

REVIEW

The application of statistical network models in disease research

Matthew J. Silk*,¹  Darren P. Croft², Richard J. Delahay³, David J. Hodgson⁴, Nicola Weber⁴, Mike Boots^{4,5} and Robbie A. McDonald*,¹

¹Environment and Sustainability Institute, University of Exeter, Penryn TR10 9FE, UK; ²Centre for Research in Animal Behaviour, University of Exeter, Exeter EX4 4QJ, UK; ³National Wildlife Management Centre, Animal and Plant Health Agency, Woodchester Park, Nympsfield, Stonehouse GL10 3UJ, UK; ⁴Centre for Ecology and Conservation, University of Exeter, Penryn TR10 9FE, UK; and ⁵Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA

Summary

1. Host social structure is fundamental to how infections spread and persist, and so the statistical modelling of static and dynamic social networks provides an invaluable tool to parameterise realistic epidemiological models.
2. We present a practical guide to the application of network modelling frameworks for hypothesis testing related to social interactions and epidemiology, illustrating some approaches with worked examples using data from a population of wild European badgers *Meles meles* naturally infected with bovine tuberculosis.
3. Different empirical network datasets generate particular statistical issues related to non-independence and sampling constraints. We therefore discuss the strengths and weaknesses of modelling approaches for different types of network data and for answering different questions relating to disease transmission.
4. We argue that statistical modelling frameworks designed specifically for network analysis offer great potential in directly relating network structure to infection. They have the potential to be powerful tools in analysing empirical contact data used in epidemiological studies, but remain untested for use in networks of spatio-temporal associations.
5. As a result, we argue that developments in the statistical analysis of empirical contact data are critical given the ready availability of dynamic network data from bio-logging studies. Furthermore, we encourage improved integration of statistical network approaches into epidemiological research to facilitate the generation of novel modelling frameworks and help extend our understanding of disease transmission in natural populations.

Key-words: contact network, epidemiology, exponential random graph model, network-based diffusion analysis, relational event model, stochastic actor-oriented model, temporal network autocorrelation model

Introduction

Direct contact is critical to the transmission of many of the most important infectious diseases and so an understanding of contact networks is integral to the epidemiology of many parasites and pathogens (Keeling & Eames 2005; Read, Eames & Edmunds 2008; Danon *et al.* 2011; Craft 2015). Populations are not completely mixed and significant population structure arises from spatial (Webb, Keeling & Boots 2007a,b) and social interactions. A growing number of empirical studies in humans (Rohani, Zhong & King 2010; Stehlé *et al.* 2011; Eames *et al.* 2012) and non-human animals (reviewed in Craft 2015; White, Forester & Craft 2017) have found important effects of social network structure on epidemiology, both at an individual and a population level. As a result, many epidemiological models now incorporate some concept of

non-random social structure that has important consequences for understanding the spread of infections (Keeling & Eames 2005; Lloyd-Smith *et al.* 2005; Craft 2015).

It may also be important to consider networks as dynamic, rather than static, structures, with changes affecting transmission over longer time-scales, particularly in endemic diseases (Funk, Salathé & Jansen 2010; Ezenwa *et al.* 2016; Silk *et al.* 2017). Not only will the temporal structure of interactions have a direct influence on transmission opportunities but social behaviour may also change in response to infection, including both the behaviour of the infected or diseased individual and the response of other individuals towards it (Bansal *et al.* 2010; Croft *et al.* 2011a). Furthermore, these changes in behaviour have been shown to alter contact network structure, with implications for transmission (Tunc & Shaw 2014; Lopes, Block & König 2016). Therefore, accounting for the co-dynamics of network structure and infection is key to improving our understanding of disease spread and control in many systems (fig. 1; Bansal *et al.* 2010; Wang *et al.* 2010).

*Correspondence authors. E-mails: matthewsilk@outlook.com; r.mcdonald@exeter.ac.uk

An increasing number of theoretical studies have modelled disease on dynamic networks (e.g. Eames *et al.* 2012; Tunc & Shaw 2014); however, there has been relatively little use of empirical data to explore this topic (but see Rohani, Zhong & King 2010; Reynolds *et al.* 2015; Lopes, Block & König 2016). Using empirical data to test hypotheses about the relationship between sociality and disease (e.g. Drewe 2010; Weber *et al.* 2013) will substantially advance our understanding of the dynamics of infection transmission, and using the outputs of statistical models could help improve the parameterisation of predictive, analytical epidemiological models (Rohani, Zhong & King 2010; Hamede *et al.* 2012; Reynolds *et al.* 2015). Nevertheless, there are unique problems associated with applying conventional statistical modelling approaches to network datasets (Croft *et al.* 2011b; Farine & Whitehead 2015). First, and perhaps most importantly, social networks recognise the influence of community members on each other, causing non-independence that must be accounted for statistically. Second, social networks are rarely described completely. The impact of sampling process on network parameters should be accounted for in statistical models. This is a particular problem if there is variation among individuals in the completeness of sampling. While this can be an issue for interaction-based networks (here defined as networks constructed from biologically relevant interactions), it is especially problematic in association-based networks (here defined as networks constructed by connecting individuals that have shared particular groups or spatio-temporal colocations rather than directly to each other), where the extent of sampling is harder to directly assess.

A range of modelling approaches (Table 1), developed within the field of social network analysis, could be applied to study infection in contact networks. These are split broadly into models that continue to use individual traits as a dependent variable while accounting for network structure, and models that use network topology as a dependent variable. The latter could be particularly valuable by directly relating network structure to infection and transmission. Several of these approaches model networks dynamically and offer great potential to improve our understanding of the dynamics of social behaviour and disease. Here we outline these statistical network approaches and provide a guide for how they can best be applied to test a variety of hypotheses related to infection in different types of network. For a selection of modelling frameworks, we use example data from a population of European badgers *Meles meles* naturally infected with bovine tuberculosis (bTB), to illustrate how the approaches can be applied.

Models for static networks

GENERAL AND GENERALISED LINEAR MODELS AND NETWORK AUTOCORRELATION MODELS

Traditional statistical modelling frameworks offer an appealing solution to understanding how infection status and social position covary with other individual traits. In particular, the use of generalised linear models (GLMs) and generalised linear mixed models (GLMMs) can help study the relationship between

social network position and disease state in the context of other predictor traits (e.g. sex, age, physiological condition), either controlling for these traits or considering interactions with them. However, the non-independence of nodes and edges within a network complicates the use of GLMs and GLMMs (Croft *et al.* 2011b), which assume statistical independence of residuals. Also, association-based networks (especially frequent for animal networks) can lead to further biases introduced by the method of network construction (see Farine & Whitehead 2015 for a simulated example of this).

One approach to adapt these modelling techniques appropriately to network data is to use permutation approaches that rely on randomisations of the network or original datastream (see Croft *et al.* 2011b; Farine & Whitehead 2015). A key difference here emerges between interaction networks and association-based networks. The latter requires permutation of the original datastream, due to additional sampling biases (Farine & Whitehead 2015). For these types of networks, other key considerations in implementing data permutations are likely to be the size of social groups, spatio-temporal constraints on interactions, differences in detectability of particular types of individuals and differences in the probabilities of interactions within, vs. outside, social groups (Croft *et al.* 2011b). While biases generated by incomplete sampling can still occur in interaction-based networks, there is greater potential to control this within a modelling framework. For example, if incomplete sampling results from differences in the length of time each individual is observed then this can be accounted for within any model used.

The R package **asnipe** (Farine 2013) offers a range of algorithms that shuffle association-based data to randomise such networks. However, it may be most appropriate to design system-specific randomisations. One problem worth highlighting is that using a permutation-based approach to test hypotheses creates confidence intervals around the null hypothesis rather than the estimated parameter. The development of approaches that generate uncertainty around observed network data would be highly beneficial in this regard. One example of this idea is provided by Farine & Strandburg-Peshkin (2015), who created probability distributions of edge weights using Bayesian inference. If GLM or GLMM analyses are completed within a Bayesian framework then this sort of uncertainty can be incorporated into the final analysis.

An alternative approach that can be used for interaction- or contact-based networks is to incorporate network autocorrelation into the model within a GLM or GLMM framework to address the issue of covariance driven by network structure. This can be achieved using the package **tnam** incorporated within the **xergm** suite of packages (Leifeld, Cranmer & Desmarais 2016), or the function **tnam()** in the package **sna** (Butts 2014) in R. The former is discussed here as it has more comprehensive provisions for dependency structures and can incorporate non-Gaussian error distributions. Models constructed using **tnam()** offer a variety of user-defined dependency terms that control for the expectation that individuals may influence other individuals they interact with within a network (see <https://cran.r-project.org/web/packages/tnam/tnam.pdf>).

