian & Nijman 2007), with between 9.3 and 19% of species reported to hybridise (Grant & Grant, 1992; Randler, 2006 and McCarthy, 2006). A recent compilation of all reported bird hybrids has been published by McCarthy (2006), providing a comprehensive data-source with which to generate hybrid-networks. Secondly, recall that hybridisation is the most frequent between evolutionarily young taxa. A recent publication of an avian super-tree by Jetz et al. (2012) provides a visual indication of how recently and rapidly speciation has occurred in many avian clades. Using these two criteria as a guide, we chose two orders of species, the *Anseriformes* and *Galliformes*, to illustrate this approach; the *Galliformes* being a more ancient clade according to Jetz *et al.* 2012. Further, two avian families, the migratory New-World warblers *Parulidae* and the Birds of Paradise *Paradisaeidae* were chosen as a representation of hybridisation at the familial level.

Using hybrid reports to construct hybrid networks has some inherent problems, but is currently the best available form of evidence for comparative studies across numerous species. The four avian taxa chosen in this study are likely to be well observed for a number of different reasons; this is important as we can only be sure of the presence of reports as indicative of whether hybridisation has taken place between two species. The *Anseriformes* may be well observed given that they typically breed near and around water sources, which is also where a large proportions of human settlements are found (World Health Organisation 2005). Many of the *Galliformes* are commonly hunted for sport and food (Keane et al. 2005) and both the *Parulidae* and *Paradisaeidae* have diverse and diverse and colourful male plumages (see Beehler et al. 1986; Curson et al. 1994) . Given that males are the homogametic sex in avian species, conspicuous male plumage patterns may aid the identification of hybrids. Haldane’s rule dictates that hybrid offspring in birds are more likely to be male, the homogametic sex in birds (see Orr 1993 for an explanation of the causes of Haldane’s rule). The migratory species in the family *Parulidae* were chosen as they are more evolutionarily recent and breed in more densely populated areas in North America than their non-migratory counterparts in the same family which remain in Central and South America year-round.

The hybrid networks of each of the four taxa are shown in figure 4. To create the four hybrid networks, reports of natural hybridisation were taken from McCarthy (2006). The wood warbler hybrid network also includes later evidence from Griffiths et al. (2008) and Bonter and Lovette (2007) whom provided evidence for a new and a previously questionable cross. Species, the nodes of the networks, were defined using recent molecular phylogenies published on each of the four taxa (Lovette et al. 2010; Kimball et al. 2011; Gonzalez et al. 2009; Irestedt et al. 2009). Edges were weighted qualitatively based on the extent of hybridisation between species using the classifications defined in McCarthy (2006); this ranged from reports of infrequent (2), ongoing (3) to extensive-ongoing (4) hybridisation. The very weakest edges in the networks represented crosses with which there was some questionable evidence, typically just a sighting (1); crosses which could not be narrowed down to one pair of species were omitted. The networks were visualised using the "igraph" package (Csardi & Nepusz 2006) in R (R Core Team 2013). Species that had not been reported to hybridise but were present in each of the four groups, were omitted from the networks.

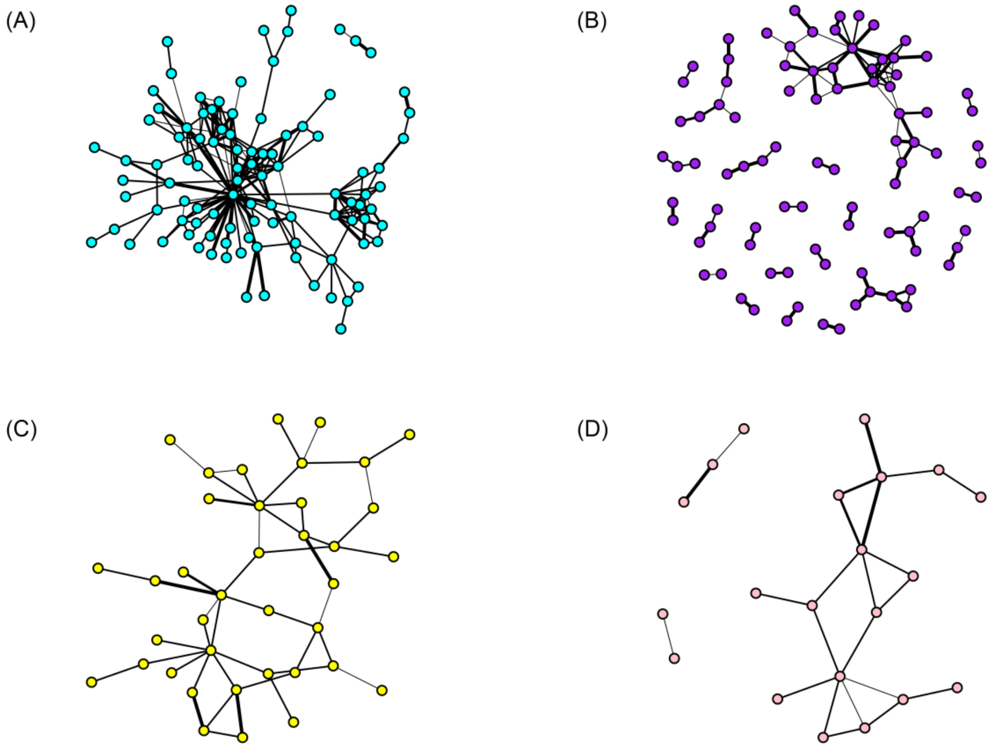


Figure 4: Examples of natural avian hybrid networks: (A) *Anseriformes*, (B) *Galliformes*, (C) *Parulidae* breeding in North America, and (D) *Paradisaeidae*. Each edge represents a hybrid specimen recorded in nature weighted by the quality of evidence for the cross as defined by McCarthy (2006). Each node represents a species: as defined by (A) Gonzalez *et al.* (2009), (B) Kimball *et al.* (2011), (C) Lovette *et al*. (2010), (D) Irestedt *et al.* (2009). For clarity, species not reported to hybridise were omitted: (A) 62, (b) 124, (C) 17 and (D) 18. The number of species in the largest connected component in each hybrid network totals: (A) 96, (B) 30, (C) 38 and (D) 16.

In figure 4, a large interconnected component is observed in each hybrid network. A component being a collection of nodes all linked to each other. This may be surprising when you consider that in these networks, only 0.4% to 3.1% of possible edges exist. We might therefore expect that there is some biological explanation behind the formation of these network components. However, the large components in each of these networks are no bigger than expected by chance, which suggests that the presence of hybrid networks may just be the product of the frequency of hybridisation. We can show this by randomly allocating hybridisation events between the same number of species for each of the four taxa and record how often random networks contain a component larger than observed to derive a p-value (see Manly 1997). The results show that there is nothing especially large about the large components in these hybrid networks (p=0.9998, p=0.9096, p=0.4438, p=0.3068 respectively). For very large collections of nodes, network theory states that a large component will form when, on average, each node has at least one edge attached to it (see Newman, 2003). In the context of hybridisation; this would translate to a large component emerging when there are many cases of hybridisation as there are defined species in taxon. In each of the hybrid networks shown in figure 4, we find that the number of hybrid reports per species in the group is: *Anseriformes* 1.317; *Galliformes* 0.415; *Parulidae* 0.855, and *Paradisaeidae* 0.590. In three of four cases, there are fewer reports per species than expected to form a large component using theory designed for large networks; this suggests that the threshold of one report per species may be overly conservative for small collections of species.

Given that the presence of hybrid networks (with large components) are an expected product of frequent hybridisation we are left asking what, if anything, is biologically interesting about these structures? Closer visual examination of the networks reveals non-trivial structural features that suggest hybridisation does not take place randomly between species. The *Parulidae* network is partitioned into two halves, one of which contains predominantly species in the genus *Setophaga* and the other *Geothlypis*. The network of *Anseriformes* is partitioned into at least three communities (see Girvan & Newman 2002) representing species belonging to the whistling ducks Dendrocygninae, dabbling ducks *Anatinae* and the swans and geese *Anserinae*. Guimera, Sales-Pardo, & Amaral (2004) showed that random networks can exhibit community structure. However, the community structure in at least two of our networks appear to be on preferential hybridisation between species of the same genus or subfamily, in line with the notion that hybridisation happens between evolutionarily recent species (see Mallet 2005). Community structure of this sort provides an indication that network structure is being influenced by homophily (see section 2.3.2) on the basis of certain traits.

Another starting point to investigate hybrid networks may be to look at the number of connections each species has. Degree is perhaps the most suitable node-based measure to consider in hybrid networks. A missed hybrid report may have far reaching consequences on measures based on the shortest paths through the network, such as betweenness, but would only alter the degree of the two species involved by one. The degree of species in the hybrid network indicates how prolifically they hybridise with other species. If degree cannot be explained by chance, we might expect that certain factors promote some species to be prolific hybridisers. Using the same edge randomisation procedure previously stated, we find that some species have higher degree (they hybridise with more species), than we would expect. In the *Anseriformes*, we would only expect to see a species hybridise with four or more others once in every five thousand random networks (equivalent of a p-value of 0.0002), but the mallard duck *Anas platyrhynchos* is reported to hybridise with over thirty in nature. In *Parulidae*, the observation of three species with six or more connections is only expected in two of a thousand random networks. Species with unexpectedly high degrees under random hybridisation also exist in both the *Galliformes* and *Paradisaeidae.* To date, no study has considered the factors associated with prolific hybridisers such as these. In the *Anseriformes* hybrid network, there is visible centralisation (see figure 4 *A*), where many nodes are connected to one central node (Freeman 1979)- the mallard duck in this case. Clearly, the null expectation we might have, that hybridisation occurs randomly in these networks, is unlikely to be true. It seems plausible that specific traits of species may affect how much they hybridise and traits shared between species may therefore affect the species that any given species hybridises with.

Assortment is the term that refers to the preferential attachment of nodes on the basis of a certain trait (Newman 2003). The non-random structural features of the four hybrid networks support the notion that there are factors which affect, or are otherwise linked to hybridisation between species in each network. Finding these factors is important as they will provide a family or order-level indication of the patterns of hybridisation early in the speciation process, and indicate how these may promote, or limit, the diversification of the taxon. Furthermore, these factors may be able to help inform conservation action through directing observation efforts to those species with compatible traits. In *Parulidae* for example, the observed northward shift in the distribution of North American birds (La Sorte and Thomson 2007) may push many new pairs of species into contact. Given that these species are evolutionarily young (Bermingham et al. 1992; Jetz et al. 2012; Willis et al. 2014) and that the population of warbler species are also likely to decline in the future (Strode, 2003), hybridisation may pose a real conservation threat in the near future.

The migratory wood warblers (family: *Parulidae*) are a good candidate system to look for assortment in a hybrid network. Sixty-nine percent of species have been reported to hybridise (McCarthy, 2006) and both species and hybrids in this group are often well observed and characterised (e.g. Curson et al. 1994). A study by Willis *et al.* (2014) (see section 4.1) suggested that hybridisation in wood warblers tends to take place between species that are evolutionarily young, have large breeding range overlap and sing similar mating songs. Migratory wood warblers, which we define on the basis of species which breed on the North American continent, are known for undergoing large scale northward migrations from South America to breed; adopting diverse and often brightly coloured male breeding plumages (see Curson et al. 1994). This migratory behaviour is thought to have developed in periods of intermittent glaciation in the Pleistocene era where rapid speciation may have occurred in the clade (Bermingham et al. 1992; Price et al. 1998). The annual spring-summer migrations predominantly take place up the more densely populated coastal regions of North America. Thus, the assumption can be made that the migratory wood warblers are well observed in their breeding season, which is attractive when the identification of hybrid offspring is required to confirm a cross. There are also comprehensive libraries of data on wood warblers with which to examine the patterns of hybridisation in the network, including distribution maps (Birdlife International 2013), song records (Borror and Gunn 1985), plumage illustrations (Curson, Quinn and Beadle 1994) and importantly, a recent molecular phylogeny (Lovette et al. 2010).

4.2.1 Hypotheses

Reproductive isolation, the barrier to interspecific hybridisation, is achieved through numerous mechanisms, some of which happen before fertilisation, known as pre-zygotic isolation, and some after, known as post-zygotic isolation (Mayr 1963). Pre-zygotic isolation mechanisms can take a number of forms. Ecological isolation mechanisms restrict the co-occurrence of two separate species either spatially or temporally. In migratory warblers (family: *Parulidae*) for instance, there is variation both in the use of space for breeding and variation in the times of breeding migrations. Behavioural isolation is another pre-zygotic isolation mechanism which maintains the integrity of species. In species with distinct male and female organisms, certain behaviours associated with finding and choosing a mate also serve as isolation mechanisms. Birdsong is a mating ritual linked to reproductive isolation (King et al. 1980; Brambilla et al. 2008) and, although not a behaviour directly, male-plumage may also be linked to species recognition and isolation (Sæther et al. 2007; Martin et al. 2015).

Breeding range and habitat

To hybridise, two species must first come into contact, achieved either through translocation or sympatry. Species with large breeding ranges may have a greater opportunity to hybridise with others through being in sympatry with more compatible species, or smaller translocations being required to encounter them. Species with larger breeding ranges may therefore have a larger degree than those with smaller breeding ranges through increased contact with each other. Those species which use the same breeding habitats may not only be more likely to come in contact with each, but may also be better adapted to build nests and raise offspring together in these habitats. Species that are generalists with respect to breeding habitat may be more prolific hybridisers than those that are specialists through having breeding habitat in common with a greater number of species.

Pairs of species which are geographically proximate will have more opportunity to hybridise, but may also have developed stronger barriers to introgression than those more isolated. In a recently radiated taxon such as the new world warblers we might expect that barriers may not be fully formed, as such there would be a tendency for species with large breeding range overlaps to hybridise and there to be resulting positive assortment with respect to this trait in the warbler hybrid network. In a similar sense, species which use the same habitat types to breed may also be more inclined to hybridise, if so, we would expect to see positive assortment in the hybrid network on the basis of this trait.

Breeding song and plumage

Morphological and behavioural factors associated with mate choice, can vary significantly between families and orders of species. In the family of the New-world warblers *Parulidae,* song and breeding plumage are important for intra-species mate choice (Weatherhead and Shutler 1990; Byers, 2007; Taff et al. 2012), whereas the same may not hold true in other taxa. In the order *Galliformes*, plumage and visual displays may be important with intra-mate choices (Madge et al. 2002; Kolm et al. 2007; Lislevand et al. 2009; Kimball et al. 2011). Mating rituals have been identified in a number of cases to be potent isolation mechanisms (King et al. 1980; Sæther et al. 2007; Brambilla et al. 2008; Martin et al. 2015) . We might expect warbler species with similar breeding plumage and song to hybridise, with these sexually selected traits only functioning as sufficient isolation mechanisms when significantly different to each other. If so, there will be positive assortment in the hybrid network for both song and male plumage traits. Song and plumage characteristics have been shown to be important factors for mate recognition in many species of birds (Searcy 1992; Bennett et al. 1997; Ballentine & Hill 2003; Hill & Mcgraw 2004; Byers & Kroodsma 2009; Toomey & McGraw 2012)

Sexual selection has been linked to hybridisation in other species, such as the Yellow Swordtails *Xiphophorus clemenciae* (Schumer et al. 2013). If sexual selection does promote hybridisation, we would expect to see more sexually selected species to have higher degrees.

Phylogeny

When pre-zygotic isolation mechanisms are underdeveloped or otherwise break down. The newly formed zygote is now subject to a number of post-zygotic isolation mechanisms which reduce the chance of a hybrid lineage being produced. Firstly, the accumulation of genetic incompatibilities between the parent species may lead to the zygote becoming inviable; this is part of the Bateson-Dobzhansky-Muller Model of speciation (Orr 1996). Species with recent common ancestors may be more likely to produce hybrid offspring, not having had enough time to accrue genetic incompatibilities (Orr 1996). We may therefore expect the hybrid network to assort such that hybridising species are often recently diverged; this would support the findings of Mallet (2005) and Willis et al. (2014). Further, we might expect that species that sit in large clades of the phylogeny will have higher degrees, having many recently diverged species with which to hybridise.

4.3 Methods

Data Collection

To look for network assortment and explain the variation in the degree of hybridising warbler species, data were collected for five traits thought to be associated with hybridisation. These traits broadly fall into the categories of phylogenetic, morphological, behavioural and biogeographic and ecological factors. In particular, data were collected on song similarity, distance in a phylogeny, plumage similarity, the extent of sympatry, the amount of shared breeding habitat, and the distance between the breeding ranges of pairs of species. A summary of this data can be observed in figure 5. Species attribute data were taken to explain the number of connections each species has; these data were: the size of the species breeding range, the size of the clade of the phylogeny the species belongs to, the difference in the male and female plumages of the species as a proxy for the amount of sexual selection, and the number of habitats used to breed as a measure of ecological specialism.

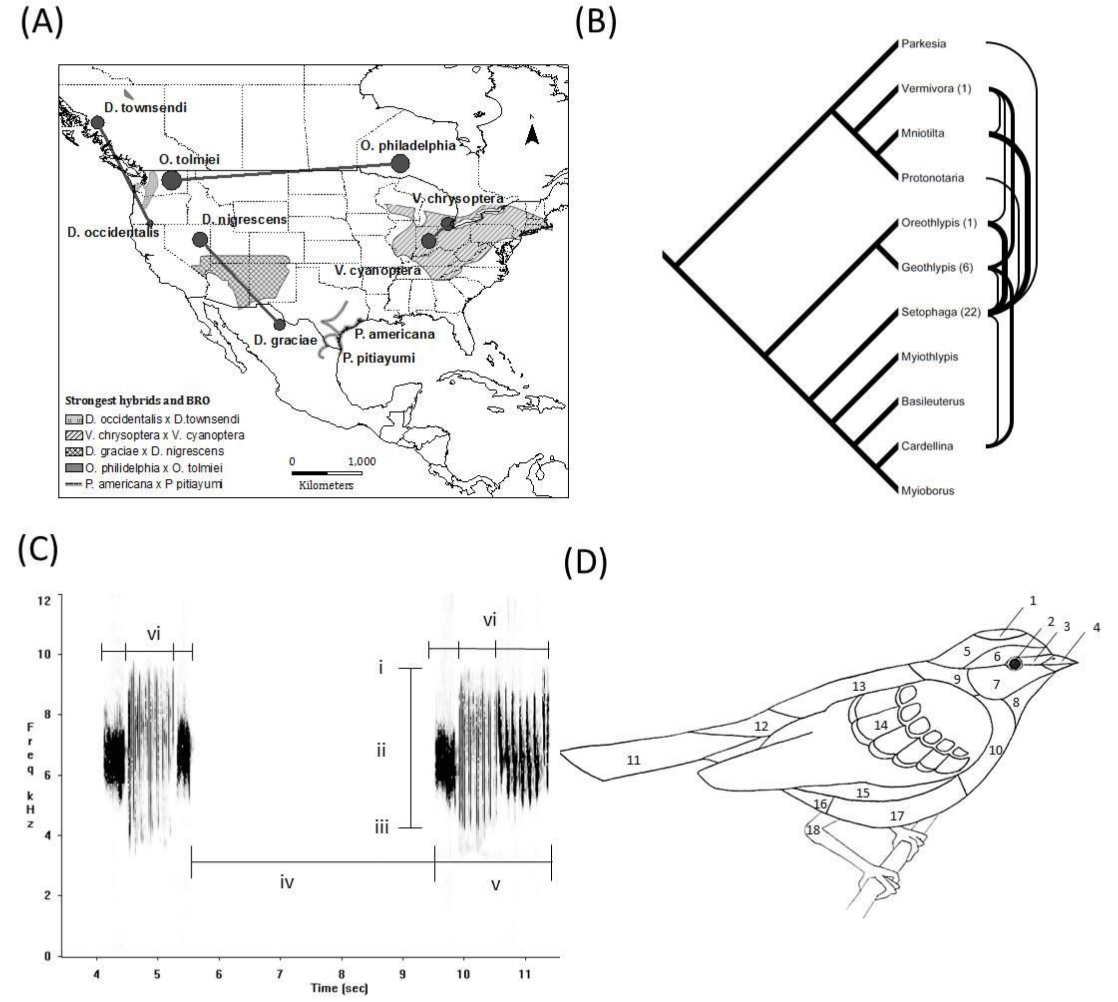


Figure 5: Examples of datasets used for network analysis of hybridisation. (A) Breeding ranges. Levels of sympatry were studied through the presence or absence of breeding range overlap. Breeding range overlaps of specific species that hybridise are displayed, with the species centroids proportional to the size of breeding range. (B) Phylogeny. Date since last common ancestor was investigated for correlation with hybridisation events. Intra-genera hybridisations are within parentheses on the cladogram and genera are connected by “links” proportional to the number of inter-genera hybridisations. (C) Birdsong. Quantitative data were collected on the following, maximum frequency (i), bandwidth (ii), minimum frequency (iii), inter-song interval (iv), strophe duration (v) and number of distinct syllables (vi). The song of the blue-winged warbler (*Vermivora cyanoptera*) was used for visualisation. (D) Plumage. Quantitative data were collected on the following, crown patch (1), eye (2), lore (3), bill (4), head (5), supercilium (6), auricular (7), throat (8), nape (9), breast (10), tail (11), rump (12), mantle (13), wing bars (14), flank (15), undertail (16), belly (17) , legs/feet (18), wing (19*). * Not depicted on diagram.

Phylogeny

It is accepted that hybridising species are often recently diverged (Mallet 2005; Willis et al. 2014). Phylogenetic data for the wood warblers were obtained from Lovette et al. (2010). The phylogeny was also used to control for the confounding effects of shared ancestry whilst assessing whether hybridising species had more similar plumage and song traits. For each pair of species, the amount of time since their last common ancestor (TLCA) was measured relative to the scale-bar provided in the phylogeny. The arbitrary scale ranged from 0 (present-day) to 1 (which links the *Parulidae* to other allied families). Given that the phylogeny was constructed using molecular markers, TLCA might be a good proxy for the amount of genetic differentiation between species. For each species, the number of migratory warbler species in the same genus was also recorded to identify whether species from more speciose genera have higher degrees.

Breeding range

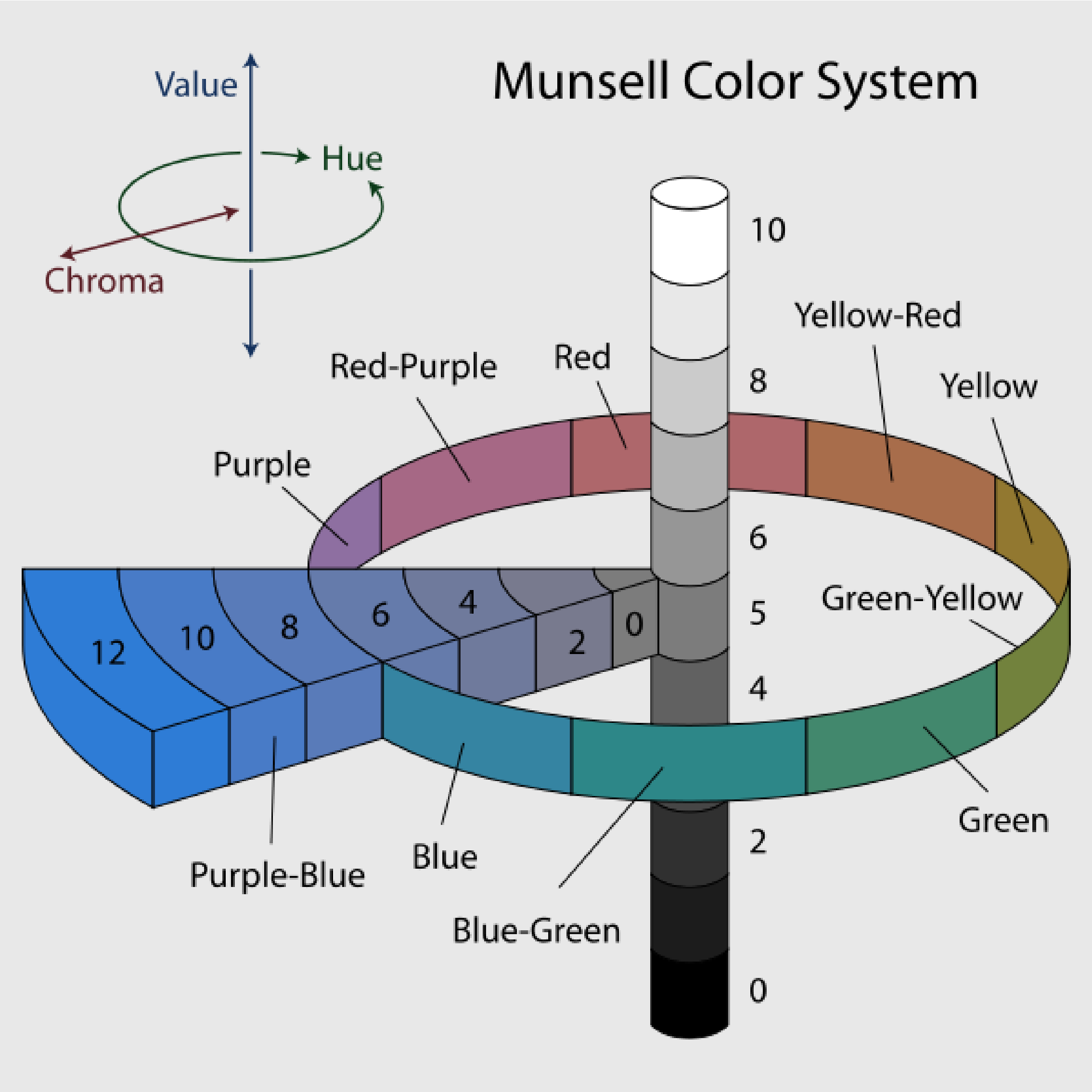
To the measure the size, amount of overlap and distance between breeding ranges, electronic distribution data were obtained from Birdlife International (NatureServe & BirdLife International 2012 http://www.birdlife.org/). Distance was measured between the centroids of each breeding range. The proportion of breeding range overlap for each pair of species was calculated using a Jaccard index (Jaccard 1912). Breeding ranges were projected with an Azimuthal equidistant projection for measurements of centroid distances and an Albers equal area projection (see: Snyder 1987) for measurements of area and proportion overlap. All measurements were conducted using the R-packages “rgeos” (Bivand et al. 2014) and “sp” (Pebesma & Bivand 2005). A diagrammatic representation of the breeding ranges of the most extensive warbler crosses is shown in figure 6 *A*.

Habitat

Species that are ecological generalists in their breeding season may hybridise with more species than those that are ecological specialists. The rationale for this hypothesis is that species which thrive in a wide array of habitats will encounter more species which also utilise the same habitat types than those which use few types of habitat. To determine whether hybridising species share a greater proportion of habitat in their breeding season, the different types of breeding habitat used by each species was obtained from Birdlife International (NatureServe & BirdLife International 2012). A Jaccard index was used to measure the proportion of shared habitat for different species pairs.

Plumage

In order to compare the plumages of large number (703) of pairs of warbler species, we developed a new methodology which only required a human observer to quantify the plumage of the individual (38) warbler species using a fixed quantitative system. The plumage profiles generated for each species could then be compared computationally. To quantitatively measure the plumage of hybridising warblers, illustrations were obtained from Curson *et al.* (1994). Artist-produced illustrations were used due to a lack of standardised photos for all of the species in the system. For each species, plumage was broken down into 19 distinct patches depicted in figure 5 *D* - an adaption of the plumage topography in Curson *et al.* (1994). The coloration of each patch was quantified using the Munsell colour system (Munsell 1912) under standard light conditions by a single observer. Matching the colour of each plumage patch to the corresponding colour-card in the New Munsell Colour Set (Long & Luke 2011) provided a measure of “value”, the light or darkness of the colour; “chroma” the saturation of colour and “hue”, the type of colour. Black white and grey patches were only assigned a value, being ascribed as “neutral colours” in the Munsell system (see insert 1).



Insert 1: A diagrammatic representation of the Munsell colour system. © 2007, Jacob Rus. The image is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

To calculate the plumage similarity between two species, the mean difference over all plumage patches were summed for each of the three Munsell colour measures. Patches which had a neutral colour in one species and not the other were only compared on the basis of their value alone. In the Munsell system, hue is depicted on a colour "wheel" with values between 0 and 100, we scored the difference in hue as the shortest distance around the wheel between the two hues. For example, the hue red would be scored as a 5, yellow-red 15 and red-purple 95. The distance between red and either other colour in this example would be 10. The mean difference over plumage patches for each of the three metrics were weighted equally and summed to make an overall similarity index which ranged between 0 and 1. In this index, a value of 0 would correspond to identical plumage and a value of 1 would correspond to the most different plumages observed with respect to the three Munsell colour metrics. This procedure was used to determine the difference between the plumages of breeding males and adult females between different species. For a measure of plumage dimorphism, breeding male and adult female plumages were compared within the same species.

Our overall method has the advantage of reducing the amount of time taken to make comparisons through the use of computation. This allows large quantitative datasets to be generated by one observer, as opposed to having many observers rank pairs directly (see Martin et al. 2015).

Song

A similar procedure was used to quantify song similarity. To see if hybridising species sing similar mating songs after controlling for ancestry, song recordings were taken from Borror and Gunn's Warbler Songs of North America (Borror & Gunn 1985), part of the Macaulay library from the Cornell Lab of Ornithology and Xeno-Canto (http://www.xeno-canto.org/ - accessed on 04/10/2012). Five separate recordings were taken for each species and 4 strophes, as defined by Darolová et al. (2012), were randomly selected from each recording. Seven traits were measured in each of the 20 selected strophes using the programmes Audacity® (Audacity Team 2008) and Syrinx (Burt 2005). These 7 traits were: highest frequency, lowest frequency, bandwidth defined as the difference in the highest and lowest frequency, maximum number of syllables as defined by Catchpole (1986), minimum number of syllables, strophe duration and the rest period between the selected and subsequent strophe in the recording (see figure 6 *C*). For each species, an average was taken for each song trait. A principal components analysis (Jolliffe 2002) was used to reduce the resulting dataset down to 4 principal components accounting for 96% of the variation in the data. The index was calculated as the sum of the absolute differences across the 4 components. The sum of differences was scaled such that a 0 would represent identical songs within the remit of the measured traits and 1 the most disparate pair of species in the dataset.

Data analysis

Hybridisation data are relational by nature, as each case involves two species. The non-independence of relational data (see section 2.3) coupled with the perceived threat of unreported hybrids means that care has to be taken with the choice of statistical analyses. As advocated in Croft et al. (2011), we opt for null model based hypothesis testing (NMBHT) of our hybrid networks (see section 3.5). Recall that in NMBHT, a test-statistic is chosen to represent a feature of interest in the observed network. To determine the statistical significance of the statistic, it is compared to a reference distribution of the same measure generated by a null model. Statistical significance is derived through examining how often the observed measure of the trait occurs in the networks produced by the null model, as outlined in Manly (1997).

To look for assortativity in the hybrid-network, a test statistic was chosen to represent the tendency for species to hybridise on the basis of a given trait; we will refer to this statistic as the median-edge value (MEV). To calculate the MEV for a given trait, a median is taken of all of the trait values attributed to the edges present in the network. A median was chosen as opposed to a mean as it makes fewer assumptions about the distribution of the data it describes. The MEV has the advantage of only looking at the edges which are present in the network rather than those which are absent, which is very important given the perceived likelihood of unreported hybrids. Given the focus of our analyses on species which have been reported to hybridise, we restricted our explanatory datasets to include just these 38 species.

To first see if recently diverged species were more likely to hybridise in the hybrid network, the MEV for Time Since Last Common Ancestor (TLCA) was taken as a test statistic. To determine if this observed MEV for TLCA was lower than we would expect if hybridisation was uncorrelated with this trait, the species labels on the warbler phylogeny (Lovette et al. 2010) were permuted whilst keeping the observed network fixed. Some pairs of species, often those that are allopatric, have diverged from their common ancestor so recently that their classification as separate species may be questionable. Recall that the whole hybrid-network may be lumped into one species under the strictest interpretation of the Biological Species Concept (Mayr 1982). To try and account for questionable species definitions, the analysis was repeated three times with the edges belonging to 5,6 and 13 pairs of species with a TLCA more recent than 0.05, 0.10, and 0.15 (relative to the time-scale bar provided in Lovette et al. (2010)) fixed. Fixing the presence of certain edges so that they occur in every null network effectively freezes out their contribution to the significance of the test statistic. Given that the edges with a weight of 1 in the network have only been identified through sightings of hybrids, there is a risk that some of these may have been allocated to the wrong pair of species or are misidentified conspecific offspring. To check that any signal observed was not reliant on edges with which there was questionable evidence, the analyses were repeated on both the complete network and a version of the network with the weakest edges removed, which we will refer to as the reduced network.

Further tests were carried out to see if hybridising species shared a greater proportion of breeding range, habitat usage, had more similar song, plumage, or were more geographically proximate than expected by chance. Using Mantel tests (Mantel 1967) on our more reliable independent variables, we revealed that both male song and plumage similarity were significantly correlated with TLCA and a trend was observed with female plumage similarity (male plumage: $R_{\text{Spearman's}}=0.1345$, $p=0.0082$; male song: $R_S=0.200$, $p=0.0106$; female plumage: $R_S=0.077$, $p=0.0998$). This correlation creates a problem, as recently diverged species have been identified to be more inclined to hybridise. Using raw plumage or song similarity data may give us a positive result based on the underlying phylogenetic signal alone. A method was needed to control for the phylogenetic signal present in these datasets. A Multiple Regression Quadratic Assignment procedure (MRQAP) (Krackhardt 1988; Dekker et al. 2007) was used to remove the phylogenetic signal from song and plumage traits. TLCA was used as the independent variable in the MRQAP and male song similarity, male plumage or female plumage the dependent variable in turn. Residuals were taken as being representative of each trait after removing the similarity contributed through shared ancestry. A MEV was taken from these sets of residuals and used in the analysis outlined above; no other traits were significantly correlated with TLCA.

In the studies of interspecific hybridisation using reports, the absence of reports is essentially missing data, providing little information that a hybrid has not at some point occurred. Hybrid offspring can be rare, in many cases only one confirmed hybrid may indicate that two species can cross and produce offspring (see McCarthy 2006). Adopting this conservative standpoint means that using a hybrid network as a response variable for model-based approaches is unwise. In the hybrid network of the new world warblers Parulidae, as few as 3.2% of possible edges were reported, meaning the other 96.8% are potentially missing data. Some network models have been developed with the ability to cope for some level of missing data (e.g. Robins et al. 2004; Koskinen et al. 2010). However, modelling approaches simply cannot function when there is no variation in the response variable, there needs to be at least some edges with which we can be confident of their absence. Using MRQAPs to partial out some of the phylogenetic signal from explanatory variables allows more complex multivariate hypotheses to be tested on networks which are unsuitable for use as a response variable in conventional modelling approaches. The approach is multivariate in the sense that the effect of one variable is evaluated using NMBHT (outlined in Croft et al. 2011) whilst controlling for one or more others. However, there are no effect sizes unlike purely model-based approaches.

An assumption of using MRQAPs to remove the phylogenetic signal in song and plumage is that the relationship between TLCA and the trait in question is constant across the whole phylogeny. It is possible that the relationship between plumage or song similarity and TLCA varies with different clades of the phylogeny. For example, most *Geophlypis* species have yellow colourations whereas species in *Setophaga* exhibit more diverse plumage colourations. We resolve this issue when testing the subsequent effect of a modified explanatory variable on the hybrid network. The null model used to test these modified variables conserves the amount of hybridisation occurring in and between different clades of the phylogeny, allowing a comprehensive control for phylogenetic effects on the traits being tested. An advantage of the approach is that factors which may otherwise have been controlled for via the development of elaborate null networks with many constraints have instead been accounted for through modelling on the explanatory variables. This allows the null model to be tailored to constrain for other biological or structural features of the focal system.

The null model developed to assess the significance of the MEVs of traits other than TLCA randomly allocated edges between species, but conserved the number of crosses between species that were: sympatric within the same genera, sympatric between different genera and allopatric within the same genera - there were no allopatric crosses involving species of two different genera. For the purposes of this study we use the term sympatric to denote any pair of species with a breeding range overlap. The phylogenetic constraint was necessary to account for the assumption of using a MRQAP to control for TLCA, which would only control for phylogenetic effects driven by patristic distance which may vary within and between different clades in the phylogeny. The constraint also ensured that null networks did not under-sample within-genera crosses, which account for the significantly high ($p_{perm}<0.0002$) proportion (65%) of edges in the observed network but only 35% of all pairs of migratory warbler species. The biogeographic constraint conserved the high proportion of sympatric pairs of species in the observed network (42 pairs, 89%) which is significantly higher ($p_{perm}<0.0002$) than expected if 47 pairs of species were chosen at random, as only 55% of pairs of migratory warblers are sympatric.

At this point, it could be argued that the null networks produced to determine the significance of traits correlated with TLCA would control for phylogeny as they contained the same number of within and between-genus crosses as the observed network and because of this the use of the MRQAP is unnecessary. However, the null networks produced were not conditionally uniform with respect to the MEV for TLCA as this would be difficult to constrain explicitly whilst controlling simultaneously for sympatry. The production of networks with high or low MEVs for TLCA would increase the variance of the test statistic under the null and affect hypothesis testing. A null network with a very low MEV for TLCA will tend to have very low plumage disparity and vice versa.