Table 1. The advantages and disadvantages of the main statistical modelling approaches to studying contact networks for disease

Model	Dependent variable	Network type	When to use	Advantages	Disadvantages	Mathematical details	Software
Generalised linear (mixed) model (GLM/GLMM)	Individual traits	Static/ Dynamic	Can be used to test a whole range of hypotheses related to network position (with appropriate randomisations) <i>E.g. Do network positions of individuals infected with PathogenX show distinct properties from those of uninfected individuals?</i>	Familiarity of researchers with well-developed methods in animal social networks Can be used with group-based or association-based methods of network construction more easily	Not specifically designed to incorporate non-independence implicit in networks System-specific randomisations required that generate uncertainty around the null hypothesis rather than the observed parameter	Croft et al. (2011b); Farine & Whitehead (2015)	lme4 (modellingigraph/asnipe (randomisations)
Temporal network autocorrelation model (TNAM)	Individual traits	Static/ Dynamic	For testing hypotheses about how individual traits change in the context of a single network or series of network snapshots. <i>E.g. How do network position, past network position and the infection status of neighbouring individuals best explain infection with pathogenX?</i>	Can be used to explicitly account for non-independence of network data Enables the direct and indirect effects of other individuals in the network to be modelled. Same modelling framework can be applied to static and dynamic (multiple network snapshots) networks	Network dependency must be defined by user and its goodness-of-fit cannot be tested Complex to include interactions between more than two variables. [It is possible if the model matrix is generated using the function tnadata()] Robustness when used in group-based or association-based networks or with randomisation-based hypothesis testing unknown.	Leenders (2002); Hays, Kachi & Franzes (2010)	xergm (tnam)
Multiple regression quadratic assignment procedure (MRQAP)	Edge values	Static	For testing hypotheses about how relational traits are affected by other dyadic variables (i.e. matrix correlations) <i>E.g. Does there tend to be a difference in interaction strength between susceptible-susceptible-infected dyads</i>	Familiar to ecologists Accessible method to implement Can be used to analyse association-based animal networks (with appropriate randomisations)	No opportunity to model dependency structure of network No standard errors estimated around model parameters Problems in sparse networks and with collinear explanatory variables	Martin (1999); Dekker, Krackhardt & Snijders (2007)	sna, asnipe

Table 1. (continued)

Model	Dependent variable	Network type	When to use	Advantages	Disadvantages	Mathematical details	Software
Exponential random graph model (ERGM)	Network topology	Static	For testing hypotheses about the properties of edges or local network topology in a single network. <i>E.g. How does pathogen X infection affect an individual's social relationships?</i>	Modelling framework accounts for conditional dependence within the network Models the edges themselves, which are often of most interest from an epidemiological perspective Can include structural effects of biological interest or control for study design/social system e.g. distance/group membership	Lack of flexibility to have interaction terms within the model (It is possible to set up user-defined terms but this is challenging) Restricted to interaction- or contact-based network in which the researcher is confident of ties (use for group-based networks untested)	Robins <i>et al.</i> (2007); Lusher, Koskinen and Robins (2013)	<code>ergm</code> , <code>ergm.count</code> , <code>GERGM</code>
Latent space model	Edge values	Static/ Dynamic	For testing hypotheses about the properties of dyads in a single network (no inclusion of network topology). <i>E.g. How does pathogen X infection affect an individual's social relationships?</i>	Modelling framework accounts for conditional dependence within the network Models the edges themselves, which are often of most interest from an epidemiological perspective Generally simpler implementation and fitting than ERGMs as dependencies estimated automatically	Hypotheses related to network topology cannot be tested as network dependencies are included in the latent space component Lack of flexibility to have interaction terms within the model Use in association-based networks untested Interpretation of coefficients can be complex if correlated with the positions of nodes in latent space User-defined definitions of latent space need to be completed with caution	Hoff, Raftery & Handcock (2002); Krivitsky <i>et al.</i> (2009)	<code>latentnet</code>
Network-based diffusion analysis (NBDA)	Transmission process	Static	For testing the hypothesis that the acquisition of trait on a static network is a social process. <i>E.g. Does the spread of pathogen X depend on contact network structure?</i>	Simple to implement with a clear hypothesis test (whether the acquisition of a trait is best explained by social or non-social processes) that is highly relevant to disease research	Lack of flexibility Only takes into account a single static network structure (cf. <code>tnam</code>)	Franz & Nunn (2009); Nightingale <i>et al.</i> (2014)	code available online (see main text) <code>spatialnbd</code>

Table 1. (continued)

Model	Dependent variable	Network type	When to use	Advantages	Disadvantages	Mathematical details	Software
Stochastic actor-oriented model (SAOM)	Network topology and individual traits	Dynamic	For testing hypotheses related to how a trait influences an individual's dynamic network position or for testing hypotheses about how a trait and an individual's social network position are inter-related <i>E.g. How does infection with PathogenX covary with social behaviour?</i>	Accounts for conditional dependence within the network Can model both the probability of edges over time and differences in rates of network change depending on structural effects, and nodal and dyadic covariates	Restricted to interaction- or contact-based network in which the researcher is confident of ties. Use for association-based or group-based networks untested. Only possible to use for binary or ordinal networks Excessive changes in network composition over time can lead to estimation problems	Snijders, Van de Bunt & Steglich (2010)	Rsiena
Temporal exponential random graph model (TERGM)	Network topology	Dynamic	For testing hypotheses about the properties of edges or local network topology in a series of network snapshots. <i>E.g. How stable are social relationships and how does infection with PathogenX affect this?</i>	Modelling framework accounts for conditional dependence within the network Models the edges themselves, which are often of most interest from an epidemiological perspective Can include structural effects of biological interest or control for study design/ social system e.g. distance, group membership Temporal covariates enable tests of interaction stability and can interact with covariates to test how this affected by dyadic covariates Able to provide user-defined functions (which can be nonlinear) for temporal change in network structure	Lack of flexibility to have interaction terms within the model (It is possible to set up user-defined terms but this is challenging) Restricted to interaction- or contact-based network in which the researcher is confident of ties (use for group-based networks untested) Relative to SAOMs, less informative about rates of network change over time Missing data have to be imputed or the individuals removed from the network	xergm (btergm)	Hanneke, Fu & Xing (2010); Leifeld, Cramer and Desmarais (2017)

Table 1. (continued)

Model	Dependent variable	Network type	When to use	Advantages	Disadvantages	Mathematical details	Software
Relational event model (REM)	Interaction or contact events	Temporally explicit Dynamic	For testing hypotheses about the timing and patterns of interactions or contacts in temporally explicit data. <i>E.g. Is the temporal pattern of social contacts different for individuals infected with PathogenX?</i>	Temporally explicit Support constraints make the framework very adaptive as to appropriate datasets Does not require individuals to be present for the entire study period	More complex implementation and interpretation Harder to test hypotheses directly related to network structure and position than other approaches; this often has intuitive appeal for disease research.	Butts (2008)	relevant (+informR), rem

For example, the `weightlag()` or `netlag()` terms can incorporate autocovariance related to network distance or the `attribsim()` can incorporate autocovariance related to shared attribute values such as group membership. These functions can incorporate additional arguments to make dependency functions more complex. For example, the `netlag()` term can include a number of network steps over which autocovariance may be expected and a mathematical description of the decay. A potential disadvantage here is that dependency structures are defined by the user, and it is necessary for them to argue that the dependencies incorporated are appropriate and sufficient for the data in question (there is no goodness-of-fit test that allows this to be tested within the model). As well as incorporating these autocorrelation terms, network autocorrelation models (NAMs) can fit effects of nodal covariates that are either individual-level network metrics (e.g. centrality metrics, clustering coefficient) or exogenous to the network (e.g. sex, age etc.), and the interactions between them (see <https://cran.r-project.org/web/packages/tnam/tnam.pdf>). There are some potential issues with negatively biased parameter estimates for `netlag()` terms that should be considered when interpreting autocovariance terms in these models (Mizruchi & Neuman 2008; Neuman & Mizruchi 2010), although these are typically only problematic in high-density networks.

Network autocorrelation model for bTB infection in badgers

We provide an example of a NAM using our badger data in the supplementary material, in which we model bTB infection status as a function of sex, age and flow centrality while accounting for autocovariance among neighbouring individuals in the network. The results are presented in Table S1, Supporting Information. This modelling approach finds a positive effect of between-group flow centrality on the probability of bTB infection, as expected from the results of Weber *et al.* (2013). We also found a strong positive correlation between within-group eigenvector centrality and bTB infection, which is of interest as this was not a metric considered by Weber *et al.* (2013). The model also revealed a weak effect of increasing within-group degree on the probability of infection but we would encourage a tentative interpretation of this given the marginal effect and as no attempt has been made to control for the duration that individuals were monitored in our example analysis. These effects of centrality occur independently of differences associated with age class (adults being more likely to be infected than yearlings) and sex (males being more likely to be infected than females). Individuals were also less likely to be infected if their interactions were biased towards infected, not uninfected, individuals (the `weightlag()` term). Two phenomena are likely to contribute to this seemingly counter-intuitive finding. First, test positive individuals were considered to be infected (test positive by serology or interferon gamma release assay; see Weber *et al.* 2013) rather than necessarily infectious (test positive by bacterial culture), thus reducing the expectation of positive network covariance in infection. Second, infected individuals were distributed evenly among the

Table 2. Details of the type of model term, what type of network to use it in and guidance on how and when to use it for a selection of standard terms to consider when using ERGMs and TERGMs

ERGM term	Network type	Term type	Use to...
edges	Binary	Structural	Similar to an intercept in a GLM – gives the probability of edges in the network relative to a random network. Density is equivalent to edges divided by $n(n-1)/2$ (where n is the number of nodes in the network)
density			
nonzero	Weighted	Structural	Zero-inflation term in weighted networks (accounts for the fact that most networks are sparse and therefore distribution of edge weights is zero-inflated)
sum	Weighted	Structural	Similar to the intercept in a GLM for weighted networks
kstar(x:y)	Binary	Structural	A statistic for each kstar between x and y . $kstar(1)$ is equivalent to edges
triangle	Binary	Structural	A statistic for the number of triangles in the network (i.e. a measure of clustering/transitivity). $localtriangle(x)$ calculates only triangles between neighbours which are given using an indicator matrix x .
localtriangle(x)			
transitiveweights()	Weighted	Structural	Both of these terms can be used to calculate triangles in weighted networks taking into account the weights of edges
cyclicalweights()			
nodefactor(x)	Both	Node-based	The effect of a categorical nodal attribute on the probability/weight of edges
nodecov(x)	Both	Node-based	The effect of a continuous nodal attribute on the probability/weight of edges
nodematch(x)	Both	Node-based	The probability/weight of edges between two individuals of the same vs. different values of a categorical nodal attribute. The argument $diff=TRUE$ can provide separate estimates for each level of the factor
absdiff(x)	Both	Node-based	The effect of the difference in values of a nodal attribute between nodes on the probability/weight of an edge formed between them
absdiffcat(x)			
edgecov(x)	Both	Dyad-based	The effect of a dyadic covariate (e.g. relatedness) on the probability/weight of edges formed. Using $dyadcov(x)$ applies directed covariates when the network itself is directed
dyadcov(x)			
memory(type="")	Both	Temporal	The stability of edges over time. Additional arguments in type can be used to test different memory effects e.g. all potential edges ('stability') or only complete edges ('autoregression')
timecov(x, transform=function(t))	Both	Temporal	Trends in edge formation over time (nature of trend given by transform argument). Can additionally include a dyadic covariate x to create an interaction effect

badger social groups in the original study, which focussed on a subsample of the wider population with high bTB incidence (fig. 1 in Weber *et al.* 2013).

PARTIAL MATRIX REGRESSIONS USING QUADRATIC ASSIGNMENT PROCEDURES

Multiple regression quadratic assignment procedures (MRQAP) facilitate multivariate regressions between matrices with complex dependencies by using permutation-based estimates of statistical significance (Martin 1999; Dekker, Krackhardt & Snijders 2007; Cranmer *et al.* 2016). Therefore, they offer great utility as a tool to explain social network structure using a set of other dyadic relationships. For an ecologist, these are most likely to represent relatedness, some measure of spatial distance, or potentially some measure of difference in individual attributes (e.g. infection status). MRQAP is an accessible method already in use by ecologists. Its direct application to hypotheses related to infection is somewhat limited because it only models dyadic correlations; however, there are some situations where it may be useful. For example, VanderWaal *et al.* (2014) used MRQAP to compare social networks and transmission networks in giraffes *Giraffa camelopardalis* while controlling for a number of other variables such as spatial overlap. They showed that social network structure better explained transmission network structure than did networks of spatial overlap.

Multiple options are available for calculating MRQAP regressions for network data. Two more familiar options for

ecologists are the `netlm()` function in R package **sna** (Butts 2014), or the `mraqap.dsp()` and `mraqap.custom.null()` functions in **asnipe** (Farine 2013) that enable MRQAP to be used alongside randomisation-based approaches for networks of associations.

EXPONENTIAL RANDOM GRAPH MODELS

Exponential random graph models (ERGMs) form a class of statistical models specific to network analysis. They are edge-based models that model the probability (Robins *et al.* 2007; Lusher, Koskinen & Robins 2013) or weight (Desmarais & Cranmer 2012; Krivitsky 2012; Wilson *et al.* 2017) of each edge as a function of network structure and the characteristics of individuals (nodes) within the network. Local structural configurations can be used alongside nodal or edge covariates to model the pattern of edges observed (see Table 2). ERGMs fit parameters that produce a distribution of networks centred on the observed network (for more details see Lusher, Koskinen & Robins 2013). Goodness-of-fit of ERGMs can then be assessed by comparing (non-fitted) metrics from the simulated networks with those from the observed network (Lusher, Koskinen & Robins 2013). The fitting of ERGMs can be complicated by the fact that many parameter combinations can result in model degeneracy (producing model fits that are either very dense or sparse networks); however, this does reduce the likelihood of misspecified models being used. ERGMs are best used with contact or interaction-based data because association- or group-based methods of network construction include

uncertainty regarding the true nature of social associations and introduce sampling biases that need to be controlled for (Croft *et al.* 2011b). It may be possible to utilise two-mode ERGMs (modelling networks in which edges can only connect between two sets of nodes) for some association-based network data, especially when the links to specific locations are of interest (i.e. modelling what drives any individual's connections to particular locations or groups rather than to each other). In general, however, a restriction to interaction-based networks will not be a major issue in epidemiological research, which typically employs interaction-based networks.

An advantage of ERGMs is the ability to simulate networks based on the parameters for the structural features, and node and edge characteristics included in the observed network with an appropriately fitted model. ERGMs can be a powerful tool for parameterising uncertainty in any epidemiological models constructed (see Welch, Bansal & Hunter 2011), and this is likely to be especially useful in understanding disease epidemiology, as small differences in network structure have the potential to substantially alter transmission dynamics. This is especially true for studies that use simulation-modelling of the spread of disease across a network (see Reynolds *et al.* 2015). ERGMs also facilitate modelling of social contacts or interactions in response to individual traits, or the properties of dyads (other relationships between individuals such as relatedness). Individual traits (e.g. sex, age, disease state) can be used to explain both the tendency to form connections, and the likelihood of interacting with similar individuals (assortativity). This offers great potential to test hypotheses about the relationship between individual traits, including disease state, and network topology. For example, infected individuals having more interactions than uninfected individuals or tending to interact more frequently with susceptible individuals will increase risk of exposure at a population level. By contrast, assortment among infected individuals would signify that they associate disproportionately and therefore that infection may be socially, and perhaps spatially, restricted in the population. The same argument applies to traits that make individuals more susceptible to infection. Using relatedness as a dyadic variable is a good illustration: related individuals may be more likely to share a genetic susceptibility to some pathogens, so the relationship between the genetic structure and social structure of the population could influence the spatio-temporal distribution of infection.

ERGMs can be constructed using the packages **ergm** (Hunter *et al.* 2008; Handcock *et al.* 2015), **ergm.count** (Krivitsky 2015) and **GERGM** (Denny *et al.* 2016) in R. The package **ergm.count** extends ERGMs to Poisson and geometrically distributed edge weights and the package **GERGM** generalises ERGMs to all types of weighted network. The latter is a new tool and its use in the type of networks used for epidemiological research is untested. We provide the most relevant terms used in **ergm** and **ergm.count** in Table 2, and a full list of possible terms is included in the help pages for these packages. The range of possible terms is more limited for **GERGM**. The most important terms to include depend on the type of network being used, any structure implicit to it,

and the questions being asked (Table 2). R code for an example ERGM is provided in the supplementary material. The **simulate()** function in these packages can then be used to generate new networks based on the modelled parameters to assess goodness-of-fit or for use in further analysis or network models. We demonstrate its use in the supplementary material.

ERGM to relate bTB infection and network topology in badgers

We provide an example of ERGM in the supplementary information that finds no significant relationship between bTB infection and number of contacts in a binary network (although the weak positive effect might be stronger if contact strengths were considered), and reveals that males tended to have more contacts than females (Table S2). By using an ERGM, we were able to control for the structure imposed by social groups, and for variation in group size and the number of individuals collared within groups, in the model structure. One might also control for other constraints in the dataset using nodal or dyadic covariates, for example detection biases caused by variation in signal strength in proximity loggers (Drewe *et al.* 2012). We also used our ERGM to simulate badger networks with the same parameters fitted in the model, and show that they are broadly similar to the observed network, albeit not fully capturing the observed network structure (Fig. S1).