Conserving the amount of crosses within and between genera, and those which were sympatric and allopatric, represented the null expectation that closely related species which come into contact with each other were the most likely to hybridise. The null model had enough degrees of freedom such that no two identical null networks were produced in 4999 random replicates. The analysis was repeated on both the complete and reduced network using both our own dataset and that used in Willis *et al.* (2014). In each case, the analysis was repeated with the edges fixed between species that had a TLCA lower than the three given thresholds.

The degree of warbler species observed to hybridise in the complete network was correlated with node attribute data to see if a species propensity to hybridise could be explained by clade size, breeding range area, habitat usage or sexual dimorphism. For each trait, a Spearman's correlation coefficient (Spearman 1904) was taken as a test statistic, as the distribution of the degrees of species was non normal, exhibiting heavy right-hand skew. The correlation coefficient between degree and each trait in question was compared to a distribution of coefficients produced by using a null model. In this case, the null model kept the network structure fixed and randomised the attribute data assigned to each node. The analysis was repeated for traits which were found to be linked to hybridisation on the reduced network to test the robustness of the results to uncertainty in our observed crosses.

4.4 Results

The warbler hybrid network

In figure 6, we have taken a closer look at the hybrid-network of wood warblers. Of the 55 species that breed in North America, 38 interbreed with 47 recorded crosses between them – 32 of these occur between species of the same genera and 42 occur between sympatric species. Although the network of hybridisation events between migratory wood warbler species is sparse, containing only 3.2% of possible edges, the network forms one connected component. The component is dominated by species in the genus *Setophaga* (blue nodes in figure 6) which account for 22 of the species in the complex. All 5 allopatric crosses occur between species of the same genus. Removing the 8 weakest edges in the network, which McCarthy defines as reports to which there is some questionable evidence, leaves a large component containing 35 species excluding the Belding's yellowthroat 3 (*Geothlypis beldingi*); Hooded warbler 21 (*Setophaga citrina*); and Kirtland's warbler 23 (*Setophaga kirtlandii*) .

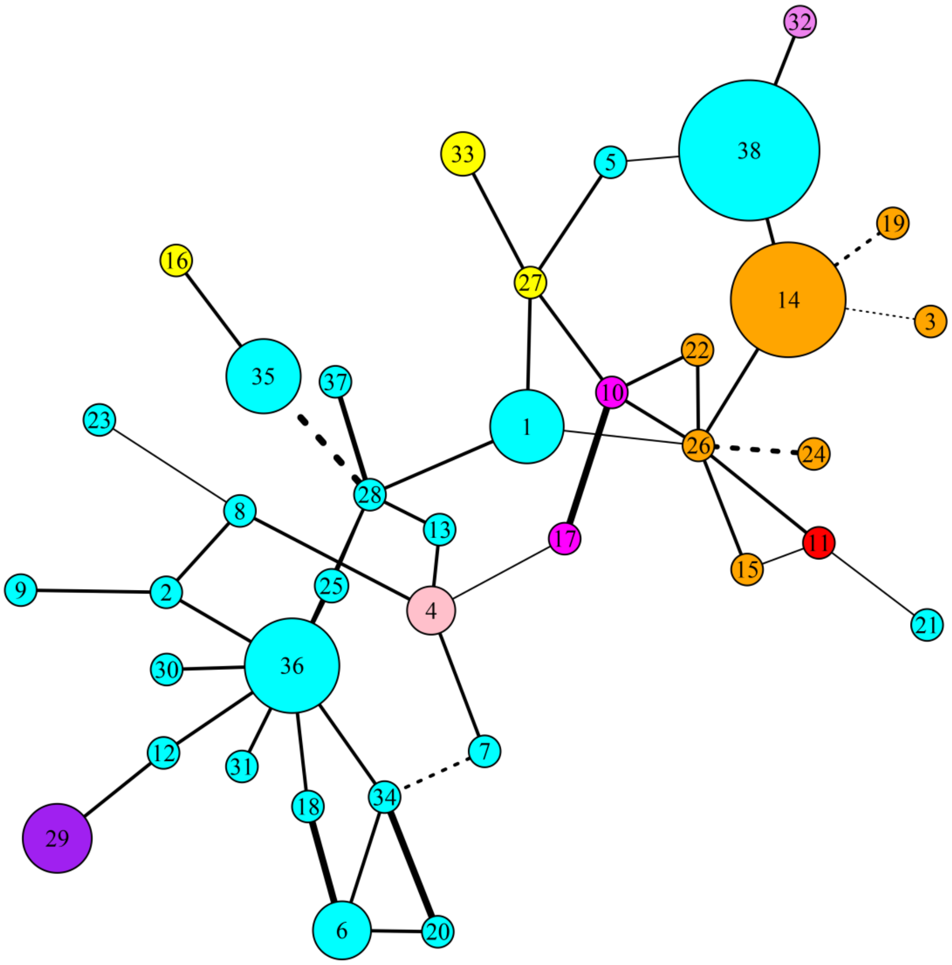


Figure 6: Network depicting cases of interspecific hybridisation in the wood warblers. Nodes depict species as defined in Lovette (2010), labelled to represent the identity of each species. The size of the nodes represents the breeding range area of species and the colour represents genus as defined in Lovette (2010). Edges are weighted based on the extent of the cross as defined in McCarthy (2006); solid lines represent crosses between sympatric species, dashed lines allopatric. Species labels: 1 American redstart (*Setophaga ruticilla*); 2 Bay-breasted warbler (*Setophaga castanea*); 3 Belding's

yellowthroat (*Geothlypis beldingi*); 4 Black-and-white warbler (*Mniotilta varia*); 5 Black-throated blue warbler (*Setophaga caerulescens*); 6 Black-throated gray warbler (*Setophaga nigrescens*); 7 Black-throated green warbler (*Setophaga virens*); 8 Blackburnian warbler (*Setophaga fusca*); 9 Blackpoll warbler (*Setophaga striata*); 10 Blue-winged warbler (*Vermivora cyanoptera*); 11 Canada warbler (*Cardellina canadensis*); 12 Cape May warbler (*Setophaga tigrina*); 13 Cerulean warbler (*Setophaga cerulea*); 14 Common yellowthroat (*Geothlypis trichas*); 15 Connecticut warbler (*Oporornis agilis*); 16 Crescent-chested warbler (*Oreothlypis superciliosa*); 17 Golden-winged warbler (*Vermivora chrysoptera*); 18 Grace's warbler (*Setophaga graciae*); 19 Gray-crowned yellowthroat (*Geothlypis poliocephala*); 20 Hermit warbler (*Setophaga occidentalis*); 21 Hooded warbler (*Setophaga citrina*); 22 Kentucky warbler (*Geothlypis formosa*); 23 Kirtland's warbler (*Setophaga kirtlandii*); 24 MacGillivray's warbler (*Geothlypis tolmiei*); 25 Magnolia warbler (*Setophaga magnolia*); 26 Mourning warbler (*Geothlypis philadelphia*); 27 Nashville warbler (*Oreothlypis ruficapilla*); 28 Northern parula (*Setophaga americana*);29 Northern waterthrush (*Parkesia noveboracensis**); 30 Palm warbler (*Setophaga palmarum*); 31 Pine warbler (*Setophaga pinus*); 32 Prothonotary warbler (*Protonotaria citrea*);33 Tennessee warbler (*Oreothlypis peregrina*); 34 Townsend's warbler (*Setophaga townsendi*); 35 Tropical parula (*Setophaga pitiayumi*); 36 Yellow-rumped warbler (Setophaga coronata); 37 Yellow-throated warbler (*Setophaga dominica*); 38 Yellow warbler (*Setophaga petechia*).

Factors associated with the degree of hybridising species

Of the factors tested, breeding range provided the best explanation of why some species hybridise with many others and some with few. The degree of a species was correlated with the number of species its breeding range intersected (R_S=0.506, p=0.0012). Species with larger breeding ranges also hybridised with a greater number of others (R_S= 0.334, p=0.0240), unsurprising given that larger breeding ranges intersect a greater number of breeding ranges (R_S=0.572, p<0.0002). There was also weak evidence that sexual selection measured through plumage dimorphism correlates with the number of hybridization events exhibited by a given species. The amount of dimorphism between the female and breeding male plumage of a given species may be weakly correlated with degree (R_S=0.212, p=0.0946). However, these correlations are reliant on crosses with weak supporting evidence; removing these 8 edges from the network reduces both the correlation between degree and the number of intersected breeding ranges R_S=0.180, p=0.1598), breeding range area (R_S=0.192, p=0.1394) and plumage dimorphism (R_S= -0.070, p=0.6646). There was no evidence to suggest that species which use a greater number of habitats to breed have higher degrees (R_S=0.1269, p=0.9397), nor do those belonging to more speciose genera (R_S=0.0341, p=0.8387).

Assortativity in the hybrid network

Hybridising species have a more recent common ancestor than expected if hybridisation was independent of phylogenetic history. The observed MEV for the TLCA of hybridising species in the complex was 0.315; this was significantly lower (p=0.0004) than expected given the null model, which produced networks with a median MEV of 0.579. The network remained significantly assorted on the basis of TLCA when species more recently diverged than a given threshold (TLCA<0.05, 0.1, 0.15) were lumped through fixing their edges in each null network (MEV=0.315, p= 0.0074; MEV=0.315, p=0.0098; MEV=0.315, p=0.0226); this was to account for uncertainty in species definitions. Upon removing the weakest edges in the network, the analysis was repeated at each stated threshold. Without lumping, the hybrid network remained significantly assorted with an MEV of 0.315 and a p-value of 0.0018. The result was stable to having the weakest edges removed and the four pairs of species with a TLCA less than 0.05 lumped (MEV=0.315, p=0.0280). Upon the removal of hybrid reports with questionable supporting evidence and lumping species which had recent ancestors more recent than 0.1 and 0.15 lumped, the result became statistically non-significant (MEV=0.315, p=0.0564; MEV=0.315, p=0.0890)

The observed hybrid network was also assorted such that hybridising species tended to be geographically close and had more similar plumage and song traits after controlling for phylogenetic history and preserving the number of within-genera, between-genera, sympatric and allopatric crosses in the null model. The results are summarised in Table1, complete results can be found in Table 2 and Table 3. Hybridising species had significantly closer breeding range centroids under all of the conditions tested. In the complete network, hybridising species had consistently more similar song and plumage traits. The reduced network was significantly assorted on the basis of male plumage, but this was reduced to a trend when any species were lumped. If the observed network was taken as given, there was a trend for species that share more breeding range to hybridise as was there a trend for species that are closer longitudinally to hybridise, whereas latitudinal distance between breeding ranges was not significant (see Table 2). The observed network was not found to be assorted with respect to TLCA in this analysis given that the number of hybridising species within the same genera and between different genera was preserved in each null network. Hybridising species did not share significantly more habitat than expected given the null model.

Table 1: Assortativity of the warbler hybrid network

	Complete network				**Reduced network**			
Threshold	0	0.05	0.1	0.15	0	0.05	0.1	0.15
Overlap	.	.	.	*	*	*	*	*
Male plumage	**	*	*	*	*	*	.	.
Female-plumage	*	*	.	*	*	.	ns	.
Song	**	**	**	***	ns	ns	ns	*
Longitudinal-distance	ns	ns	.	ns	.	.	*	.
Latitudinal-distance	ns	ns	ns	ns	ns	ns	ns	ns
Direct distance	*	*	*	*	*	*	*	*
Habitat	ns	ns	ns	ns	.	ns	ns	ns
TLCA	ns	ns	ns	ns	ns	ns	ns	ns

Notes: Table 1 depicts the traits with which the warbler hybrid network was tested for assortativity on the complete and reduced hybrid network as defined in the methods. Tests were conducted on the complete network and the reduced network with species lumped with a TLCA lower than the stated threshold. Statistical significance indicated by period and asterisk: ns $p \geq 0.1$; . $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; or *** $p < 0.001$.

Table 2: P-values of the assortativity analyses depicted in Table 1.

	Complete network				**Reduced network**			
Threshold	0	0.05	0.1	0.15	0	0.05	0.1	0.15
Overlap	0.0740	0.0702	0.0666	0.0434	0.0302	0.0232	0.0288	0.0152
Male-plumage	0.0044	0.0198	0.0284	0.0320	0.0112	0.0404	0.0558	0.0672
Female-plumage	0.0250	0.0376	0.0772	0.0358	0.0438	0.0822	0.1052	0.0524
Song	0.0086	0.0026	0.0010	0.0006	0.2404	0.1466	0.1018	0.0374
Longitudinal-distance	0.1274	0.1124	0.0796	0.1048	0.0752	0.0536	0.0460	0.0532
Latitudinal-distance	0.3596	0.1970	0.1516	0.1568	0.2930	0.1366	0.1060	0.1268
Direct distance	0.0402	0.0186	0.0168	0.0232	0.0456	0.0204	0.0158	0.0180
Habitat	0.1176	0.1406	0.1958	0.3414	0.0860	0.1112	0.1602	0.3290
TLCA	0.3694	0.5552	0.7222	0.8252	0.8042	0.9174	0.9478	0.9790

Table 3: MEVs of the assortativity analyses depicted in Tables 1 and 2

	Complete network	**Reduced network**
Overlap (%)	20.0	22.0
Male- plumage	-0.0381	-0.0458
Female- plumage	-0.0458	-0.0492
Song	-0.0775	-0.0445
Longitudinal- distance (km)	574.49	540.27
Latitudinal- distance (km)	1296.63	1296.63
Direct distance (km)	1453.10	1453.10
Habitat (%)	40	40
TLCA	0.3150	0.3150

Comparison with Willis et al. (2014)

We obtained similar findings using the same American wood warbler hybridization dataset, consisting of only post-1980 hybrid reports, as that presented in Willis et al. (2014) using our analytical approach in place of Willis et al's use of Mantel (Mantel 1967) and partial Mantel tests (Legendre & Legendre 2012). By lumping most recent evolved species (TLCA<0.05) and by assuming complete trust in the quality of the evidence for each cross, we find that hybridising species tend to have more similar song (MEV=-0.039, p=0.0644), but this trend was lost either when the reports with the most questionable evidence were removed (MEV=0.019, p=0.2628), or when none of the species in the dataset were lumped (MEV=-0.039, p=0.1066). Similar to our cross list, when weakly supported hybridisation events are omitted from the dataset, there is only a significant result upon lumping species with a TLCA less than 0.15, 7 of which are reported to hybridise (MEV =-0.019, p=0.0146). Full results using our methodologies on the list of hybrid crosses used in Willis *et al*. (2014) can be found in Tables 4 and 5.

Table 4: P-values of the assortativity analyses conducted on the hybridisation dataset in Willis et al. (2014)

	Complete network				**Reduced network**			
Threshold	0	0.05	0.1	0.15	0	0.05	0.1	0.15
Overlap	0.5696	0.4004	0.3482	0.3480	0.6070	0.4482	0.3920	0.3882
Male-plumage	0.0396	0.1458	0.1572	0.2292	0.0982	0.3574	0.3914	0.4948
Female-plumage	0.1530	0.2672	0.3128	0.1992	0.1090	0.1984	0.2248	0.1276
Song	0.1044	0.0614	0.0254	0.0058	0.2410	0.2720	0.0984	0.015
Longitudinal-distance	0.1396	0.1340	0.0956	0.1224	0.2032	0.2146	0.1620	0.2176
Latitudinal-distance	0.4088	0.2050	0.2082	0.1712	0.3666	0.1482	0.1678	0.1246
Direct distance	0.1248	0.0644	0.0706	0.0732	0.2176	0.1054	0.1424	0.1698
Habitat	0.2926	0.3824	0.5466	0.7448	0.3210	0.4328	0.6146	0.8478
TLCA	0.0002	0.0018	0.0046	0.0470	0.0002	0.0008	0.0010	0.0590

Table 5: MEVs of the assortativity analyses (Table 4) conducted on the hybridisation dataset in Willis et al. (2014)

	Complete Network	**Reduced Network**
Overlap (%)	15.5	16.0
Male- plumage	-0.04361	-0.0482
Female- plumage	-0.0287	-0.0387
Song	-0.0390	-0.0194
Longitudinal- distance (km)	405.29	412.20
Latitudinal- distance (km)	1251.02	1296.63
Direct distance (km)	1358.63	1494.88
Habitat (%)	50	50
TLCA	0.2677	0.2677

4.5 Discussion

Interspecific hybridisation is an important evolutionary process with profound implications to biodiversity, but to better understand its consequences we also need to understand its causes. If we know which traits makes species prone to hybridise, then perhaps we can learn more about the evolutionary trajectories of those traits early in the speciation process. Perhaps the traits that show the most disparity in a family or order of species are those which bear the closest links to hybridisation in their evolutionary histories. Finding definitive causes of natural hybridisation has remained a challenge due to the inherent difficulties in performing manipulative experimentation. Instead we are left in the realms of comparative biology to gain insight, further complicated by incomplete hybrid data-sets, interlinked hybridisation events in recently diverged taxa and species which are linked through shared ancestry. Studies have often been able to tackle one or two of these issues at a time; we have presented an approach which allows all three to be addressed simultaneously, opening up some of the richest hybrid data-sets we currently have; such as those documented in McCarthy (2006).

Identifying pairwise correlates of hybridisation requires large datasets of explanatory variables. Every pair of species needs to be compared so that appropriate null or permutated datasets can be constructed. Automating some of the data generating process serves both to save time and reduce bias. Martin et al. (2015) required multiple human observers to score plumage similarity between pairs of New World Warblers. Our novel approach requires one human observer to quantify the coloration of plumage patches of individual species using a quantitative approach based around the Munsell Colour system (Munsell 1912; Long and Luke 2011). The species are then compared computationally; producing quantitative measures of similarity and reducing the number of human assessments by (n-1)/2 where n is the number of species in the study system. Although direct computational measures of plumage may provide increased accuracy at measuring colour saturation, hue or brightness than a human observer, finding standardised photos for multiple species can prove difficult; our approach only requires illustrations. Studies show that human observations correlate highly with spectrophotometer readings, and that although there is individual variation in human vision, our colour discrimination abilities are usually superior to technology (Bowers 1956, Armenta et al. 2008).

Using a network-based approach, we have illustrated for the first time, the interconnected nature of hybridisation between species providing new insight into how studies of hybridisation in evolutionarily young collections of species can be structured. We have demonstrated that large network components are an expected feature in taxa when every species hybridises, on average, with at least one other; the significance of hybrid networks may hinge on how freely genetic information can flow through them. Undirected networks may however be misleading when it comes to representing genetic flow through a hybrid network; mtDNA analysis shows that, in the blue-winged/golden-winged warbler cross, blue-winged warbler genes are moving into golden-winged populations, but not vice versa (Gill 1994). As more molecular data becomes available, it may be possible to make and analyse these directed networks using network modelling approaches, such as those outlined in Snijders (2011). Regardless, networks may provide a useful framework with which to study hybridisation in systems where it is frequent; dyadic approaches have often required the removal of large quantities of data to maintain the assumption that one dyad is independent of another. In the wood warblers, we have started to identify the factors that may dictate or at least correlate with structural aspects of the hybrid network. In doing so, we have found some behavioural and ecological factors which may promote hybridisation between a pair of species and cause some species to hybridise more than others.

Mallet (2005) found that evolutionarily young species were more likely to hybridise, not having had enough time to develop complete reproductive barriers (see Seehausen et al., 2014). Our study supports Mallet as the warbler hybrid-network was assorted such that hybridising species had recent TLCAs on the whole. Mallet made a general inference about the relationship between ancestry and hybridisation across many taxa, whereas our finding is taxon-specific. We have been able to sample many cases from a recently radiated family of species (Jetz et al. 2012) and show that even within such an evolutionarily young collection of species, there is still a further tendency for hybridisation to occur between the most recently diverged species within.

The warbler hybrid network was assorted by breeding range. Hybridising species were close together and species with larger breeding ranges also had higher degrees. This may be unsurprising as two individuals of different species have to come into contact to mate, which is more likely if they are geographically close. However, the large number of sympatric hybrid pairs in the hybrid network may be surprising given the belief that sympatric species are supposed have more developed isolation mechanisms than their allopatric counterparts as result of the “Wallace effect” (Sawyer and Hartl 1981; Gillespie 1991). Given the observed northward shift in the distribution of North American birds (La Sorte and Thomson 2007), we might expect previously allopatric species to become sympatric and a subsequent increase in the incidence of hybridisation. An increase in the incidence in hybridisation in the warblers is supported by the forecast that the population of warbler species are likely to decline in the future (Strode, 2003) and species pairs where one is endangered have been found to produce more hybrids in other taxa (Randler 2006). Rhymer & Simberloff (2012) highlight that hybridisation can pose a credible threat of extinction to rare species.

Willis (2013) made a call for a greater understanding of the behavioural factors associated with hybridisation to better understand the process of diversification and inform conservation action. Behavioural factors associated with mate choice can be taxa-specific. In the Parulidae family, song and breeding plumage is important for intra-species mate choice (Weatherhead and Shutler, 1990; Taff et al. 2012; Byers 2007), whereas the same may not hold true in other taxa. In the *Galliformes*, plumage and visual displays may be important with intra-mate choices (Madge et al. 2002; Kolm et al. 2007; Lislevand et al. 2009; Kimball et al. 2011). The network-based approach has allowed us to study the behavioural factors associated with hybridisation in warblers specifically. We found that species that hybridise sing similar songs and have more similar male plumage after accounting for shared ancestry. Furthermore, the trend for species with greater plumage dimorphism, a proxy for sexual selection, to hybridise more is also likely to be system specific, as we see degree variation in taxa such as the *Accipitriformes*, which often show little, if any plumage dimorphism. Further work is needed to clarify if species with greater levels of sexual selection do indeed hybridise more, using different systems and different measures as a proxy for sexual selection in birds.

There are multiple pitfalls associated with analysing hybrid networks. Inference can hinge on how we choose and define species, which could be a contentious issue when analysing crosses between recently diverged species. The likelihood of missing edges in network data restricts us from performing certain analyses and looking at network measures that are sensitive to the addition of edges (see section 2.5 or James et al. 2009). The qualitative weightings of edges restrict us from taking more informative, weighted network measures and asking detailed questions about the factors that affect the extent of hybridisation between species. Further, some of these crosses are so rare that some may argue that they have no evolutionary value; many of our findings currently depend on trusting rare crosses with questionable evidence. The warbler hybrid network was tested in scenarios where recently diverged species were lumped and the least reliable edges in the network removed. When numerous species in the network are lumped together and weak edges are removed, many of the findings of this study became statistically non-significant. Further, although we have focussed our analyses on the presence rather than absence of edges, the results have not been tested against the simulated inclusion of missed or newly reported hybrids nor have they been tested against situations where the species needed to be split instead of lumped. With the increasing availability of genomic data, it will be easier to identify and quantitatively weight hybridisation through genetic introgression and define nodes more accurately.

The approach of Willis et al. (2014) could further our understanding of the links between behavioural and ecological factors and hybridisation. However, the methods they used have requires the absence of evidence to be equivalent to evidence of absence, which is a questionable assumption for less observed pairs species. Through comparing our results to that of a previous study attempting to address some of the same questions (Willis et al. 2014), we suggest that their assertion that song similarity is linked to hybridisation was likely incorrect given their data (see Table 4). There is a temptation to use analytical approaches which have been developed for high quality data-sets beyond that which is currently available across multiple species pairs. We chose to avoid approaches which bear undue emphasis on the absence of hybrid reports until molecular measures of hybridisation are available for all species in a given group. Treating absent reports as "true" negatives could potentially bias results given the nature of report data. We applied our analytical methods (see: data analysis), which were weighted solely on the presence of hybrid reports, to the same hybridisation dataset as used in Willis et al. (2014). A parallel can be observed in the study of animal social networks, where the field has historically had to cope with data with similar issues as those observed in hybrid report data. Akin to the increasing use of molecular methods to improve hybridization datasets, the field of animal social behaviour is undergoing a revolution in the quality of data available through the adoption of remote-sensing technologies (Krause et al. 2013). The change in data quality is being reflected in the choice of analyses often employed in such studies; providing a hotbed of analytical methods for future studies of hybrid networks.

Our conservative approach of choosing to ignore all hybrid pairs is not without its own fault. We assume that absent hybrid reports are equivalent to absence of evidence, whereas we might more logically assume that there has been sampling effort directed at observing most if not all hybrid pairs. There is a centre ground between the study presented in this chapter and that of Willis et al. (2014) where a proxy of sampling effort for hybrids can be constructed from sightings data on each species. For instance, if two warbler species both have a very incidence of sightings, then we might assume that there is also a high sampling effort for hybrids between those two species. Conversely, if two species are seldomly sighted, then we might assume the sampling effort for hybrids of those two species is lower. Entering that sampling index into a logistic model (such as the *netlogit* logistic MRQAP by Butts 2014) along with other factors of interest will help control for variable sampling whilst evaluating the contribution of biologically interesting factors.

The novelty of the work presented in this chapter is that it is the first study to implicitly link male plumage traits to hybridisation across a whole family of species. In doing so it builds on a growing body of literature (such as Martin 2015) which builds on the notion of hybridisation shaping plumage evolution in some taxa. Secondly, the study is the first to directly test the biology behind hybrid network structure, perhaps with trivial results, finding that the most widely distributed species are also those which tend to hybridise with the most others- acting as hub individuals in the hybrid network. Thirdly, our study is the first to take a network modelling approaches, the MRQAP, and use it to partial out phylogenetic signal from explanatory variables linked to ancestry, allowing network assortment on the basis plumage and song traits to be evaluated more or less independently of ancestry using NMBHT. Using MRQAPs to partial out nuisance signals from variables are now also being used in animal social network studies (see James et al. 2015). Using a tailored null model in combination with the MRQAP has allowed the phylogenetic signal to be removed from associated variables not just through partialling out the signal explained by patristic distance, but also accounting for phylogenetic structure through the null model used in NMBHT. Overall, we have presented a new approach to model networks where we can be confident in the quality of our explanatory variables, but have to explain a response variable (the observed network) which may for a number of reasons be unsuitable for a generic network modelling approach.

To conclude, taking a network-based approach to hybridisation allows multiple cases of hybridisation to be compared simultaneously in a biologically distinct group of species. The method has yielded significant findings which we may interpret as indications given the quality of our hybrid-data, helping further our understanding of the formation and structure of hybrid networks, which may be viewed as a form of species complex in light of their connectivity. The details of all reported avian crosses are available in McCarthy (2006), which can be used as source to create hybrid networks. Data with which to compare hybrid networks can often be found freely available to download from various sources (see methods). Both R and the associated packages used to create and analyse the networks in this study are freely available from http://www.r-project.org/.There are many opportunities to expand and develop the approach presented in this study. There are many other avian systems where hybridisation between species is likely to form complexes. A fast route to identifying these systems is to look for those where there are at least as many hybridisation events as species. Further studies should address sampling effort to avoid both the limitations of our own study and that of Willis et al. (2014).

III. Multivariate analysis of remotely sensed network data

5. Inter-nest drifting in Paper wasps *Polistes canadensis*

5.1 Abstract

Stable social groups usually consist of families. However, recent studies have revealed higher-level social structure, with interactions between family groups across different levels of social organisation. These interactions seemingly challenge our notions of kin selection as a driver of eusociality as workers appear to help non-natal nests over their own natal nest. In this chapter, we use automated radio-tagging data and social network analyses to uncover the adaptive value of interactions across levels of social organisation in the primitively eusocial paper wasp *Polistes canadensis*. We detected three levels of social organisation (nest, aggregation, community) which exchange 'drifter' individuals within and between levels. Using a variation of the MRQAP, stratified to account for study design, networks of drifter movements were explained by distance between nests, group size of donor nests and the worker to brood ratios on both donor and recipient nests. Drifters responded to experimental manipulation of the worker to brood ratios on natal nests, by drifting preferentially to smaller sized nests. These findings suggest an adaptive role for multi-level social interactions.

5.2 Introduction

The abundance of inter-nest drifting in paper wasps *Polistes canadensis* poses a major challenge to kin-selection, the predominant theory used to explain many aspects of social evolution, such as eusociality. In a leading study of wild populations of paper-wasps, "drifting" of workers from natal to non-natal nests was detected in 94% of nests monitored, with 56% of female paper-wasps identified to drift (Sumner et al. 2007), a thirty-fold increase on previous reports of drifting in a natural eusocial insect population (Paar et al. 2002). Under kin-selection theory, workers would be expected to only help their natal colony, receiving indirect fitness benefits through helping to raise closely related brood. Understanding the mechanisms which underpin inter-nest drifting can thus reveal more about both the evolution and social organisation of eusocial societies.

Sociality in animals can be described as the tendency for individuals in animal populations to associate into social groups and form cooperative societies (Croft et al. 2008), regarded by Maynard Smith & Szathmary (1995) as one of the major evolutionary transitions of life. Eusociality is a highly organised form of sociality where some individuals in a cooperative society forfeit their direct reproductive success to help raise the reproductive success of closely related kin. Eusocial societies are also defined by: the use of a common nest site; the presence of parents, offspring and overlapping adult generations in the same nest; cooperative care of young; the reproductive division of labour, and a caste system – where individuals are specialised into groups which perform certain roles in the colony (Wilson 1971; Crespi & Yanega 1995; Wilson & Hölldobler 2005; Nowak et al. 2010).

Eusociality is an evolutionarily interesting phenomenon, as many of the individuals involved in eusocial societies do not pass on their genes directly through the generation of offspring. Given this, one might ask how the genes encoding worker-like behaviours could spread across populations given the apparent costs to the individual. There are a number of hypotheses which have been proposed explaining how a system like eusociality could evolve, many of which are mutually compatible. All social societies are also affected by the benefits and costs associated with group-living. Advantages may include reduced predation through mobilised defence strategies and faster predator detection, what Hamilton refers to as the “selfish herd” as each individual has lower mortality risk (Hamilton 1971). Information sharing may lead to increased foraging efficacy- Many eusocial organisms have strategic advantage when acquiring food in groups (Clark & Mangel 1986; Valone 1989). In areas where predation is high or resources are hard to come by, there may be more merit in staying in the natal nest with the chance of inheriting it, than risking dispersal (e.g. Lovegrove 1991). These benefits however must offset some common costs of living in groups such as increased risk of disease and parasites, and easier detection of the group by predators (see Krause & Ruxton 2002). However, group-living alone does not explain the occurrence of eusociality within a group (see Crespi 1994).

The widespread incidence of eusociality in Hymenoptera, which includes paper wasps, may be attributed to a haplodiploid sex determination system. In Hymenoptera, males are haploid, having one copy of each chromosome, and arise from the development of unfertilised eggs. In contrast, females are diploid having chromosome pairs, like most eukaryotic organisms developing from fertilised eggs. The consequences of a haplodiploid sex determination system is that sisters are very closely related to one another, always containing the same set of chromosomes from the father and having a fifty percent chance of receiving the same complementary chromosomes from the mother. The haplodiploidy hypothesis proposes that the high relatedness amongst full haplodiploid sisters is responsible for the frequency of evolution of eusocial behaviour in hymenopterans (Foster et al. 2006). In kin selection, genes may increase in frequency when $Br>C$, where r is the genetic relatedness of the recipient to the actor, B the reproductive benefit gained by the recipient of the altruistic act is greater than the reproductive cost C to the individual performing the act (Hamilton 1964). The high relatedness between haplodiploid sisters facilitates the adoption of behaviours more costly to the individual. However, the haplodiploid hypothesis alone has been shown not to be in itself sufficient to explain the occurrence of eusociality given the presence of eusociality in non-haplodiploid species (Wilson & Hölldobler 2005; Nowak et al. 2010).

The concept of inclusive fitness is a popular, yet debated (see Nowak et al. 2010), explanation of how eusociality can exist as a strategy (Bourke 2011). Inclusive fitness, the defining feature of kin selection theory, states that as related individuals share genes, acts carried out by one individual to increase the fitness of a related individual will lead to more of that individual’s genes being passed into the next generation, even if it does not breed itself. This can be more formally described as Hamilton’s rule (Hamilton 1964), as outlined above. Increasing the relatedness between individuals in a colony may facilitate the development of eusociality, allowing more costly acts in terms of an individual’s direct fitness to aid the reproductive success of closely related kin. In hymenoptera, haplodiploidy facilitates higher relatedness than observed in many other species. For both haplodiploid and diploid organisms, if the queen is monogamous for the duration of her lifetime, her progeny will be equally related to each other as well as any offspring they have, which means that natural selection will favour strategies to raise siblings if they are more efficient than those directed at raising offspring (Boomsma 2007;

Boomsma 2009). In eusocial hymenoptera, lifelong monogamy is a ubiquitous ancestral feature (Hughs et al. 2008).

The observation of nest-drifting, where workers visit nests other than their natal nest, challenges many of the notions associated with eusociality. Firstly, if we consider that eusocial organisms often live in patchy resource environments, with often hostile conditions outside of the nest, there is substantial risk associated with leaving the nest (see Perrin & Lehmann 2001). Eusocial thrips, aphids, shrimps, mole-rats, beetles and some ant and termite species all manage to harvest many of the resources required for reproduction within the confines of the natal nest. Individuals who do leave these nests to found new colonies may receive huge fitness benefits which may offset some of the risk of dispersal. Yet drifting behaviour has been detected in social wasps, such as vespine wasps (Akre et al. 1976), stenogastrine wasps (Cervo et al. 1996) and polistine wasps (e.g. Sumner et al. 2007); honeybees (Paar et al. 2002; Chapman et al. 2010; Neumann et al. 2000), sweat bees (Ulrich et al. 2009), and bumblebees (Birmingham et al. 2004; Lopez-Vaamonde et al. 2004; Takahashi et al. 2010; O'Connor et al. 2013; Zanette et al. 2014). In all of the examples above, individuals leave the nest to forage, which may provide some indication that these environments are less hostile relative to the adaptions of these individuals. However, it does not explain why individuals drift to other nests, apparently having no fitness benefits for themselves or their close kin. Visiting other nests may also the risk of bringing parasites and disease back to the natal nest (see Schmid-Hempel (1995) for review on the parasites of social insects). Without apparent fitness payoffs to kin, the presence of nest drifting would contradict inclusive fitness theory, as it carries potential costs to the individual. Further, under group-selection, groups with nest drifters would be less efficient than those whose workers performed tasks solely for the good of the nest (Nowak et al. 2010).

In some cases, inter-nest drifting has been attributed to discrimination errors by the individuals involved. In *Apis* and *Bombus* bee species drifting has been attributed to social parasitism, where workers lay eggs in a non-natal nest but do not help with raising young or have any positive role in the colony. The parasitic offspring may take advantage of colonies without a queen or reproductively out-compete resident workers (Lopez-vaamonde et al. 2004; Nanork et al. 2005; Chapman et al. 2009; Takahashi et al. 2010).

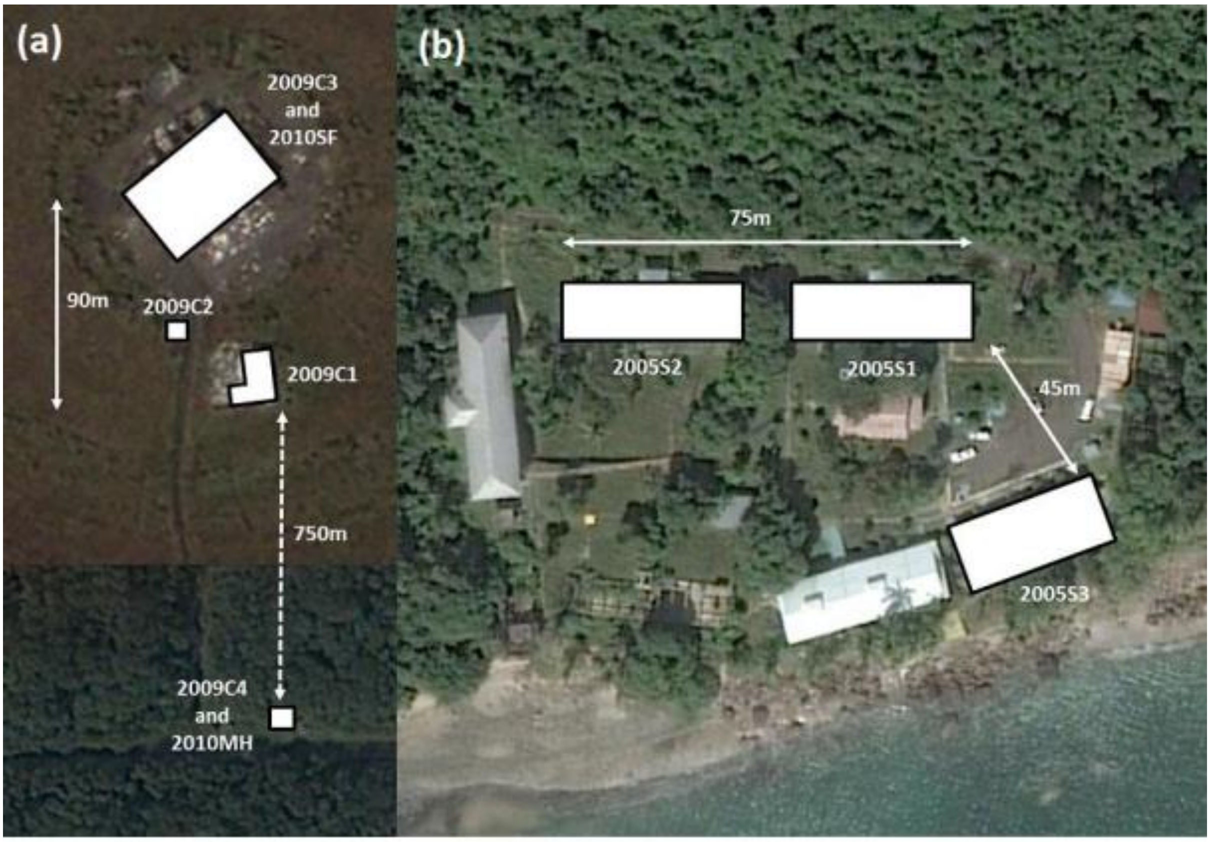
However, in some species the payoffs of inter-nest drifting are unclear; this includes paper-wasps (Birmingham et al. 2004; Sumner et al. 2007; Blacher et al. 2013). Other explanations of nest drifting in primitively eusocial insects, such as *Polistes*, where all females have the potential to lay eggs, include attempts to usurp a queen or queue as what is known as a “reproductive hopeful” (Nonacs & Reeve 1995). However, in our focal species, *P.canadensis,* females that engage in nest drifting behaviour, were all found to be reproductively inactive. It was suggested that workers appeared to gain indirect fitness benefits by helping on several related colonies in a viscous population structure, being observed to perform helping behaviours such as delivering food (Sumner et al. 2007). This structure is created by nest-founding through a process of budding or fission, where a reproductively mature female moves a short distance from a natal nest to found a new nest, often in abandoned buildings or near human settlement (Strassmann 1981; Reeve 1981; Lengronne et al. 2012), often it is not clear whether these high-level social interactions are a general, persistent phenomenon, and whether they are driven by adaptive traits, such as fitness payoffs.