LATENT SPACE NETWORK MODELS

Latent space models offer an alternative method to ERGMs for the modelling of relational data. They effectively act as GLMs for edge values while controlling for network dependence by placing nodes in k-dimensional space according to their social network distance (Cranmer *et al.* 2016). Covariates can then include relational/dyadic properties (such as relatedness, or differences in a particular attribute) or an attribute of either node represented as a matrix with the same dimensions as the network. This means that the range of nodal and dyadic covariates is very similar to those for ERGMs (Cranmer *et al.* 2016). The potential applications to hypothesis testing in epidemiological studies are therefore broadly similar to ERGMs, but hypotheses about local network dependencies cannot be tested. Furthermore, interpretation of model coefficients can be complicated if the position of nodes in latent space covaries with values of nodal attributes (Cranmer *et al.* 2016).

Latent space models can be fitted in R using the package **latentnet** (Krivitsky & Handcock 2008, 2015). Latent space models can model weighted edges with a number of predefined error distributions. It is possible to use terms from the **ergm** package as explanatory variables in latent space models. However, these are limited to the binary variants of model terms, and do not include terms that induce dyadic dependence (such as those incorporating transitivity) as **latentnet** only fits models with dyadic independence. The other possible terms that can be included in the model are provided in the **latentnet**

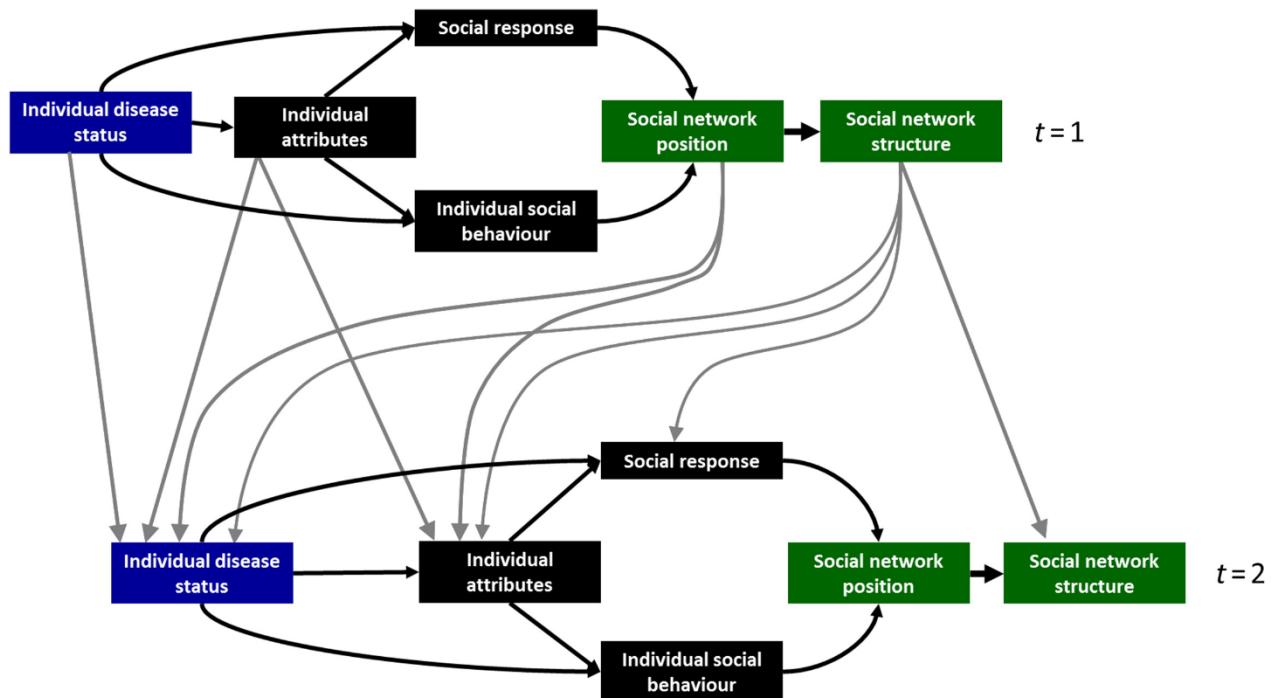


Fig. 1. The dynamics of social interactions and disease across two time points ($t = 1$ and $t = 2$). Models of static networks can only explore correlations at one point in time; by incorporating dynamic modelling approaches, it is possible to explore causation. Individual attributes in this graph refer to both fixed phenotypic traits such as sex, and conditional traits such as physiological stress, immunocompetence and condition. Social response represents the social behaviour of other individuals towards a focal individual.

manual (<https://cran.r-project.org/web/packages/latentnet/latentnet.pdf>).

2013; Aplin *et al.* 2015) or online (<http://lalandlab.st-andrews.ac.uk/freeware/>).

NETWORK-BASED DIFFUSION ANALYSIS

Network-based diffusion analysis (NBDA) compares the likelihood of explaining the spread of a trait through a population for two individual-based models; one assuming purely asocial acquisition of a trait, and the other purely social acquisition of a trait (Franz & Nunn 2009). This tests the extent to which social transmission is responsible for explaining the spread of that novel trait through a population. It requires that a single (static) social network and the specific timing of trait acquisition in each individual is known, although the latter can be order-based or timing-based (Hoppitt *et al.* 2010). Subsequent developments in the models have enabled Bayesian inference (Nightingale *et al.* 2014). This approach would be particularly valuable in determining the role of contact networks for the transmission of diseases that may have alternative hosts or be spread indirectly via the environment. This is because it tests the hypothesis that a trait spreads through a network, using asocial transmission as the null hypothesis. The use of NBDA in real-world populations may be slightly limited, however, by the requirement to know at least the order in which individuals acquired infection.

Lack of data on the order of infection precludes us from providing a badger case study; however, R Code to complete NBDA is available in the relevant literature (e.g. Allen *et al.*

Models for dynamic networks

Incorporating a dynamic view of population social structure will greatly enhance applications of social networks to epidemiology. Both social structure and infection are dynamic traits that interact at population and individual levels (Fig. 1; White, Forester & Craft 2017). Two categories of approaches have been suggested: (i) modelling the changes in a series of aggregated static networks using GLMMs, stochastic actor-oriented models (Snijders, Van de Bunt & Steglich 2010) and temporal ERGMs (Hanneke, Fu & Xing 2010), or (ii) using relational event models (Butts 2008) to model temporally explicit contact data. Both of these approaches, especially the latter, require high-resolution temporal data on social interactions (and to capture co-dynamics similar resolution data on infection), and so may be limited to detailed datasets.

GENERALISED LINEAR MIXED MODELS AND TEMPORAL NETWORK AUTOCORRELATION MODELS

Both randomisation-based GLMM and NAM approaches can be used to study a set of aggregated networks or network snapshots with, in the latter case, the models becoming temporal network autocorrelation models (TNAMs). Randomisation-based GLMM approaches can be extended to network snapshots by including individual as a random effect in a

model that relates social network position and disease state (alongside other variables of interest). It is also possible to incorporate change in values of network metrics over time as an additional variable to improve the extent to which these models capture the importance of social dynamics. When GLMMs are used to model a temporal series of networks, the simplest way to design appropriate randomisations would be to permute or randomise the network or association data within the sampling period used to construct each network snapshot (Farine & Whitehead 2015).

TNAMs can incorporate temporal autocorrelation by using the `lag` argument for each model term. This is equally applicable to the response variable re-fitted as a time-lagged covariate, e.g. an individual's disease state being dependent on its disease state in preceding time-steps; other covariates, e.g. an individual's disease state depending on body condition at a previous time-step as well as the current one; and network features, e.g. disease state could depend on the disease state of neighbouring individuals in the network at the current and preceding time-steps. For cases in which changes in disease state are regularly observed, this approach offers great potential to better appreciate the temporal scale over which social relationships influence acquisition of infection. The rate of change in observed bTB infection in badgers is too low relative to our 1-year sample of contact network data for it to be possible to provide a badger example, but the implementation of TNAMs in R (also using `tnam/xergm`) is very similar to that of NAMs.

STOCHASTIC ACTOR-ORIENTED MODELS

Stochastic actor-oriented models (SAOMs) use an individual-based approach to model how network structure changes through time, and can link these changes to structural features of the network, individual traits or dyadic covariates (Snijders, Van de Bunt & Steglich 2010; Fisher *et al.* 2017). Model terms (structural terms, and individual or dyadic covariates) can be used to explain both the rate that an individual has an opportunity to change to its network position (the 'rate' function) and the probability that it does so when the opportunity arises (the 'objective' function) (Snijders, Van de Bunt & Steglich 2010; Ripley, Snijders & Preciado 2011). Both individual and dyadic covariates can remain fixed (e.g. sex in our example) or change over time, but act only as explanatory variables (e.g. bTB infection in our example). Individual traits can also coevolve with network structure and form part of the response.

SAOMs are most appropriate for use with interaction- or contact-based networks, due to the similar constraints described for ERGMs (i.e. the uncertainty over the true nature of interactions and data structure in association-based networks). However, similar to ERGMs, it is possible to control for structural features in interaction- or contact-based data using covariates, e.g. distance effects or shared group effects (Fisher *et al.* 2017). SAOMs can currently model only binary or ordered networks, so are best used in cases where the presence/absence of an edge is more informative than its weight, or when network snapshots are constructed over relatively short

time windows (Fisher *et al.* 2017). However, being able to incorporate ordered networks does at least enable relationships of different strengths to be modelled separately (see <http://www.stats.ox.ac.uk/~snijders/siena/RscriptSienaOrdered.R>), which may be important for particular diseases or social systems.

A major advantage of using SAOMs is the ability to model the 'co-dynamics' of social strategy and infection status. This would enable better understanding of what drives the correlation between network position and infection status, especially important for research on endemic infections. For example, individuals with more contacts may be more at risk of infection, but it is equally possible that increases in social contacts are caused directly by infection or disease. Additionally, SAOMs enable the modelling of the influence of disease state and other variables (e.g. sex) on both the probability of individuals forming particular interactions and the rate at which they change these interactions. This helps disentangle how different social strategies influence susceptibility to disease. Finally, an extension of the SAOM framework enables a response variable, for example immunity, to be fixed once it is acquired, i.e. no return is possible to the original state (Ripley, Snijders & Preciado 2011; Greenan 2015), and this may facilitate the addition of immunity into hypothesis testing in real-world contact networks.

SAOMs are implemented in R using the package **RSiena** (Ripley, Boitmanis & Snijders 2013). Models are best constructed in a stepwise manner (see Supporting Information), starting with basic structural terms and adding in more complex structural terms, and then behavioural terms, once the current model converges and fits the data at each step (Ilany, Booms & Holekamp 2015; Fisher *et al.* 2017). The data requirements, as well as details on tests for model convergence, goodness-of-fit and significance, are provided elsewhere (Ripley, Snijders & Preciado 2011; Ilany, Booms & Holekamp 2015; Fisher *et al.* 2017). However, we highlight two important considerations of direct relevance to disease research. First, it is possible to include individuals that were not present at all time points by incorporating structural zeroes into the association matrices (Ripley, Snijders & Preciado 2011), meaning that individuals that enter or leave a population during the study period can be included. Second, if a trait is intended to coevolve with network structure in the model, it must be a binary or ordinal variable. In disease modelling, this is likely to be equivalent to classifying individuals as uninfected or infected, or to using numbers that reflect progressive disease states. For example, multiple classes used to describe bTB infection states in European badgers (e.g. Graham *et al.* 2013) could be coded ordinally.

Using a SAOM to examine seasonal changes in badger interactions

We use an SAOM to explore badger social network dynamics from summer through winter, showing that there is no evidence for bTB increasing either the probability of

interactions or the rate at which interactions change for a binary network of all interactions (potentially as a result of using a binary contact network, and the reduced subset of individuals included; $n = 36$, cf. $n = 51$ for the ERGM). However, there are interesting differences in the rate of network change between the sexes, with males changing their interactions faster than females between summer and winter. Differences such as this may provide a behavioural explanation for males being more likely to acquire infection than females in this system (Graham *et al.* 2013). Furthermore, the significant effects of distance between burrows and shared group membership reveal the importance of spatial behaviour in structuring the badger social system, and highlight the importance of accounting for data structure when using this sort of statistical model.

TEMPORAL EXPONENTIAL RANDOM GRAPH MODELS

Temporal ERGMs (TERGMs) represent a generalisation of the ERGM framework to a temporal series of static networks (Hanneke, Fu & Xing 2010; Leifeld, Cranmer & Desmarais 2017). TERGMs assume that a network in one time-step is dependent on network structure in the preceding time-steps, with the number of previous time-steps used determined by a parameter within the model.

The ability to simulate networks in longitudinal datasets is a particular advantage of using TERGMs. Studies that use network models of disease in animals often encompass change in network structure over time, for example in response to seasonal changes (Reynolds *et al.* 2015). Therefore, TERGMs offer an ideal framework to simulate networks into the future, based on a set of network snapshots. In terms of hypothesis testing, the incorporation of temporal dependencies can enable (i) the role of disease in network topology to be estimated while accounting for variation in interaction stability over time or (ii) the role of disease state in influencing temporal changes in interactions to be estimated (if disease state of two individuals is included as a dyadic covariate).

TERGMs can be fitted using the package **btergm**, part of the **xergm** package suite (Leifeld, Cranmer & Desmarais 2016) in R. The TERGM framework can handle changes in network size between time-steps if row or column labels are provided in the matrix. This can be achieved by removing these nodes or by incorporating them as structural zeroes. However, within a time-step, individuals must possess a full set of network information and covariate values. If this is problematic, it is possible to impute values either for covariates or network data (e.g. Koskinen *et al.* 2013). Basic imputation can be done within the **xergm** package.

The **btergm()** function enables models containing time-dependent covariates (`timecov()` argument) and effects of tie stability (`memory()` argument) and delayed reciprocity (`delrecip()` argument for directed networks) to be fitted alongside conventional ERGM terms (Table 2; Leifeld, Cranmer & Desmarais 2017). The parameter k defines the number of preceding time-steps which affect the current time-step. It is possible for k to take values greater than 1, but as k increases,

the number of time-steps remaining to model reduces, placing a constraint upon the user. The `timecov()` argument enables interactions between dyadic covariates and temporal trends in edge formation (with the exact nature of the temporal trend provided as a function by the user) so is likely to be especially useful in understanding differences in interactions linked to infection status. The provision of a user-defined temporal pattern of interactions requires some careful thought from the researcher when implementing the model, but provides a more flexible tool for defining temporal change in network structure than available in SAOMs. Furthermore, other dyadic covariates can vary through time if they are provided as a list of matrices. This is likely to be particularly relevant to individual-level variables that also vary temporally.

Example TERGMs for badger-TB epidemiology

We provide some basic examples of the fitting of TERGMs to our dataset in the supplementary material, making use of the same subset of data used for the SAOM example. While only using a temporal series of three networks restricted us to simple model constructs, we show how the different terms can be used to test hypotheses about temporal changes in network structure alongside hypotheses related to individual-level covariates. The first example model shows that there is greater stability in badger contact networks than expected by chance (Table S4), while the second shows that there is a decline in the probability of contacts between summer and winter (Table S5). There is no consistent pattern between models for the effects of bTB infection and sex, suggesting the use of binary network data might be limiting the power of detecting these effects. These example models are also used to show how to use goodness-of-fit tests for TERGMs (Fig. S3). For further information we refer readers to Leifeld & Cranmer (2015) and Leifeld, Cranmer & Desmarais (2017).

RELATIONAL EVENT MODELS

Relational event models (REMs) provide a modelling framework capable of analysing data on contacts, interactions or associations that have not been aggregated, remain temporally explicit and are instantaneous events without measurable duration (Butts 2008; Tranmer *et al.* 2015). The concept is similar to event models used in survival analysis, and estimates a hazard function for the rate of interaction events conditional on covariates measured on either individuals or events, and also on patterns of these interactions in the past (Tranmer *et al.* 2015). Within a ‘relational’ framework, it is possible to additionally estimate coefficients for the influence of network effects on these events such as transitivity – a tendency to interact with ‘friends of friends’ (Butts 2008). It is now possible to incorporate a decay function so that events that have happened more recently have a greater effect (Lerner *et al.* 2013). In addition, another recent extension of the REM framework can be used to make them applicable to two-mode networks (Brandenberger 2016), in which edges can only connect between two independent sets of nodes. This could extend their use to