In Sumner et al. (2007), to determine whether or not drifting in paper-wasps was due to discrimination errors, a hypotheses was constructed that stated that accidental drifting would result in a relationship between drifting and the geographic distance between nests, but not between drifting and the relatedness between nests. Drifting wasps would supposedly be most likely to mistake a nest close to their natal nest as their own. If drifting was not due to discrimination errors, there would be a relationship between drifting and the relatedness between drifters and the nest they drifted to. Sumner et al. found that drifters were both more likely to visit nearby nests (Mantel test, $r = -0.442$, $p < 0.001$) and that drifters were more closely related to the adults on nests that they visited than those on nests they did not visit ($t = 3.39$, $p = 0.005$). However, the study also reported that wasps on neighbouring nests tend to be closely related (Mantel test, $r = -0.138$, $p < 0.05$); known as a viscous population structure. Given that no control for distance was made when testing relatedness in this study, and the identified correlation between distance and relatedness, it is difficult to accept the paper’s conclusions that drifting is not due to discrimination errors. The signal that wasps are drifting to nests with which they are closely related could just be a by-product of the relationship between the distance and relatedness between nests. Instead the viscous population structure selecting for drifting behaviour through indirect fitness, it may just weaken the selection on accurate nest recognition through reducing the costs. Therefore, a new study was needed to further probe the mechanisms which promote drifting between nests to try and discern whether drifting in wasps is indeed adaptive. If drifting was adaptive, we would expect to see consistent and non-random patterns of drifting, with drifting preferentially take place from nests with the smallest cost of drifting to those that would receive the largest benefit. Further, we would expect to see wasps drift to nests with which they were closely related.

5.3 Methods

Monitoring nest drifting using RFID

To investigate the factors which influence patterns of drifting in paper wasps, inter-nest drifting was monitored in a different site consisting of four abandoned buildings near Colón, Panama (insert 2 *a*) in both 2009 and 2010. This data was combined with drifting data collected in a separate field site in 2005 by Sumner et al. (2007). Each of the 2005 and 2009/2010 field sites consisted of a number of abandoned buildings which contained numerous paper-wasp nests. These collections of buildings were located far away from other similar buildings (see insert 2), which provided a means to which to set boundaries for the study population and the subsequent drifting networks produced.



Insert 2: Two separate field sites (a) the 2009 and 2010 field site situated near Punta Galeta near Colón (9°24'08.28"N, 79°52'19.41"W) and (b) the 20056 field site situated near Panama city (8°54’17.42”N, 79°34’35.41”W), Panama. These images were acquired from Lengronne (2013) who modified them from images taken from Google Earth.

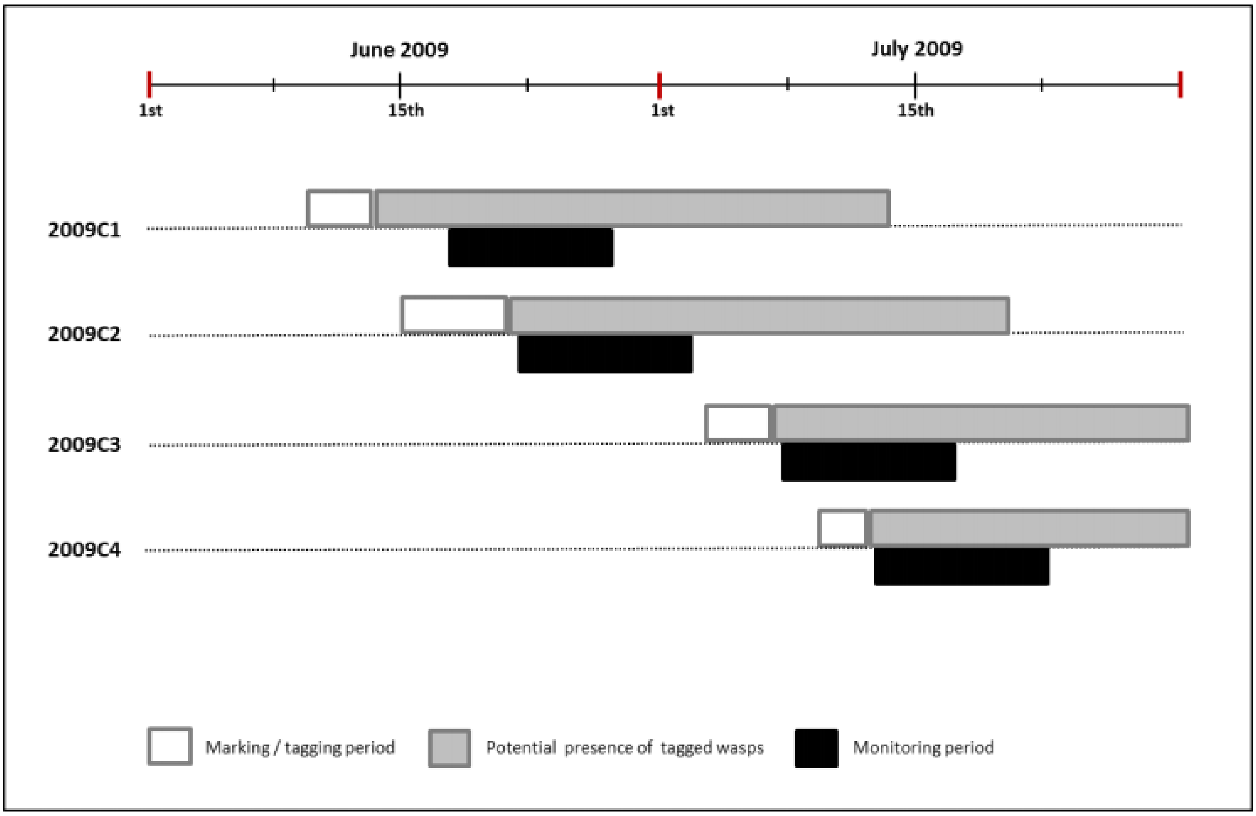
In order to remotely monitor drifting between paper-wasp nests, individuals were captured and fitted with passive integrative transponder (PIT) tags. PIT tags are cylindrically shaped, small (between 8-32 mm long, and 1-4 mm in diameter) tags which act as a unique identifier for an individual. Originally developed to track fish (see Gibbons & Andrews 2004), the size of PIT tags makes them an attractive option for tracking large insects such as paper-wasps. The small size is possible because PIT tags are dormant until activated and therefore do not require any internal source of power throughout their lifespan. To activate the tag, a low-frequency radio signal is emitted by a scanning device or an antenna that generates a close-range electromagnetic field (Smyth & Nebel 2013). Upon encountering this field, the tag sends a unique alpha-numeric code back to the reader (Keck 1994). The readers attached to antennae have an internal clock to log the date and time that different PIT tags (GiS TS-Q5Bee Tags) come within a given proximity of an antenna. PIT tags were mounted on the backs of adult paper-wasps using an adhesive as shown in insert 3.

Insert 3: A paper wasp with a PIT tag mounted on the back of the abdomen. This photo is taken from Lengronne (2013).

1599 individuals from 93 nests were tagged across the three studies. Given that paper-wasps have no fixed nest entrances due to the lack of a nest-envelope, the nests were partially enclosed in acetate sheeting, leaving one open face where the antennae of a reader were placed (insert 4). This step was taken to ensure that the antennae of readers could efficiently detect drifters, with 80% of the remaining entrance within range of detection (see Lengronne et al. 2012). Given that wasps could be detected upon entering or exiting a non-natal nest, 80% coverage would result in an estimated 96% chance of detection (using binomial probability). There was no reason to believe that the presence of missing drifting events would bias our analyses as the coverage was more or less constant across nests. Continuous automated RFID monitoring was conducted from 8am to 6pm (the main foraging period) for each day of each study period.

Insert 4: A paper wasp nest with two copper antennae placed over the nest entrance. This image is taken from Lengronne (2013).

In 2005, all 33 nests across 3 buildings in which nests were present were monitored using RFID simultaneously. In 2009, there were substantial periods of overlapping monitoring (insert 5- black coloured blocks), with the potential presence for previously tagged wasps to drift from nests in previously monitored buildings to nests currently being monitored (insert 5- indicated by grey coloured blocks. This is because pit tags were not actively removed from individuals once the study period had ended. In 2010, nests from different buildings were observed at different times with no overlap, thus drifting between buildings was impossible to detect given the study design.



Insert 5: Monitoring periods in the 2009 study. This image is taken from Lengronne (2013)

There are some additional issues with sampling drifting interactions. The orientation of entrances may favour drifting to and from certain nests, perhaps. Upon tagging individuals, the nest the individual is present on at the time of tagging is assumed to be its natal nest. However, wasps may be tagged whilst at a non-natal nests; as such the individual may then be recorded to have a very high number of supposed drifting events with a nest that is in fact its true natal nest. Further, paper-wasps were observed to sit on the antennae in front of nests, providing multiple readings. Given this, the analyses of drifting data were conducted using networks where the number of unique drifting wasps was used to weight edges as opposed to the raw number of drifting logs recorded. However, for the purposes of network visualisation, the raw number of drifting events has been used.

Visualising drifting networks

To create visualisations of networks depicting drifting between nests, the “igraph” package (Csardi & Nepusz 2006) in R (R Core Team 2013) was used. Nests formed the nodes of the networks and the edges represented the presence of drifting wasps recorded between them. Edges were directed with arrows pointing from the donor to the recipient nest and weighted based on the number of times drifting wasps were detected. Figure 7 shows the drifting networks of both the 2005 and 2009 populations where there was simultaneous or overlapping monitoring of nests between different buildings.

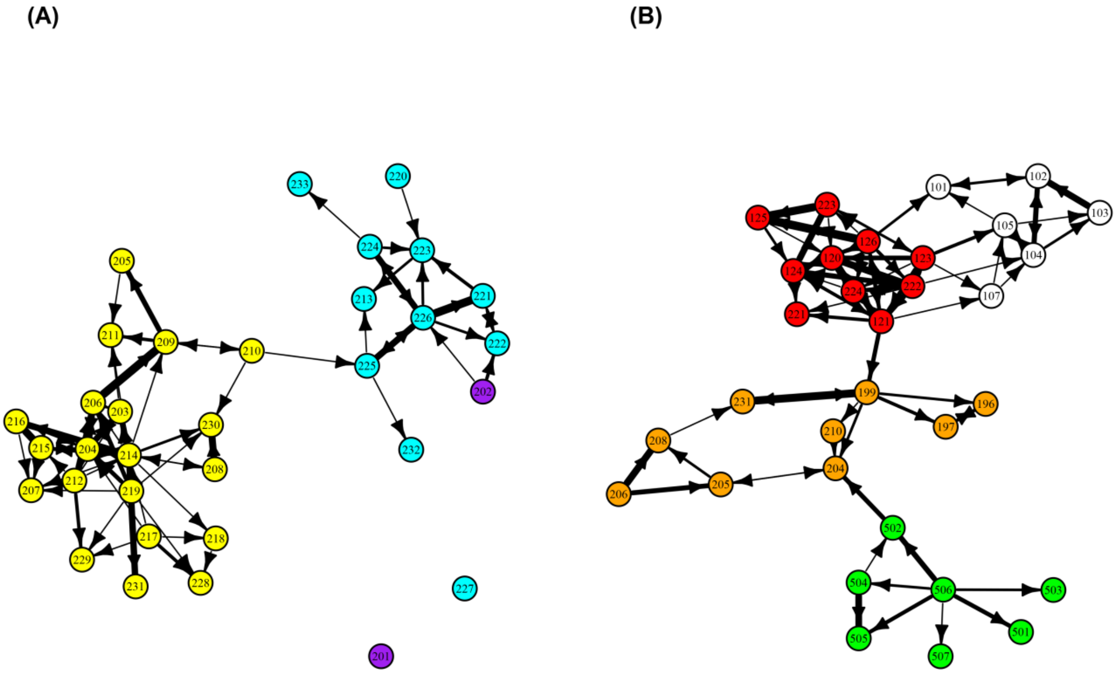


Figure 7: Networks depicting the number of drifting events recorded between nests in 2005 (A) and 2009 (B). Nodes represent nests and the edges (connecting lines) between them represent drifting between nests. Nodes are shaded to reflect the building the nest was in and edges were weighted to represent the number of drifting events recorded.

Drifting partitioned by native building

The visually striking aspect of the 2005 drifting network, where nests from different buildings were monitored simultaneously (figure 7 A), is that it is partitioned into communities on the basis of the abandoned buildings present at the study site. To test if this partitioning was statistically significant, a mixing matrix was created from a binarised version (whether there was drifting from nest A to nest B) of the network; this contained the relative proportion of edges within each building and between each pair of buildings. From the mixing matrix, a Newman's assortativity coefficient r was calculated (Newman 2003) using the R package "igraph" (Csardi & Nepusz 2006); this provided a measure of the amount of within-building drifting relative to between-building drifting.

The statistical significance of r was determined by jackknifing the network as described in (M. E. J. Newman 2003). Jackknifing provided a variance from which the standard deviation was derived. Significance was inferred through how many standard deviations the observed r value was away from 0 - the value expected if there was no partitioning of the network on the basis of building. If the observed value was greater than 2 standard deviations away from 0, it would be significant at a 95% confidence level. This test was also conducted on the 2009 data (figure 7 B) where nests from different buildings were not monitored simultaneously, but had substantial periods of overlapping monitoring. For the 2010 data, it was not possible to examine between- aggregation drifting since both 2010SF and 2010MH were monitored at 2 consecutive, but non-overlapping, time periods.

In contrast to 2005, nests were not monitored simultaneously in 2009. Given this, we also used a null model to check our results were robust to the fact that drifting in certain directions could not occur between buildings that did not have overlapping monitoring periods. We compared the trace of the directed mixing matrix Tr(e) (proportion of within-building drifting) to an ensemble of traces under the null. In each of 4999 null networks, the out-degree or number of edges leaving each nest was fixed, but the destinations of those edges were randomised within given constraints. These constraints were that edges can go both ways between C1 and C2 nests and C3 and C4 as these had periods of simultaneous monitoring. However, edges were only allowed to go from C1 or C2 to C3 or C4, as C3 and C4 were monitored later in time than C1 and C2 with no overlap, but wasps were still present with RFID tags from C1 and C2. C4 was monitored for a day when tagged wasps from C1 could have been present; given this, the large distance between these aggregations (750m) and the absence of any observed drifters between these aggregations, no drifting events were randomly allocated between these aggregations. A p-value was calculated by examining the proportion of null networks with a trace higher than the observed network.

Social differentiation

Heterogeneous and consistent drifting patterns are the expected features of drifting networks which have underlying factors which promote the occurrence of drifting as opposed to it being a random process. Social differentiation refers to one edge-weight based measure of how different a given network is compared to what is expected by chance.

If drifting within each aggregation was a random process, the resulting networks would be homogeneous, with each nest typically receiving an equal share of drifting wasps leaving each donor nest. We examined how heterogeneous the drifting networks were in each aggregation from the 2005, 2009 and 2010 studies though comparison with networks generated by a null model. For this analysis, edges were weighted representing the number of unique drifting wasps between nests. In the null model, the total number of wasps “donated” by each nest (the out-strength) was preserved, but the recipient nest for each individual wasp was randomised. Out-strength was preserved to control for the tendency of more populous nests to send more drifters (R_S=0.33, p_{perm}=0.0008). For the observed and null networks, a measure of social differentiation, S, was taken from Whitehead (2008) as a measure of heterogeneity:

$$S = \frac{\sum_{i}^{N}\sum_{j}^{N}(O_{ij} - E_{ij})^2}{N(N-1)}$$

Where O_{ij} is the observed number of unique wasps that drifted from nest i to nest j, E_{ij} is the expected number if the network was homogeneous, and N is the total number of nests in the aggregation.

Under the null hypothesis, the expected value E for each edge would be the out-strength of the donor nest divided by the number of recipient nests in the aggregation. S therefore measures how heterogeneous the network is through taking a sum of squares of the difference between the observed numbers of unique drifting wasps from one nest to another and the expected numbers if drifting patterns are homogeneous. Given that S sums the raw differences between the observed and expected number of wasps observed to drift in each dyad, aggregations with many recorded drifting events will tend to have higher values of S compared to those aggregations where drifting is less common. Therefore, S should not be used to compare how socially differentiated two or more aggregations are relative to each other, only how differentiated an aggregation is relative to its reference distribution which is generated using a null model where the incidence of drifting is preserved. If the observed aggregation produced a value of S greater than 95% of the values of S in the reference distribution (generated by measuring S from null networks), the observed network was deemed to be significantly socially differentiated. The observed network was always counted as a network which could have arisen as result of the null model, which meant that p-values could never be zero.

Temporal consistency of drifting patterns

Consistent patterns of drifting between nests over time would be indicative of stable factors that regulate drifting and infer that the snapshots of these systems that we have observed are generalisable. To identify if there were fewer transient and a greater number of consistent or recurring edges than we would expect if wasps drifted without preference, the number of edges that were present between the same nests for greater or equal to one, two, three and all four of the observation periods were measured. Data collected from the 2005 population was used, which consisted of two main nest aggregations (2005S1, 2005S2) monitored simultaneously for drifting over four consecutive periods of five days, although the spatial structure within aggregations may still shape the non-randomness of the system.

The significance of the number of edges recorded at each of four consistency thresholds was determined via comparison with an ensemble of networks generated by a null model. For each observation period, edges were randomly allocated between nests, preserving the total number of edges found within and between aggregations, whilst also preserving the out-degree of each nest. The number of transient and recurring edges in each resulting null network was measured. This process was repeated many times to generate a distribution of null frequencies for each consistency threshold with which the observed figures could be assessed for significance. A p-value was derived based on the proportion of null networks which had an equal or more extreme number of edges at the given threshold than the observed network.

Factors which might drive patterns of drifting

Given the seemingly non-random and consistent nature of drifting between *P.canadensis* nests, the next step taken was to determine if drifting could be modelled in terms of relevant explanatory factors, in particular those pertaining to nest need for help.

In addition to the native building of nests, the size of individual nests combined with the distance, relatedness, and difference in worker-brood ratios between nests were suspected to influence patterns of drifting between nests. If drifting was an adaptive behaviour, drifters may direct their help not only to close nests, with which they are also related, but also those with the greatest need for help, having fewer workers to look after brood or being smaller in size and more liable to perturbation events.

Censuses of the numbers of wasps were performed every 3 days at night to estimate the group size (number of adults) and number of brood; these were used to provide a weekly estimate of nest size and worker to brood ratio for each nest in each aggregation. Wasps from both sites (Panamá City and Colón) were collected at the end of the experiment for molecular analyses (except for 2010). Samples from 2009 were genotyped for estimating relatedness using 7 specific markers (Pcan01, Pcan05, Pcan09, Pcan15, Pcan16, Pcan23 and Pcan24 (Lengronne et al. 2012)). Relatedness was calculated by Lengronne using the program RELATEDNESS 5.0.8 and weighting nests equally (Goodnight & Queller 1999). Standard errors were estimated by jackknifing over loci. Relatedness data for 2005 dataset was obtained from (Krieger et al. 2000).

The number of adult wasps which belonged to a given nest was taken as its size. The difference in worker-brood ratios between nests was calculated as the difference in the number of workers divided by the number of brood for each nest in the dyad - this was a directional measure. Relatedness data was not available for the 2010 data set and some missing data were present in 2009. As missing data poses a considerable problem for many network modelling approaches (James et al. 2009), distance was used as a proxy for relatedness given both the significant inverse correlation between the two (R_S= -0.554, p_{perm}=0.001, n=290) and the findings of Sumner et al. (2007). This correlation was identified through comparing the relatedness and distance between pairs of nests for which there was complete data. To determine the significance of the coefficient, the relatedness values between nests were permuted within each aggregation, holding the position of nests constant; this was repeated 4999 times. The distribution of null correlation coefficients produced as the product of this null model was used to determine significance.

Experimental manipulations of the nests were carried out by Sumner and Lengronne to determine which cues were used by drifters when choosing nests to visit; these manipulations were carried out prior to any of the analyses regarding the factors associated with drifting presented in this chapter. The two likely variables that Sumner and Lengronne suspected would affect patterns of drifting were: the distance between nests, group size and need for help (brood number). It was impossible for them to re-locate nests with wasps within the home range of a nest, as any re-located wasps will re-orientate back to their original nest location (Sumner pers. obs.). They therefore focused on manipulating group size (i.e. number of female adults) and need for help (i.e. brood number). To determine the effect of group size on drifting, they permanently removed 30% of the foragers on 14 nests (3 nests in 2009 (1 in 2009C1, 1 in 2009C2 and 1 in 2009C3) and 11 in 2010 (6 in 2010SF and 5 in 2010MH)). To determine the effect of brood number on drifting they permanently removed 30% of the brood on 9 nests (3 nests in 2009 (1 in 2009C1, 1 in 2009C2 and 1 in 2009C3) and 6 in 2010 (6 in 2010MH)). Large brood (at least 60% of medium and large larvae) were preferentially removed as they represent the most valuable brood, and require the greatest helping effort to rear. Cells that had contained the removed brood were also removed to prevent wasps perceiving empty cells as a decrease in “queen quality” or nest quality.

To identify which of the explanatory variables served to be significant predictors of drifting, a logistic Multiple Regression Quadratic Assignment Procedure (MRQAP) was used (Krackhardt 1988; Butts & Carley 2001) where the drifting network was the response variable with distance, nest-size and worker-brood ratio the explanatory variables of interest. A logistic model was chosen over a linear model due to the sparseness of drifting data, which also exhibited little variation in non-zero numbers. The effectiveness of the MRQAP has been shown to degrade under conditions of extreme skewness (Dekker et al. 2007). To control for an identified tendency for larger nests to send more drifters, the nest size of both the donor and recipient nest was included as explanatory variables in the model. Due to the various kinds of dependency inherent in network data (Snijders 2011), the significance of each individual factor and the model as a whole was determined using a randomisation procedure. Due to a slight, but significant correlation between a nest’s size and its difference in worker-brood ratio with other nests ($R_S=\pm0.18$, $p_{perm}<0.0002$), the randomisation procedure chosen was a Double Semi Partialling Quadratic Assignment Procedure (QAP-DSP), which has been demonstrated to be robust to the effects of autocorrelation and co-linearity when determining significance (Dekker et al. 2007). QAP-DSP was a departure from the out-strength constrained edge randomisations used thus far. However, the QAP-DSP approach is well established for use in modelling networks. The inclusion of nest-size as a variable in the model not only allowed its effect on drifting to be characterised, but also compensated for the lack of this constraint in the QAP-DSP. The same analysis was conducted on the drifting data following the manipulation experiments (excluding the 2005 aggregations) as that applied to the pre-manipulation data. In this dataset, there was a stronger correlation between a nest's size and the difference in worker brood ratio with other nests ($R_S=\pm0.25$, $p_{perm}=0.0002$).

The use of a QAP-DSP null model in response to an observed correlation between explanatory variables of at most 0.25 may seem overly-conservative. Indeed, correlation does not necessarily imply collinearity (Alin 2010). However, given that the wasp drifting network can be assumed to be evenly-sampled, we believe little is lost using a null model common to the social sciences as opposed to using the same null model where the recipient nests of unique drifting wasps are randomised. This second choice of null model might behave as a typical response variable permutation as outlined in Dekker et al. (2007), in which case we would expect that any correlation between the response variable and an explanatory variable caused by the presence of a third collinear explanatory variable would lead to a high rate of type 1 error. However, this null procedure is not a QAP as in Dekker et al. (2007) and may therefore behave differently.

The logistic MRQAP "netlogit" function in the "sna" (Butts 2014) package in R(R Development Core Team. 2008) was modified to include stratification, or a block structure; this was so that data from all aggregations across all years could be included into one model whilst partitioning the QAP-DSP to take place only within blocks that could exchange wasps and not across blocks that could not interchange wasps due to substantial separation in distance and/or time. Through including all aggregations in one model in this way, each aggregation was treated as independent from each other in some respect and the analysis was restricted just to look at drifting within-aggregation. This step was justified given the strong partitioning of the wasp drifting networks on the basis of building in the 2005 and 2009 data. However, one limitation of partitioning the randomisation approach underlying the model is that aggregation-wide effects cannot be tested.

A significant advance in our approach is through the inclusion of a block-structure into the MRQAP, without which alternatives might include having to construct a unique model for each of the eight aggregations and combining p-values using a Fisher's combination test (Fisher 1948) or the weighted-Z method (Whitlock 2005), for example. Two severe drawbacks of this approach are as follows: firstly, maximum likelihood estimation (used in the logistic MRQAP) requires more data than other approaches, such as Ordinary Least Squares (OLS), for accurate fitting of the model- a general rule of thumb being that at least ten events (the least frequent category in the dependent variable i.e. drifting present from nest *a* to *b*) are required per variable (Peduzzi et al. 1996). Splitting the dataset into aggregations reduces the number of events per variable considerably, with nearly all aggregations having too few edges (events) for the number of variables we include in the model. Secondly, all of the aforementioned combination tests assume each "test" (aggregation-based model in this case) is independent of others. Given that some study sites are used repeatedly in 2009 and 2010, with the potential for the same physical nests to be present in both study periods, such an assumption of independent tests might be questionable. The results of the analyses for the various factors thought to affect drifting are shown in table 8.

5.4 Results

Drifting partitioned by native building

The null model identified that the drifting networks of wasps were significantly partitioned (p=0.0002) on the basis of native building with an observed Tr(e)= 0.9125, where the mean null Tr(e) was 0.3367175.

Jacknifing the observed networks also confirmed that both the networks in 2005 and 2009 were significantly partitioned on the basis of buildings into what we describe as "aggregations", with significant assortment with respect to building identity (Newman's assortativity coefficient r for both 2005 (r=0.893) and 2009 (r=0.867); jackknifing revealed assortativity values to be 15.1 and 17.8 standard deviations away from 0 respectively). This partitioned the data into eight different aggregations of nests (two aggregations in the 2005 population (2005S1 and 2005S2) (Building 3 had only two nests), four aggregations in the 2009 population (2009C1 to 2009C4), and a further two aggregations in 2010, (2010SF and 2010MH). The number of nests per aggregation ranged from six to 20 (11.3±1.7, mean± standard error). The mean distance between nests within an aggregation (7.7±0.3m) was significantly lower than the distance between aggregations (298±127m). Inter-nest drifting was detected in all eight aggregations (pre-manipulation data from 2009C1 to 2009C4, 2010SF, 2010MH, 2005S1 and 2005S2). On average 40.4±3.9% of the wasps (n=403 wasps, 93 nests) were drifters. At the nest level, 92.7±0.1% (66.7-100%) of nests in each aggregation received or produced drifters (85.8±0.2% of the nests received and 71.1±0.3% of these nests produced drifters), indicating that drifting is a general phenomenon and not restricted to particular nests. 19.3% of nests show extreme levels of drifting, with >60% of records from drifters. Finally, we found rare interactions between aggregations within populations. In each case these inter-aggregation edges were attributed to a single drifter which visited only one nest in the other aggregation (i.e. unidirectional and never reciprocated), albeit multiple times (2005: 1.7±0.7 visits (1-3 visits per drifter); 2009: 3.0±0.9 visits (1-6 visits per drifter)). Weak links are known to play potentially important roles in social networks, providing information flow between otherwise segregated populations (Granovetter 1973).

The partitioning of drifting networks into aggregations creates at least three possible levels of social organisation (figure 1). The lowest level was the nest. Most wasps leaving a nest returned to the same nest; these nest departures were not drifting events. A median of 96% of nest departures return directly to the natal nest; this was 100% in a third of nests. Only 25% of nests had fewer than 70% of departures return directly to the natal nest. These wasps interact with close relatives (r (adult natal wasps)=0.69±0.02; n=129 wasps from 26 nests, from the same 2009 study population (Lengronne et al. 2012). The second level was between nests forming aggregations (figure 2) and third, identified in this study, between aggregations in the population. Social organisation reaches beyond the unit of the immediate family group. To our knowledge, our data are the first quantitative evidence of interactions across more than two levels of social organisation in a eusocial insect. Higher-order interactions influence lower-order interactions, such that the traditional 'social unit' (i.e. the nest) is not independent of neighbouring units.

Social differentiation

Within seven of the eight aggregations, the patterns of drifting observed were more heterogeneous than expected if drifting was a random process (i.e. the observed social differentiation S was significantly greater than expected under a null model; Table 6; figure 2). The heterogeneous patterns of drifting suggest that drifting is unlikely to be the result of accidental events due to discrimination errors. However, the underlying heterogeneity could still be driven by the spatial structure within nest aggregations and as such, discrimination errors can still not be ruled out entirely.

Table 6: Heterogeneity of drifting in each aggregation. “Mean null S” is the mean social differentiation observed across 4999 null networks. P-values state the probability that a value of S equal to or higher than the observed could be produced by the null model.

Aggregation	**Mean null S**	**Observed S**	**P-value**
2005S1	0.067	0.108	0.002
2005S2	0.196	1.111	<0.0002
2009C1	0.955	9.560	<0.0002
2009C2	0.741	1.881	<0.0002
2009C3	0.354	0.988	<0.0002
2009C4	0.295	0.385	0.1884
2010SF	0.469	1.063	<0.0002
2010MH	0.186	0.544	<0.0002

Temporal consistency of drifting patterns

The non-random, persistent nature of interactions between nests suggests there may be some adaptive significance to drifting. There was strong evidence of temporal persistency in the patterns of drifting between nests, with fewer transient and more persistent and recurring links than we would expect if drifting were random or performed only fleetingly (Table 7). Over the four consecutive monitoring periods, six pairs of nests had persistent edges (8.2% of observed edges). We also found 12 edges (16% of observed edges) that were persistent for at least three of the four 5-day monitoring periods and 22 edges (30% of observed edges) that were persistent for at least two of the four monitoring periods. The remaining 51 edges were more transient with detections occurring during only one monitoring period. There were significantly fewer transient edges and significantly more recurring edges at each threshold than we would expect compared to null versions of the dataset (Table 7). Persistent edges correspond to a higher level of drifting between nests than transient edges, with on average 4.3±0.7 drifting events in any pair of nests (compared with 2.4±0.6 for more transient edges; Mann-Whitney U, W= 2236, p_{perm}<0.0002), and up to 31 drifting events for a pair of nests in a single monitoring period (max value=31 drifting events for persistent edges and 18 drifting events for transient edges).

Table 7: Drifting rates are consistent over time. “Observed (cumulative)” shows the number of edges observed to meet each of the four consistency thresholds. “Mean null” shows the mean number of edges which met each threshold taken from 4999 null versions of the dataset, with standard deviation included “σ null”. “p-value” represents the proportion of null datasets produced with equal to or more edges meeting the given threshold than observed. [1] represents the proportion of null datasets produced with equal to or fewer transient edges than observed.

Number of observation periods	Observed (cumulative)	Mean null	σ null	p-value
>=1	73	96.66	2.74	<0.0002[1]
>=2	22	16.52	2.48	0.0284
>=3	12	2.44	1.26	<0.0002
=4	6	0.17	0.41	<0.0002

Factors which might drive patterns of drifting

The output of the logistic MRQAP suggested that a number of factors influence drifting in *P.canadensis* within aggregations. Firstly, drifters were more likely to drift from larger nests (Table 8, p=0.003). In general, pay-off per unit of helping effort diminishes with group size in eusocial insects (Michener 1964; Krieger et al. 2000); thus drifters appear to be maximising the value of their helping effort by drifting away from large natal nests in order to help out other nests, where their efforts have higher per unit payoffs. There was no significant relationship between the size of the recipient nest and its tendency to receive drifters (Table 8).

Secondly, drifters were more likely to drift from nests with higher worker-brood ratios to nests with a smaller worker-brood ratio (Table 8, p=0.046). Group size (number of wasps) is not necessarily indicative of brood number, and so if drifting is a strategy to increase fitness by responding to need for help, wasps are expected to visit nests where there are few workers relative to brood. Our analyses were restricted to medium and large larvae, which are fed the bulk of the forage, indicating that drifters appear to be responding to provisioning needs rather than any other form of help.

Finally, drifting was more likely to take place between nests that were close together than far apart (Table 8, p<0.0002). Nests in aggregations show significant genetic viscosity such that nests close to each other are more closely related than nests further away (Sumner et al. 2007). In support of this, we identified a strong inverse correlation in the distance between nests and their relatedness (R_S=-0.554, p_{perm}=0.001, n=290). Wasps are therefore drifting to closely related nests (R_S=0.25±0.08 (Sumner et al. 2007)), where the indirect fitness benefits of helping are greatest and they appear to use nest proximity as a reliable cue for where to drift, however, to confirm this a manipulative experiment focussing on altering the locations of nests would be necessary.

Table 8: The results of the logistic model on the pre-manipulation drifting data. Pr(<=β) is used to determine the significance of factors thought to inversely affect the likelihood of drifting, whereas Pr(>=β) is used for factors which are thought to have a positive effect on the likelihood of drifting. Exp(β) is the exponent of the estimated beta for the given factor. The bold font indicates the p-values linked to our given hypotheses.

Coefficients	**Estimate**	**Exp(β)**	**Pr(<=β)**	**Pr(>=β)**
(intercept)	-0.609	0.544	1.0000	0.0000
Worker-brood ratio	1.394	0.248	**0.0464**	0.9536
Distance	-0.003	0.997	**0.0000**	1.0000
Donor nest size	0.017	1.017	0.9966	**0.0034**
Recipient nest size	-0.011	0.989	**0.6004**	0.3996

Lengronne and Sumner experimentally manipulating the worker: brood ratios of nests to test the importance of nest proximity and group size as mechanistic cues to workers for drifting. The same logistic MRQAP procedure was performed on the drifting data recorded in the five days following these manipulations. As shown in table 9, drifters did not respond to the change in worker-brood ratio (p=0.998). However, distance remained a significant predictor of drifting in the network (p=0.013) as did the tendency for larger nests to send more drifters (p=0.0492). Interestingly, small nests became significantly more likely to receive wasps despite no significant effect being recorded before the manipulation (Table 9).

Table 9: The results of the logistic model on the post-manipulation drifting data.

Coefficients	**Estimate**	**Exp(β)**	**Pr(<=β)**	**Pr(>=β)**
(intercept)	-0.896	-0.025	0.0000	1.0000
Worker-brood ratio	-0.113	0.893	**0.9984**	0.0016
Distance	-0.003	0.997	**0.0126**	0.9874
Donor nest size	0.023	1.023	0.9508	**0.0492**
Recipient nest size	-0.025	0.975	**0.0028**	0.9972

5.5 Discussion

Social organisation

The findings presented in this chapter reveal social organisation in *P.canadensis* similar to that observed in other (non-insect) animal societies, providing the first quantitative evidence of interactions across more than two levels of social organisation in a eusocial insect. Higher-order interactions are likely to influence lower-order interactions, such that the traditional ‘social unit’ (i.e. the nest) is not independent of neighbouring units. Interactions across multiple levels of social organisation have previously been identified in a diverse range of group-living vertebrates, for example in the elephant *Loxodonta* *africana (Wittemyer et al. 2005),* gelada *Theropithecus gelada (Dunbar & Dunbar 1975),* hamadryas (Kummer 1968), orca *Orcinus orca (Ford et al. 2000),* Galapagos sea lions *Zalophus wollebaeki* (M. Wolf et al. 2007), noctule bats *Nyctalus lasiopterus (An 2008),* Colombian ground squirrels *Spermophilus columbianus (Manno 2008)* and white-fronted bee-eater birds *Merops bullockoides (Hegner et al. 1982).*

Interestingly, the analyses in this chapter reveals an intriguing contrast between social organisation in social vertebrates and eusocial insect societies. Hill et al. (2008) noted that the sizes of hierarchical social groups of several species of mammal exhibit a common scaling factor, each social layer of a society being between three and four times the size of the preceding (smaller) grouping level or "Horton order". Hill et al (2008) proposed that this general scaling factor may reveal how societies originated. The relatively low value of the scaling factor (3-4) may be due to cognitive constraints (e.g. animals can only manage a small number of interactions at a given level (Kudo & Dunbar 2001); or social time constraints (e.g. the 'handling time' needed to maintain relationships (Dunbar 1996)). In *P. canadensis,* the mean scaling factor was 8.4, with rather more variation between levels than Hill *et al.* (2008) found for mammalian societies (size ratios 16:14:2:1 (2005 data: 420 individuals in 27 nests in two aggregations in one population), ~20:8:4:1. (2009 data: 665 individuals in 32 nests in four aggregations in one population) and 536:28:2:1 (2010 data; 536 individuals in 28 nests in two aggregations in one population (figure 8)). These comparisons suggest that the common patterns of social organisation in social mammals may not be generalised to more complex societies, like eusocial insects. A possible explanation for this may be that social organisation in eusocial insects is not constrained in the same way as it is in social vertebrates. For example, eusocial insects have evolved complex modes of communication (e.g. queen pheromones that regulate worker reproduction) which means that social time constraints are less relevant. Moreover, the evolution of decentralised, self-organised processes which constitute a defining feature of eusocial insects (Krieger et al. 2000) may also enable them to circumvent cognitive constraints that limit the complexity of animal societies. Figure 8 shows the scaling between successive levels of organisation in *P.canadensis* (black) and orca *Orcinus orca* and Elephants *Loxodonta Africana* as reported in Hill et al. (2008).