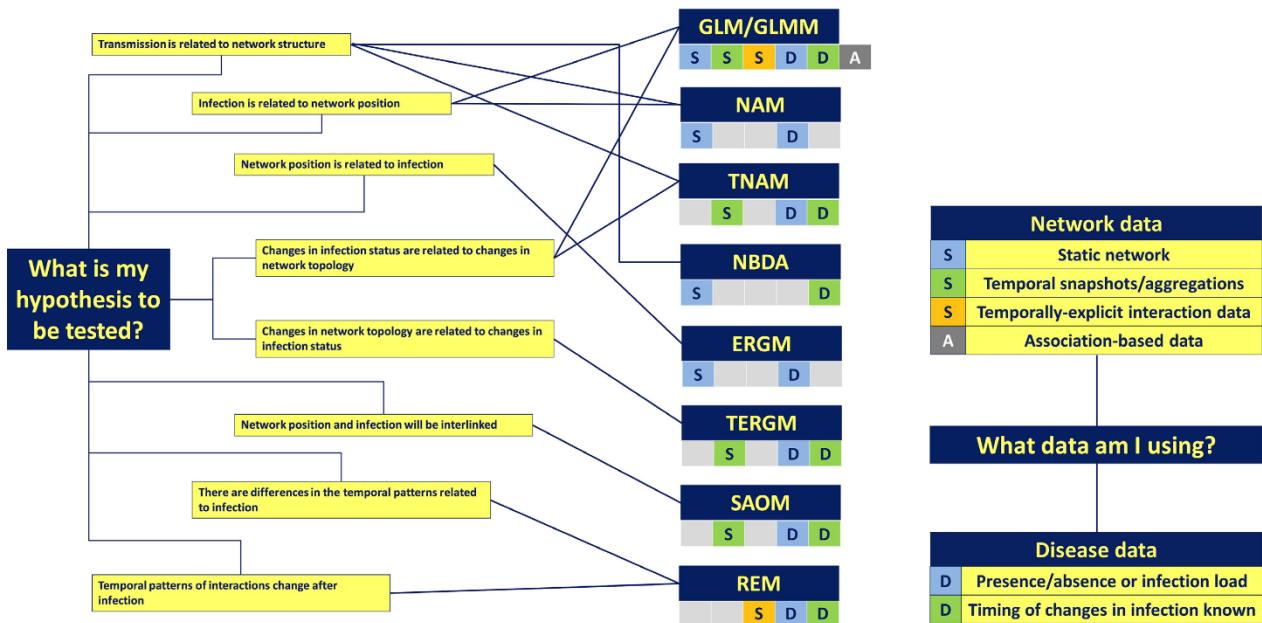


Fig. 2. A guide to statistical model use to test hypotheses about the relationship between social contacts/interactions and disease for the most appropriate models to test hypotheses about networks and disease. GLM, generalised linear model; GLMM, generalised linear mixed model; NAM, network autocorrelation model; TNAM, temporal network autocorrelation model; ERGM, exponential random graph model; NBDA, network-based diffusion analysis; SAOM, stochastic actor-oriented model; TERGM, temporal exponential random graph model; REM, relational events model.

association-based networks in which individuals are connected to particular groups or locations rather than directly to each other.

The potential applications of REMs to wildlife disease research are manifold, especially given the growing number of studies in this field that use temporally explicit data from proximity loggers (e.g. Hamede *et al.* 2009; Cross *et al.* 2012; Weber *et al.* 2013). This framework could be highly informative in understanding how the acquisition or progression of an infection influences the likelihood of repeat social contacts with uninfected individuals, or the persistence of an individual's social associations (Fig. 1). Additionally, for populations in which social structure represents an important barrier to the spread of infection, REMs would facilitate the modelling of differences between the dynamics of intra-group and inter-group interactions. The temporal structure of inter-group interactions would be expected to have a substantial effect on disease spread and previous interactions within a dyad, especially those in the recent past, could increase the likelihood of further interactions occurring. Finally, differences in these parameters between the sexes or for individuals of different ages might explain patterns of age- or sex-biased infection.

REMs can be fitted in R using the package **rem** (Brandenberger 2016) or using the package **relevent** (Butts 2008), with prior data manipulation requiring the package **informR** (Marcum & Butts 2015). This includes the addition of support constraints (additional binary indicators within the model that restrict which actions or events are possible) that can help account for elements of the study design, and therefore are likely to be particularly beneficial in studies of animals (Tranmer *et al.* 2015). For example, support constraints could

inform a model when individuals are collared in a contact network study, or to indicate whether two individuals are on different sides of a geographical barrier (e.g. a river) and therefore unable to interact. Extensions to incorporate weightings on temporal dependencies among events are incorporated in the **rem** package.

Choosing a model

With such a wealth of approaches, it may not be immediately clear which offers the most appropriate tool to test a particular hypothesis. In Table 1, we outline the advantages and disadvantages of using all of the modelling frameworks outlined here. In Fig. 2, we provide a data- and question-driven approach to selecting the most suitable statistical tool. For further comparisons between statistical models of networks, and guidance to their usage, we refer readers to recent reviews in other subject areas (Hunter, Krivitsky & Schweinberger 2012; Leifeld & Cranmer 2015; Cranmer *et al.* 2016). In addition to using statistical network models, it may also be possible to use statistical models of contact rates to test hypotheses relating disease and social behaviour, especially within social groups (Cross *et al.* 2012).

There are a few important general rules to consider when selecting a modelling framework. The first of these is how the network data are obtained. Networks constructed using group-based (or association-based) approaches contain data structure and biases that on current knowledge require randomisation-based approaches that employ GLMs or GLMMs. For networks constructed from defined social contacts or interactions, then any approach could be useful depending on the question of interest. If data are temporally

explicit (time-ordered) then the use of REMs offers the most powerful analytical approach by facilitating the use of temporal patterns of contacts in addition to their structure. However, these models are complex to construct and so for answering simpler questions it might be appropriate to aggregate data into a temporal series of networks and use simpler approaches. It may even be that for some questions aggregating all network data into a single static network still enables the relevant hypotheses to be tested.

When selecting between network-focussed statistical models – (T)ERGMs, (T)NAMs and SAOMs – a fundamental first consideration is whether the hypotheses being tested are related to properties of relational data or the properties of nodes. For hypotheses related to network topology, (T)ERGMs and SAOMs are most appropriate, while for nodes, (T)NAMs are best (or alternatively GLMMS with randomisations). Many hypotheses revolving around the topic of social behaviour and disease are most suitable for testing using models of network topology. For example, any question asking whether diseased individuals show different patterns of social behaviour to non-diseased individuals, or asking how social behaviour changes as infection state changes are ‘network topology’ questions. (T)NAMs are especially useful in testing hypotheses linking change in infection status to the network position of an individual and the infection status of individuals surrounding it in the network (alongside any other individual-level fixed effects). Thus, modelling how network structure influences the probability of acquiring infection should be considered a ‘node-based’ question.

Missing information and hypothesis testing in networks

Many network studies of disease transmission are likely to contain missing information, either because they are based on a subsample of the total population or record only a subset of the interactions that occur among individuals. Few studies have investigated the impact of missing information on network analysis (but see e.g. Lee, Kim & Jeong 2006; Smith & Moody 2013; Silk *et al.* 2015; Smith, Moody & Morgan 2017), and none has gone on to test how different types and levels of missing information affect hypothesis testing approaches. As a result, we would currently urge caution in applying these methods where networks are constructed using only a small proportion of individuals within a study population. An alternative option when there are high levels of missing information is to model contact rates independently of network structure, for example the methods outlined in Cross *et al.* (2012). If statistical network methods are influenced in different ways by the subsampling of network data, then the choice of model might also depend on the level of sampling in the network of interest. For example, Shalizi & Rinaldo (2013) suggested that an ERGM based on a sampled network is unlikely to reflect population-level parameters, although how this might affect the testing of hypotheses is unclear. Conversely, Páez, Scott

& Volz (2008) found that the power of NAMs to detect network effects remained high until a majority of edge information was missing. Developing an improved understanding of how different modelling approaches are affected by sampling of a network will be a valuable area of future methodological research.

Network approaches and epidemiological modelling

A natural end point of applying social network analytical methods to the study of disease is in helping to construct and parameterise epidemiological models and there are numerous advantages of this approach. First, uncertainty can be incorporated more easily – any estimates for structural effects or individual differences from ERGMs, SAOMs or REMs will include standard errors, which can be included to test the robustness of the conclusions drawn from the model. Second, statistical models (especially ERGMs) facilitate the easy simulation of large number of networks with equivalent expected properties to the observed network, useful for simulation-modelling of disease. Third, the use of dynamic statistical models (SAOMs, temporal ERGMs) makes it easier to incorporate information on network dynamics into any constructed models. For SAOMs in particular, the ability to estimate the co-dynamics of social strategy and disease could have major implications (e.g. the inclusion of avoidance behaviour in epidemiological models: Shaw & Schwartz 2008; Tunc & Shaw 2014). As a result, the incorporation of these statistical network models alongside epidemiological models offers great potential to develop stronger links between empirical data and disease modelling, especially in models of endemic diseases, for which the co-dynamics of social systems and infection are likely to be more important.

Conclusions and future directions

There is considerable scope to extend current modelling frameworks and it would be highly beneficial for epidemiological researchers to become more involved in their continued development. For example, many of these methods are rather poor at dealing with missing data, and integrating elements from Bayesian population models (using state-space/multi-state models to address the issue of missing data and hidden states: Kéry & Schaub 2012) and models of network topology could make substantial advances in dealing with this issue.

Developments in hypothesis testing in networks will enable important progress in understanding the links between individuals, social structure and infection. This is especially true for endemic infections, such as with our worked examples of bTB in badgers, where the longer time-scales involved will mean that understanding the dynamic interaction between social behaviour and disease is that much more important. Furthermore, implementing statistical approaches specifically designed to model networks can facilitate more detailed parameterisation of epidemiological models and provide an

idea of uncertainty around key parameters. Together this means that statistical models of networks can offer a powerful tool in linking empirical data on population social structures with theoretical models of disease.