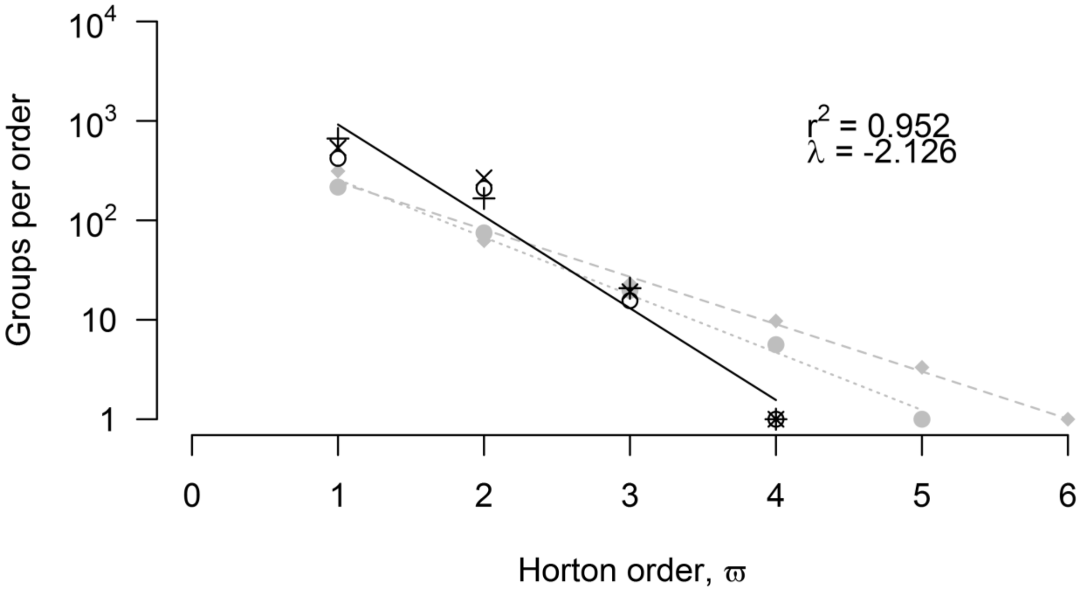


Figure 8. The mean number of individual wasps in successive social layers or Horton orders, from sub-population (1), aggregation (2), nest (3) to individual (4). Open circles depict data from 2005; crosses, 2009 and diagonal crosses, 2010. The lambda of the line of best fit has been calculated using the natural log of the three wasp data-sets combined, and is shown in black. Included in grey are examples from Hill et al. (2008), elephants are shown as filled diamonds and orcas as filled circles. The lambda for the line of best fit for the elephants (dashed) is -1.096 and for the orcas (dotted) -1.334.

Factors which explain drifting patterns

The multivariate analysis of the drifting network revealed that the social organisation observed in paper-wasps is the result of inter-nest drifting that either confers a fitness advantage to individuals, or has such a low cost that any selection pressure against it is negligible. The response in drifting patterns observed following the manipulative experiment, albeit not in the manner hypothesised, supports the notion that drifting is on the whole not accidental, being the result of discrimination error.

The pre and post-manipulation analyses indicate that wasps are using simple cues in deciding which nests to interact with, and that this mechanism is driven primarily by nest proximity but also group size. Nest location is fixed, and so this cue is likely to be a first order rule that over-rules any other cues. Group size changes steadily as nest size (number of cells) increases, presenting an easily perceived visual cue for need for drifting. Brood number and composition on the other hand changes rapidly over time as nests move through the colony cycle, and requires careful inspection of cells to determine brood stage or absence of brood. The significant effect of worker: brood ratios on drifting effort observed before the perturbation may reflect an equilibrium state that is achieved over time frames longer than our post-manipulation monitoring period of 5 days. Drifters respond to rapid changes in group size, but not brood number. Further, the removal of workers from nests may have elicited an alarm response, being viewed as an attack by the wasps. *P.canadensis* venom is known to act as an alarm pheromone which stimulates alarm behaviour wing buzzing and jerky movements which may attract other wasps (Eberhard & West 1969; Sledge et al. 1999) directly attracts other members and lowers the threshold for attack (Richards 1971; Jeanne 1982; Bruschini et al. 2006). Over such short time frames, these cues may still be present and skewing drifting behaviour, although more focussed investigations would be needed to confirm this.

Admittedly, there are a number of other statistical approaches which could have been applied to hypotheses three and four (the explanatory factors that explain drifting pre and post manipulation), such as ERGMs (Holland & Leinhardt 1981), which now have the ability to be used on data with block structures (Hunter et al. 2013) and multilevel P2 models (Zijlstra et al. 2006). An MRQAP based analytical approach was chosen as the QAP-DSP null model used is known to be robust under a number of data-conditions such as skewness and collinearity is present. Furthermore, at this point of time, we were only interested in explaining the external explanatory factors which may affect drifting to resolve hypotheses three and four (such as the distance between nests) whilst controlling for network dependencies, not the means in which network dependencies such as reciprocation or triad closure may affect network structure. However, that is not to say that network dependencies in drifting networks such as reciprocation would not be biologically meaningful - reciprocity is one of the key tenets underlying cooperative and altruistic behaviours (Trivers 1971). The main advantage of the MRQAP is perhaps pragmatic. In the “sna” package (Butts 2014) in R (R Core Team 2013), the randomisation procedure underlying the MRQAP can be readily extracted and, with care, modified to address issues common in animal social network studies.

Overall Conclusions

Overall, the analyses in this chapter has revealed that social organisation in eusocial insects can be highly complex, functioning at different interaction levels and that these high-level interactions are likely to be adaptive, conferring a fitness advantage to individuals. Using an automated real-time monitoring system of individually tagged paper wasps, the interactions within and between levels of societal living were quantified to reveal three novel insights into the importance of multi-level social interactions in a eusocial insect. Four levels of social organisation were identified, namely the individual, nest, aggregation, and community (figure 8). High-level interactions (e.g. at levels above the family group/nest) were found to be highly structured and non-random. Furthermore, drifters appeared to apportion their help in respect to the needs of the recipient nest for help, and that the mechanism for ‘where to drift’ is likely to be nest proximity and group size rather than any fine-scale fluctuations for need on individual nests.

The presence of drifting in *P.canadensis* is not surprising when we consider the concept of inclusive fitness and Hamilton’s rule: that altruistic behaviours can exist when relatedness multiplied by benefit is greater than the cost of the behaviour. In *P. canadensis,* drifters help raise brood on the nests they visit (Sumner et al. 2007), and this tends to be nests close-by with which they have a high relatedness due to the spatial viscosity created through the “budding” process of nest founding. This close distance potentially also facilitates a reduction in the cost of drifting both in terms of risk of predation, metabolic costs associated with the activity, and the loss of help for the native nest. The apportioning of drifting as to help nests with the greatest level of need, maximises the fitness benefits to recipients of drifters and minimises the relative fitness costs to nests which "donate" drifters, which tend to have more workers to provision brood.

The wider implications of the results presented in this chapter is that drifting behaviour is not necessarily maladaptive occurring by mistake, or a selfish behaviour driven by "reproductive hopefuls" undermining the stability and persistence of eusocial societies. Instead, it appears drifting may provide a system-wide fitness benefits through directing help at related nests which need it most.

6. Variable performance of radio proximity loggers: implications for multivariate analyses.

6.1 Abstract

The use of animal-borne tracking technologies has facilitated the remote collection of animal network data. The datasets produced using such methods can be rich and free from many of the biases associated with directly observing the interactions or use-of-space of free-living animals. However, these technologies also introduce new sampling issues and biases which, if not controlled for, will affect network structure and bias subsequent analyses. Using a herd of dairy cattle as a case study, issues regarding the variable performance of Sirtrack® radio-proximity loggers are highlighted and effects characterised on both static and temporal networks. Two correction procedures are presented for both static and temporal networks to address biases in the data. The results of this chapter highlight the need for appropriate pre-deployment testing of technologies used to automatically gather network data and associated corrections to the data before the point of modelling.

6.2 Introduction

Over the last decade, there has been a transition in the way that animal social network data are collected. This has been facilitated by the miniaturisation of tracking technologies to the point where devices can be mounted upon animals directly to monitor their use-of-space (Ropert-Coudert & Wilson 2005; Krause et al. 2011; Krause et al. 2013). Collecting spatial-data remotely provides a significant advance over "traditional" methods of directly observing associations between animals (Krause et al. 2013). Animal-borne remote-sensing devices can produce datasets which are more reliable, contain less uncertainty, and are more rich than directly-observed data; this has led some believe that these datasets are suitable for multivariate modelling (Krause et al. 2013; Pinter-Wollman, Hobson, et al. 2014). However, a number of issues regarding the performance of some types of spatial-proximity logging devices have been reported (Prange et al. 2006; Watson-Haigh et al. 2012; Drewe et al. 2012; Rutz et al. 2015; Boyland et al. 2013; Bettaney et al. 2015). In this chapter, we add to these issues through the discovery of persistent logging biases in Sirtrack® radio-proximity loggers. We show that these biases can affect both static and temporal network structure. Two corrections are proposed to correct for the biased sampling of loggers so that the data collected can be used in either a static network model, such as the MRQAP (Krackhardt 1988), or a temporal model, such as the relational events model (Butts 2008a ; Patison et al. 2015).

In traditional animal social network studies, humans observe associations between animals directly. There can be advantages with using direct observation (DO) to identify associations. The nature of an association and the context in which it took place can be recorded in detail. DO can allow the researcher distinguish whether associations were socially motivated, or driven by ecological constraints. Furthermore, animals may not need to be captured if readily identifiable (Croft et al. 2008). However, the advantages of DO are often outweighed by substantial drawbacks. Firstly, DO can be expensive and labour-intensive. The number of undirected relationships that can exist between n individuals is $0.5(n^2-n)$, a large amount of data is thus required to determine whether each of these relationships exists. The presence of a human observer may also disturb the natural behaviour of the animals being observed; the observer may, for example, be treated as a potential predator (Frid & Dill 2002; Beale & Monaghan 2004; Krause et al. 2011). Species which live in hostile habitats for humans or are hard to individually identify may also present challenges to using DO (Drewe & Perkins 2014). Further, it is often only practical to carry out DO opportunistically, or at intervals, so social associations and confounding events can be missed if they occur outside of an observation period (Croft et al. 2008). Therefore, a concern with traditional studies of animal social dynamics is whether the quantity and quality of field data is sufficient and can adequately describe population-level processes (Krause et al. 2014). Furthermore, most traditional studies identify patterns from data collected over long timeframes through repeated observations, ignoring short-term temporal dynamics (Croft et al. 2008; Krause et al. 2013).

Remote-sensing (RS) refers to a set of techniques used to acquire information about an object or phenomenon without having to observe it directly. In this sense, RS has the potential to remedy many of the issues associated with DO. RS could be used to describe instruments set up in field-sites to automatically collect climatic data (e.g. Gillies & Carlson 1995), the analysis of satellite imagery to determine habitat types and change (e.g. Skole & Tucker 1993), or the use of devices to monitor anthropogenic or animal activity in an area, an example being the use of passive sonar devices to measure oceanic traffic or the songs of cetaceans (e.g. Zimmer et al. 2003; Ogden et al. 2011). Animal-attached RS, or bio-logging devices allow various types of data to be collected from an animal's perspective, without the presence of a human observer. Animal-borne devices have been around for a long time; perhaps the first example of the use of animal-borne RS was Eliassen (1960), who used a device that transmitted the heart and wing-beat rates of ducks. The increasing sophistication and miniaturisation of animal-borne devices have extended their use to a wide range of species to measure a multitude of factors with high temporal resolution. Many early uses of animal-borne RS devices focussed on physiological traits or the location of animals (Ropert-Coudert & Wilson 2005), but have since been used to record data about an individual's external abiotic environment, i.e. the medium through which the animal is moving, rather than simply concentrating on the carrier itself (Boehlert et al. 2001). One of the latest uses of animal-borne RS technologies is to measure the social environment of animals through identifying social relationships though means of association via spatial coincidence (use of same area) or proximity (Krause et al. 2011; Krause et al. 2013).

The RS of animal spatial associations in terrestrial species is often, but not exclusively, conducted using any of the three following tools: Global Position System (GPS) trackers, video surveillance and radio-proximity loggers. In aquatic species, acoustic systems tend to be used over radio based systems due to more efficient signal propagation compared to radio-waves (Krause et al. 2013). GPS-based systems have the advantage of providing information on not only the distance between individuals, which can be used to infer relationships between them, but also where individuals are in space. Given that the focal individuals can first be captured to fit loggers, GPS systems are especially suited to capturing movement data on long-ranged or cryptic species. One limitation of GPS trackers is that they have a point accuracy of, at best, 3.5m (U.S. Air Force 2014) in unobstructed terrain; this restricts their applicability to certain collections of organisms where socially meaningful distances between individuals may be in the tens of metres (e.g. Haddadi et al. 2011). Video-recording of animal social interactions has the advantages of being unobtrusive; especially if individuals do not require marking. Video footage can provide detailed information as to the precise nature and context of an interaction (e.g. was it aggressive? did it take place whilst feeding?). Automated video tracking and identification of animal contact interactions have become increasingly prominent in laboratory studies of insects (e.g. Otterstatter & Thomson 2007; Mersch et al. 2013). However, unless video cameras are affixed directly to individuals in wild populations of animals (perhaps in combination with GPS trackers), the location of the animals needs to be predicted ahead of time so that cameras can be set-up in suitable locations; this restricts the applicability of these technologies to animals with predictable movement patterns. Further, the analysis of videos can be labour intensive for free-living populations of animals (Drewe & Perkins 2014). Radio-based systems are perhaps becoming the most popular tool to study free-living populations of animals which inhabit predictable geographic areas (e.g. Böhm et al. 2008; Hamede et al. 2009; Rutz et al. 2012).

The use of radio frequency identification (RFID) to identify social interactions can be broken down into “passive” tag-to-reader and “active” tag-to-tag systems. In passive systems, the tags affixed to animals can be very small (8-32mm long, 1mm to 4mm diameter (Smyth & Nebel 2013)) as the contents of the tag consist of only an integrated circuit chip, a capacitor, and an antenna coil, which may be encased in glass (Roussel et al. 2000). The small size of these tags means they can be fitted to a very wide number of species, including those too small for active RFID devices (see chapter 5 for an example). Upon a tag passing through the antenna of a fixed reader, a low-frequency radio signal emitted by the reader generates a close-range electromagnetic field. The field excites the tag which broadcasts a unique signal back to the reader (Keck 1994) which is logged along with the time of the interaction. PIT tags thus have the advantage of not requiring a battery, a limitation of other technologies. However, the data captured using PIT tags is limited to spatial coincidence, where tagged animals are read by the same fixed reader at a given site. In chapter 5, we used PIT tags to study the movement of paper-wasps *Polistes canadensis* between nests.

In active RFID systems the predictable decay of radio signals from a point source is exploited to infer the distance between loggers. Loggers both emit a unique radio signal whilst being open to receive the signal of other loggers, the strength of a signal received from another logger can be stored explicitly or binarised to represent a contact when a given threshold strength has been exceeded. The strength of the transceiver (a radio transmitter and receiver) in the loggers can be adjusted to change the distance required to initiate and terminate contact logs. In active RFID systems, instances of close proximity between individuals can be captured in real-time, as loggers contain their own internal clocks. The advantages of these systems are that they can provide high resolution spatial and temporal data, with the timing and duration of contacts recorded. However, there are drawbacks associated with RFID based technologies; animals must first be captured to fit tags or loggers which may also require retrieval to collect data and RFID technologies (in general) cannot provide information regarding the nature or reasons behind a contact event (Krause et al. 2014).

The active RFID technology we use in this chapter to remotely sense close proximity between animals is the spatial proximity loggers manufactured by Sirtrack Tracking Solutions® (Havelock North, New Zealand). These loggers can be attached to individuals via an ear-tag, collar, directly using glue, or alternatively can be placed in the field to monitor use of space or aggregation around a particular resource. The loggers contain a radio-transceiver, which transmits and receives radio-waves between 148 to 174 MHz in frequency, an internal clock, and internal memory capable of recording up to 17000 logs, each up to 65535 seconds. Each logger pulses a unique signal up to 40 times per minute that identifies itself to other loggers, whilst simultaneously being open to receive the unique signal of any other loggers. A contact begins when a logger detects a signal from another logger within a proximity chosen by adjusting the strength of the transceiver. The contact ends upon losing that signal for user-controlled amount of time between 1 and 255 seconds (we call this the “separation time”); at which point, the start-time, contact duration and identification of the encountered logger are recorded (we call this a “contact log”). To adjust detection distance, the transceiver settings can be altered on a scale between 0 and 62, the lower half (0-31) alters the signal strength while the upper half (32-62) affects the detection sensitivity, meaning that only one of these can be controlled at any one time. Up to 250 loggers can be deployed at once, and an individual logger can log up to 8 others simultaneously; making these devices attractive for the real-time study of sociality in many group-living mammalian species (Krause et al. 2014). Sirtrack® proximity loggers have, for example, been used to study interactions between badgers and cattle to help better understand transmission of Bovine Tuberculosis (Böhm et al. 2009; Weber et al. 2013); discover seasonal variability in the social behaviour of Tasmanian Devils, helping characterise the epidemiology of devil facial tumour disease (Hamede et al. 2009); examine spatial and temporal heterogeneities in the contact behaviour of rabbits (Marsh et al. 2011); show that racoon social structure is independent of genetic relatedness between individuals (Hirsch et al. 2013a); and study the changes in temporal and spatial associations between pairs of cattle during the process of familiarisation (Patison et al. 2010).

Previous studies have found issues with the reliability of Sirtrack® spatial proximity loggers. Prange et al. (2006) and Drewe et al. (2012) demonstrated that continuous contacts between two loggers placed within both of their detection ranges can be represented by one or both loggers as a series of intermittent contacts; known as “broken contacts”. Broken contacts can affect contact duration and frequency; increasing the frequency and decreasing the duration of contacts. Drewe et al. (2012) suggested annealing contacts which occur within a set timeframe, or “amalgamation window”, of each other as a solution to this problem. In the same study, Drewe et al. (2012) also noted that the detection distances of loggers reduce over time, potentially as battery life decreases. Another issue with Sirtrack® loggers is the frequent recording of one second contacts, dubbed “phantom contacts”, thought to occur when loggers are just outside of their detection ranges (Prange et al. 2006). It has been suggested that phantom contacts are best removed early in data processing (Drewe et al. 2012; Watson-Haigh et al. 2012). There is some debate as to the extent to which loggers misidentify each other or record erroneous logs, with Prange (2006) stating that loggers were correctly identified in all cases in their study, but Watson-Haigh et al. (2012) stating that logs can be erroneously attributed to loggers which were not deployed. The logs collected from Sirtrack® loggers can also show poor corroboration, or “reciprocal agreement” over time (Hamede et al. 2009; Watson-Haigh et al. 2012). Reciprocal agreement (RA) referring to instances where loggers concurrently record each other as opposed to only one logger recording the presence of another. The recommendation made by Watson-Haigh et al. (2012) was to remove pairs of loggers with poor RA, whereas Hamede (2009) opted for using an approach which Watson-Haigh et al. (2012) refers to as “pseudo-contacts”; where a contact is said to be occurring when one or both of the loggers involved are recording a contact at a given time.

Watson-Haigh et al. (2012) left open the question as to what the causes of poor RA could be, although they did suggest that a lack of date/time synchronisation, different detection distances or a lack of free memory in a logger would bias the use of pseudo-contacts (Hamede et al. 2009) as a means to solve issues with poor RA. The varying height of loggers off the ground (we will call this “elevation”) could be described as a contributing factor to poor RA, with loggers close to the ground performing less effectively than those raised (Böhm et al. 2009). A recent potential cause of poor RA was made by Rutz et al. (2015), who highlighted the effect that the relative orientation of a pair of loggers might have on signal propagation. Rutz et al. (2015) stated that the antennae fixed to these devices would have an anisotropic (doughnut-shaped) radiation pattern, radiating (and receiving) more power to (and from) some directions than others. Little radiation would be produced (or received) along the axis of the antenna, and the direction of maximum power would be perpendicular to this axis. The study by Rutz et al. (2015) was largely orientated towards another radio-proximity based system known as “Encounternet” (Encounter Net project; http://encounternet.net/). However, their findings are supported by the observation of Prange (2006) that Sirtrack® loggers detect a greater range when antennae are vertical as opposed to horizontal.

In this chapter, Sirtrack ® loggers are tested on two separate herds of dairy cattle to identify sampling issues inherent in the technology and to develop subsequent corrections. Dairy cattle are an economically important livestock species in the UK. In 2013, the dairy industry contributed £4.27 billion to the UK economy, with milk accounting for 16.1% of all agricultural output (Baker 2015). In a typical UK dairy herd, 25% of individuals are replaced per year (DairyCo industry data). On average, dairy cows undergo 3.6 lactation cycles in their lifetimes, short of the economic optimum of 4.9 (Stott 1994). One of the main reasons why lifespan is shorter than the economic optimum is that 50% of culling is conducted not on the basis of milk yields, but instead in response to poor welfare traits (e.g. excessive lameness) (DairyCo industry data). Given that many production animals are reared in social environments and that pairwise social interactions have been linked to production and welfare traits (Bijma et al. 2007; Ellen et al. 2008; Bergsma et al. 2008), understanding and managing the social structure in a herd of dairy cattle may provide a route to improving welfare, lifespan and productivity.

In many modern dairy herds, management practises mean that cows are often held in barns and other confined spaces for some, if not all of the day. As such, proximity-based measures of social structure are likely to produce saturated networks (where every cow has recorded contact logs with every other). The saturated structure of contact networks means that little to no insight is gained regarding the effects of social structure through looking at the presence or absence of social ties. Instead, the strength of ties (i.e. the association times) between individuals can be used and social structure examined in terms of the heterogeneity in the weighting of edges (Croft et al. 2008). However, the weights of edges are likely to be influenced by the detection distances of the loggers affixed to the individuals. In this study, we are left in a position where we might trust any absence of interactions between cows as reliable in such confined settings, but could not trust the weightings of the contacts which have been recorded as taking place; this is a stark contrast to the problem inherent in hybrid networks (chapter 4) where we could trust the presence, but not trust the absence of edges.

In the following section of this chapter, it is shown that variation in detection distances is indeed a contributing factor to the poor RA of loggers. In addition, we characterise and address the wider issue of variability in logger performance, which may also include factors such as subtle differences in transmitter components, antenna configurations or both (Rutz et al. 2015). In particular, we demonstrate how the variable performance of loggers might affect the structure of social networks produced from logger data. The first correction proposed to account for the variable performance of loggers is for static networks- we will call this the "static correction". The static correction corrects the total association times recorded by loggers over a given time period. Given that radio-proximity loggers provide time-stamped contact logs; temporal information is lost when these logs are condensed into association times. In the final section of this chapter, we propose an extension to the static correction in Boyland et al. (2013) which corrects contact logs in a manner which preserves temporal resolution; this allows the resulting data to be used in multivariate models of temporal networks (e.g. Patison et al. 2015). We explore the impact of clock-drift and propose a more informed approach to increasing RA over time than either pseudo-contacts outlined by Hamede et al. (2009), or the amalgamation window approach proposed by Drewe et al. (2012)- we call this the "temporal correction". The temporal correction is more informed than other approaches in that it takes into account whether the loggers of the individuals involved tend to over or under-sample interactions as well as correcting for clock-drift as recommended by Watson-Haigh et al. (2012).

6.3 Static Correction

6.3.1 Methods

Standardised test for detection distance

To test the detection distance of 20 randomly sampled Sirtrack collar-mounted loggers (model E2C 181C), we modified the experimental procedure outlined in Drewe et al. (2012). Each collar was mounted, in turn, on plastic stands set at a fixed height (50cm) above the ground. Antennae were fixed in a vertical position and loggers were set to a fixed setting (43 - a medium strength detection sensitivity) with a separation time set to 10 seconds to reduce the time waiting for termination of logs before repeating the experiment. A “base-station” (a free standing logger) was mounted at the same elevation and moved towards the fixed collar in 10cm increments from a starting point outside of both the logger and base-station’s detection range (6m separation). After each increment, the base station was held at that distance for 5 seconds to observe whether a signal has been picked up, indicated by the flash of the LED of one or both of the loggers. One-off flashes were observed to occur outside of the “consistent” detection range (this had also been noted in Prange et al. (2006)), so initiation was recorded as starting once the LEDs on both the collar and base-station had flashed consecutively five times. The distance at which contact was established, the ‘initiation distance’, was noted for both collar and base station. The base-station was then moved away from the collar in 10cm increments, with a five second pause after each increment, until one or both of the LEDs stopped flashing. The distance at which contact was lost, the ’termination distance’, was recorded for both logging devices. The base-station was then moved back to its starting position, well outside of the detection range of the collar until the log had been stored (indicated by a prolonged LED flash). The process was repeated for eight times with each collar with a mean taken of both the initiation and termination distances - the base-station used remained the same throughout the experiment.

Field deployment

The same twenty proximity collars as used in the standardised test were deployed for three weeks on 20 dairy cows, within a herd of 120, on a commercial dairy farm in Cullompton, Devon, UK. The herd was housed in a 45x30m barn with a robotic milking system. Low-yielding cows were given access to pastures, whereas high-yielding cows spent their entire day in the barn. Collars were turned on and then affixed to high-yielding cows whilst they were temporally immobilised, either during milking or routine husbandry procedures. Following the data-collection period, loggers were removed opportunistically during milking. The detection strength setting of the loggers was kept the same as in the standardised test, however, to avoid loggers recording too many logs and filling up prior to the end of the deployment, the separation time required for a contact to be logged was increased to 120 seconds.

Data were collected from three days after collars were affixed to individuals for the following three weeks. The three-day period was to allow cows to become used to the presence of the loggers. After removing the logs that took place during the first three days following collar attachment and those that occurred after the start of collar removal, one–second logs were removed as recommended by (Drewe et al. 2012). Following these steps, an association matrix depicting the total durations each logger had recorded other loggers over the deployment was created; this matrix was not symmetrical as two loggers within a dyad may have recorded a different duration of contact with each other. To calculate the pairwise biases of individual loggers, the percentage difference in total contact duration B was calculated for each dyad using the formula: $B_{ij} = 100\,\frac{D_{ij} - D_{ji}}{D_{ij} + D_{ji}}$, where D is the duration from the association matrix for loggers i and j. For example, if logger i recorded logger $j = D_{ij}$ = 5000 seconds, and the duration that logger j recorded logger $i = D_{ji}$ = 4000 seconds then B_{ij} = +11.1% and B_{ji} = -11.1%. Performing this calculation across all dyads created what we will refer to as the pairwise bias matrix B. The average logging bias each logger had with all other loggers in the deployment was taken; this provided a measure of how each logger performed comparative to all others, hereafter referred to as the logging bias.

The link between detection distance and logging bias

To determine the role that the observed variation in logger detection distances might have on the logging biases observed in the field deployment, we first correlated the measured detection distances with logging biases.

To probe the link between detection distances and logging biases, a simple analytical model (we will refer to this as "*A*") was proposed by James. In this model, each logger *i* has a single distance r_i (set as the logger initiation distance from the standardised test) within which it detects any other logger *j*. The model assumes that individuals use space uniformly, so the time *t* that logger *i* detects any other logger *j* is proportional to the square of *i's* initiation distance: $t_i = \alpha r_i^2$. The constant α is the same for all loggers, depending only on the size of the area used, and the total duration of the observations. Using the above model, a matrix of pairwise logger biases A_{ij} was constructed using the derived formula:

$$A_{ij} = 100\,\frac{r_i^2 - r_j^2}{r_i^2 + r_j^2}$$

In this formula, the percentage difference A_{ij} for each dyad is taken as the difference in the square of each logger's initiation distance (r_i and r_j respectively) relative to the average of both of the logger's initiation distances. The A_{ij} bias matrix produced using this formula was correlated to the B_{ij} matrix observed in the field deployment using a Mantel test (Mantel 1967) with a Spearman's rank correlation coefficient in the package "vegan" (Dixon & Palmer 2003) in R (R Development Core Team. 2008).

The B_{ij} bias matrix from the field data was significantly correlated with the A_{ij} biases created using the analytical model (R_S=0.52, p=0.008); this further confirms that variation in the detection distances of loggers does indeed influence logging biases. However, the R^2 (of the correlation) suggests that it only explains about 25% of the variation in the ranking of logging biases suggesting factors other than initiation distance affects pairwise logging biases.

Given that A was simplified to consider only the initiation distances of loggers, a computational model (we will refer to this as "*C*") was used to check that this simplification was justified. In this model, 20 agents representing cows wearing loggers were allowed to roam around a hypothetical arena (area: 100m^2) for 15,000 time steps. The simulated loggers affixed to agents in the model had the same initiation and termination distances as those observed in the standardised test. Upon an agent entering the initiation distance of another logger, a log was started, upon the agent subsequently exiting the termination distance of that same logger, the log was finished and the contact duration (in terms of time steps) recorded. C was used to produce a pairwise bias C_{ij} matrix using the same procedure as with the field deployment. The C_{ij} matrix was correlated with the A_{ij} matrix created by the analytical model using a Mantel test with a Spearman's rank correlation coefficient.

The effect of logging biases on network production

To characterise the consequences of variable logger performance on the construction of social networks, we calculated the in and out degree-strength of each of the 20 collared individuals in the field deployment (i.e. the total time an individual's logger recorded or was recorded by other loggers respectively). Degree strength could be taken as a measure of the overall sociality of different individuals and how central they are in the social structure of the herd (Croft et al. 2008).

Static Correction Procedure

One solution to control for the biasing effect of loggers might be to include variables which account for this source of bias in a multivariate model. Two options may be considered at the point of modelling. The first would be to include logger ID as a categorical, or "dummy" (Suits 1957) variable when modelling social structure; this would ensure that each logger's performance is accounted for when determining the effect size and significance of other variables in the model. However, a severe drawback of this approach is that it potentially requires a unique explanatory variable for every logger; this may lead to both an over-parameterised model and one where individual differences in sociality are masked by the logger ID variable. An alternative approach might be to have two explanatory variables in the model such as “Focal logger detection area” and “Encountered logger detection area”; the advantage of this approach would be that only two variables are required to capture the variability in logger performance, so a more parsimonious model could be proposed. However, as we show in section 6.3, detection distances are a significant, but marginal factor in determining the variability in logger performance; explaining about 25% of the variation in logger biases. Therefore, detection distances may not account for suspected latent causes of variability in logger performance and may fail to sufficiently represent logger bias.

To analyse networks constructed from contact logs collected by radio-proximity loggers, an approach is needed which corrects for the variability in logger performance without being overly reliant on detection distances or requiring logger-specific explanatory variables in a model. With this in mind, we developed a correction to address logger biases in static networks (which we refer to as the “static correction”).

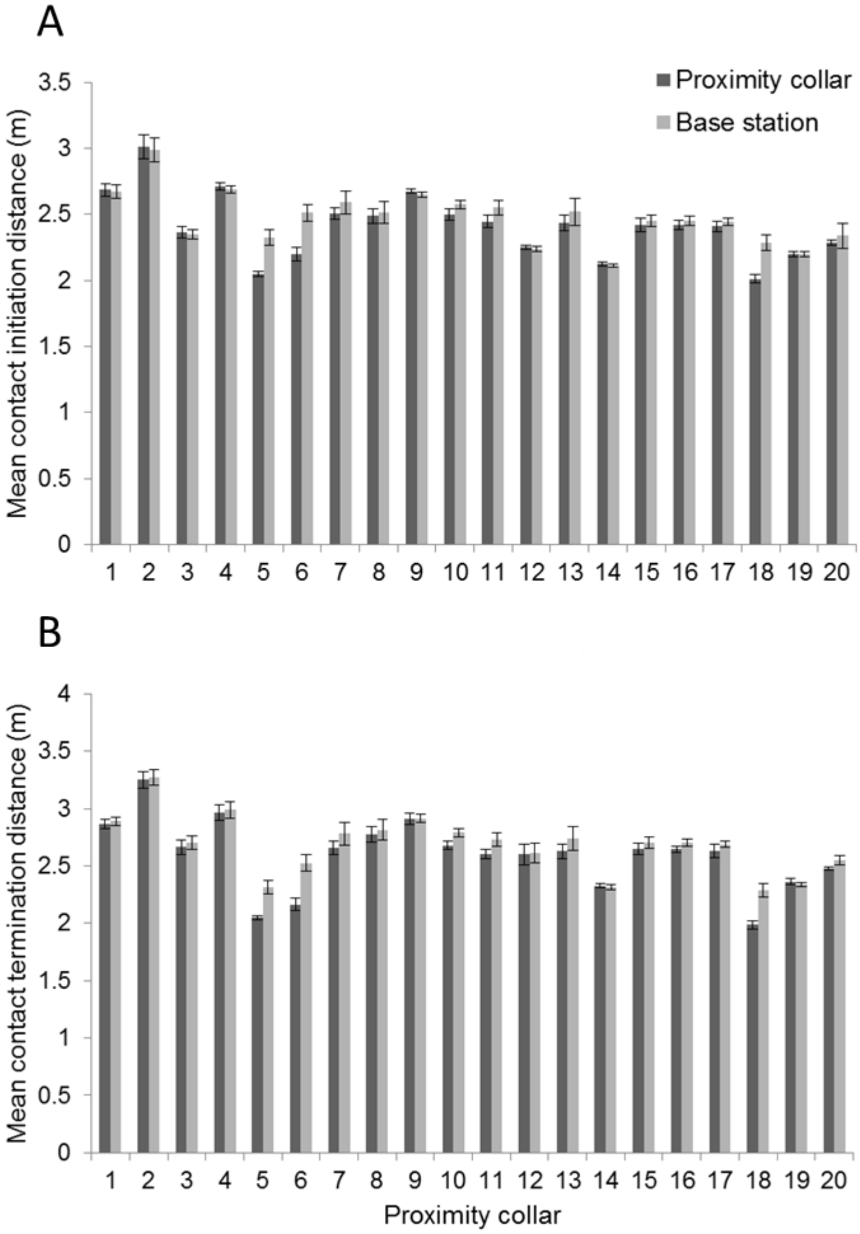
To provide a measure of the reciprocity between loggers in each dyad prior to the correction, a Spearman’s rank correlation coefficient was calculated between the durations recorded in dyads ij and ji over all values of i and j – we will call this measure the “dyad reciprocity”. Prior to the correction being applied to the association matrix from the field deployment, the dyad reciprocity was 0.76. If there was perfect reciprocity in the deployment (i.e. each logger in a dyad recorded the other for an identical duration) then the dyad reciprocity would be 1.

To correct for the variability of logger performance, we proposed that the durations recorded by loggers in the association matrix are reduced proportional to how much the focal individual’s logger was identified to oversample interactions relative to the weakest logger in the deployment. The justification for this approach is that it is conservative, removing rather than creating data, and intuitive, as the weakest logger in the deployment would provide a representation of the sampling effort. To do this, all logging biases of loggers were first recalculated relative to the least recorded logger in the deployment. In our field test, this logger had a bias of -14.8%, so 14.8% was added to all logger biases, meaning the strongest logger in the deployment (previously with a bias of +22.1%) over-sampled interactions by +36.9% relative to the weakest logger. To continue with this example, all durations recorded by the strongest logger would be reduced by 36.9%, whereas the durations recorded by the weakest logger would not be changed. Note that this procedure is reliant on calculating pairwise logger biases to identify the relative logging bias of deployed loggers, as such, the approach would not be suitable for sparse datasets where few loggers have encountered each other over the duration of the deployment. In such cases, the analytical model could be used to provide logging biases from the A_{ij} matrix if the distances of loggers have been measured.

6.3.2 Results

Standardised test for detection distance

The standardised tests revealed that there was significant variation in the mean initiation and termination distances of loggers, with mean initiation distances ranging from 2.01 to 3.01m (mean: 2.41m± 0.05m standard error), and mean termination distances ranging from 1.99 to 3.25m (mean: 2.59m± 0.07m standard error). The mean initiation distance of loggers was found to have a Pearson's correlation of 0.96 with the mean termination distances ($p<0.0001$). A Pearson's correlation coefficient was used as the datasets show no significant deviations from normality (initiation: $W = 0.9619$, p-value = 0.5826; termination: $W = 0.9512$, p-value = 0.3864) and have homogenous variances ($F = 0.6313$, $df = 19$, $p= 0.3245$). In insert 6, the initiation distances (*A)* and termination distances (*B)* are shown against for each collar against the fixed base-station. The detection distances of loggers were consistent. Across the 8 measurements made for each logger the average variance in initiation distance was 1.4cm, four times lower than the variance between loggers of 5.6cm.



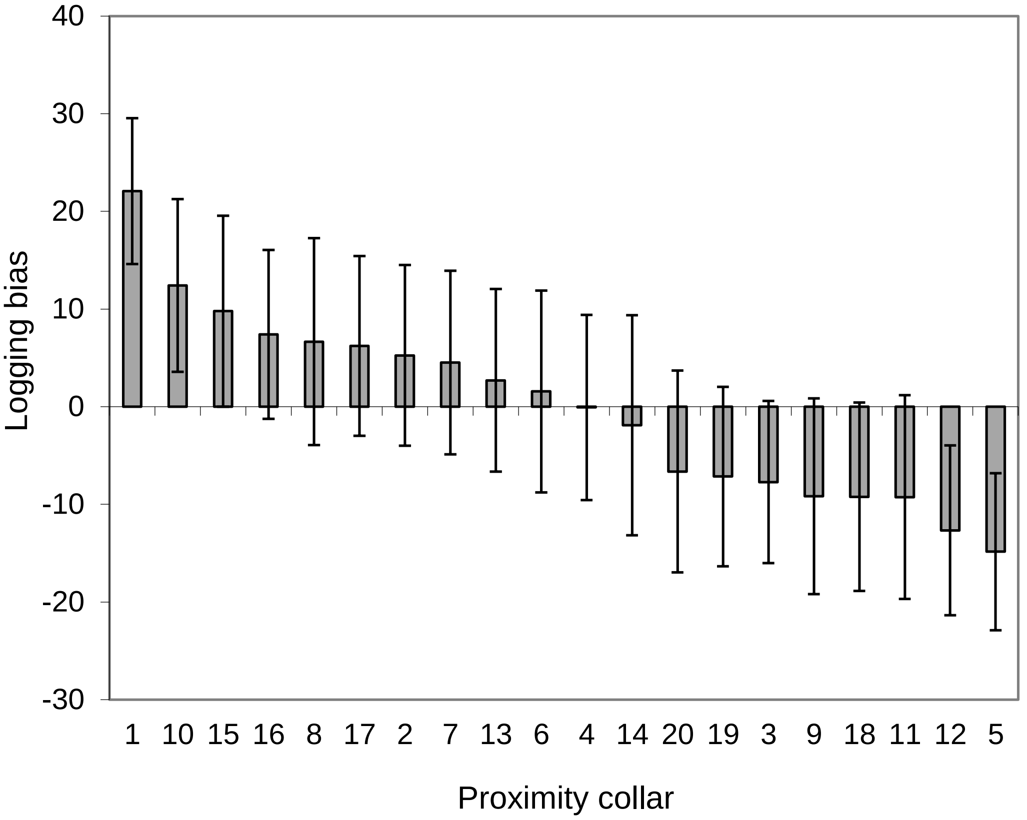
Insert 6: *A* - The mean contact initiation distance (±standard error) for each of the 20 loggers tested and the corresponding distances recorded on the base-station. *B* - The mean contact-termination distance (±standard error) for each of the 20 loggers tested and the corresponding distances recorded on the base-station. These figures have been adapted into a panel-figure with the permission of Natasha Boyland.

The standardised test thus indicated that the strongest logger in the deployment (~3m range) could detect contacts 1.5 times as far away as the weakest logger (~2m range). If we consider that these loggers sample an area for other the presence of other loggers, the strongest logger would sample an area 2.25 times as large as the weakest logger in the deployment (~3m^2/ 2m^2).

Field deployment

The field deployment revealed substantial variation in logging bias, ranging from -14.8% to +22.1% (see insert 7). To see if these biases were repeatable and thus consistent for each logger, the deployment was split into three one-week observation periods and the biases recalculated for each collar using the same procedure as for the whole deployment. Repeatability was measured as the variance between loggers divided by the sum of the variance between loggers and the residual variance (see Nakagawa & Schielzeth (2010) and Schuett et al. (2011) for details). Ninety-five percent confidence intervals (CI) for repeatability estimates (r) were obtained from parametric bootstrapping (N=1000 simulation iterations) (see Nakagawa and Schielzeth (2010) for details).

The proximity loggers showed a very high degree of consistency in logging bias over the three time periods (r=0.992, 95% CI: 0.982 to 0.996).



Insert 7: The logging bias of each logger in the deployment shown in descending order. This figure was used with permission from Natasha Boyland.

The link between detection distance and logging bias

The logging bias observed in the field study was significantly correlated with the mean initiation (R_S=0.457, n=20, p=0.043) and termination (R_S=0.49, n=20, p=0.028) distances of the base station (distance at which the base station detected collars). Similarly, the logging bias was significantly correlated with the mean initiation distances (R_S=0.460, n=20, p=0.041) of the collars (distance at which the collars detected the base station). The mean termination distances of collars were not significantly correlated with logging bias, though the correlation was in the same direction (R_S=0.368, n=20, p=0.111).

The analytical and computational model were highly consistent in the calculation of logging biases (R_S=0.99, p<0.0001); this suggested that *A* could indeed be used as the most parsimonious way to explain logging biases.

The effect of logging biases on network production

We identified a significant relationship between the in-degree strength of an individual and logging bias of associated logger. Loggers which tended to oversample other loggers had a higher in-degree strength (linear regression with permutation test, n=10,000 permutations, adjusted -R^2 =0.546, F=2.58, p<0.001). However, there was no significant relationship between the weighted out-degree strength and logging bias (linear regression with permutation test, adjusted -R^2 =-0.095, F=0.17, p=0.705). Given this, logging bias of an individual's logger could affect how central we believe an individual is in a social network. Individuals fitted with strong loggers will appear more central, whereas those fitted with weaker loggers will appear more peripheral.

Static Correction Procedure

After, the correction was performed, dyad reciprocity rose from R_S=0.76 to 0.99 using the B_{ij} matrix, and 0.84 using the Aij matrix. The increases in reciprocity indicate that the correction resolved much of the pairwise discrepancies in contact duration observed in dyads prior to the correction. In this study, the B_{ij} matrix provided the most effective correction. However, producing the B_{ij} matrix requires every logger to have come into contact with each other to function effectively. In captive populations, such as livestock, this may be common place, but in free-living animal populations the association matrices produced may be too sparse to calculate a B_{ij} matrix rich enough to determine accurate logging biases. The A_{ij} matrix provides a suitable alternative for situations where the association matrix is sparse, as it only requires the detection distances of loggers to be quantified prior to deployment.

6.4 Temporal Correction

The static correction outlined in section 6.3 was directed towards adjusting an association matrix containing the total durations that loggers had observed each other over a certain period of time. One mentioned limitation of the static correction using B_{ij} is that these association matrices need to be rich, with all loggers having encountered each other for the correction to perform well. Although A_{ij} does not have this specific limitation, it has only thus been applied to static networks. Both methods have not had to consider which contacts, or parts of contacts, should be removed over continuous time to account for logging bias. In this section, a procedure is outlined to increase the reliability of contact logs without reducing the temporal resolution of the data. The five principles behind the temporal correction are as follows:

1. To synchronise the timing of logs to the logger with the most stable clock for the duration of the deployment - we specifically refer to this part of the overall correction as the “synchronisation procedure”.
2. To assume longer contacts are more reliable than shorter contacts.
3. To assume two contacts between the same loggers are more reliable if they are close in time.
4. To assume that the contacts recorded by under-sampling loggers are more reliable indicators of an interaction taking place than the logs of over-sampling loggers.
5. To assume that times when both loggers in a dyad record each other concurrently are more reliable than times where only one logger records the other.

The temporal correction uses these criteria to produce one set of logs for each dyad, weighted according to the strength of evidence attributed to whether the contact had indeed taken place. The data set can then be filtered to remove all logs less reliable than a given threshold. We recommend setting this threshold at a level where all of the logs greater than one second in duration from the weakest pair of loggers in the deployment are retained and no higher. In this sense, the maximum amount of data is retained whilst removing logs driven by oversampling loggers.

The synchronisation procedure first increases the alignment of contact logs recorded within each dyad through correcting for clock-drift that has taken place over the course of the deployment. Secondly, although the weighting of contact logs never generates "new data" like the "amalgamation window" approach (Drewe et al. 2012; Bettaney et al. 2015), it does allow intermittent interactions by one logger to be annealed if the missing interaction time is logged by the other logger in the dyad. However, in contrast to the use of pseudo-contacts (Hamede et al. 2009), interactions will not always be annealed under a given reliability threshold. If the two loggers involved heavily oversample interactions, then interactions missed by one may instead indicate that interactions are towards the edge of the detection range of already oversampling loggers, and thus should be removed given the desire for unbiased sampling. An advantage of our approach is that it both considers the biased performance of loggers, shown to be correlated with detection distance, and clock-drift when correcting collar logs. Watson-Haigh et al. (2012), highlighted that these two factors would bias existing approaches.

6.4.1 Methods

Field deployment

To demonstrate this correction procedure in full and characterise some of the factors which lead to the poor reciprocal agreement of contact logs. A second deployment of loggers was carried out in 2014. Ninety-five collar-mounted Sirtrack® radio proximity loggers were attached to individuals in a herd of ~120 Holstein dairy cattle situated at Evershot Farm, Dorchester in October 2014 for a total of 10 days. Before the deployment, new batteries were fitted and all internal clocks of loggers were synchronised. The loggers were set to detect contacts from at least 2 metres, which corresponded to them being set at UHF 40 in all but 12 cases, which required settings between 1 and 5 to sample at 2m. The management system at the farm throughout the duration of the deployment was as follows: cattle were held in barns until 8am, at which point they were milked and then allowed to graze outdoors in paddocked fields until 3pm, when they were milked and returned to the barn until the next morning. Collars were both fixed and removed from cows through running the whole herd through a "crush", a device used for restraining cows, usually for routine husbandry procedures.

Pre correction data-processing

The downloaded logs from the 95 loggers were amalgamated into one dataset, which was subsequently trimmed to remove any logs that occurred before 7am on day 3 and after 7pm on day 9 were omitted; logs that crossed these times were truncated accordingly. Data were removed from the first three days of the deployment to allow time for the cows to adjust to the presence of the collars. The proportion of logs with a duration of one second was recorded, but in contrast to the recommendations of Drewe et al. (2012) these logs were not removed from the dataset, as these may provide evidence that loggers are near the edge of their detection range (Prange et al. 2006). Further, the proportion of logs that exceeded the memory limit of the loggers (65535 seconds) was recorded, as these have been observed in previous deployments. Contacts where the duration exceeds the single-log memory limit of the loggers are recorded as two or more logs with simultaneous starting points, with all but the last of these logs having duration of 65535 seconds. Although far less common than one second logs, these split logs will artificially reduce average contact duration and increase contact frequency if not annealed.

Before implementing the correction algorithm, the total time that each pair of loggers were in reciprocal agreement across the deployment was recorded. Recall that reciprocal agreement is defined as both loggers in a given dyad logging each other simultaneously. The amount of time each logger recorded the other without reciprocation was also recorded for each dyad; these measures served as a benchmark to evaluate logger performance before correction.

To identify potential causes of unreciprocated contacts, all instances where these occurred within dyads were identified. First, the logging biases of each collar were used to explain lack of reciprocation. Logging biases were determined from a B_{ij} matrix (as outlined in Boyland et al. (2013) and section 6.3). Logging biases could provide a measure of how often unreciprocated logs were caused by a contact being recorded by just the stronger logger in the dyad; this could potentially be the result of cows being at a distance from each other outside of the detection radius of weak logger, but within the distance of the strong logger. Secondly, unreciprocated contacts could be the result of clock-drift over the course of the deployment, which could lead to one set of logs being recorded at an earlier or later time than the other. To quantify this, the total amount of reciprocal contact time for each dyad was compared before and after logger's clocks were synchronised using the synchronisation procedure.

Temporal correction algorithm

The first part of the temporal correction is the synchronisation procedure. The first step of which is to identify how offset the clocks of loggers are from each other. We will call the time offset matrix produced by this process the T_{ij} matrix. Using the T_{ij} matrix, the logger with the most stable clock in the deployment can be identified (see: "i. Time synchronisation"). The clocks of all other loggers can then be synchronised to this logger using the corresponding offsets in the T_{ij} matrix. After synchronisation, all contacts then receive a weight based on their duration and proximity in time to the nearest contact with the same logger - under the assumption that contacts with a long duration are more reliable than those with a shorter duration and that an isolated contact with a given logger is less trustworthy than one close in time to another with the same logger. The weights assigned to contacts are then scaled down by the logging bias of the logger relative to the weakest logger in the deployment. In dense, captive populations, this logging bias would most likely be calculated on the B_{ij} matrix, but if the association matrices at the end of the deployment are sparse, then we have demonstrated (in section 6.3) that the A_{ij} matrix can also provide informative logging biases. The last step in the correction is to combine the two sets of logs for each dyad into one set of weighted contacts; when this step is applied, the weights of logs are added together at times when there is reciprocal agreement between loggers- as reciprocal agreement is assumed to increase the reliability of logs. We now consider the approach in more detail:

1. Time synchronisation

For each dyad, the total duration of time that both loggers concurrently registered contact with each other was summed over the course of the deployment. The logs from one logger were then shifted forwards and backwards in time against the logs from the other in one second intervals. After each shift, the new duration of reciprocated contact was summed. Logs were shifted until the displacement in time required to generate the first local maximum in the sum duration of reciprocated contact was identified. This time-displacement was taken as indicative of how offset the clocks of two loggers in a dyad were relative to each other. Applying this method across all dyads generated a matrix of observed time displacements between the loggers in every dyad. In figure 9, the method used to identify the offset of the clocks of two loggers are shown in two examples: *A*- depicts two loggers which have clocks that are well aligned, whereas *B* shows two of the least synchronised clocks present in the deployment.

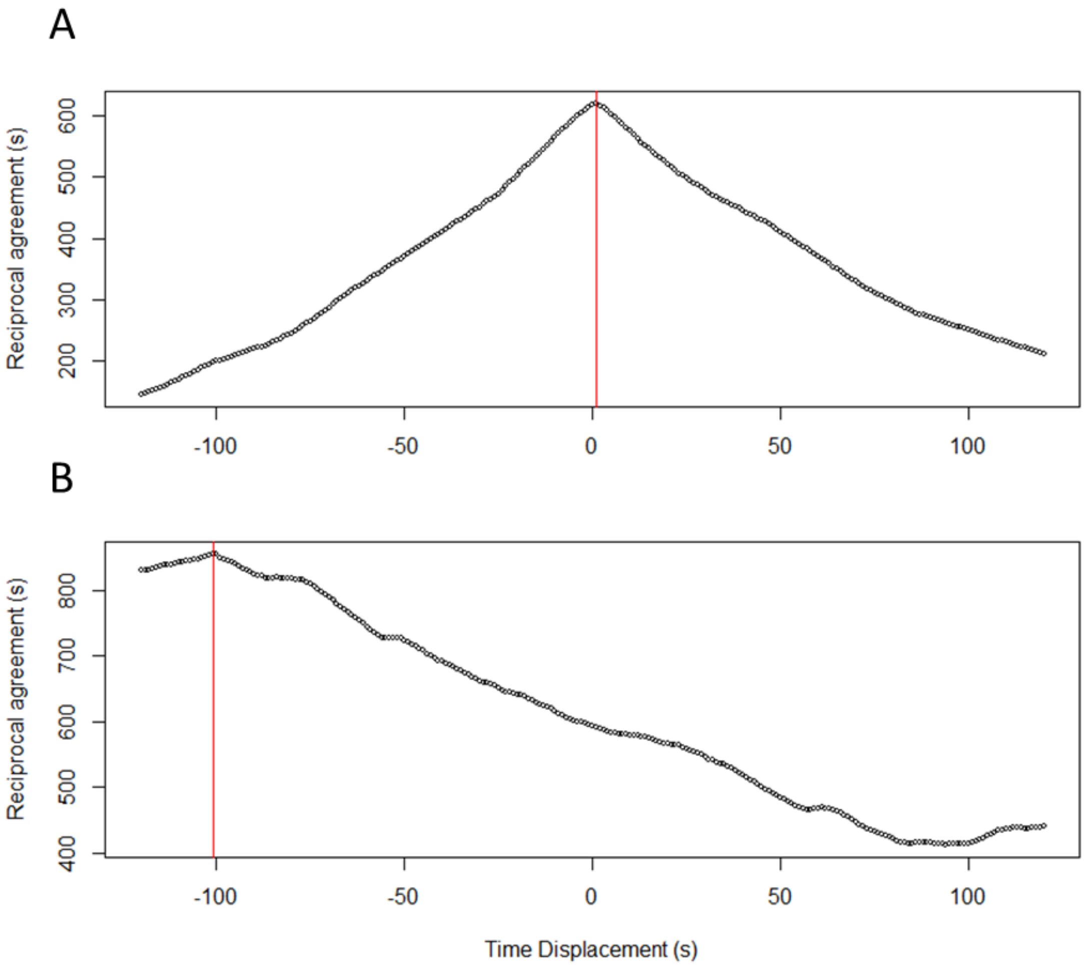


Figure 9: *A*- Time-shift plot showing an identified time-displacement of -8.5 seconds (shown in red) between loggers 12 and 5. *B*- time-shift plot showing an identified time-displacement of -101s (shown in red) between loggers 86 and 6.

The observed time-displacements between loggers were not perfectly consistent across dyads (i.e. $T_{ab}=x$ and $T_{ac}=y$ then T_{bc} may not always equal $y-x$), which suggested the clocks of some loggers may have drifted relative to others over the course of the deployment. An alternative source of inconsistency may have been cases where the synchronisation process identified a local maximum which was not representative of the true offset between the clocks of two loggers. Consequently, the T_{ij} matrix alone could not provide consistent adjustments required to synchronise the logs from different loggers in the deployment. Instead, the clocks of each logger were corrected to the clock of one logger whose offsets with other loggers in the T_{ij} matrix could most accurately predict the offsets between all loggers in the deployment. To determine which logger to use, a "predicted" time-displacement (P_{ij}) matrix was constructed for each logger using the time-displacements in the T_{ij} matrix. For instance, in logger *A*'s P_{ij} matrix, if logger *A* was found to be 3 seconds ahead of logger *B* ($T_{AB}=3$) and 4 second ahead of logger *C* ($T_{AC}=4$), then, according to *A*, the clock of *B* would be 1 second ahead of the clock of *C* ($P_{BC}=1$). The mean of the differences between the $\mathrm{T_{ij}}$ matrix and each $\mathrm{P_{ij}}$ matrix was calculated - which we will refer to as the "residual time" (see figure 10). The clock of the logger that provided the lowest residual time (we will call this logger θ) was chosen to synchronise the clocks of the other loggers to; this was achieved via shifting the logs recorded by other loggers forwards or backwards in time according to the offset that they were identified to have with θ in the T_{ij} matrix.

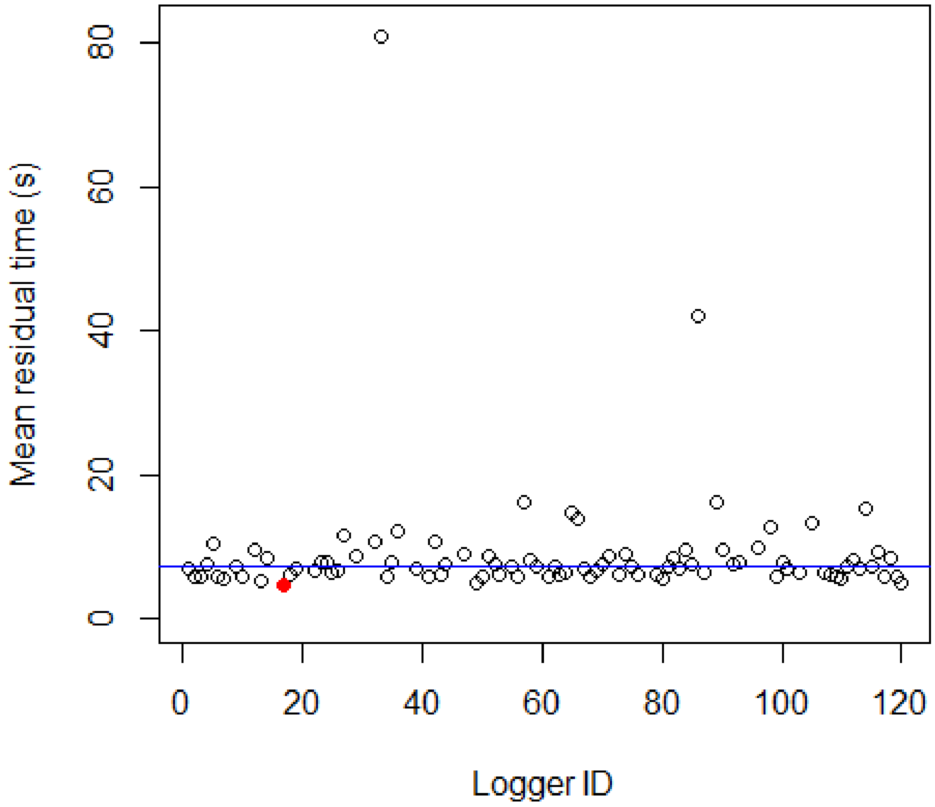


Figure 10: Mean residual times of the 95 loggers deployed. The blue line indicates the average offset between loggers identified through the local maxima in the time-shift plots for every dyad (7.32seconds). Logger 17 (red) provided the lowest residual time of 4.75 seconds.

2. Weighting contact logs

Following the synchronisation process, the contacts recorded by loggers were weighted in a four-step process- the criteria used to weight logs was chosen heuristically. The first step was to weight assign each log a weighting between 0 and 1 on the basis of their duration. The second step was to add a weight between 0 and 1 to logs based on how close in time they occurred to other logs between the same loggers. The third step was to scale-down the combined weights of logs (which could now range between 0 and 2) relative to how much the logger they were recorded on was identified to oversample interactions. The fourth step was to combine the logs recorded by both loggers in a given dyad into one set, at times where both loggers recorded a contact, their weights would be added together. At the end of this process logs could have a weighting between 0 and 4. The rationale behind this process is that the weights of logs would represent how much we could trust that they represented a real contact between the animals to which the loggers were affixed. Thresholding out the logs with the lowest reliability would, due to the scaling of weights, preferentially remove unreliable logs from oversampling loggers, thus removing some of this sampling bias and increasing the suitability of the resulting data for a multivariate model.

Contact logs were weighted based on their duration under the assumption that contact logs with a long duration were more reliable than those with a short duration. The weight assigned to each log was calculated as the proportion of logs recorded with a shorter length in the deployment. Therefore, the contact log which had the median length of time for the deployment would receive a weight of 0.5, the contact with the longest duration in the deployment would receive a weight of 1 and the shortest, 0.

Contacts were then weighted based on how close in time they were to another log between the same focal and encountered-logger, we call this measure the shortest interval time (SIT). The assumption was that isolated contacts with large SITs were less reliable than contacts close to others with small SITs. The weight assigned to each log was the proportion of SITs in the deployment greater than that of the log being weighted; this provided a range of weights where the most isolated contact received a 0 and the least isolated, 1. The weights generated by SITs and contact durations were then summed to give each contact log a weighting between 0 and 2.

Using the method proposed by Boyland et al. (2013) (section 6.3), logging biases were calculated for each logger in the deployment from a B_{ij} matrix. The weights of the contacts logged by each logger were then reduced relative to how much the given logger oversampled interactions compared to the weakest (most conservative) logger in the deployment. If the weakest logger in the deployment sampled at 100%, then a logger with a logging bias 20% higher than this was sampling at 120% - we call this the comparative sampling effort (CSE) of that logger. The weights of all contacts recorded by a logger *i* would be multiplied by $100/\mathrm{CSE}_{\mathrm{i}}$. For example, if logger *i* logged interactions 10% more than the most conservative logger, then the logs of logger *i* represent a CSE of 110%. In this example, the weights of logs recorded by logger *i* would be multiplied by 100/110 to reduce the weightings of logs recorded by *i*.

The scaled contact weights were combined to produce one set of logs for each dyad in the deployment (figure 11). At times where two loggers logged each other concurrently the weights were added from each logger in the dyad under the assumption that corroborative contact between loggers was strong evidence that a contact had taken place. Addition was chosen over multiplication to combine logs so that the combined contacts would not have their weight reduced to zero in instances where only one logger logged an interaction within a given dyad. The final set of logs were then filtered to remove those with a weight less than the least reliable log recorded between the two weakest loggers in the deployment, excluding 1-second (phantom) contacts.

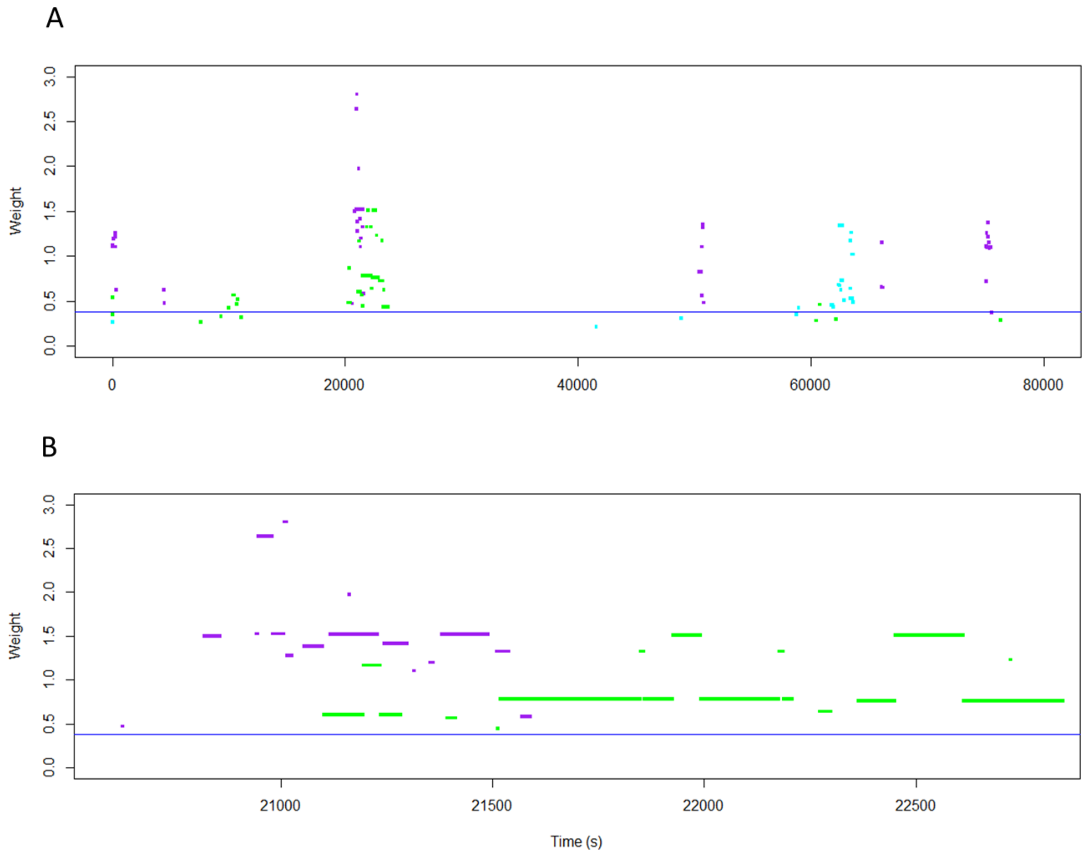


Figure 11: *A*- weighted contact logs of two weak loggers (purple), two strong loggers (green) and two medium sampling loggers (cyan) over time in seconds. The filter threshold (0.3862) is shown by the blue horizontal line. *B*- weighted contact logs of the same pair of weak loggers (purple), strong loggers (green) and medium sampling loggers (cyan) magnified between 20000 and 23000 seconds. The same filter threshold is shown.

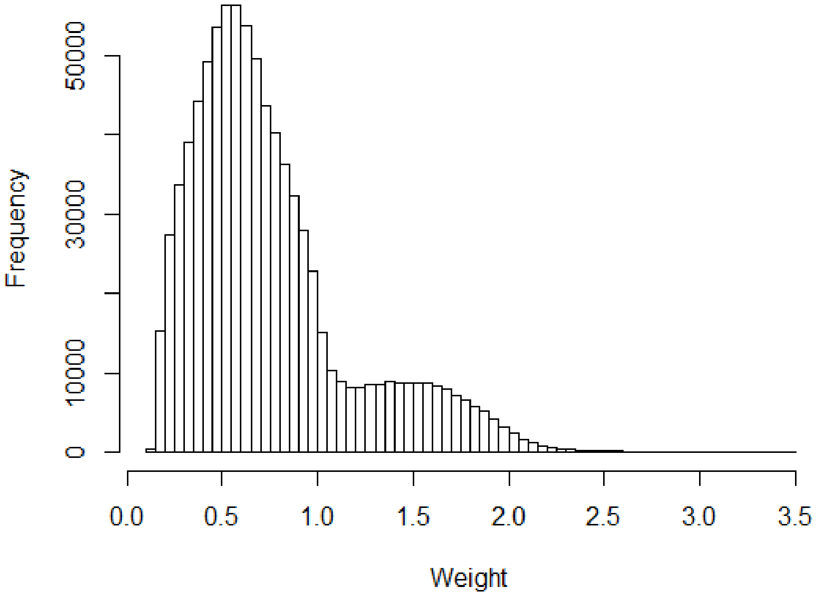


Figure 12: A histogram showing the distribution of weights assigned to logs in the deployment.

In-field Validation of the correction procedure

The rationale behind taking videos of cows affixed with loggers was to provide an external, independent means to validate the temporal correction procedure. The contact logs provided by the loggers could be compared to video footage to ensure that contacts that received a low reliability weighting from the temporal correction procedure were indeed those on the edge of the detection range of loggers and that logs which received high weightings corresponded with instances where loggers were indeed close together.

Video footage was taken to record the position and movements of selected cows from 14-10-2015 to 17-10-2015 of the deployment using a Polaroid ID2020 Full HD Camcorder. A cow was selected at random and recorded for 10 minutes with the goal of identifying potential contacts with other cows which the loggers may record. Each cow was identified either by its logger ID painted on the collar strap, or its freeze-brand/ ear-tag number. If the selected cow became isolated from others, or became difficult to video without disturbing other cows, then neighbouring accessible cows would be recorded for the remainder of the 10 minutes. The videos were than analysed, with data being collected on the position of cows (standing/lying), their orientation relative to each other, and the estimated distances between loggers (assuming a Holstein Cow's body length is 2.5m) over time.

Detailed observations were recorded from 15 randomly-selected ten minute videos; assisted greatly by two project students at Exeter University, Primrose Manning and Ruth Shen. These videos provided 519 complete data entries, where both of loggers involved in the interaction could be identified along with the distance separating loggers and the time of day that the loggers were separated at a given distance. A statistic *V* was chosen to represent how well the loggers performed; this statistic consisted of multiplying the weighting of a log by the inverse (1/distance) of the distance loggers were observed to be apart in the video (+0.1m to prevent division by 0) and then summing this score for every second logger and video contacts corroborated. Using this statistic, highly weighted logs assigned to instances where cows were observed to be very closely together would contribute very highly to the final test statistic.

The first test of the weighting procedure was to see if the observed *V* was greater than expected if weights were assigned to contacts randomly- preserving the same distribution of contact weights. The null model used to test this involved permuting the contact weights assigned to contact logs 4999 times; this generated a distribution of *V* under a null where the logs of loggers had already been synchronised. We had recommended setting a weighting threshold that filters the phantom contacts of the weakest pair of loggers in the deployment and no more - which worked out as 0.3862 in this deployment. The second test conducted was to check that removing logs with a weight less than the threshold of 0.3862 did not remove close-contacts identified by the video footage. To test whether this was significant, a null model was constructed which randomised contact weights and then removed all contacts below the same threshold.

6.4.2 Results

Quantifying the effects of the temporal correction

Prior to performing the temporal correction procedure, a number of erroneous logger IDs were recorded belonging to loggers which weren't deployed. Twenty-nine logger IDs were observed to occur above the range of IDs which were deployed, supporting the previous observations of Watson-Haigh et al. (2012). It is possible that more erroneous logger IDs are present in the data, but have been missed as they exist within the range of logger IDs deployed. It was observed that the majority of erroneous logs were recorded for duration of only 1 second (91.5%), with the highest duration recorded being 5 seconds. Logger IDs which were clearly erroneous were removed prior to the temporal correction. Erroneous logger IDs which may exist within the range of deployed logger IDs, and as such cannot be easily identified, will be most likely removed during the temporal correction procedure on account of their short duration.

Over the whole deployment, 31.9% of logs had a 1 second duration (or were "phantom contacts" (Prange et al. 2006)), with loggers recording a range of 26.5 to 54.0% of interactions. 40.1% of all interactions in the deployment lasted less than 5 seconds. In this deployment of loggers, there were no contacts greater than 65535 seconds (which results in split logs starting at the same starting time). Prior to the temporal correction being performed, there was a mean reciprocal agreement between dyads of 48.7% ± 2.2% (mean ± standard error) - measured as the amount of time both loggers were on divided by the amount of time either or both loggers were on. A mean of 39.7% ± 2.4% of logs consisted of the strong logger recording the weak logger in the dyad without reciprocation. A mean of 11.4% ± 0.9%   of logs consisted of the weak logger recording the strong logger in the dyad without reciprocation. Before the temporal correction, it was observed that loggers with higher logging biases recorded both a higher frequency of interactions ($R_{Pearson's}=0.399$, $p=0.0008$), and interactions which lasted a longer duration ($R_P=0.53$, $p<0.0002$). The dyad reciprocity in the association matrix for the whole deployment (as defined in section 6.3) was $R_S=0.62$.

The T_{ij} matrix was highly correlated with differences determined through using the Sirtrack® logger software, which provides the amount of time each loggers had drifted since the last data download (Mantel test: R_S=0.74, p<0.0002). The most stable logger in the deployment was logger 17, which on average had a residual time of 4.75 seconds i.e. the average amount of time the observed displacement from loggers was different to that predicted by the logger's interaction with logger 17. Uncorrected, loggers were on average, 7.32 ± 0.18 seconds out-of-synch with each other (figure 10). The amount each logger was suggested to drift was modelled against the times returned upon downloading Sirtrack logs (after removing two outliers, logger ID 68 and 86) (see figure 13). The linear model suggested the clock-drift identified by the synchronisation procedure Y could be predicted by the clock-drift suggested by the Sirtrack software X, ($Y \sim -12.38 + 0.313X$, Adjusted R^2 =0.763, p<0.0002).

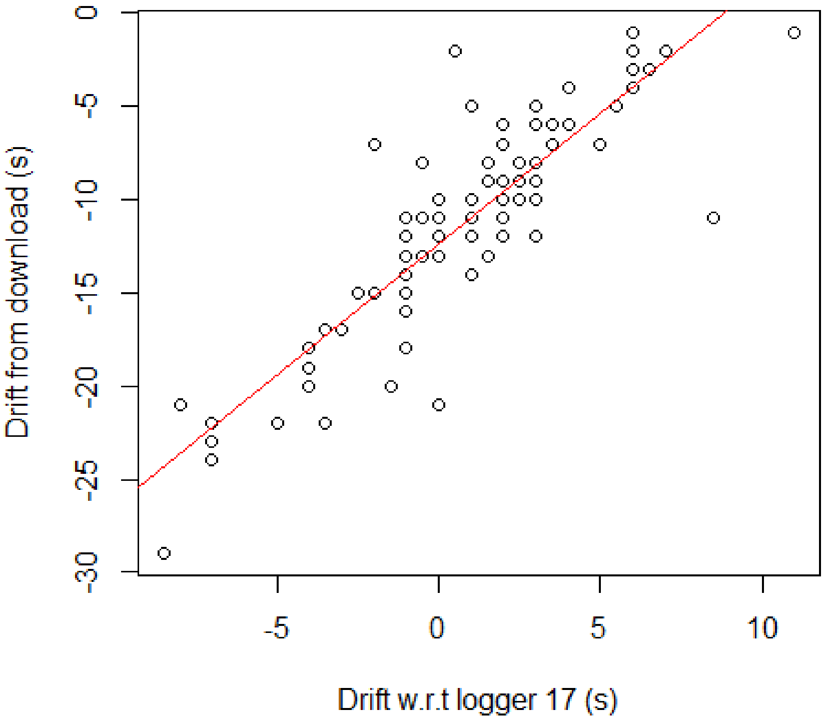


Figure 13: A plot showing the correlation between the clock-drift of loggers identified by logger 17 and those identified by the Sirtrack logger software. In red is the line of best fit from the regression. Two outliers have been omitted (logger 68 and 86).