Authors' contributions

M.J.S. conceived the manuscript, with all authors contributing to developing its full scope. M.J.S. conducted example analyses. N.W., R.J.D. and R.A.M. were involved in the original data collection. All authors contributed to drafts and gave final approval for publication.

Acknowledgements

M.J.S. is funded by a NERC standard grant (NE/M004546/1) awarded to R.A.M., D.P.C., D.J.H. and M.B., with the APHA team at Woodchester Park, UK (lead scientist is R.J.D.) as project partners. Data used in example analyses were collected for NW's PhD funded by Defra. We thank two anonymous reviewers for useful comments on the manuscript.

Data accessibility

Full r code for example models are provided in the supplementary information. The data analysed are also provided as supplementary files.

References

- Allen, J., Weinrich, M., Hoppitt, W. & Rendell, L. (2013) Network-based diffusion analysis reveals cultural transmission of lobtail feeding in humpback whales. *Science*, **340**, 485–488.
- Aplin, L.M., Farine, D.R., Morand-Ferron, J., Cockburn, A., Thornton, A. & Sheldon, B.C. (2015) Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, **518**, 538–541.
- Bansal, S., Read, J., Pourbohloul, B. & Meyers, L.A. (2010) The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, **4**, 478–489.
- Brandenberger, L. (2016) *Rem: Relational Event Models*. R package version 1.1.2. Available at: <https://cran.r-project.org/web/packages/rem/index.html>. (accessed 29 March 2017).
- Butts, C.T. (2008) A relational event framework for social action. *Sociological Methodology*, **38**, 155–200.
- Butts, C.T. (2014) sna: Tools for Social Network Analysis. R package version 2.3-2. Available at: <https://CRAN.R-project.org/package=sna>. (accessed 29 March 2017).
- Craft, M.E. (2015) Infectious disease transmission and contact networks in wildlife and livestock. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **370**, 20140107.
- Crammer, S.J., Leifeld, P., McClurg, S.M. & Rolfe, M. (2016) Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science*, **61**, 237–251.
- Croft, D.P., Edenbrow, M., Darden, S.K., Ramnarine, I.W., van Oosterhout, C. & Cable, J. (2011a) Effect of gyrodytlyid ectoparasites on host behaviour and social network structure in guppies Poecilia reticulata. *Behavioral Ecology and Sociobiology*, **65**, 2219–2227.
- Croft, D.P., Madden, J.R., Franks, D.W. & James, R. (2011b) Hypothesis testing in animal social networks. *Trends in Ecology & Evolution*, **26**, 502–507.
- Cross, P.C., Creech, T.G., Ebinger, M.R., Heisey, D.M., Irvine, K.M. & Creel, S. (2012) Wildlife contact analysis: emerging methods, questions, and challenges. *Behavioral Ecology and Sociobiology*, **66**, 1437–1447.
- Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V. & Vernon, M.C. (2011) Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, **2011**, 284909.
- Dekker, D., Krackhardt, D. & Snijders, T.A.B. (2007) Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, **72**, 563–581.
- Denny, M.J., Wilson, J.D., Crammer, S., Desmarais, B.A. & Bhamidi, S. (2016) GERGM: Estimation and fit diagnostics for generalized exponential random graph models. R package version 0.10.0. Available at: <https://CRAN.R-project.org/package=GERGM>. (accessed 29 March 2017).
- Desmarais, B.A. & Crammer, S.J. (2012) Statistical inference for valued-edge networks: the generalized exponential random graph model. *PLoS ONE*, **7**, e30136.
- Drewe, J.A. (2010) Who infects whom? Social networks and tuberculosis transmission in wild meerkats. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 633–642.
- Drewe, J.A., Weber, N., Carter, S.P., Bearhop, S., Harrison, X.A., Dall, S.R., McDonald, R.A. & Delahay, R.J. (2012) Performance of proximity loggers in recording intra- and inter-species interactions: a laboratory and field-based validation study. *PLoS ONE*, **7**, e39068.
- Eames, K.T.D., Tildon, N.L., Brooks-Pollock, E. & Edmunds, W.J. (2012) Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Computational Biology*, **8**, e1002425.
- Ezenwa, V.O., Archie, E.A., Craft, M.E., Hawley, D.M., Martin, L.B., Moore, J. & White, L. (2016) Host behaviour-parasite feedback: an essential link between animal behaviour and disease ecology. *Proceedings of the Royal Society of London B: Biological Sciences*, **283**, 20153078.
- Farine, D.R. (2013) Animal social network inference and permutations for ecologists in R using asnipe. *Methods in Ecology and Evolution*, **4**, 1187–1194.
- Farine, D.R. & Strandburg-Peshkin, A. (2015) Estimating uncertainty and reliability of social network data using Bayesian inference. *Open Science*, **2**, 150367.
- Farine, D.R. & Whitehead, H. (2015) Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, **84**, 1144–1163.
- Fisher, D., Ilany, A., Silk, M. & Tregenza, T. (2017) Analysing animal social network dynamics: the potential of stochastic actor-oriented models. *Journal of Animal Ecology*, **86**, 202–212.
- Franz, M. & Nunn, C.L. (2009) Network-based diffusion analysis: a new method for detecting social learning. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1829–1836.
- Funk, S., Salathé, M. & Jansen, V.A.A. (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, **7**, 1247–1256.
- Graham, J., Smith, G.C., Delahay, R.J., Bailey, T., McDonald, R.A. & Hodgson, D. (2013) Multi-state modelling reveals sex-dependent transmission, progression and severity of tuberculosis in wild badgers. *Epidemiology and Infection*, **141**, 1429–1436.
- Greenan, C.C. (2015) Diffusion of innovations in dynamic networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **178**, 147–166.
- Hamede, R.K., Bashford, J., McCallum, H. & Jones, M. (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**, 1147–1157.
- Hamede, R., Bashford, J., Jones, M. & McCallum, H. (2012) Simulating devil facial tumour disease outbreaks across empirically derived contact networks. *Journal of Applied Ecology*, **49**, 447–456.
- Handcock, M., Hunter, D., Butts, C., Goodreau, S., Krivitsky, P. & Morris, M. (2015) _ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks_. The Statnet Project (<http://www.statnet.org>). R package version 3.5.1, Available at: <https://CRAN.R-project.org/package=ergm>. (accessed 29 March 2017).
- Hanneke, S., Fu, W. & Xing, E.P. (2010) Discrete temporal models of social networks. *Electronic Journal of Statistics*, **4**, 585–605.
- Hays, J.C., Kachi, A. & Franzes, R.J. Jr (2010) A spatial model incorporating dynamic, endogenous network interdependence: a political science application. *Statistical Methodology*, **7**, 406–428.
- Hoff, P.D., Raftery, A.E. & Handcock, M.S. (2002) Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, **97**, 1090–1098.
- Hoppitt, W., Kandler, A., Kendal, J.R. & Laland, K.N. (2010) The effect of task structure on diffusion dynamics: implications for diffusion curve and network-based analyses. *Learning & Behavior*, **38**, 243–251.
- Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M. & Morris, M. (2008) ergm: a package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, **24**, 1–29.
- Hunter, D.R., Krivitsky, P.N. & Schweinberger, M. (2012) Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, **21**, 856–882.
- Ilany, A., Booms, A.S. & Holekamp, K.E. (2015) Topological effects of network structure on long-term social network dynamics in a wild mammal. *Ecology Letters*, **18**, 687–695.
- Keeling, M.J. & Eames, K.T.D. (2005) Networks and epidemic models. *Journal of the Royal Society Interface*, **2**, 295–307.
- Kéry, M. & Schaub, M. (2012) *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. Academic Press, Waltham, MA, USA.