Despite the high correlation between the offsets identified from the synchronisation procedure and those identified from the Sirtrack® software, adjusting the timing of logs by the software-offsets reduced the reciprocal agreement of loggers by 1.6% across the deployment. Given that the downloaded logs represent the amount of drifting that has occurred over the whole deployment, a temptation might be to take half of the software-offsets to approximate the amount of drift at the mid-point of the deployment. For example, if the Sirtrack® software indicates that a logger had drifted 20seconds over the course of the deployment, its logs would be adjusted by 10 seconds. However, taking mid-points still serves to reduce reciprocal agreement, by 0.8%. Adjusting the timings of logs relative to how much the focal logger was identified to drift by the most stable logger in the deployment increased reciprocal agreement by 0.7%, a total of 168,000 seconds; this may seem trivial, but the loggers were only deployed for 10 days following the fixture of new batteries and a reset of the clocks all to one laptop. In the earlier deployments of the loggers, it was noted that synchronising loggers to multiple laptops (which may be a necessity in larger deployments) could create large discrepancies in logger synchronicity.

One drawback of the current approach is that it assumes that the time-offset of loggers is constant throughout the course of the deployment, and thus corrects the timings of all logs from a given logger by a single value. A more intuitive approach would be to treat all logs as being synchronised at the beginning of the deployment (an option provided by the Sirtrack® software), and calculate the offsets between loggers over successive time-frames. Using this approach, a linear function could be fitted to represent the rate at which the clock of each logger has drifted since the beginning of the deployment. Using this function, the logs of loggers could be adjusted differently over time, perhaps leading to better alignment then the current method. One limitation of this approach is that, like the B_{ij} matrix, the T_{ij} matrix requires loggers to have encountered each other enough to find local maxima which reflect the time-displacement between loggers. Over short time-intervals, there may not be rich enough data to accurately estimate the T_{ij} matrix. One solution might be to incorporate the time offsets from the Sirtrack® software, loggers could be set up and the drift over time recorded at intervals, through repeated connections with a laptop containing the Sirtrack® software. This could reveal more about the patterns of clock drift over time (e.g. whether it is linear, a function of logger activity etc.), to the extent where we might be able to accurately fit functions to describe clock drift using just the data returned upon retrieving the loggers post-deployment.

In-field Validation of the correction procedure

Sixty-seven of the 95 cows were recorded in the videos. 195 of the 519 data entries were logs of cows estimated to be less than 2.5m apart in the videos - involving 43 different loggers. The average distance between cows observed in the video was 3.35m±0.09m standard error. The temporal correction algorithm was validated against this dataset. The observed V for the deployment was 10247.54.

The mean null value of V from the null model was 5197.64, with the upper 95% CI 6175.46. The p-value of our observed V of 10247.54 was less than 0.0002 given the null; this indicates that our weighting strategy is significantly better than chance. Removing the weakest logs in the deployment reduced our test statistic V by 0.097% (C=10237.54), whilst reducing the total number of logs in the deployment by 17.294%. The null model produced a mean null value of V of 8484.68, with a 95% critical value 9741.01, the probability that our observed value of V could have occurred as the product of the null model was 0.0032.

Overall, both the time synchronisation procedure and the weighting procedure appear to increase the reliability of contact logs. The procedure can be used to generate a symmetrical set of contact logs suitable to create both temporal networks. However, the B_{ij} matrix used to inform this correction requires that the loggers are deployed for a duration of time long enough to produce rich association matrices, as the method to calculate logger biases is based on Boyland et al. (2013). However, if the data are sparse, the Aij matrix from the detection distances of loggers could provide a replacement as demonstrated in 6.3. In this study, the A_{ij} and B_{ij} matrices showed a similar level of correlation to the A_{ij} and B_{ij} matrices in the 20-logger deployment in section 6.3 with an R_S of 0.41(compared to 0.52) and a p-value of 0.052.

6.5 Discussion

In this chapter, we have discovered that using Sirtrack® proximity logger uncorrected in a multivariate modelling approach will introduce biases that would affect the interpretation of the model. In static networks, the association matrices which may be use to define the strength of a relationship between individuals will be confounded by the loggers affixed to animals. We have demonstrated that the variable performance of loggers introduces logging biases which in turn create false variation in sociality: oversampling loggers will have inflated degree-strengths and pairs of over-sampling loggers would record stronger associations than pairs of under-sampling loggers. Without correction prior to analyses, it would be difficult to tease apart *true* social patterns from those created by the loggers.

In temporal network studies, the problems caused by logging biases are magnified further. The variable performance across loggers not only leads to over and under-sampling of association times, but also significantly affects the frequency and duration of recorded contacts. That is not to say that we are the first to discover reliability issues in Sirtrack® loggers (see Prange et al. 2006; Watson-Haigh et al. 2012; Drewe et al. 2012). Drewe et al. (2012) stated that loggers were less reliable at consistently determining the true frequency and duration of contacts. We have shown that this property varies between loggers and is linked to their logging bias. Thus, one of the conclusions in Drewe et al. (2012) that the length of the contact recorded may be a more accurate parameter to use than frequency should be taken with great caution, as the lengths of contact recorded are significantly correlated with the logging bias of loggers, and to a greater extent than frequency.

Overall, the issues and corrections presented in this chapter support an ongoing trend in research highlighting the need for appropriate testing of new remote-sensing technologies, that extend far beyond Sirtrack ® radio proximity loggers (see James et al. 2015 for an example). The risk of the sampling method biasing the networks produced is not a new one for animal social network studies and the advent of remote sensing as a means to sample social network data has brought with it many substantial advantages over previous methodologies.

IV. General discussion and conclusions

7. General discussion and conclusions

7.1 Preface

The objective of this thesis has been to explain the structure of animal networks in terms of a number of "external" explanatory variables. Recall that we use the term "external" to describe traits which affect network structure that are not the result of the pre-existing network structure itself- which we refer to as "structural" variables (see section 2.3.1). In many senses we have achieved this goal; we have found multiple factors linked to the structure of the hybridisation network of the new world warblers (chapter 4) and multiple factors associated with inter-nest drifting by *P.canadensis* workers (chapter 5). Through the course of these chapters, multivariate statistical modelling has been used to evaluate the effect of each given explanatory factor on a given network whilst controlling for the combined effect of other factors. There are a number of network models, many of which have been developed in the social sciences, which allow us to perform multivariate statistical analyses on relational datasets such as networks (see Snijders 2011 for a review). However, through performing robust statistical analyses in chapters 4 to 6, we have demonstrated the risks associated with applying these statistical modelling techniques to animal networks without careful consideration for both the system studied and protocol used to collect network-data.

In this final chapter, I will argue that further developments in MRQAP methodologies provide an avenue to understanding more advanced questions regarding the structure and functional consequences of animal networks. In each substantive chapter of this thesis, we have presented a different method to allow multivariate analyses on animal networks whilst addressing issues associated with animal network data. The first method involved using modelling approaches to remove the effect of controlled variables on each explanatory variable, and then using each of these modified explanatory variables in null model based hypothesis testing (NMBHT) on the observed network (see Croft et al. 2011). This approach would be useful for studies where explanatory variables could be measured accurately, but the network itself may contain issues which prevent its inclusion in a model-based approach- such as a high incidence of missing data. The second method involved tailoring aspects of one multiple regression quadratic assignment procedure (the MRQAP) to incorporate a more suitable null model for the system being studied; this prevented the model statistics being tested against an irrelevant reference distribution containing measures from impossible network combinations. The third approach considered ways that data collected using radio-based proximity loggers could be corrected to remove sampling biases; given that models in the social sciences assume even sampling, the resulting data would be more suitable for these approaches.

The goal of this chapter is to show that further developments to the MRQAP provides a promising avenue to testing more advanced questions regarding the structure and functional consequences of animal networks. We will first revisit some of the issues encountered in each of the three systems studied and the potential consequences of failing to consider them at the point of analysis. We then consider in more detail how the approach taken in chapter 5, to modify the null model behind the MRQAP could perhaps provide a general solution to many of sampling and reliability issues that arise when trying to perform multivariate analyses on animal networks. The potential merit in replacing the null model behind a MRQAP to deal with issues inherent in animal network data was first highlighted in Croft et al. (2011). More recently, Farine and Whitehead (2015) suggested replacing the QAP in an MRQAP with data-stream permutation methods often used in NMBHT with gambit of the group (GoG) data (see Bejder, Fletcher, and Brager 1998; Croft, James, and Krause 2008; Farine 2013; Croft et al. 2011). In this thesis, we have been more concerned with analysing networks inferred through observing dyadic interactions (see chapter 5). Through developing a thorough understanding of the MRQAP and NMBHT, we can contribute to Farine and Whitehead’s (2015) idea of adding data-stream permutation methods to the QAP by suggesting some of the potential issues which may present themselves using non-standard (e.g. raw-data permutation) approaches and how they might be avoided if they are found.

The issues we cover fit into a wider debate regarding the use of multivariate models in the field of studying animal social networks, which can be traced to at least as far back as Croft et al. (2008). Although we are concerned with animal networks in general (not necessarily social), many of the issues raised pertain to networks created from unevenly-sampled data, which unites both hybridisation networks (using currently available data) with social networks. One recent protagonist of this debate regarding modelling approaches on animal social networks was a review by Pinter-Wollman et al. (2014), who led a call for animal social network studies to incorporate more predictive analytical approaches - such as those developed in the social sciences. The article received a lot of commentary, some of which rightly centred on the idea that just because we can use these models on animal social networks, it doesn’t mean that the outputs will be interpretable in light of uneven-sampling and uncertainty in animal data (see Rendell and Gero 2013; Krause et al. 2014). These responses to Pinter-Wollman et al. (2014) prompted a reply from the authors, which was in part a concession to some of these issues and in (greater) part a pointer to where these models could be used as they stand; namely, in closed laboratory settings and with the use of remote sensing technologies (see Krause et al. 2013). Although the positives of remote sensing technologies have been demonstrated in chapter 5, we have shown that remote sensing technologies by no means “close the sampling gap” as expected by Pinter-Wollman et al. (2014) (see chapter 6). Our findings support a broader trend of researchers finding issue with remote-sensing technologies (see Prange et al. 2006; Böhm et al. 2009; Drewe et al. 2012; Watson-Haigh, O’Neill, and Kadarmideen 2012; Boyland et al. 2013; Rutz et al. 2015) and highlight the need for appropriate testing prior to deployment.

The debate regarding modelling may have also prompted the very recent “how-to” review by Farine and Whitehead (2015) titled: “constructing, conducting and interpreting animal social network analysis”; this review provides, amongst other things, suggestions and demonstrations as to how modelling approaches may be applied to animal networks whilst addressing sampling issues associated with animal data. The suggestions in this review are well-considered. Above, we mention that one particular overlap between this thesis and their review is that we both see the merit in modifying MRQAPs to increase their suitability for animal network studies. In this thesis, we make some more detailed suggestions as to how the MRQAP might be developed and issues that might arise in development and their potential solutions should they arise.

7.2 The consequences of ignoring data-quality issues

The stance currently adopted in this thesis is similar to Rendell and Gero (2013) and Krause et al. (2014), that it is dangerous to apply modelling approaches developed for human social networks to many animal networks. Throughout this thesis, the consequences of ignoring issues with animal social network data have been profound, with both multivariate and univariate analyses providing qualitatively different results to analyses which have accounted for issues inherent in animal network data. We use empirical cases to show where results have differed, but some of the consequences of failing to consider the nature of animal network data in statistical analyses have now also been shown via simulation in Farine and Whitehead (2015). In this section, a recap of the issues encountered in this thesis and the potential consequences of ignoring them are highlighted.

In chapter 4, we stated that the absence of reported hybrids between a pair of species may not be indicative of absent hybridisation, as only one hybrid report over a large timeframe would be required to confirm the presence of an edge between wide-ranging species. One recommendation to escape an over-dependence on absent reports is to use weighted edges (James et al. 2009). However, the majority of hybrid data are qualitative reports, with the extent of crosses ranging from “single” to “ongoing” to “extensive ongoing” (see McCarthy 2006). Willis, Symula, and Lovette (2014) performed a partial Mantel test (Mantel 1967; Legendre & Legendre 2012) on binarised report data, placing equal trust in the presence and absence of hybrid reports. Through using a partial Mantel test, they identified that hybridising species had significantly more similar songs after shared ancestry was accounted. However, in our analysis, which was based only upon the presence of reports in the same hybrid dataset, we found that species did not have more similar song than expected after accounting for shared ancestry. It may not be possible to know which analysis is correct, but it appears to be the case that the conclusions of Willis, Symula, and Lovette (2014) are dependent on a trust in unreliable data.

The new method we presented in chapter 4 used MRQAPs to modify the explanatory variables used in our study, so that their unique contributions could be assessed on the response variable (the hybrid network). In this sense, the goal of the procedure was similar to a principal components analysis (PCA), to remove the covariance between explanatory variables. However, our approach was tailored to account for specific confounding issues in the data, such as shared phylogenetic history affecting song and plumage similarity between species. Furthermore, our method produced outputs more readily interpretable (such as the "plumage similarity not explained by ancestry") than the "loadings" provided from a PCA (see Jolliffe 2002). The explanatory variables were more suitable for the MRQAP as they contained no missing data, were sampled evenly and, as far as possible, measured without bias. In contrast, the response variable contained a high perceived likelihood of false negatives. Therefore, we believe that using modelling approaches which assume that both the presence and absence of edges are equally reliable would not be suitable given the response variable. Instead, we sacrificed the ability to produce a predictive model with effect sizes so that NMBHT could be used (see Krause, Wilson, and Croft 2011) on the observed network. A test statistic which focussed only on the presence of reports and a null model was developed around an edge randomisation (which is also a raw-data permutation in this instance). Edge randomisations are useful for animal network data as they place less confidence in the observed network structure than node-label permutations (see Croft et al. 2011). Molecular methods will, in time, allow us to quantify the extent of meaningful hybridisation across species through measuring the amount of genetic introgression between species. Once datasets of this variety become widely available, it will be possible to use methods such as the MRQAP directly on the response variable (the hybrid network). In the meantime, using sightings of individual species to as a measure of sampling efforts for hybrids may provide a useful extension to some of the approaches presented in chapter 4.

In chapter 5, the pattern of inter-nest drifting in wasps was found to be correlated with distance, nest-size, and the difference in the worker-brood ratio of nests. In this chapter, a traditional social science null model was used in the MRQAP: the double-semi partialling QAP (QAP-DSP) (Dekker et al. 2007). The justification for a "traditional" null model was that there was a significant correlation between explanatory variables, perhaps indicative of collinearity to which the QAP-DSP has been demonstrated robust (Dekker et al. 2007). Further, in this study, we believed there was even sampling of drifting using passive RFID systems. A block-structure was incorporated into the MRQAP in a similar way to which block structures have been added to ERGMs (Hunter et al. 2013). Using our modified model, we found a significant effect of worker-brood ratio on drifting patterns. However, this effect was lost following the experimental removal of some of the workers and brood from nests. The loss of this effect highlights the importance of carrying out manipulative experiments on networks to ascertain the causes of network structure, as highlighted in Farine and Whitehead (2015), Rands (2014) and in the response of Pinter-Wollman et al. (2014). In the analysis of this manipulation in Lengronne (2013-chapter 3), a chi-squared test was used to see if drifting patterns changed in hypothesised directions. This test firstly ignored the dependency in network data when assessing significance (see section 2.3). Secondly, the analysis ignored the potential the effect that removing workers would have on the number leaving each nest regardless of drifting. Using a network-based analytical approach, the MRQAP, contradicted the findings of Lengronne (2013), suggesting that the manipulation caused a loss in the tendency for drifting patterns to respond to the worker-brood ratio difference of nests; although other factors pertaining to nest-level need were identified. In the chi-squared test, the alternative hypothesis was accepted that the manipulation caused an adaptive change in the structure of the drifting network.

Through adding a block-structure to the MRQAP, we have made the first steps in incorporating biologically realistic null models (see Croft et al. 2008; Croft et al 2011) to network modelling approaches. Through doing so, the significance of coefficients and a model as a whole have been evaluated against a reference distribution created only from null networks which could possibly occur in nature as opposed to those which could never occur. For example, an unconstrained QAP procedure would essentially allow wasps drifting back in time and/or over huge distances if not for the block-structure. Clearly, this step was necessary or the null model would not have accurately reflected the null hypothesis and reference distribution not relevant to the question being tested. That is not to say we can just approach any set of data and tailor a bespoke null model to underpin a MRQAP without further consideration. The approach taken in this thesis has relied on a well-characterised QAP-DSP null model (Dekker et al. 2007), which has been shown to be robust to a range of issues that can be present in the data; other null-models may be more sensitive to some of these issues. We discuss this idea further in section 7.3.

In chapter 6, issues with Sirtrack® loggers, one of many remote sensing technologies available to collect animal association data would have resulted in the production of networks with artificial variation in sociality if not corrected. Given that Pinter-Wollman et al. (2014) suggested that these technologies might close the sampling gap between animal and human network studies, it is clear that the gap is perhaps larger than initially anticipated at the time of Krause et al. (2013). The data we obtained using Sirtrack® radio-proximity loggers were not suitable for direct use in a MRQAP, as the variable performance of loggers introduced bias to the association matrices (Boyland et al. 2013). Logging biases were correlated with the frequency and duration of contacts logged - potentially affecting temporal network measures. Our finding supports Rendell and Gero's (2014) statement: “just because we *can* run an analysis, it does not mean that we can *interpret* it correctly” in that we conclude that animal network data may often not be suitable for modelling approaches developed in the social sciences. It should be noted that a lot of studies into disease transmission have used Sirtrack® loggers (e.g. Hamede et al. 2009; Böhm et al. 2009) and given that centrality is very important to spreading dynamics across networks (Pastor-Satorras & Vespignani 2001; Newman et al. 2002), we can only hope that their proximity loggers sampled more consistently than our own.

The first correction for Sirtrack® loggers presented in chapter 6 was for static networks (Boyland et al. 2013), which reduced the time recorded by each logger relative to how much each logger oversampled interactions compared to the weakest in the deployment. Following the use of this correction, the data might be more suitable to for use in a modelling procedure like the MRQAP. In the one of two subsequent submitted manuscripts, this static correction is used as a prerequisite to the multivariate modelling of the social structure of dairy cattle (*Bos taurus*). The multivariate analysis in this manuscript uses a Bayesian mixed effects model (Hadfield 2014) with resampling of the response variable to determine the significance of both variables and the model.

The second correction for Sirtrack® loggers was for temporal networks. The procedure seemed to perform well, preferentially removing unsubstantiated interactions as shown through comparison of the logs with video footage. Without the use of this correction, temporal network models based on Sirtrack® logger data, such as the relational events model in Patison et al. (2015), may be confounded by the performance of the loggers affixed to individual cows. Automated data-collection methods perhaps provide the best candidate datasets for models coming across from the social sciences. However, sampling must be made consistent across loggers, either through improving the hardware in these technologies, or developing post-deployment procedures such as those presented in this thesis (and also in Rutz et al. 2015) to correct for characterised issues with these technologies. Many authors have highlighted that thresholding data, particularly on the basis of edge values, can cause high incidence of both type 1 and type 2 error (Butts 2009; Langer et al. 2013; Farine 2014; Farine & Whitehead 2015). In some sense, the temporal correction we have developed does remove contacts which receive a reliability-score lower than a given threshold. However, the reliability-score of contacts are scaled down relative to how much each logger oversamples interaction relative to the weakest logger in the deployment. In this sense, the contacts removed would likely have never existed if all loggers performed equal to the weakest in the deployment.

It may be likely that some steps taken in this thesis are overly- conservative. In chapter 4, most of the pairs of species that have not been reported to hybridise may truly not hybridise; as such, there may be some useful information buried in the absent reports which we have broadly treated as missing data. Perhaps a modelling approach could be used on a dataset consisting of just the pairs of species which hybridise frequently. Under this definition, we might be more confident that the absence of observed frequent hybridisation might be genuine, but we would also have to re-evaluate the biology of what we are studying. Removing reported hybrids for reasons other than uncertainty might amount to a form of network thresholding, which we have mentioned above carries its own risks.

In chapter 5, a significant correlation between nest size and worker-brood ratio difference meant that a QAP-DSP was chosen as the null model in the MRQAP. However, the correlations were only weak ($r \leq 0.25$) and may not have been indicative of a linear relationship. The QAP-DSP as a null model is a good choice due to its robustness to skewed and spurious data (Dekker et al. 2007), but choosing this well-characterised null model hindered us from exploring the use of more biologically interpretable null models such as randomising the destinations of drifting wasps. To advance the use of MRQAP-type procedures in animal networks, two things are needed: a greater understanding of how sampling and reliability issues in animal data affect these network modelling approaches (see Farine and Whitehead 2015 for a review of where we currently stand)

and a better understanding how animal-network null models (see Croft et al. 2011) perform in light of issues which standard social science null models have been tested against. With this information, it will be possible to make more informed null model choices.

Despite the potential for our approaches to be overly-conservative, we have discovered new biology in both of the systems we have studied. In chapter 4, we are the first to find a correlative link between male breeding plumage similarity and hybridization in the new world warblers (family: *Parulidae*). The implications of this finding and other findings such as (Martin et al. 2015) suggest that the diverse colourful plumage of wood warblers might be driven through selection for reproductive isolation in combination to more traditional explanations such as sexual selection (Shutler & Weatherhead 1990). In chapter 5, we were the first to identify that the network of inter-nest drifting in the paper-wasp *Polistes canadensis* is partitioned into what we called “aggregations” on the basis of the native buildings of nests. Incorporating this structure into a modified MRQAP, we identified correlates of drifting in *Polistes canadensis* which suggest that the drifting networks of these species shows an adaptive response in the face of catastrophic events (i.e. small nests start attracting more help).

7.3 Adapting the MRQAP: issues and future directions

In this section, the potential merit in adapting the MRQAP is outlined, along with the considerations which should take place in this process. In a MRQAP, many of the assumptions which affect the fitting of a multiple linear regression (MLR) will apply - as both techniques use the same estimation procedure to fit the model. In section 2.3.1, we cover these issues in detail, which include: linearity, multi-normality, negligible error in the explanatory variables, and independent explanatory variables (no collinearity). However, issues which concern significance testing in an MRQAP are different to those in a typical MLR, as an MRQAP uses a node-label permutation (QAP) to generate a reference distribution as opposed to using distributional assumptions when evaluating the significance of t-statistics, F statistics, and partial-correlation coefficients. Recall that the reason for this is that relational data are not independent; values along rows and columns may be more similar to each other due to the presence of a shared actor (known as row or column autocorrelation).

One of the key messages in this thesis developed from Croft et al. (2011) is that there is scope to replace the generic QAP null model which underpins an MRQAP with one which can better account for aspects of the sampling protocol used to produce animal social networks (see Croft et al. 2011 for examples). In chapter 5, we make a cautious first step in demonstrating this approach through incorporating a block-structure into the MRQAP. All of the networks analysed in this thesis (chapters 4 and 5) are examples where edges have been inferred through dyadic interactions or associations as opposed to through group co-occurrence (using "gambit of the group") (see Whitehead and Dufault 1999). As such, we have not had the reason to apply a data-stream permutation procedure to a MRQAP as recommended by Farine and Whitehead (2015). In chapter 5, where the unreliability of the sampling procedure is known (the 80% cover of the antennae), a QAP could be replaced with a null model which reflects that there is an 80% chance a drifter will be recorded either entering or leaving a non-natal nest. However, this null model might be excessive given the 96% chance that drifting wasps are detected at another nest on either entry or exit. In situations where coverage at each nest could vary, a null model which accounts for the differences in antennae coverage could be very useful.

A general note regarding the replacement of the QAP with other null models is that we might re-expose ourselves to some of the issues which the QAP and variations of the QAP (such as residual-permutation based methods (Dekker, Krackhardt, and Snijders 2007)) have been developed to control. For instance, the first thing we would likely sacrifice is the preservation of structural (row-column) autocorrelation (as outlined in Krackhardt (1988)); this will mean that the observed dataset may be compared to a reference distribution generated from datasets with less structural autocorrelation. How much of a problem this will pose is currently a grey area. It may very well depend on the dataset, but given that OLS estimation is unbiased by heteroscedasticity (which will be caused by structural autocorrelation), there may be circumstances where this is not a problem. Testing the addition of a new null model's susceptibility to autocorrelation in a similar way to Krackhardt (1988) could perhaps confirm this.

When working with the MRQAP we should bear in mind that the MRQAP was developed by Krackhardt (1988) under a very restrictive set of conditions; using simulated datasets where there was no correlation among *any* of the variables. Dekker, Krackhardt, and Snijders (2007) tested the performance of the MRQAP using various different versions of the QAP procedure (Y-permutation, X-permutation, residual permutations) on datasets which exhibited some non-desirable qualities: collinearity, spuriousness (a confounder effect) and skewness; the term "spuriousness" referring to correlation between a response variable and an explanatory variable caused by the presence of another explanatory variable. Their tests showed that residual permutation procedures (such as the double semi-partialling (DSP) used in chapter 5) were the most robust to spuriousness and skewness in the data, but that all QAP-based methods (including residual-based methods) performed poorly with highly skewed and spurious data. They also found that a QAP permutation of the response matrix performed well when the test statistic used (e.g. a t-statistic) was pivotal. Recall that a statistic is defined to be pivotal when the distribution of the statistic under the null hypothesis is independent of the "nuisance" parameters (a t-statistic and a partial correlation coefficient are pivotal statistics for a multivariate normal distribution). "Pure" QAP permutations of explanatory variables performed poorly, with high rates of type 1 error. Dekker et al. (2007) suggested that this is, in part, because permuting explanatory variables breaks down the dependency that exists between them when collinearity is present, violating the ancillarity principle (Welch 1990; Ter Braak 1992; Legendre & Anderson 1999). Most animal network based null models are conducted on the response variable, as such we would not expect to violate the ancillarity principle, but may want to consider how susceptible the new null model might be to spuriousness and skewness in the data.

A MRQAP with a different null model would require appropriate naming. Examples might include a MRRDP in cases where a raw-data permutation is used in place of a QAP as a null model, as recommended in Croft et al. (2011). A MRRDP and other novel variations of the MRQAP could be sensitive to the aforementioned issues highlighted by Dekker et al. (2007). The QAP has been shown to have reduced efficacy when there is significant skewness in the data, the test statistic for the procedure is non-pivotal (affected by factors outside of the model), or if there is collinearity between independent variables.

In these situations, the DSP procedure is a robust alternative (Dekker et al. 2007). However, given suitable data (no skewness, collinearity etc.), there is no reason why alternative permutation procedures could not be applied to the response variable. In fact, there is no particular reason why residual-permutation procedures, such as the DSP, should just be restricted to QAPs and not applied to null models tailored for animal social networks.

One particularly promising residual permutation procedure which has similar robustness to the DSP is the Freedman-Lane semi partialling (FLSP) procedure (Freedman & Lane 1983; Dekker et al. 2007); this procedure is based on the response variable, which would be the most likely target of new null models (as advocated in Manly 1997; Bejder, Fletcher, and Bräger 1998; Croft et al. 2011; Farine 2013). However, caution and testing would be advised as the findings of Dekker, Krackhardt, and Snijders (2007) are based only on performing the DSP and FLSP in conjunction with a QAP procedure. It is not certain whether using a residual-based permutation procedure on top of, let's say, a raw-data permutation would have the same favourable qualities as when they are used with a QAP. In this sense, pressure-testing the incorporation of novel permutation procedures using similar approaches as those outlined in Dekker et al. (2007) might be advised.

A more trivial issue which concerns the MRQAP (and any modifications we may seek to make) are missing data. If any data in the explanatory or response matrices are missing, then the corresponding elements (where actor and recipient are the same) in all other matrices may be removed before the model is fitted (as in Butts 2014). If missing values are present, then by default, they will be permuted in the QAP procedure, leading to the random omission of data points at the fitting stage where one or more variables contain missing data; this means that the permuted datasets can vary in size and composition. Although the effect of evaluating the fit of a model on null datasets of varying size and composition has not explicitly been characterised, we can safely assume that this issue will reduce the trust that can be placed in the reference distribution created by the QAP. Solutions to this problem could be to adjust the permutation so that all elements corresponding to missing values are fixed across all explanatory and response matrices, as such the dataset will always contain the same set of values after each QAP permutation – with the same data being removed every time. However, the implementation of MRQAP in the “sna” package in R currently just provides a warning if any elements in the response or explanatory matrices are left empty (see Butts 2014). If the QAP was to be replaced, it should be considered how non-sampled or missing data in the square matrices will be dealt with in the new null model.

A final minor note of caution regards degrees of freedom. In chapter 5, the degrees of freedom were not the same as in a typical MRQAP, as the dataset was not a square matrix anymore, instead being restricted to the lead diagonal of in a block structure. In any larger replacement of a null model, the degrees of freedom in the analyses may be affected by the potential values which can exist and are subject to the permutation procedure and those which are otherwise disallowed, perhaps being physically or biologically impossible. However, the value behind knowing the degrees of freedom in the model is lessened given that these are not used in calculations of significance when permutation procedures are used to make reference distributions.

7.4 A note on Exponential random graph models (ERGMs)

One modelling approach not mentioned in Farine and Whitehead (2015) which may prove fruitful for modification to suit animal networks are the exponential random graph family of network models (ERGMs). Perhaps the best introductory text for the non-statistician on this class of model is Robins et al. (2007). ERGMs are powerful in the fact that they allow structural variables to be tested meaningfully given the assumption that the network is generated by a stochastic process. In this process, relational ties come into being in a way that may be shaped by the presence or absence of other existing ties (and possibly node-level attributes). This assumption replicates how might think of a network forming over time and thus allows meaningful testing of the endogenous feedback events that take place in the process (such as ties being reciprocated) of network formation. Another potentially attractive quality of ERGMs is that they treat the observed network as a structure which contains uncertainty - a common feature of many animal social networks (Lusseau et al. 2008). ERGMs have already been used in some animal network studies (e.g. Wey and Blumstein 2010; Dey and Quinn 2014).

That is not to say that ERGMs do not have their limitations, Cranmer and Desmarais (2011) highlight their inability to handle missing data and issues with "degeneracy", where poor model specifications prevent model fitting. A past limitation to ERGMs has been their inflexibility to different types of network data. Until recently, ERGMs were restricted to binary static networks without any sort of block structure (i.e. individuals being sampled from different classes). However, in the last few years, expansions have been proposed to ERGMs which have facilitated weighted edges (as counts) (Krivitsky et al. 2012), dynamic networks (known as a tERGM) (Hanneke, Fu, and Xing 2010), dynamic networks with weighted edges (known as stERGM) (Krivitsky and Handcock 2014), hierarchical designs (known as a hERGM) (Schweinberger and Schweinberger 2015) and designs with a block structure (known as a blkERGM) (Hunter, Goodreau, and Handcock 2013).

It is this rapid advancement of the ERGM family of models which is both testament to their potential and flexibility. As it stands, the most direct route to using these models on animal network data may be through rigorous data-cleaning prior to analysis (such as the emphasis in chapter 6), or perhaps trying to account for aspects of sampling bias through the use of covariates. One common assumption used in ERGMs is "*homogeneity of isomorphic network configurations"*, which means that each node will respond homogenously to the relative factors in the model (i.e. males will be as likely reciprocate ties as females); this assumption is used to create robust models with few parameters. However, the assumption is optional and dropping it may provide an avenue to represent some sampling issues. For example, if the females of a species are more conspicuous than males and, as a result, ties observed between females may be under-sampled, then allowing ties between females to be parameterised differently to ties between males or mixed ties may be useful.

Making modifications or extensions to ERGMs for use with animal network datasets will require a greater depth of statistical understanding than with the MRQAP. Therefore, to echo the sentiments of Krause et al. (2014); the most exciting developments with respect to ERGMs would come from the close collaboration of those who have developed null models for animal networks (e.g. Bejder, Fletcher, and Brager 1998; Croft et al. 2011; Farine and Whitehead 2015) and those who have worked on extending and developing

ERGMs (e.g. Holland and Leinhardt 1981; Wasserman and Pattison 1996; Snijders 2002; Krivitsky et al. 2012; Hunter, Goodreau, and Handcock 2013).

7.5 Future Directions

The future role of the MRQAP in animal network studies may be limited as ERGMs can now account for both “external” and “structural” variables, whereas an MRQAP will only ever be able to consider the former. However, the beauty of the MRQAP is that it is simple and well-characterised; this makes it an ideal foundation for incorporating new null models which can control for issues in animal network studies. Although we have referred to these models as MRQAPs, the use of “pure” QAPs in these models may have been confined to history since the introduction of residual-based methods such as the QAP-DSP. In the “sna” package (Butts 2014) in R (R Development Core Team. 2008) the default implementation is now a QAP-DSP. In specific cases where we are very confident in the structure of the network (see Croft et al. 2011) and are not interested in the effect of structural variables, the MRQAP is and will continue to be a trusted model choice. Examples might include captive populations of animals which are both held in confined spaces, which might lead to saturated networks, and can be observed intensively through the use of video recording or other automated technologies.

The development which will spell the end of the MRQAP as we know it will be its extension to include different data types (count-data, binary data, binomial data etc.) through the use of generalised linear models (see McCullagh and Nelder 1989). In static networks, the ever-growing adoption of Bayesian modelling approaches (such as the mixed effects model by Hadfield (2014)) have just began to be applied seriously to network data, some initial studies have coupled the approach with resampling (permutation) procedures to determine significance in light of network dependencies (e.g. Willis, Symula, and Lovette 2014), some have not (e.g. Rushmore et al. 2013). Basic resampling methods are broadly untested and may yet fall short of the complexity required to account for issues in animal data. Upon the use of sender and receiver random effects in the place of permutation (e.g. Rushmore et al. 2013), a questionable assumption is made that none of the structural dependencies outlined in section 2.3.2 and Snijders (2011) exist in the data. One lesson we can learn from the MRQAP is that its name and structure places emphasis rightly on the null model used for significance testing. The reappearance of ill-considered, unconstrained resampling procedures on the front of a Bayesian mixed effects model sends us backwards to the “straw-man” null models which the likes of Bejder et al. (1998), Croft et al. (2008) and Croft et al. (2011) argued against and led us beyond in past years. Thankfully, Farine and Whitehead (2015) have recently demonstrated how more sophisticated null models can be used with these approaches to test model statistics whilst controlling for the sampling issues in the data, and we hope that these are considered by many in the field.

The future of network analysis may go in a number of directions, it is clear that the remote sensing of animal interactions still has a lot of contributions to make and, despite some early issues, will only improve in efficacy either through technological innovation and/or through the way we interpret the data; this will result in a greater adoption of multivariate modelling approaches. There will likely be a large increase in temporal network studies facilitated by the new datasets generated- which will also prompt the continued development of new temporal statistical procedures. Behaviour observed in other types of dynamic systems such as synchronisation, resonance and damping (see Newman 2008) may all show themselves in studies of animal temporal networks and become categorised as structural factors similar to those described in Snijders (2011) for static networks. Further, we currently know little about the driving mechanisms behind temporal networks, and future directions may include asking why do interactions and relationships between animals happen when they do.

It is clear that the predictive, multivariate approaches coming out of the social sciences are powerful multivariate predictive tools. However, powerful tools are only useful if used correctly and can be all the more dangerous when used inappropriately. Thus, despite the large amount of investigative work needed, we need to familiarise ourselves with how sampling issues which affect animal networks can impede the performance of modelling procedures and how these issues might be remediated. Currently, it is a shame that the study of animal networks often requires intensive or expensive sampling procedures to know something about the network structure, but then the statistical analysis which can be safely applied to this data often has to be less ambitious due to a current inability to address issues pertaining to the modelling of animal networks. The field continues to move fast, in the last three years there has been much energy placed into moving away from descriptive approaches. However, this energy has perhaps been misplaced and needs to be channelled into rigorous testing and development of network models rather than just thinking about their biological applications. Ultimately, the hopes of Krause et al. (2014) are shared: that, in future, modifications of network models will be truly robust to sampling issues, should they still be as prevalent, whilst retaining test power. It is also agreed that the most exciting developments in this field will be made not by behavioural ecologists on their own, but in conjunction with statisticians and those involved with the development of these modelling procedures (see Pinter-Wollman et al. 2014; Krause et al. 2014).

Bibliography

Abouheif, E. & Wray, G.A., 2002. Evolution of the gene network underlying wing polyphenism in ants. *Science (New York, N.Y.)*, 297(5579), pp.249–252.

Akre, R.D. et al., 1976. Behavior and colony development of Vespula pensylvanica and V. atropilosa (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, pp.63–84.

Albert, R. & Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), p.47.

Alberts, B. et al., 2002. *Molecular Biology of the Cell* 4th ed., New York: Garland Science.

Aliabadian, M. & Nijman, V., 2007. Avian hybrids: incidence and geographic distribution of hybridisation in birds. *Contributions to Zoology*, 76, pp.59–61.

Alin, A., 2010. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), pp.370–374.

Allen, F. & Babus, A., 2010. Networks in Finance. *Network-based Strategies and Competencies*, pp.1–20.

Amblard, F. & Deffuant, G., 2004. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 343(1-4), pp.725–738.

Amitai, G. et al., 2004. Network Analysis of Protein Structures Identifies Functional Residues. *Journal of Molecular Biology*, 344(4), pp.1135–1146.

An, C.A.R.I.B., 2008. Highly structured fission e fusion societies in an aerial-hawking , carnivorous bat. *Animal Behaviour*, 75(Symington 1990), pp.471–482.

Andersen, R.M. & May, R.M., 1988. Epidemiological parameters of HI V transmission. *Nature*, 333(6173), pp.514–519.

Anderson, R.M., May, R.M. & Anderson, B., 1992. *Infectious diseases of humans: dynamics and control*, Wiley Online Library.

Aplin, L.M. et al., 2013. Individual personalities predict social behaviour in wild networks of great tits (Parus major). *Ecology letters*, 16(11), pp.1365–1372.

Arnold, M.L., 1997. *Natural hybridization and evolution*, Oxford University Press.

Audacity Team, 2008. Audacity, ver 2.02. [Computer program]. Available at: http://audacityteam.org/ (accessed 10th October 2012).

Baack, E.J. & Rieseberg, L.H., 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, 17(6), pp.513–518.

Bailey, P.J. et al., 2001. Electronic identification and tracking of cattle in Victoria, Australia. In *Performance recording of animals: state of the art, 2000. Proceedings of the 32nd Biennial Session of ICAR, Bled, Slovenia, 14-19 May, 2000.* Wageningen Pers, pp. 327–330.

Baker, C., 2015. Dairy industry in the UK: statistics. *House of Commons Library: Social & General Statistics Section*, (January).

Ballentine, B. & Hill, G.E., 2003. Female mate choice in relation to structural plumage coloration in blue grosbeaks. *BioOne*, 105(3), pp.593–598.

Barabási, A.-L., Gulbahce, N. & Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics*, 12(1), pp.56–68.

Barbarossa, S. & Scutari, G., 2007. Bio-Inspired Sensor Network Design. *IEEE Signal Processing Magazine*, 24(3), pp.26–35.

Barnett, J.L. et al., 1986. Effects of social environment on welfare status and sexual behaviour of female pigs. I. Effects of group size. *Applied Animal Behaviour Science*, 16(3), pp.249–257.

Barton, N.H., 2001. The role of hybridization in evolution. *Molecular Ecology*, 10(3), pp.551–568.

Bascompte, J. et al., 2003. The nested assembly of plant--animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16), pp.9383–9387.

Beale, C.M. & Monaghan, P., 2004. Behavioural responses to human disturbance: a matter of choice? *Animal Behaviour*, 68(5), pp.1065–1069.

Beaver, R.A. et al., 1989. Insect-fungus relationships in the bark and ambrosia beetles. *Insect-fungus interactions*, pp.121–143.

Beehler, B.M. et al., 1986. *Birds of New Guinea*, Princeton University Press Princeton.

Bejder, L., Fletcher, D. & Brager, S., 1998. A method for testing association patterns of social mammals. *Animal Behavior*, 56, pp.719–725.

Bejder, L., Fletcher, D. & Bräger, S., 1998. A method for testing association patterns of social animals. *Animal Behaviour*, 56(3), pp.719–725.

Belevitch, V., 1962. Summary of the History of Circuit Theory. *Proceedings of the IRE*, 50(5), pp.848–855.

Bell, M. & Iida, Y., 1997. *Transportation Network Analysis*, Hoboken: John Wiley & Sons, Incorporated.

Bennett, a T. et al., 1997. Ultraviolet plumage colors predict mate preferences in starlings. *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), pp.8618–8621.

Bennett, N.C., Faulkes, C.G. & Molteno, A.J., 1996. Reproductive suppression in subordinate, non-breeding female Damaraland mole-rats: two components to a lifetime of socially induced infertility. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1376), pp.1599–1603.

Bergstrom, T.C., 2002. Evolution of social behavior: individual and group selection. *Journal of Economic Perspectives*, pp.67–88.

Berlow, E.L. et al., 2004. Interaction strengths in food webs: issues and opportunities. *Journal of animal ecology*, 73(3), pp.585–598.

Bermingham, E. et al., 1992. Vicariance biogeography in the Pleistocene and speciation in North American wood warblers: a test of Mengel's model. *Proceedings of the National Academy of Sciences of the United States of America*, 89(July), pp.6624–6628.

Berry, W.D. & Feldman, S., 1985. *Multiple regression in practice*, Sage.

Bert, T.M. et al., 1993. High frequency of gonadal neoplasia in a hard clam (Mercenaria spp.) hybrid zone. *Marine Biology*, 117(1), pp.97–104.

Best, E.C. et al., 2013. Social preference influences female community structure in a population of wild eastern grey kangaroos. *Animal Behaviour*, 86(5), pp.1031–1040.

Bettaney, E.M. et al., 2015. Processing and visualising association data from animal-borne proximity loggers. *Animal Biotelemetry*, 3(1), pp.1–11.

Biggs, N., Lloyd, E.K. & Wilson, R.J., 1976. *Graph Theory, 1736-1936*, Oxford University Press.

Bijma, P. & Wade, M.J., 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *Journal of evolutionary biology*, 21(5), pp.1175–1188.

Billingham, R.E., Poole, H.K. & Silvers, W.K., 1961. Transplantation Immunity, Immunological Tolerance, and Chicken× Turkey Interspecific Hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 47(7), pp.1039–1043.

Birmingham, A.L. et al., 2004. Drifting bumble bee (Hymenoptera: Apidae) workers in commercial greenhouses may be social parasites. *Canadian Journal of Zoology*, 82(12), pp.1843–1853.

Bivand, R. et al., 2014. Package "rgeos."

Blacher, P. et al., 2013. Drifting behaviour as an alternative reproductive strategy for social insect workers. *Proceedings of the Royal Society B: Biological Sciences*, 280(1771), p.20131888.

Blair, A.P., 1941. Variation, isolating mechanisms, and hybridization in certain toads. *Genetics*, 26(4), p.398.

Blonder, B. et al., 2012. Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, 3(6), pp.958–972.

Blüthgen, N., Menzel, F. & Blüthgen, N., 2006. Measuring specialization in species interaction networks. *BMC ecology*, 6, p.9.

Boehlert, G.W. et al., 2001. Autonomous pinniped environmental samplers: using instrumented animals as oceanographic data collectors. *Journal of atmospheric and oceanic technology*, 18(11), pp.1882–1893.

Böhm, M. et al., 2009. Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. *PLoS One*, 4(4), pp.e5016–e5016.

Böhm, M. et al., 2008. Dynamic interactions among badgers: implications for sociality and disease transmission. *Journal of Animal Ecology*, 77(4), pp.735–745.

Bolland, J.M., 1988. Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, 10(3), pp.233–253.

Bollobás, B., 1998a. *Modern graph theory*, Springer Science & Business Media.

Bollobás, B., 1998b. *Random graphs*, Springer.

Bollobás, B., 2001. The Evolution of Random Graphs—the Giant Component. In *Random Graphs*. Springer. New York.

Bonabeau, E. et al., 1997. Self-organization in social insects. *Trends in Ecology & Evolution*, 12(5), pp.188–193.

Bonavita-Cougourdan, A. et al., 1991. Cuticular hydrocarbons, social organization and ovarian development in a polistine wasp: Polistes dominulus Christ. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 100(4), pp.667–680.

Boomsma, J.J., 2007. Kin selection versus sexual selection: why the ends do not meet. *Current Biology,* 17(16), pp.673-683.

Boomsma, J.J., 2009. Lifetime monogamy and the evolution of eusociality. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 364(1533), pp.3191-3207.

Borgatti, S.P. et al., 2009. Network analysis in the social sciences. *Science (New York, N.Y.)*, 323(5916), pp.892–895.

Borgatti, S.P., Everett, M.G. & Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis. *Harvard Analytic Technologies*, 2006, p.SNA Analysis software.

Borror, D.J. & Gunn, W.W.H., 1985. *Songs of the warblers of North America*, Cornell Laboratory of Ornithology.

Bourke, A.F.G., 2011. The validity and value of inclusive fitness theory. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723), pp.3313–3320.

Bouttier, J., Di Francesco, P. & Guitter, E., 2003. Geodesic distance in planar graphs. *Nuclear Physics B*, 663(3), pp.535–567.

Box, G.E.P. & Watson, G.S., 1962. Robustness to non-normality of regression tests. *Biometrika*, 49(1-2), pp.93–106.

Boyd, R. & Richerson, P.J., 1989. The evolution of indirect reciprocity. *Social Networks*, 11(3), pp.213–236.

Boyland, N.K. et al., 2013. Spatial proximity loggers for recording animal social networks: consequences of inter-logger variation in performance. *Behavioral Ecology and Sociobiology*, 67(11), pp.1877–1890.

Ter Braak, C.J.F., 1992. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and related techniques*. Springer, pp. 79–85.

Brambilla, M. et al., 2008. Song perception among incipient species as a mechanism for reproductive isolation. *Journal of evolutionary biology*, 21(3), pp.651–657.

Brauer, F., Castillo-Chavez, C. & Castillo-Chavez, C., 2001. *Mathematical models in population biology and epidemiology*, Springer.

Brown, W.L. & Wilson, E.O., 1956. Character displacement. *Systematic zoology*, 5(2), pp.49–64.

Bruschini, C., Cervo, R. & Turillazzi, S., 2006. Evidence of alarm pheromones in the venom of Polistes dominulus workers (Hymenoptera: Vespidae). *Physiological Entomology*, 31(3), pp.286–293.

Burda, H. et al., 2000. Are naked and common mole-rats eusocial and if so, why? *Behavioral Ecology and Sociobiology*, 47(5), pp.293–303.

Burt, J., 2005. Syrinx-PC: a Windows program for spectral analysis, editing, and playback of acoustic signals, ver. 2.4. Available at: www.syrinxpc.com (accessed 10th October 2012).

Butlin, R., 1987. Speciation by reinforcement. *Trends in Ecology & Evolution*, 2(1), pp.8–13.

Butts, C.T., 2008a. A relational event framework for social action. *Sociological Methodology*, 38(1), pp.155–200.

Butts, C.T., 2009. Revisiting the foundations of network analysis. *science*, 325(5939), p.414.

Butts, C.T., 2014. sna: Tools for Social Network Analysis. R package version 2.3-2. Available at: http://cran.r-project.org/package=sna.

Butts, C.T., 2008b. Social network analysis with sna. *Journal of Statistical Software*, 24(6), pp.1–51.

Butts, C.T. & Carley, K.M., 2001. *Multivariate methods for inter-structural analysis*, CASOS Working Paper, Carnegie Mellon University.

Byers, B.E., 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, 18(1), pp.130–136.

Byers, B.E. & Kroodsma, D.E., 2009. Female mate choice and songbird song repertoires. *Animal Behaviour*, 77(1), pp.13–22.

Cairns, S.J. & Schwager, S.J., 1987. A comparison of association indices. *Animal Behaviour*, 35(5), pp.1454–1469.

Carreras, B.A. et al., 2002. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos*, 12(4), pp.985–994.

Carreras, I. et al., 2007. BIONETS: Bio-inspired networking for pervasive communication environments. *IEEE Transactions on Vehicular Technology*, 56(1), pp.218–229.

Carter, G.G. & Wilkinson, G.S., 2013. Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1753), p.20122573.

Casares, P., 2008. Genética del aislamiento reproductivo. *Genética de poblaciones y evolutiva.*, 17.

Case, T.J., 1990. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proceedings of the National Academy of Sciences*, 87(24), pp.9610–9614.

Castagnone-Sereno, P., 2006. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity*, 96(4), pp.282–289.

Castellano, C., Fortunato, S. & Loreto, V., 2009. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2), p.591.

Castells, M., 2011. *The rise of the network society: The information age: Economy, society, and culture*, John Wiley & Sons.

Catchpole, C.K., 1986. Song repertoires and reproductive success in the great reed warbler Acrocephalus arundinaceus. *Behavioral Ecology and Sociobiology*, 19(6), pp.439–445.

Cayley, A., 1875. On the analytical forms called trees, with application to the theory of chemical combinations. *Rep. Brit. Assoc. Advance. Sci*, 45, pp.257–305.

Cervo, R., Dani, F.R. & Turillazzi, S., 1996. Nestmate recognition in three species of stenogastrine wasps (Hymenoptera, Vespidae). *Behavioral ecology and sociobiology*, 39(5), pp.311–316.

Chapman, N.C., Beekman, M. & Oldroyd, B.P., 2010. Worker reproductive parasitism and drift in the western honeybee Apis mellifera. *Behavioral Ecology and Sociobiology*, 64(3), pp.419–427.

Christakis, N.A., 2004. Social networks and collateral health effects. *BMJ (Clinical research ed.)*, 329(7459), pp.184–185.

Ciszek, D., 2000. New colony formation in the “highly inbred” eusocial naked mole-rat: outbreeding is preferred. *Behavioral Ecology*, 11(1), pp.1–6.

Clark, C.W. & Mangel, M., 1986. The evolutionary advantages of group foraging. *Theoretical Population Biology*, 30(1), pp.45–75.

Clauset, A., 2011. *A brief primer on probability distributions*. Santa Fe Institute.

Copeland, K.T. et al., 1977. Bias due to misclassification in the estimation of relative risk. *American Journal of Epidemiology*, 105(5), pp.488–495.

Costa, J.T., 2006. *The other insect societies*, Harvard University Press.

Coviello, N. & Munro, H., 1997. Network relationships and the internationalisation process of small software firms. *International Business Review*, 6(4), pp.361–386.

Coyne, J.A., Crittenden, A.P. & Mah, K., 1994. Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. *Science*, 265(5177), pp.1461–1464.

Cracraft, J., 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. *Speciation and its Consequences*, pp.28–59.

Crespi, B.J., 1992. Eusociality in Australian gall thrips. *Nature*, 359, pp.724-726.

Crespi, B.J., 1994. Three conditions for the evolution of eusociality: are they sufficient? *Insectes Sociaux*, 41(4), pp.395–400.

Crespi, B.J. & Yanega, D., 1995. The definition of eusociality. *Behavioral Ecology*, 6(1), pp.109–115.

Creswick, N. & Westbrook, J.I., 2006. The medication advice-seeking network of staff in an Australian hospital renal ward. *Studies in health technology and informatics*, 130, pp.217–231.

Croft, D.P. et al., 2011. Hypothesis testing in animal social networks. *Trends in Ecology & Evolution*, 26(10), pp.502–507.

Croft, D.P. et al., 2006. Social structure and co-operative interactions in a wild population of guppies (Poecilia reticulata). *Behavioral Ecology and Sociobiology*, 59(5), pp.644–650.

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

Croft, D.P., Krause, J. & James, R., 2004. Social networks in the guppy (Poecilia reticulata). *Proceedings of the Royal Society B: Biological Sciences*, 271 Suppl(Suppl_6), pp.S516–S519.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), pp.297–334.

Crucitti, P., Latora, V. & Marchiori, M., 2004. A topological analysis of the Italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1), pp.92–97.

Csardi, G. & Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal*, Complex Sy, p.1695. Available at: http://igraph.org.

Curson, J., Quinn, D. & Beadle, D., 1994. *New world warblers*, A&C Black.

Darolová, A. et al., 2012. Song complexity in male marsh warblers: does it reflect male quality? *Journal of Ornithology*, 153(2), pp.431–439.

Von Dassow, G. et al., 2000. The segment polarity network is a robust developmental module. *Nature*, 406(6792), pp.188–192.

Davidson, R. & MacKinnon, J.G., 1993. Estimation and inference in econometrics. *OUP Catalogue*.

Dawley, R.M. & Bogart, J.P., 1989. *Evolution and ecology of unisexual vertebrates*, Albany, NY: University of the State of New York, State Education Dept., New York State Museum.

Dekker, D., Krackhardt, D. & Snijders, T. a B., 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, 72(4), pp.563–581.

Dey, C.J. & Quinn, J.S., 2014. Individual attributes and self-organizational processes affect dominance network structure in pukeko. *Behavioral Ecology*, 25 (6): pp.1402-1408.

Dixon, A.F.G., 1985. Structure of aphid populations. *Annual review of entomology*, 30(1), pp.155–174.

Dixon, P. & Palmer, M.W., 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6), pp.927–930.

Dobzhansky, T.G., 1937. *Genetics and the Origin of Species*, Columbia University Press.

Dolphin, K. & Quicke, D.L.J., 2001. Estimating the global species richness of an incompletely described taxon: an example using parasitoid wasps (Hymenoptera: Braconidae). *Biological Journal of the Linnean Society*, 73(3), pp.279–286.

Dormann, C.F. et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), pp.27–46.

Dormann, C.F. et al., 2009. Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal*, 2(1) pp.7-24.

Dressler, F., 2005. Efficient and scalable communication in autonomous networking using bio-inspired mechanisms - An overview. *Informatica (Ljubljana)*, 29(2), pp.183–188.

Dressler, F. et al., 2005. Self-Organization in Sensor Networks using Bio-Inspired Mechanisms. *In ARCS Workshops*, pp.139–144.

Drewe, J. a. et al., 2012. Performance of proximity loggers in recording intra- and inter-species interactions: A laboratory and field-based validation study. *PLoS ONE*, 7(6), e39068.

Drewe, J. & Perkins, S.E., 2014. Disease transmission in animal social networks. In J. Krause et al., eds. *Animal Social Networks*. Oxford University Press, pp. 94–109.

Drewe, J.A. et al., 2013. Patterns of direct and indirect contact between cattle and badgers naturally infected with tuberculosis. *Epidemiology and infection*, 141(07), pp.1467–1475.

Drews, C., 1993. The concept and definition of dominance in animal behaviour. *Behaviour*, 125(3), pp.283–313.

Duffy, J.E., 2003. The ecology and evolution of eusociality in sponge-dwelling shrimp. *Genes, behaviors, and evolution in social insects (eds T. Kikuchi, S. Higashi & N. Azuma)*, pp.217–252.

Duffy, J.E. & Macdonald, K.S., 2010. Kin structure, ecology and the evolution of social organization in shrimp: a comparative analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1681), pp.575–584.

Duffy, J.E. & Morrison, C.L., 2000. Multiple origins of eusociality in sponge-dwelling shrimps (Synalpheus). *Evolution*, 54(2), pp.503–516.

Dunbar, R. & Dunbar, P., 1975. Social dynamics of gelada baboons. *Contributions to primatology*, 6(6), pp.1–157.

Dunbar, R.I.M., 1996. *Determinants of group size in primates: A general model.*, Oxford University Press.

Durbin, J. & Watson, G.S., 1950. Testing for serial correlation in least squares regression. I. *Biometrika*, 37(3-4), pp.409–428.

Durbin, J. & Watson, G.S., 1951. Testing for serial correlation in least squares regression. II. *Biometrika*, pp.159–177.

Dyo, V. et al., 2010. Evolution and sustainability of a wildlife monitoring sensor network. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, pp. 127–140.

Eberhard, M. & West, J., 1969. *The social biology of polistine wasps.* Miscellaneous publications, Musuem of Zoology, University of Michigan no. 140.

Edwards, J.S. & Palsson, B.O., 2000. Robustness analysis of the Escherichia coli metabolic network. *Biotechnology Progress*, 16(6), pp.927–939.

Eggleton, P., 2011. An introduction to termites: biology, taxonomy and functional morphology. In *Biology of termites: a modern synthesis*. Springer, pp. 1–26.

Ehrman, L., 1964. Courtship and mating behavior as a reproductive isolating mechanism in Drosophila. *American zoologist*, pp.147–153.

Eliassen, E., 1960. *A method for measuring the heart rate and stroke/pulse pressures in birds by normal flight*, Norwegian Universities Press.

Emmerson, M. & Yearsley, J.M., 2004. Weak interactions, omnivory and emergent food-web properties. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1537), pp.397–405.

Erdös, P. & Rényi, A., 1959. On random graphs. *Publ Math*, 6, pp.290–297.

Faloutsos, M., Faloutsos, P. & Faloutsos, C., 1999. On power-law relationships of the Internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4), pp.251–262.

Farine, D.R., 2013. Animal social network inference and permutations for ecologists in R using asnipe. *Methods in Ecology and Evolution*, 4(12), pp.1187–1194.

Farine, D.R., 2014. Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Animal Behaviour*, 89, pp.141–153.

Farine, D.R. & Milburn, P.J., 2013. Social organisation of thornbill-dominated mixed-species flocks using social network analysis. *Behavioral Ecology and Sociobiology*, 67(2), pp.321–330.

Farine, D.R. & Whitehead, H., 2015. Constructing, conducting and interpreting animal social network analysis S. Altizer, ed. *Journal of Animal Ecology*, 84(5), pp.1144–1163.

Farrar, D.E. & Glauber, R.R., 1967. Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, pp.92–107.

Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist*, pp.1–15.

Fine, P., Eames, K. & Heymann, D.L., 2011. "Herd immunity": A rough guide. *Clinical Infectious Diseases*, 52(7), pp.911–916.

Fisher, R.A., 1948. Answer to Question 14 on Combining independent tests of significance. *The American Statistician*, 2(30).

Fitze, P.S. et al., 2007. Carotenoid-Based Plumage Colors and Immune Function: Is There a Trade-Off for Rare Carotenoids? *the american naturalist*, 169(S1), pp.S137–S144.

Fletcher, P., Hoyle, H.B. & Patty, C.W., 1991. *Foundations of discrete mathematics* Internatio., Boston: PWS-KENT Publishers.

Ford, J.K.B., Ellis, G.M. & Balcomb, K.C., 2000. *Killer whales: the natural history and genealogy of Orcinus orca in British Columbia and Washington*, UBC Press.

Foster, K.R., Wenseleers, T. & Ratnieks, F.L.W., 2006. Kin selection is the key to altruism. *Trends in Ecology & Evolution*, 21(2), pp.57–60.

Franks, D.W. et al., 2008. Extremism propagation in social networks with hubs. , 44(0), pp.1–32.

Franks, D.W., Ruxton, G.D. & James, R., 2010. Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology*, 64(3), pp.493–503.

Franks, D.W., Ruxton, G.D. & James, R., 2010. Sampling animal association networks with the gambit of the group. *Behavioral ecology and sociobiology*, 64(3), pp.493–503.

Fraser, H.B. et al., 2002. Evolutionary rate in the protein interaction network. *Science*, 296(5568), pp.750–752.

Freedman, D. & Lane, D., 1983. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, 1(4), pp.292–298.

Freeman, L., 1979. Centrality in social networks conceptual clarification. *Social networks*, 1(1978/79), pp.215–239.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. *Sociometry*, pp.35–41.

Frid, A. & Dill, L., 2002. Human-caused disturbance stimuli as a form of predation risk. *Ecology and Society*, 6(1), p.11.

Gai, P. & Kapadia, S., 2010. Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120), pp.2401–2423.

Gamboa, G.J. et al., 1996. Kin recognition pheromones in social wasps: combining chemical and behavioural evidence. *Animal Behaviour*, 51(3), pp.625–629.

Gante, H.F., Collares-Pereira, M.J. & Coelho, M.M., 2004. Introgressive hybridisation between two Iberian Chondrostoma species (Teleostei, Cyprinidae) revisited: New evidence from morphology, mitochondrial DNA, allozymes and NOR-phenotypes. *Folia Zoologica*, 53(4), pp.423–432.

Garlaschelli, D. & Loffredo, M.I., 2004. Patterns of link reciprocity in directed networks. *Physical Review Letters*, 93(26), p.268701.

Garlaschelli, D. & Loffredo, M.I., 2005. Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications*, 355(1), pp.138–144.

Gauvin, L. et al., 2013. Activity clocks: spreading dynamics on temporal networks of human contact. *Scientific reports*, 3(3099).

Gentleman, R. et al., 2006. *Bioinformatics and computational biology solutions using R and Bioconductor*, Springer Science & Business Media.

Gentleman, R., 2008. *R programming for bioinformatics*, CRC Press.

Gero, S., Engelhaupt, D. & Whitehead, H., 2008. Heterogeneous social associations within a sperm whale, Physeter macrocephalus, unit reflect pairwise relatedness. *Behavioral Ecology and Sociobiology*, 63(1), pp.143–151.

Gibbons, R., 1992. *A Primer in Game Theory*, Harvester Wheatsheaf.

Gibbons, W.J. & Andrews, K.M., 2004. PIT tagging: simple technology at its best. *Bioscience*, 54(5), pp.447–454.

Gilbert, E.N., 1959. Random graphs. *The Annals of Mathematical Statistics*, pp.1141–1144.

Gill, F.B., 1994. Rapid, asymmetrical introgression of mitochondrial DNA in hybridizing populations of the blue-winged and golden-winged warblers. In *Proc XXI Int, Ornithol Congress, August*. pp. 20–25.

Gillespie, J.H., 1991. The Burden of Genetic Load. *Science*, 254(5034), p.1049.

Gillies, R.R. & Carlson, T.N., 1995. Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *Journal of Applied Meteorology*, 34(4), pp.745–756.

Girvan, M. & Newman, M.E.., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp.7821–7826.

Glejser, H., 1969. A new test for heteroskedasticity. *Journal of the American Statistical Association*, 64(325), pp.316–323.

Godambe, V.P., 1982. Ancillarity principle and a statistical paradox. *Journal of the American Statistical Association*, 77(380), pp.931–933.

Godfrey, S.S. et al., 2009. Network structure and parasite transmission in a group living lizard, the gidgee skink, Egernia stokesii. *Behavioral Ecology and Sociobiology*, 63(7), pp.1045–1056.

Goldenberg, A. et al., 2010. A survey of statistical network models. *Foundations and Trends{\textregistered} in Machine Learning*, 2(2), pp.129–233.

Goldfeld, S.M. & Quandt, R.E., 1965. Some tests for homoscedasticity. *Journal of the American statistical Association*, 60(310), pp.539–547.

Goldstein, R., 1993. Conditioning Diagnostics: Collinearity and Weak Data in Regression. *Technometrics*, 35(1), pp.85–86.

Gómez, A. & Carvalho, G.R., 2000. Sex, parthenogenesis and genetic structure of rotifers: microsatellite analysis of contemporary and resting egg bank populations. *Molecular Ecology*, 9(2), pp.203–214.

Gonzalez, J., Düttmann, H. & Wink, M., 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *Journal of Zoology*, 279(3), pp.310–318.

Good, P.I., 2012. *Introduction to statistics through resampling methods and R/S-PLUS*, John Wiley & Sons.

Goodnight, K.F. & Queller, D.C., 1999. Computer software for performing likelihood test of pedigree relationships using genetic markers. *Molecular Ecology*, 8, pp.1231–1234.

Granovetter, M.S., 1973. The Strength of Weak Ties. In *Social Stratification: Class, Race, and Gender in Sociological Perspective*. p. 576.

Grant, K.A. & Grant, V., 1964. Mechanical isolation of Salvia apiana and Salvia mellifera (Labiatae). *Evolution*, pp.196–212.

Grant, P.R. & Grant, B.R., 1992. Hybridization of bird species. *Science (New York, N.Y.)*, 256(5054), pp.193–197.

Grant, P.R., Grant, B.R. & Petren, K., 2005. Hybridization in the recent past. *The American Naturalist*, 166(1), pp.56–67.

Grant, V., 1994. Modes and origins of mechanical and ethological isolation in angiosperms. *Proceedings of the National Academy of Sciences*, 91(1), pp.3–10.

Grant, V., 1963. *The origin of adaptations*, Columbia University Press.

Grant, V., 1966. The selective origin of incompatibility barriers in the plant genus Gilia. *American Naturalist*, pp.99–118.

Grossman, J.W., 2002. *Erdos Number Project*, Jerrold W. Grossman.

Guimera, R., Sales-Pardo, M. & Amaral, L., 2004. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2).

Haddadi, H. et al., 2011. Determining association networks in social animals: choosing spatial–temporal criteria and sampling rates. *Behavioral Ecology and Sociobiology*, 65(8), pp.1659–1668.

Hadfield, J., 2014. *MCMC glmm course notes*. CRAN.

Hagan, M.T., Demuth, H.B. & Beale, M.H., 1996. *Neural Network Design*, Pws Pub. Boston.

Hałaj, G. & Kok, C., 2013. Assessing interbank contagion using simulated networks. *Computational Management Science*, 10(2-3), pp.157–186.

Haldane, J.B.S., 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics*, 12(2), pp.101–109.

Hall, B., 2008. *Strickberger's evolution*, Jones & Bartlett Learning.

Hamede, R.K. et al., 2009. Contact networks in a wild Tasmanian devil (Sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters*, 12(11), pp.1147–1157.

Hamilton, W.D., 1971. Geometry for the selfish herd. *Journal of theoretical Biology*, 31(2), pp.295–311.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. I. *Journal of theoretical biology*, 7(1), pp.1–16.

Handcock, R.N. et al., 2009. Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing. *Sensors*, 9(5), pp.3586–3603.

Hanneke, S., Fu, W. & Xing, E.P., 2010. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4, pp.585–605.

Hanneman, R.A. & Riddle, M., 2005. *Introduction to social network methods*. University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)

Harper, F.M. & Hart, M.W., 2005. Gamete compatibility and sperm competition affect paternity and hybridization between sympatric Asterias sea stars. *The Biological Bulletin*, 209(2), pp.113–126.

Harrell, F.E., 2013. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, Springer Science & Business Media.

Harrison, R.G., 1990. Hybrid zones: windows on evolutionary process. *Oxford surveys in evolutionary biology*, 7, pp.69–128.

Harrison, R.G., 1993. Hybrids and hybrid zones: historical perspective. *Hybrid zones and the evolutionary process*, pp.3–12.

Harvey, P.H., Pagel, M.D. & others, 1991. *The comparative method in evolutionary biology*, Oxford university press, Oxford.

Hastings, A. & Powell, T., 1991. Chaos in a three-species food chain. *Ecology*, pp.896–903.

Hayden, B. et al., 2010. Hybridisation between two cyprinid fishes in a novel habitat: genetics, morphology and life-history traits. *BMC evolutionary biology*, 10, p.169.

Hegner, R.E., Emlen, S.T. & Demong, N.J., 1982. Spatial organization of the white-fronted bee-eater. *Nature*, 298(5871), pp.264–266.

Hemsworth, P.H. et al., 1986. Effects of social environment on welfare status and sexual behaviour of female pigs. II. Effects of space allowance. *Applied Animal Behaviour Science*, 16(3), pp.259–267.

Hill, G.E. & Mcgraw, K.J., 2004. Correlated changes in male plumage coloration and female mate choice in cardueline finches. *Animal Behaviour*, 67(1), pp.27–35.

Hill, R.A., Bentley, R.A. & Dunbar, R.I.M., 2008. Network scaling reveals consistent fractal pattern in hierarchical mammalian societies. *Biology letters*, 4(6), pp.748–751.

Hill, T.L., 1963. *Thermodynamics of small systems*, Courier Corporation.

Hinde, R.A., 1976. Interactions, relationships and social structure. *Man*, pp.1–17.

Hirsch, B.T. et al., 2013a. Genetic relatedness does not predict racoon social network structure. *Animal Behaviour*, 85(2), pp.463–470.

Hirsch, B.T. et al., 2013b. Raccoon social networks and the potential for disease transmission. *PLoS One*, 8(10), p.10.

Holden, H. & Risebro, N.H., 1995. A mathematical model of traffic flow on a network of unidirectional roads. *SIAM Journal on Mathematical Analysis*, 26(4), pp.999–1017.

Holland, P. & Leinhardt, S., 1976. Local structure in social networks. *Sociological methodology*, 7, pp.1–45.

Holland, P.W. & Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373), pp.33–50.

Holme, P. & Saramäki, J., 2012. Temporal networks. *Physics reports*, 519(3), pp.97–125.

Howard, D.J., 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. *Hybrid zones and the evolutionary process*, pp.46–69.

Huber, P.J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. pp. 221–233.

Hubert, L. & Schultz, J., 1976. Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29(2), pp.190–241.

Hughes, W. O. H.; Oldroyd, B. P.; Beekman, M.; Ratnieks, F. L. W. (2008). "Ancestral Monogamy Shows Kin Selection is Key to the Evolution of Eusociality". *Science* 320 (5880): pp.1213–1216.

Hull, D.L. & Ruse, M., 1998. *The philosophy of biology*. Oxford University Press.

Hultgren, K.M. & Duffy, J.E., 2012. Phylogenetic community ecology and the role of social dominance in sponge-dwelling shrimp. *Ecology letters*, 15(7), pp.704–713.

Hunter, D.R., Goodreau, S.M. & Handcock, M.S., 2013. ergm. userterms: A Template Package for Extending statnet. *Journal of statistical software*, 52(2), p.2.

Huyse, T. et al., 2013. Hybridisation between the two major African schistosome species of humans. *International Journal for Parasitology*, 43(8), pp.687–689.

Iida, Y., 1999. Basic concepts and future directions of road network reliability analysis. *Journal of Advanced Transportation*, 33(2), pp.125–134.

Inkpen, A.C. & Tsang, E.W.K., 2005. Social capital, networks, and knowledge transfer. *Academy of management review*, 30(1), pp.146–165.

Innes, D.J. & Hebert, P.D.N., 1988. The origin and genetic basis of obligate parthenogenesis in Daphnia pulex. *Evolution*, pp.1024–1035.

Iori, G. et al., 2008. A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1), pp.259–278.

Irestedt, M. et al., 2009. An unexpectedly long history of sexual selection in birds-of-paradise. *BMC evolutionary biology*, 9, p.235.

Jaccard, P., 1912. The distribution of the flora in the alpine zone. *New phytologist*, 11(2), pp.37–50.

James, R., Croft, D.P. & Krause, J., 2009. Potential banana skins in animal social network analysis. *Behavioral Ecology and Sociobiology*, 63(7), pp.989–997.

Jarvis, J.U.M. & Bennett, N.C., 1993. Eusociality has evolved independently in two genera of bathyergid mole-rats—but occurs in no other subterranean mammal. *Behavioral Ecology and Sociobiology*, 33(4), pp.253–260.

Jeanne, R.L., 1982. Evidence for an alarm substance in Polistes canadensis. *Experientia*, 38(3), pp.329–330.

Jeanne, R.L., 1980. Evolution of social behavior in the Vespidae. *Annual review of entomology*, 25(1), pp.371–396.

Jenelius, E., Petersen, T. & Mattsson, L.-G., 2006. Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice*, 40(7), pp.537–560.

Jeong, H. et al., 2001. Lethality and centrality in protein networks. *Nature*, 411(6833), pp.41–42.

Jetz, W. et al., 2012. The global diversity of birds in space and tim. *Nature*, 491(7424), pp.1–5.

Jolliffe, I.T., 2002. *Principal component analysis,* John Wiley and Sons Ltd.

Judge, G.G., Hill, R.C., Griffiths, W., Lutkepohl, H. and Lee, T.C., 1988. *Introduction to the theory and practice of econometrics*. Wiley

Keane, A., Brooke, M. de L. & McGowan, P.J.K., 2005. Correlates of extinction risk and hunting pressure in gamebirds (Galliformes). *Biological Conservation*, 126(2), pp.216–233.

Keck, M.B., 1994. Test for detrimental effects of PIT tags in neonatal snakes. *Copeia*, pp.226–228.

Keeling, M.J. & Eames, K.T.D., 2005. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4), pp.295–307.

Kennedy, D.M., Braga, A. a. & Piehl, A.M., 1997. The (un)known universe: Mapping gangs and gang violence in Boston. *Crime Mapping and Crime Prevention*, 02138, pp.219–248.

Kent, D.S. & Simpson, J.A., 1992. Eusociality in the beetle Austroplatypus incompertus (Coleoptera: Curculionidae). *Naturwissenschaften*, 79(2), pp.86–87.

Kermack, W.O. & McKendrick, A.G., 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772), pp.700–721

Kimball, R.T., Mary, C.M.S. & Braun, E.L., 2011. A macroevolutionary perspective on multiple sexual traits in the phasianidae (galliformes). *International journal of evolutionary biology*, 2011, p.423938.

King, A.P., West, M.J. & Eastzer, D.H., 1980. Song structure and song development as potential contributors to reproductive isolation in cowbirds (Molothrus ater). *Journal of Comparative and Physiological Psychology*, 94(6), p.1028.

Kisaka, H. et al., 1997. Production and analysis of plants that are somatic hybrids of barley (Hordeum vulgare L.) and carrot (Daucus carota L.). *Theoretical and Applied Genetics*, 94(2), pp.221–226.

Kolm, N. et al., 2007. Can sexual selection drive female life histories? A comparative study on Galliform birds. *Journal of Evolutionary Biology*, 20(2), pp.627–638.

Kormondy, E.J., 1969. *Concepts of ecology*. Prentice-Hall.

Koskinen, J.H., Robins, G.. L. & Pattison, P.E., 2010. Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Statistical Methodology*, 7(3), pp.366–384.

Krackhardt, D., 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(4), pp.359–381.

Krause, A.E. et al., 2003. Compartments revealed in food-web structure. *Nature*, 426(6964), pp.282–285.

Krause, J. et al., 2014. *Animal social networks*, Oxford University Press.

Krause, J. et al., 2013. Reality mining of animal social systems. *Trends in Ecology and Evolution*, 28(9), pp.541–551.

Krause, J., Lusseau, D. & James, R., 2009. Animal social networks: an introduction. *Behavioral Ecology and Sociobiology*, 63(7), pp.967–973.

Krause, J. & Ruxton, G.D., 2002. *Living in groups*, Oxford University Press.

Krause, J., Wilson, A.D.M. & Croft, D.P., 2011. New technology facilitates the study of social networks. *Trends in Ecology & Evolution*, 26(1) pp.5-6.

Krebs, V.E., 2002. Mapping Networks of Terrorist Cells. *Connections*, 24(3), pp.43–52.

Krieger, M.J., Billeter, J.B. & Keller, L., 2000. Ant-like task allocation and recruitment in cooperative robots. *Nature*, 406(6799), pp.992–995.

Krivitsky, P.N. et al., 2012. Package "ergm. count." *Journal of Statistics*, 6, pp.1100–1128.

Krylov, D.M. et al., 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome research*, 13(10), pp.2229–2235.

Kudo, H. & Dunbar, R.I.M., 2001. Neocortex size and social network size in primates. *Animal Behaviour*, 62(4), pp.711–722.