- Koskinen, J.H., Robins, G.L., Wang, P. & Pattison, P.E. (2013) Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, **35**, 514–527.
- Kravitzky, P.N. (2012) Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, **6**, 1100.
- Kravitzky, P. (2015) `ergm.count`: Fit, Simulate and Diagnose Exponential-Family Models for Networks with Count Edges_. The Statnet Project (<http://www.statnet.org>). R package version 3.2.0, Available at: <http://CRAN.R-project.org/package=ergm.count>. (accessed 29 March 2017).
- Kravitzky, P.N. & Handcock, M.S. (2008) Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, **24**, 1–23.
- Kravitzky, P. & Handcock, M. (2015). `latentnet`: Latent Position and Cluster Models for Statistical Networks_. The Statnet Project (<http://www.statnet.org>). R package version 2.7.1, Available at: <http://CRAN.R-project.org/package=latentnet>. (accessed 29 March 2017).
- Kravitzky, P.N., Handcock, M.S., Raftery, A.E. & Hoff, P.D. (2009) Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, **31**, 204–213.
- Lee, S.H., Kim, P.-J. & Jeong, H. (2006) Statistical properties of sampled networks. *Physical Review E*, **73**, 016102.
- Leenders, R.T.A.J. (2002) Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, **24**, 21–47.
- Leifeld, P. & Cranmer, S.J. (2015) A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *arXiv preprint arXiv:1506.06696*.
- Leifeld, P., Cranmer, S.J. & Desmarais, B.A. (2017) Temporal Exponential Random Graph Models with `xergm`: estimation and Bootstrap Confidence Intervals. *Journal of Statistical Software* (accepted).
- Leifeld, P., Cranmer, S.J. & Desmarais, B.A. (2016) `xergm`. Extensions for Exponential Random Graph Models. R package version 1.7.0.
- Lerner, J., Bussmann, M., Snijders, T.A. & Brandes, U. (2013) Modeling frequency and type of interactions in event networks. *Corvinus Journal of Sociology and Social Policy*, **4**, 3–32.
- Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E. & Getz, W.M. (2005) Superspreaders and the effect of individual variation on disease emergence. *Nature*, **438**, 355–359.
- Lopes, P.C., Block, P. & König, B. (2016) Infection-induced behavioural changes reduce connectivity and the potential for disease spread in wild mice contact networks. *Scientific Reports*, **6**, 31790.
- Lusher, D., Koskinen, J. & Robins, G. (2013) *Exponential Random Graph Models for Social Networks*. Cambridge University Press, Cambridge, UK.
- Marcum, C.S. & Butts, C.T. (2015) Constructing and modifying sequence statistics for relevant using `informR` in R. *Journal of Statistical Software*, **64**, 1–36. Available at: <http://www.jstatsoft.org/v64/i05/>. (accessed 29 March 2017).
- Martin, J.L. (1999) A general permutation-based QAP analysis approach for dyadic data from multiple groups. *Connections*, **22**, 50–60.
- Mizrahi, M.S. & Neuman, E.J. (2008) The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, **30**, 190–200.
- Neuman, E.J. & Mizrahi, M.S. (2010) Structure and bias in the network autocorrelation model. *Social Networks*, **32**, 290–300.
- Nightingale, G., Boogert, N.J., Laland, K.N. & Hoppitt, W. (2014) Quantifying diffusion in social networks: a Bayesian approach. *Animal Social Networks*, (eds J. Krause, R. James, D. Franks & D. Croft), pp. 38–52. Oxford University Press, Oxford, UK.
- Páez, A., Scott, D.M. & Volz, E. (2008) Weight matrices for social influence analysis: an investigation of measurement errors and their effect on model identification and estimation quality. *Social Networks*, **30**, 309–317.
- Read, J.M., Eames, K.T.D. & Edmunds, W.J. (2008) Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface*, **5**, 1001–1007.
- Reynolds, J.J.H., Hirsch, B.T., Gehrt, S.D. & Craft, M.E. (2015) Raccoon contact networks predict seasonal susceptibility to rabies outbreaks and limitations of vaccination. *Journal of Animal Ecology*, **84**, 1720–1731.
- Ripley, R.M., Snijders, T.A.B. & Preciado, P. (2011) *Manual for RSiena*. University of Oxford, Department of Statistics, Nuffield College, Oxford, UK.
- Ripley, R., Boitmanis, K. & Snijders, T.A.B. (2013) RSiena: Siena - simulation investigation for empirical network analysis. R package version 1.1-232. Available at: <http://CRAN.R-project.org/package=RSiena>.
- Robins, G., Pattison, P., Kalish, Y. & Lusher, D. (2007) An introduction to exponential random graph (p*) models for social networks. *Social Networks*, **29**, 173–191.
- Rohani, P., Zhong, X. & King, A.A. (2010) Contact network structure explains the changing epidemiology of pertussis. *Science*, **330**, 982–985.
- Shalizi, C.R. & Rinaldo, A. (2013) Consistency under sampling of exponential random graph models. *Annals of Statistics*, **41**, 508–535.
- Shaw, L.B. & Schwartz, I.B. (2008) Fluctuating epidemics on adaptive networks. *Physical Review E*, **77**, 66101.
- Silk, M.J., Jackson, A.L., Croft, D.P., Colhoun, K. & Bearhop, S. (2015) The consequences of unidentifiable individuals for the analysis of an animal social network. *Animal Behaviour*, **104**, 1–11.
- Silk, M.J., Croft, D.P., Delahay, R.J., Hodgson, D.J., Boots, M., Weber, N. & McDonald, R.A. (2017) Using social network measures in wildlife disease ecology, epidemiology, and management. *BioScience*, **67**, 245–257.
- Smith, J.A. & Moody, J. (2013) Structural effects of network sampling coverage I: nodes missing at random. *Social Networks*, **35**, 652–688.
- Smith, J.A., Moody, J. & Morgan, J.H. (2017) Network sampling coverage II: the effect of non-random missing data on network measurement. *Social Networks*, **48**, 78–99.
- Snijders, T.A.B., Van de Bunt, G.G. & Steglich, C.E.G. (2010) Introduction to stochastic actor-based models for network dynamics. *Social Networks*, **32**, 44–60.
- Stehlé, J., Voisin, N., Barrat, A. et al. (2011) Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, **9**, 1.
- Tranmer, M., Marcum, C.S., Morton, F.B., Croft, D.P. & de Kort, S.R. (2015) Using the relational event model (REM) to investigate the temporal dynamics of animal social networks. *Animal Behaviour*, **101**, 99–105.
- Tunc, I. & Shaw, L.B. (2014) Effects of community structure on epidemic spread in an adaptive network. *Physical Review E*, **90**, 22801.
- VanderWaal, K.L., Atwill, E.R., Isbell, L.A. & McCowan, B. (2014) Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*). *Journal of Animal Ecology*, **83**, 406–414.
- Wang, B., Cao, L., Suzuki, H. & Aihara, K. (2010) Epidemic spread in adaptive networks with multitype agents. *Journal of Physics A: Mathematical and Theoretical*, **44**, 35101.
- Webb, S.D., Keeling, M.J. & Boots, M. (2007a) Host-parasite interactions between the local and the mean-field: how and when does spatial population structure matter? *Journal of Theoretical Biology*, **249**, 140–152.
- Webb, S.D., Keeling, M.J. & Boots, M. (2007b) Spatially extended host-parasite interactions: the role of recovery and immunity. *Theoretical Population Biology*, **71**, 251–266.
- Weber, N., Carter, S.P., Dall, S.R.X., Delahay, R.J., McDonald, J.L., Bearhop, S. & McDonald, R.A. (2013) Badger social networks correlate with tuberculosis infection. *Current Biology*, **23**, R915–R916.
- Welch, D., Bansal, S. & Hunter, D.R. (2011) Statistical inference to advance network models in epidemiology. *Epidemics*, **3**, 38–45.
- White, L.A., Forester, J.D. & Craft, M.E. (2017) Using contact networks to explore mechanisms of parasite transmission in wildlife. *Biological Reviews*, **92**, 389–409.
- Wilson, J.D., Denny, M.J., Bhamidi, S., Cranmer, S.J. & Desmarais, B.A. (2017) Stochastic weighted graphs: flexible model specification and simulation. *Social Networks*, **49**, 37–47.

Received 9 January 2017; accepted 24 February 2017

Handling Editor: Jessica Metcalf

Supporting Information

Details of electronic Supporting Information are provided below.

Data S1. The application of statistical network models in disease research: Word document containing a description of and results from the four example analyses used in the paper, together with the annotated R code for implementing these examples.

Data S2. (Ages.csv): Age data for use in network autocorrelation model and exponential random graph model examples.

Data S3. (Complete Membership.csv): Social community membership for use in network autocorrelation and exponential random graph model examples.

Data S4. (indivsexes.csv): Sex data for use in network autocorrelation model and exponential random graph model examples.

Data S5. (overallnetwork.csv): Network data for use in network auto-correlation model and exponential random graph model examples.

Data S6. (TBstatsF.csv): bTB infection data for use in network auto-correlation model and exponential random graph model examples.

Data S7. (autumnmatrix.csv): binary autumn network for use in stochastic actor-oriented model and temporal exponential random graph model examples.

Data S8. (summermatrix.csv): binary summer network for use in stochastic actor-oriented model and temporal exponential random graph model examples.

Data S9. (wintermatrix.csv): binary winter network for use in stochastic actor-oriented model and temporal exponential random graph model examples.

Data S10. (grouplocsSAOM.csv): group location data for use in the stochastic actor-oriented model example.

Data S11. (MembershipSAOM.csv): Social community membership data for use in the stochastic actor-oriented model example.

Data S12. (SAOMsexes.csv): Sex data for use in the stochastic actor-oriented model example.

Data S13. (SAOMTBstats.csv): bTB infection data for use in the stochastic actor-oriented model example.

Data S14. (MembershipTERGM.csv): Social community membership data for use in the stochastic actor-oriented model example.

Data S15. (TERGMsexes.csv): Sex data for use in the stochastic actor-oriented model example.

Data S16. (TERGMBstats.csv): bTB infection data for use in the stochastic actor-oriented model example.