Kukuk, P.F., 1994. Replacing the terms "primitive" and "advanced": new modifiers for the term "eusocial." *Animal Behaviour*, 47(6), pp.1475–1478.

Kummer, H., 1968. *Social Organization of Hamadryas Baboons*. Karger Publishers.

Langer, N., Pedroni, A. & Jäncke, L., 2013. The problem of thresholding in small-world network analysis. *PloS one*, 8(1), p.e53199.

Legendre, P. & Anderson, M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological monographs*, 69(1), pp.1–24.

Legendre, P. & Legendre, L.F.J., 2012. *Numerical ecology*, Elsevier.

Lemke, N. et al., 2004. Essentiality and damage in metabolic networks. *Bioinformatics*, 20(1), pp.115–119.

Lengronne, T. et al., 2012. Little effect of seasonal constraints on population genetic structure in eusocial paper wasps. *Ecology and Evolution*, 2(10), pp.2615–2624.

Lengronne, T., 2013. *Social Dynamics in Paper Wasps : The case of nest-drifting in Polistes canadensis*. University of Lausanne.

Lenz, L.W., 1959. Hybridization and speciation in Pacific Coast Irises. *Aliso*, 4, pp.237–309.

Lessios, H.A. & Cunningham, C.W., 1990. Gametic incompatibility between species of the sea urchin Echinometra on the two sides of the Isthmus of Panama. *Evolution*, pp.933–941.

Levine, L., 1979. *Biología del gen*, Barcelona: Ed. Omega.

Lhuilier, G., 1813. Géométrie. Mémoire sur la polyédrométrie; contenant une démonstration directe du théorème d'Euler sur les polyèdres, et un examen des diverses exceptions auxquelles ce théorème est assujetti. *Annales de Gergonne*, 3, pp.169–189.

Liljeros, F. et al., 2001. The web of human sexual contacts. *Nature*, 411(6840), pp.907–908.

Liptack, M.K. & Druehl, L.D., 2000. Molecular evidence for an interfamilial laminarialean cross. *European Journal of Phycology*, 35(02), pp.135–142.

Lislevand, T., Figuerola, J. & Székely, T., 2009. Evolution of sexual size dimorphism in grouse and allies (Aves: Phasianidae) in relation to mating competition, fecundity demands and resource division. *Journal of Evolutionary Biology*, 22(9), pp.1895–1905.

Logan, M., 2011. *Biostatistical design and analysis using R: a practical guide*, John Wiley & Sons.

Long, J. & Luke, J.T., 2011. *The new Munsell student color set* 3rd Ed., Fairchild Publications.

Lopez-Vaamonde, C. et al., 2004. Social parasitism by male-producing reproductive workers in a eusocial insect. *Nature*, 430(6999), pp.557–560.

Lovegrove, B.G., 1991. The evolution of eusociality in molerats (Bathyergidae): a question of risks, numbers, and costs. *Behavioral Ecology and Sociobiology*, 28(1), pp.37–45.

Lovette, I.J. et al., 2010. A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). *Molecular phylogenetics and evolution*, 57(2), pp.753–70.

Luce, R.D. & Perry, A.D., 1949. A method of matrix analysis of group structure. *Psychometrika*, 14(2), pp.95–116.

Ludlow, A.M. & Magurran, A.E., 2006. Gametic isolation in guppies (Poecilia reticulata). *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1600), pp.2477–2482.

Lusseau, D., 2003. The emergent properties of a dolphin social network. *Proceedings. Biological sciences / The Royal Society*, 270 Suppl (November 2014), pp.186–188.

Lusseau, D., Whitehead, H. & Gero, S., 2008. Incorporating uncertainty into the study of animal social networks. *Animal Behaviour*, 75(5), pp.1809–1815.

MacCarthy, T., Seymour, R. & Pomiankowski, A., 2003. The evolutionary potential of the Drosophila sex determination gene network. *Journal of theoretical biology*, 225(4), pp.461–468.

Macdonald, K.S., Ríos, R. & Duffy, J.E., 2006. Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. *Diversity and Distributions*, 12(2), pp.165–178.

Magder, L.S. & Hughes, J.P., 1997. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, 146(2), pp.195–203.

Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20(5), pp.229–237.

Manly, B.F.J., 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology* 2nd ed., New York: Chapman & Hall.

Manna, K.B. and G.M. and J.S. and K.K. and S.S., 2008. The International Trade Network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(02) p.02002

Manno, T.G., 2008. Social networking in the Columbian ground squirrel, Spermophilus columbianus. *Animal Behaviour*, 75(4), pp.1221–1228.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), pp.209–220.

Marsh, M.K. et al., 2011. Spatial and temporal heterogeneities in the contact behaviour of rabbits. *Behavioral ecology and sociobiology*, 65(2), pp.183–195.

Martin, P.R., Montgomerie, R. & Lougheed, S.C., 2015. Color Patterns of Closely Related Bird Species Are More Divergent at Intermediate Levels of Breeding-Range Sympatry. *The American Naturalist*, 185(4), pp.443–451.

Martins, C.R., 2008. Mobility and Social Network Effects on Extremist Opinions. *Physical Review E*, 78(3), p. 036104.

Masi, G. De et al., 2011. An Analysis of the Japanese Credit Network. *Evolutionary and Institutional Economics Review*, 7(2), pp.209–232.

Matute, D.R., 2010. Reinforcement of gametic isolation in Drosophila. *PLoS Biol*, 8(3), p.e1000341.

May, R.M., 2001. *Stability and complexity in model ecosystems*, Princeton University Press.

May, R.M. & Lloyd, A.L., 2001. Infection dynamics on scale-free networks. *Physical Review E*, 64(6), p.66112.

Maynard Smith, J. & Szathmary, E., 1995. The major evolutionary transitions. *Nature*, 374, pp.227–232.

Mayr, E., 1963. *Animal species and evolution*, Belknap Press of Harvard University Press Cambridge, Massachusetts.

Mayr, E., 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*, Harvard University Press.

Mayr, E., 1982. *The growth of biological thought: diversity, evolution, and inheritance*, Harvard University Press.

McCarthy, E.M., 2006. *Handbook of avian hybrids of the world*, Oxford University Press New York.

McCullagh, P. & Nelder, J.A., 1989. *Generalized linear models*, CRC press.

McDonald, D.B., 2007. Predicting fate from early connectivity in a social network. *Proceedings of the National Academy of Sciences of the United States of America*, 104(26), pp.10910–10914.

McNaught, M.K. & Owens, I.P.F., 2002. Interspecific variation in plumage colour among birds: species recognition or light environment? *Journal of Evolutionary Biology*, 15(4), pp.505–514.

McPhail, J.D., 1994. Speciation and the evolution of reproductive isolation in the sticklebacks (Gasterosteus) of south-western British Columbia. *The evolutionary biology of the threespine stickleback*, pp.399–437.

McPherson, J.M. & Smith-Lovin, L., 1987. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American sociological review*, pp.370–379.

McPherson, M., Smith-Lovin, L. & Cook, J.M., 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pp.415–444.

Meloun, M. et al., 2002. Crucial problems in regression modelling and their solutions. *Analyst*, 127(4), pp.433–450.

Mendelson, T.C., 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). *Evolution; international journal of organic evolution*, 57(2), pp.317–327.

Von Mering, C. et al., 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, 417(6887), pp.399–403.

Miranda-de la Lama, G.C. & Mattiello, S., 2010. The importance of social behaviour for goat welfare in livestock farming. *Small Ruminant Research*, 90(1), pp.1–10.

Mittler, R. et al., 2004. Reactive oxygen gene network of plants. *Trends in plant science*, 9(10), pp.490–498.

Mohar, B., 1997. *Some applications of Laplace eigenvalues of graphs*, Springer.

Molm, L.D., Schaefer, D.R. & Collett, J.L., 2007. The value of reciprocity. *Social Psychology Quarterly*, 70(2), pp.199–217.

Montoya, J.M., Pimm, S.L. & Solé, R. V, 2006. Ecological networks and their fragility. *Nature*, 442(7100), pp.259–264.

Moreno, J.L. & Jennings, H.H., 1934. *Who shall survive?Foundations of Sociometry, Group Psychotherapy and Sociodrama.* Beacon. New York.

Morris, M., 1993. Epidemiology and social networks: Modeling structured diffusion. *Sociological Methods & Research*, 22(1), pp.99–126.

Muenchen, R.A., 2012. The popularity of data analysis software. *r4stats. com, http://r4stats. com/articles/popularity/, last accessed*, 28.

Munsell, A., 1912. A pigment color system and notation. *The American Journal of Psychology*, 23(2), pp.236–244.

Nakagawa, S. & Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85(4), pp.935–956.

Nanork, P. et al., 2005. Entomology: Asian honeybees parasitize the future dead. *Nature*, 437(7060), p.829.

Natarajan, M., 2006. Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data. *Journal of Quantitative Criminology*, 22(2), pp.171–192.

NatureServe & BirdLife International, 2012. species distribution maps of the world, version 2.0. Available at: Available from: http://www.biodiversityinfo.org/spcdownload/u7g0 5/downloadBOTW.htm. (accessed: 1st November 2013).

Neter, J. et al., 1996. *Applied linear statistical models*, Irwin. Chicago.

Neumann, P., MORITZ, R.F. & Mautz, D., 2000. Colony evaluation is not affected by drifting of drone and worker honeybees (Apis mellifera L.) at a performance testing apiary. *Apidologie*, 31(1), pp.67–80.

Neutel, A.-M., Heesterbeek, J. a P. & De Ruiter, P.C., 2002. Stability in real food webs: weak links in long loops. *Science (New York, N.Y.)*, 296(5570), pp.1120–1123.

Newman, M., 2008. The physics of networks. *Physics today*, 61(11), pp.33–38.

Newman, M.E.J., 2002a. Assortative mixing in networks. *Physical review letters*, 89(20), p.208701.

Newman, M.E.J., 2003. Mixing patterns in networks. *Physical Review E*, 67(2), p.26126.

Newman, M.E.J., 2002b. Spread of epidemic disease on networks. *Physical review E*, 66(1), p.16128.

Newman, M.E.J., 2003. The structure and function of complex networks. *SIAM review*, 15(3), pp.247–262.

Newman, M.E.J., 2003. The structure and function of complex networks. *Society for Industrial and Applied Mathematics*, 45, pp.167–256.

Newman, M.E.J., Forrest, S. & Balthrop, J., 2002. Email networks and the spread of computer viruses. *Physical Review E*, 66(3), p.35101.

Nonacs, P. & Reeve, H.K., 1995. The ecology of cooperation in wasps: causes and consequences of alternative reproductive decisions. *Ecology*, pp.953–967.

Nowak, M.A., Tarnita, C.E. & Wilson, E.O., 2010. The evolution of eusociality. *Nature*, 466(7310), pp.1057–1062.

Nowak, R.M., 1999. *Walker's Mammals of the World*, JHU Press.

Nowotarski, M., 2011. Don ' t Steal My Avatar ! Challenges of Social Networking Patents. , pp.1–9.

O'Brien, T.G., 1993. Allogrooming behaviour among adult female wedge-capped capuchin monkeys. *Animal behaviour*, 46(3), pp.499–510.

O'Connor, S., Park, K.J. & Goulson, D., 2013. Worker drift and egg dumping by queens in wild Bombus terrestris colonies. *Behavioral Ecology and Sociobiology*, 67(4), pp.621–627.

O'Donnell, S. & Jeanne, R.L., 2002. The nest as fortress: defensive behavior of Polybia emaciata, a mud-nesting eusocial wasp. *Journal of Insect Science*, 2(1), p.3.

O'Riain, M.J., Jarvis, J.U.M. & Faulkes, C.G., 1996. A dispersive morph in the naked mole-rat. *Nature*, 380(6575), pp.619–621.

Obayashi, T. et al., 2009. ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Research*, 37(SUPPL. 1).

Ogden, G.L. et al., 2011. Extraction of small boat harmonic signatures from passive sonar. *The Journal of the Acoustical Society of America*, 129(6), pp.3768–3776.

Ohno, S. et al., 1964. Chromosomal uniformity in the avian subclass Carinatae. *Chromosoma*, 15(3), pp.280–288.

Ohtsuki, H. et al., 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092), pp.502–505.

Oldham, K.T.S., 2008. *The Doctrine of Description: Gustav Kirchhoff, Classical Physics, and the" purpose of All Science" in 19th-century Germany*, ProQuest.

Orr, H.A., 1993. Haldane's rule has multiple genetic causes. *Nature*, 361 (6412): 532–533.

Orr, H.A., 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics*, 144(4), p.1331.

Osborne, J. & Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2).

Osborne, J.W., 2001. A new look at outliers and fringeliers: Their effects on statistic accuracy and Type I and Type II error rates. *Unpublished manuscript, Department of Educational Research and Leadership and Counselor Education, North Carolina State University*.

Ostojic, S., 2014. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature neuroscience*, 17(4), pp.594–600.

Otte, E. & Rousseau, R., 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), pp.441–453.

Paar, J. et al., 2002. Drifting of workers in nest aggregations of the giant honeybee Apis dorsata. *Apidologie*, 33(6), pp.553–562.

Papp, B., Pál, C. & Hurst, L.D., 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, 429(6992), pp.661–664.

Pardi, L., 1948. Dominance order in Polistes wasps. *Physiological Zoology*, pp.1–13.

Pastor-Satorras, R. & Vespignani, A., 2001. Epidemic Spreading in Scale-Free Networks. *Physical Review*, 86(14), pp.3200–3203.

Pastor-Satorras, R. & Vespignani, A., 2002. Immunization of complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 65(3).

Paterson, H.E.H., 1985. The recognition concept of species. In E. Vrba (ed.), Species and Speciation. Transvaal Museum Monograph No. 4. Pretoria.

Patison, K.P. et al., 2010. Changes in temporal and spatial associations between pairs of cattle during the process of familiarisation. *Applied Animal Behaviour Science*, 128(1), pp.10–17.

Patison, K.P. et al., 2015. Time is of the essence: an application of a relational event model for animal social networks. *Behavioral ecology and sociobiology*, 69(5), pp.841–855.

Patten, S.B. & Arboleda-Flórez, J. a., 2004. Epidemic theory and group violence. *Social Psychiatry and Psychiatric Epidemiology*, 39(11), pp.853–856.

Pearson, S.F. & Rohwer, S., 2000. Asymmetries in male aggression across an avian hybrid zone. *Behavioral Ecology*, 11(1), pp.93–101.

Pebesma, E.J. & Bivand, R.S., 2005. sp: classes and methods for spatial data. R package version 0.9-44.

Peck, J.R., 1994. A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, 137(2), pp.597–606.

Pedhazur, E.J., 1997. *Multiple regression in behavioral research: Explanation and prediction.* Wadsworth.

Peduzzi, P. et al., 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, 49(12), pp.1373–1379.

Perdeck, A.C., 1958. The isolating value of specific song patterns in two sibling species of grasshoppers (Chorthippus brunneus Thunb. and C. biguttulus L.). *Behaviour*, 12(1), pp.1–75.

Perrin, N. & Lehmann, L., 2001. Is sociality driven by the costs of dispersal or the benefits of philopatry? a role for kin-discrimination mechanisms. *The American Naturalist*, 158(5), pp.471–483.

PewResearch, 2014. Social Networking Fact Sheet. *PewResearch Internet Project*. Available at: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/. [Accessed July 7, 2015].

Pickett, K.M. & Wenzel, J.W., 2004. Phylogenetic analysis of the New World Polistes (Hymenoptera: Vespidae: Polistinae) using morphology and molecules. *Journal of the Kansas Entomological Society*, 77(4), pp.742–760.

Pindyck, R.S. & Rubinfeld, D.L., 1998. *Econometric models and economic forecasts*, Irwin/McGraw-Hill Boston.

Pinter-Wollman, N., Hobson, E.A., et al., 2014. Response to comments on the dynamics of network dynamics. *Behavioral Ecology*, 25(2), pp.260–261.

Pinter-Wollman, N., Hobson, E.A., et al., 2014. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology*, art.047.

Plowes, N., 2010. An Introduction to Eusociality. *Nature Education Knowledge*, 3(10:7).

Pons, P. & Latapy, M., 2006. Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2), pp.191–218.

Porter, A. & Rafols, I., 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), pp.719–745.

Prange, S. et al., 2006. New radiocollars for the detection of proximity among individuals. *Wildlife Society Bulletin*, 34(5), pp.1333–1344.

Price, T. et al., 1998. Different timing of the adaptive radiations of North American and Asian warblers. *Proceedings of the Royal Society B: Biological Sciences*, 265(1409), pp.1969–1975.

Promislow, D.E.L., 2004. Protein networks, pleiotropy and the evolution of senescence. *Proceedings of the Royal Society of London-B*, 271(1545), pp.1225–1234.

Proulx, S.R., Promislow, D.E.L. & Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, 20(6 SPEC. ISS.), pp.345–353.

Qin, J. et al., 2005. Analyzing Terrorist Networks : A Case Study of the Global Salafi Jihad Network. *Intelligence and security*, pp.287–304.

Quade, D. et al., 1980. Effects of misclassifications on statistical inferences in epidemiology. *American Journal of Epidemiology*, 111(5), pp.503–515.

Queller, D.C. & Strassmann, J.E., 1998. Kin selection and social insects. *Bioscience*, pp.165–175.

R Core Team, 2013. *R: A language and environment for statistical computing.* Available at: http://www.r-project.org/.

R Development Core Team., 2008. *R: A language and environment for statistical computing.*, Vienna: R Foundation for Statistical Computing, Vienna, Austria.

Radil, S.M., Flint, C. & Tita, G.E., 2010. Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles. *Annals of the Association of American Geographers*, 100(2), pp.307–326.

Ralston, J., Ermacor, D.A. & Kirchman, J.J., 2015. A new wood warbler hybrid (Oreothlypis celata x O. ruficapilla) from the Adirondack mountains, New York. *The Wilson Journal of Ornithology*, 127(1), pp.21–28.

Ramos-Fernández, G., Boyer, D. & Gómez, V.P., 2006. A complex social structure with fission–fusion properties can emerge from a simple foraging model. *Behavioral Ecology and Sociobiology*, 60(4), pp.536–549.

Randler, C., 2002. Avian hybridization, mixed pairing and female choice. *Animal Behaviour*, 63(1), pp.103–119.

Randler, C., 2006. Behavioural and ecological correlates of natural hybridization. *Ibis*, 148, pp.459–467.

Randler, C., 2004. Frequency of bird hybrids: does detectability make all the difference? *Journal of Ornithology*, 145(2), pp.123–128.

Rands, S.A., 2014. We must consider dynamic changes in behavior in social networks and conduct manipulations: comment on Pinter-Wollman et al. *Behavioral Ecology*, art.129.

Rau, P., 1940. Co-Operative Nest-Founding by the Wasp, Polistes Annularis Linn. *Annals of the Entomological Society of America*, 33(4), pp.617–620.

Reeve, H., 1981. Polistes. In K. G. Ross & R. W. Matthews, eds. *The Social Biology of Wasps*. Ithaca, NY: Cornell University Press, pp. 99–148.

Rendell, L. & Gero, S., 2013. The behavioral ecologist's essential social networks cookbook--comment on Pinter-Wollman et al. *Behavioral Ecology*, 25(2), pp.257–258.

Ressler, S., 2006. Social Network Analysis as an Approach to Combat Terrorism : Past , Present , and Future Research. *Homeland Security Affairs*, 11(2), pp.1–10.

Rhoten, D., 2004. Interdisciplinary Research: Trend or Transition. *Items and Issues: Social Science Research Council*, 5, pp.6–11.

Rhymer, J.M. & Simberloff, D., 1996. Extinction by hybridisation and introgression. *Annual Review of Ecology and Systematics*, 27(1), pp.83–109.

Rhymer, J.M., Williams, M.J. & Braun, M.J., 1994. Mitochondrial analysis of gene flow between New Zealand mallards (Anas platyrhynchos) and grey ducks (A. superciliosa). *The Auk*, pp.970–978.

Richards, O.W., 1971. The biology of the social wasps (Hymenoptera, Vespidae). *Biological Reviews*, 46(4), pp.483–528.

Richardson, B., 2005. Sustainable transport: analysis frameworks. *Journal of Transport Geography*, 13(1), pp.29–39.

Richardson, T., 2003. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1), pp.145–157.

Robertson, H.M. & Paterson, H.E.H., 1982. Mate recognition and mechanical isolation in Enallagma damselflies (Odonata: Coenagrionidae). *Evolution*, pp.243–250.

Robins, G. et al., 2007. An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2), pp.173–191.

Robins, G., Pattison, P. & Woolcock, J., 2004. Missing data in networks: Exponential random graph (p*) models for networks with non-respondents. *Social Networks*, 26(3), pp.257–283.

Robinson, T.J., Morris, D.J. & Fairall, N., 1991. Interspecific hybridization in the bovidae: Sterility of Alcelaphus buselaphus\times Damaliscus dorcas F1 progeny. *Biological conservation*, 58(3), pp.345–356.

Rohwer, S., Bermingham, E. & Wood, C., 2001. Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution*, 55(2), pp.405–422.

Rohwer, S. & Wood, C., 1998. Three Hybrid Zones between Hermit and Townsend's Warblers in Washington and Oregon. *The Auk*, 115(2), pp.284–310.

Rombach, M. et al., 2014. Core-periphery structure in networks. *SIAM Journal on Applied Mathematics*, 74(1), pp.167–190.

Ropert-Coudert, Y. & Wilson, R.P., 2005. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, 3(8), pp.437–444.

Ropert-Coudert, Y. & Wilson, R.P., 2005. Trends and Perspectives in Animal-Attached Remote Sensing. *Frontiers in Ecology and the Environment*, 3(8), p.437.

Ross, K.G. & Matthews, R.W., 1991. *The social biology of wasps*, Cornell University Press.

Von Rueden, C., Gurven, M. & Kaplan, H., 2010. Why do men seek status? Fitness payoffs to dominance and prestige. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20102145.

Rushmore, J. et al., 2013. Social network analysis of wild chimpanzees provides insights for predicting infectious disease risk. *Journal of Animal Ecology*, 82(5), pp.976–986.

Rutz, C. et al., 2015. Calibrating animal-borne proximity loggers. *Methods in Ecology and Evolution*, 6(6), pp.656–667.

Ryder, T.B. et al., 2012. Proximity data-loggers increase the quantity and quality of social network data. *Biology letters*, rsbl20120536.

Sade, D.S. et al., 1988. Sociometrics of Macaca mulatta: II. Decoupling centrality and dominance in rhesus monkey social networks. *American journal of physical anthropology*, 77(4), pp.409–425.

Sæther, S.A. et al., 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *science*, 318(5847), pp.95–97.

Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. *Science*, 308(5722), pp.648–652.

Schluter, D., 2001. Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7), pp.372–380.

Schmid-Hempel, P., 1995. Parasites and social insects. *Apidologie*, 26, p.245.

Schuett, W. et al., 2011. Personality variation in a clonal insect: the pea aphid, Acyrthosiphon pisum. *Developmental psychobiology*, 53(6), pp.631–640.

Searcy, W.A., 1992. Song reportoire and mate choice in birds. *American Zoologist*, 32, pp.71–80.

Seehausen, O., 2006. Conservation: losing biodiversity by reverse speciation. *Current Biology*, 16(9), pp.R334–R337.

Seehausen, O. et al., 2014. Genomics and the origin of species. *Nature reviews. Genetics*, 15(3), pp.176–92.

Seehausen, O., 2004. Hybridization and adaptive radiation. *Trends in ecology & evolution*, 19(4), pp.198–207.

Seehausen, O.L.E. et al., 2008. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, 17(1), pp.30–44.

Sen, P. et al., 2003. Small-world properties of the Indian railway network. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 67(3 Pt 2), p.036106.

Seppä, P., Queller, D.C. & Strassmann, J.E., 2002. Reproduction in foundress associations of the social wasp, Polistes carolina: conventions, competition, and skew. *Behavioral Ecology*, 13(4), pp.531–542.

Sherman, P.W., Jarvis, J.U.M. & Alexander, R.D., 1991. Biology of the Naked Mole-rat. *Monographs in behavior and ecology (USA).*

Shirley, M.D.F. & Rushton, S.P., 2005. The impacts of network topology on disease spread. *Ecological Complexity*, 2(3), pp.287–299.

Short, R. V, 1997. An introduction to mammalian interspecific hybrids. *Journal of Heredity*, 88(5), pp.355–357.

Shutler, D. & Weatherhead, P.J., 1990. Targets of sexual selection : song and plumage of wood warblers. *Evolution*, 44(8), pp.1967–1977.

Silver, N., 2012. *The signal and the noise: Why so many predictions fail-but some don't.* Penguin. London.

Skole, D. & Tucker, C., 1993. Tropical deforestation and habitat fragmentation in the Amazon. Satellite data from 1978 to 1988. *Science(Washington)*, 260(5116), pp.1905–1910.

Sledge, M.F. et al., 1999. Venom induces alarm behaviour in the social wasp Polybioides raphigastra (Hymenoptera: Vespidae): an investigation of alarm behaviour, venom volatiles and sting autotomy. *Physiological entomology*, 24(3), pp.234–239.

Slikker, M. & den Nouweland, A., 2012. *Social and economic networks in cooperative game theory*, Springer Science & Business Media.

Smith, D.A. & White, D.R., 1991. Structure and dynamics of the global economy: Network analysis of international trade 1965-1980. *Soc. F.*, 70(4), p.857.

Smocovitis, V.B., 1996. Unifying biology: The evolutionary synthesis and evolutionary biology. *Journal of the History of Biology*, 25(1), pp.1-65.

Smyth, B. & Nebel, S., 2013. Passive Integrated Transponder (PIT) Tags in the Study of Animal Movement. *Nature Education Knowledge*, 4(3)(3).

Snijders, T., 2011. Statistical Models for Social Networks. *Annual Review of Sociology*, 37(1), pp.131–153.

Snijders, T.A.B., 2002. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), pp.1–40.

Snijders, T.A.B., 2011. Statistical Models for Social Networks. *Annual Review of Sociology*, 37(1), pp.131–153.

Snyder, J., 1987. *Map Projections: A Working Manual*, U.S. Geological Survey Professional Paper 1395.

Sokal, R.R. & Rohlf, F.J., 1987. *Introduction to biostatistics*, New York.

Solé, R. V & Montoya, J.M., 2001. Complexity and fragility in ecological networks. *Proceedings. Biological sciences / The Royal Society*, 268(1480), pp.2039–2045.

De Solla Price, D., 1976. A Generai Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American society for Information science*, p.293.

Soltis, P.S. & Soltis, D.E., 2009. The role of hybridization in plant speciation. *Annual review of plant biology*, 60, pp.561–588.

Southon, R.J. et al., 2015. Long live the wasp: adult longevity in captive colonies of the eusocial paper wasp Polistes canadensis (L.). *PeerJ*, 3, p.e848.

Stanton, I. & Pinar, A., 2011. Sampling Graphs with a Prescribed Joint Degree Distribution Using Markov Chains. In *ALENEX*. SIAM, pp. 151–163.

Sterba, S.K., 2009. Alternative Model-Based and Design-Based Frameworks for Inference From Samples to Populations: From Polarization to Integration. *Multivariate Behavioural Rsearch*, 44(6), pp.711–740.

Stern, D.L. & Foster, W.A., 1996. The evolution of soldiers in aphids. *Biological Reviews*, 71(1), pp.27–79.

Stott, A.W., 1994. The economic advantage of longevity in the dairy cow. *Journal of Agricultural Economics*, 45(1), pp.113–122.

Strassmann, J.E., 1981. Evolutionary implications of early male and satellite nest production in Polistes exclamans colony cycles. *Behavioral Ecology and Sociobiology*, 8(1), pp.55–64.

Suits, D.B., 1957. Use of dummy variables in regression equations. *Journal of the American Statistical Association*, 52(280), pp.548–551.

Sumner, S. et al., 2007. Radio-Tagging Technology Reveals Extreme Nest-Drifting Behavior in a Eusocial Insect. *Current Biology*, 17(2), pp.140–145.

Sumner, S., Pereboom, J.J.M. & Jordan, W.C., 2006. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, Polistes canadensis. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1582), pp.19–26.

Swain, D.L. et al., 2015. Pregnant cattle associations and links to maternal reciprocity. *Applied Animal Behaviour Science*. 168, pp.10-17.

Szymura, J. & Barton, N.H., 1986. Genetic Analysis of a Hybrid Zone Between the Fire-Bellied Toads , Bombina bombina and B . variegata , Near Cracow in Southern Poland Published by : Society for the Study of Evolution Stable URL : ht. *Evolution*, 40(6), pp.1141–1159.

Tadić, B., 2001. Dynamics of directed graphs: The world-wide Web. *Physica A: Statistical Mechanics and its Applications*, 293(1-2), pp.273–284.

Takahashi, J. ichi et al., 2010. Male production by non-natal workers in the bumblebee, Bombus deuteronymus (Hymenoptera: Apidae). *Journal of Ethology*, 28(1), pp.61–66.

Tanabe, T. & Sota, T., 2008. Complex copulatory behavior and the proximate effect of genital and body size differences on mechanical reproductive isolation in the millipede genus Parafontaria. *The American Naturalist*, 171(5), pp.692–699.

Templeton, A.R., 1989. The meaning of species and speciation: a genetic perspective. *The units of evolution: Essays on the nature of species*, pp.159–183.

Thébault, E. & Fontaine, C., 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993), pp.853–856.

Thiele, T.N., 1880. Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en “systematisk”Karakter. *Det Kongelige Danske Videnskabernes Selskabs Skrifter-Naturvidenskabelig og Mathematisk Afdeling*, pp.381–408.

Thorne, B.L. et al., 1999. Reproductive dynamics and colony structure of subterranean termites of the genus Reticulitermes (Isoptera Rhinotermitidae): a review of the evidence from behavioral, ecological, and genetic studies. *Ethology Ecology & Evolution*, 11(2), pp.149–169.

Thorne, B.L., Grimaldi, D.A. & Krishna, K., 2000. Early fossil history of the termites. In *Termites: evolution, sociality, symbioses, ecology*. Springer, pp. 77–93.

Toomey, M.B. & McGraw, K.J., 2012. Mate choice for a male carotenoid-based ornament is linked to female dietary carotenoid intake and accumulation. *BMC Evolutionary Biology*, 12(1), p.3.

Tracey, J.P., Lukins, B.S. & Haselden, C., 2008. Hybridisation between mallard (Anas platyrhynchos) and grey duck (A. superciliosa) on Lord Howe Island and management options. *Notornis*, 55(1), pp.1–7.

Trivers, R., 1971. The evolution of reciprocal altruism. *Quarterly review of biology*, 46(1), pp.35–57.

Trudeau, R.J., 2013. *Introduction to graph theory*, Courier Corporation.

Tsang, J., Zhu, J. & van Oudenaarden, A., 2007. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Molecular Cell*, 26(5), pp.753–767.

Tyler, F., Rodríguez-Muñoz, R. & Tregenza, T., 2013. Fertilisation and early developmental barriers to hybridisation in field crickets. *BMC evolutionary biology*, 13, p.43.

U.S. Air Force, 2014. GPS Accuracy. Available at: gps.gov [Accessed August 12, 2015].

Uetz, P. et al., 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403(6770), pp.623–627.

Ulrich, Y., Perrin, N. & Chapuisat, M., 2009. Flexible social organization and high incidence of drifting in the sweat bee, Halictus scabiosae. *Molecular Ecology*, 18(8), pp.1791–1800.

Vallerand, R.J. & Bissonnette, R., 1992. Intrinsic, extrinsic, and amotivational styles as predictors of behavior: A prospective study. *Journal of personality*, 60(3), pp.599–620.

Valone, T.J., 1989. Group foraging, public information, and patch estimation. *Oikos*, pp.357–363.

VanderWaal, K. et al., 2013. Network structure and prevalence of Cryptosporidium in Belding’s ground squirrels. *Behavioral Ecology and Sociobiology*, 67(12), pp.1951–1959.

Vázquez, D.P. et al., 2007. Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116(7), pp.1120–1127.

Wagner, A., 2005. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays*, 27(2), pp.176–188.

Wasserman, S. & Faust, K., 1994. *Social network analysis: Methods and applications*, Cambridge university press.

Wasserman, S. & Pattison, P., 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika*, 61(3), pp.401–425.

Watson-Haigh, N.S., O’Neill, C.J. & Kadarmideen, H.N., 2012. Proximity loggers: data handling and classification for quality control. *Sensors Journal, IEEE*, 12(5), pp.1611–1617.

Watts, D.J. & Strogatz, S.H., 1998. Collective dynamics of small-world networks. *Nature*, 393(6684), pp.440–442.

Webb, B., 2002. Swarm Intelligence: From Natural to Artificial Systems. *Connection Science*, 14(2), pp.163–164.

Weber, N. et al., 2013. Denning behaviour of the European badger (Meles meles) correlates with bovine tuberculosis infection status. *Behavioral Ecology and Sociobiology*, 67(3), pp.471–479.

Welch, W.J., 1990. Construction of permutation tests. *Journal of the American Statistical Association*, 85(411), pp.693–698.

West-Eberhard, M.J., 1986. Dominance relations in Polistes canadensis (L.), a tropical social wasp. *Monitore Zoologico Italiano-Italian Journal of Zoology*, 20(3), pp.263–281.

West-Eberhard, M.J., 1983. Sexual selection, social competition, and speciation. *Quarterly review of biology*, pp.155–183.

Wey, T. et al., 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour,* 75, pp.333–344.

Wey, T.W. & Blumstein, D.T., 2010. Social cohesion in yellow-bellied marmots is established through age and kin structuring. *Animal Behaviour*, 79(6), pp.1343–1352.

Wheeler, D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, 39(10), p.2464.

Whitehead, H., 2008. *Analyzing animal societies: quantitative methods for vertebrate social analysis,* University of Chicago Press. London.

Whitehead, H., 1995. Investigating structure and temporal scale in social organizations using identified individuals. *Behavioral Ecology*, 6(2), pp.199–208.

Whitehead, H., Bejder, L. & Andrea Ottensmeyer, C., 2005. Testing association patterns: issues arising and extensions. *Animal Behaviour*, 69(5), p.e1.

Whitehead, H. & Dufault, S., 1999. Techniques for Analyzing Vertebrate Social Structure Using Identified Individuals: Review. *Advances in the Study of Behavior*, 28, p.33.

Whitehead, H. & James, R., 2015. Generalized affiliation indices extract affiliations from social network data. *Methods in Ecology and Evolution*. 6(7), pp.836-844.

Whitlock, M.C., 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of evolutionary biology*, 18(5), pp.1368–1373.

Wiegand, K.M., 1935. A taxonomist's experience with hybrids in the wild. *Science*, 81(2094), pp.161–166.

Willis, P.M., 2013. Why do animals hybridize? *Acta Ethologica*, 16(3), pp.127–134.

Willis, P.M., Symula, R.E. & Lovette, I.J., 2014. Ecology, song similarity and phylogeny predict natural hybridization in an avian family. *Evolutionary Ecology*, 28(2), pp.299–322.

Wilson, E.O., 2008. One giant leap: how insects achieved altruism and colonial life. *BioScience*, 58(1), pp.17–25.

Wilson, E.O., 1971. The insect societies. *Belknap Press of Havard University Press.* Cambridge.Massachusetts.

Wilson, E.O. & Hölldobler, B., 2005. Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, 102(38), pp.13367–13371.

Winer, B.J., Brown, D.R. & Michels, K.M., 1971. *Statistical principles in experimental design*, McGraw-Hill, New York.

Wittemyer, G., Douglas-Hamilton, I. & Getz, W.M., 2005. The socioecology of elephants: Analysis of the processes creating multitiered social structures. *Animal Behaviour*, 69(6), pp.1357–1371.

Wold, S. et al., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), pp.735–743.

Wolf, J.B.W. et al., 2007. Social structure in a colonial mammal: unravelling hidden structural layers and their foundations by network analysis. *Animal Behaviour*, 74(5), pp.1293–1302.

Wolf, M. et al., 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature*, 447(7144), pp.581–584.

Wolfe, K.H., Sharp, P.M. & Li, W.-H., 1989. Mutation rates differ among regions of the mammalian genome. *Nature*, 337(6204), pp.283–285.

World Health Organisation, 2005. *Ecosystems and human well-being*, Island Press Washington, DC.

Wu, C., 1995. Sexual Isolation in Drosophila melanogaster: A Possible Case of Incipient Speciation. *Proceedings of the National Academy of Sciences*, 92(7), pp.2519–2523.

Wu, C.-I. & Davis, A.W., 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *American Naturalist*, pp.187–212.

Xu, J.J. & Chen, H., 2005. Criminal Network Analysis And Visualization. *Communications of the ACM*, 48(6), pp.100–107.

Zanette, L.R.S. et al., 2014. Bumble bee workers drift to conspecific nests at field scales. *Ecological Entomology*, 39(3), pp.347–354.

Zbinden, J.A. et al., 2011. Migratory dichotomy and associated phenotypic variation in marine turtles revealed by satellite tracking and stable isotope analysis. *Marine Ecology Progress Series*, 421, pp.291–302.

Zhang, B. & Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1).

Zijlstra, B.J.H., Van Duijn, M.A.J. & Snijders, T.A.B., 2006. The multilevel p2 model. *Methodology*, 2(1), pp.42–47.

Zimmer, W.M.X. et al., 2003. Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (Physeter macrocephalus). *Oceanic Engineering, IEEE Journal of*, 28(1), pp.13–28.

Zuur, A. et al., 2009. *Mixed effects models and extensions in ecology with R*, Springer Science & Business Media.